Vasco Manuel Jorge Soares Mantas

Remote Sensing of Oceanic fronts off Portugal (2003-2014). An integrated environmental approach

Tese de doutoramento em Geologia, ramo de Recursos Geológicos e Ambiente orientada pelo Prof. Doutor Alcides José Sousa Castilho Pereira e Prof. Doutor João Carlos Marques e apresentada no Departamento de Ciências da Terra da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Dezembro de 2016

· U C ·

Universidade de Coimbra

Remote Sensing of Oceanic fronts off Portugal (2003-2014). An integrated environmental approach

Detecção remota de frentes oceânicas ao largo da costa Portuguesa (2003 – 2014).

Uma perspectiva ambiental integrada

Vasco Manuel Jorge Soares Mantas

Tese de doutoramento em Geologia, ramo de Recursos Geológicos e Ambiente orientada pelo Prof. Doutor Alcides José Sousa Castilho Pereira e Prof. Doutor João Carlos Marques e apresentada no Departamento de Ciências da Terra da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Dezembro de 2016

· U C ·

UNIVERSIDADE DE COIMBRA

Imagem de capa: Space Shuttle *Challenger* sobrevoa Portugal durante a missão STS-51-F. Imagem obtida a 30 de Julho de 1985. Fotografia NASA.

“Nothing like a perfect zoological division of the earth is possible. The causes that have led to the present distribution of animal life are so varied, their action and reaction have been so complex, that anomalies and irregularities are sure to exist which will mar the symmetry of any rigid system.”

Wallace, 1876

This work was supported by the Fundação para a Ciência e Tecnologia through the grant SFRH/BD/89972/2012.

Agradecimentos

Qualquer obra resulta não apenas do esforço do autor mas das condições reunidas para a sua elaboração, sendo de particular relevância a influência positiva de muitos daqueles que nos rodeiam.

Agradeço assim ao Prof. Doutor Alcides Pereira e Prof. Doutor João Carlos Marques, meus orientadores, por todo o inestimável apoio, aconselhamento, confiança e amizade com que sempre me brindaram quando necessário.

Ao Laboratório de Radioactividade Natural, por onde uma parte significativa da minha carreira passou. A todos os que nele trabalham ou colaboram, hoje ou no passado, os meus sinceros agradecimentos. Em particular deixo uma palavra de agradecimento ao Prof. Doutor Luís Neves, pelo apoio dado ao longo dos anos e constante simpatia.

Ao MARE e a todos os seus membros (especialmente a Gabi) por tornarem tudo mais fácil, garantindo que para todos os obstáculos havia uma solução.

A Zhong Liu pelo apoio científico, conversas e sugestões que se tornaram indispensáveis para a concretização de diferentes tarefas integrantes deste trabalho.

Muitos foram também aqueles que através de conversas e emails contribuíram com informação e encorajamento académico, ajudando de muitas formas para a obra que agora apresentamos.

Ao Paulo Santos, que durante estes anos foi um amigo sempre presente e que nunca vacilou na sua camaradagem.

À Sofia Grade, amiga de longa data e de inestimável valor. Constante fonte de inspiração, tantas vezes necessária para ultrapassar o inultrapassável.

Aos meus pais, por tudo. É difícil e limitante enumerar todas as facetas do apoio que os meus pais desde sempre me deram. As oportunidades que tive ao crescer, os valores que me incutiram, a liberdade para escolher o caminho que desejava são apenas alguns dos elementos fundamentais que contribuíram para que enfrentasse todas as etapas que culminam neste momento. A minha gratidão não pode portanto ser adequadamente expressa por palavras, restando-me assim sublinhar o meu sincero agradecimento.

Às minhas irmãs agradeço o apoio, motivação e exemplo.

À Ana, pilar fundamental que me permitiu embarcar neste desafio com maior tranquilidade e segurança. Pelo apoio e compreensão, mesmo nos momentos mais desafiantes, o meu obrigado.

List of Acronyms

AIC: Akaike Information Criterion (see BIC)

AMO: Atlantic Multidecadal Oscillation

BC: Biogeographic Classification

BIC: Bayesian Information Criterion (see AIC)

BU: Biogeographic Unit

CMS: Content Management System

CTD: Conductivity, Temperature, and Depth

CTZ: Coastal Transition Zone

CUE: Canary Upwelling Ecosystem

DTW: Dynamic Time Warping

EBSA: Ecologically and Biologically Significant Areas

ENACW: Eastern North Atlantic Central Water

ENSO MEI: Multivariate El Niño/Southern Oscillation Index

EO: Earth Observation

EOF: Empirical Orthogonal Function

EROS: Earth Resources Observation and Science

ESA: European Space Agency

FPAR: Fraction of Photosynthetically Active Radiation

FTP: File Transfer Protocol

GDAL: Geospatial Data Abstraction Library

GERM: General Empirical Relation Model

GOES: Geostationary Operational Environmental Satellite

HAB: Harmful Algal Bloom

ICF: Instance Configuration Files

IMERG: Integrated Multi-satellitE Retrievals for GPM

IPC: Iberian Poleward Current

IT: Internal Tides

ITCZ: Intertropical Convergence Zone

IW: Internal Wave

JEBAR: Joint Effect of Baroclinicity and Relief

LME: Large Marine Ecosystems

LOD: Limit of Detection

MODIS: Moderate Resolution Imaging Spectroradiometer

MPA: Marine Protected Area

MSS: Multispectral Scanner System

NAC: North Atlantic Current

NAO: North Atlantic Oscillation

NBR: Normalized Burn Ratio

NDMI: Normalized Difference Moisture Index

NDVI: Normalized Difference Vegetation Index

nLw: Normalized Water-leaving Radiance

NOAA: National Oceanic and Atmospheric Administration

OC: Chlorophyll a concentration

PCA: Principal Component Analysis

PC: Principal Component

PCCC: Portugal Coastal Counter Current (synonym for Iberian Poleward Current, IPC)

PFD: Probability of Front Detection

PFZ: Potential Fishing Zone

PoC: Portugal Current

POD: Probability of Detection

SAR: Synthetic Aperture Radar

SEOF: Standard Empirical Orthogonal Function

SNR: Signal-to-Noise Ratio

SRE: Satellite Rainfall Estimates

SST: Sea Surface Temperature

SWODDIES: Slope Water Oceanic Eddies

TEPS: TRMM Explorer Processing Script

TIRS: Thermal Infrared Sensor

TMPA: Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis

TRMM: Tropical Rainfall Measuring Mission

TSG: Thermosalinograph

TZCF: Transition Zone Chlorophyll Front

UCF: User Configuration Files

UI: Upwelling Index

UI_{sst}: Upwelling Index from Sea Surface Temperature

UPGMA: Unweighted Pair Group Method with Arithmetic Mean

USGS: United States Geological Survey

VISSR: Visible and Infrared Spin-scan Radiometer

VHRR: Very High Resolution Radiometer

WCSS: Within-cluster Sum of Squares

WIWiFs: Western Iberia Winter Fronts

WIP: Western Iberian Peninsula

WIBP: Western Iberian Buoyant Plume

Index

Abstract

Oceanic thermohaline fronts are natural boundaries separating water masses with different characteristics including temperature, salinity, current direction and velocity, and productivity. Fronts can be described as discontinuities and ecosystems, promoting the development of complex and vibrant communities of marine life, from microscopic phytoplankton to large marine mammals and birds. Fronts are also the natural and measurable boundaries delineating regions highlighted by the analysis of biophysical variables at different scales.

Biogeographic units, and the fronts separating these natural regions, are influenced by a complex set of variables either local or remote. Phenomena such as upwelling jets, the interaction of different currents (e.g. Iberian Poleward Current and the Portugal Current) or river discharge plumes and the buoyant lenses formed by their action, can all generate strong and sometimes seasonally recurrent fronts. As such, understanding water masses and fronts, the object of many biogeographic studies, requires a keen understanding of oceanic dynamics and watershed processes, as suggested by the Large Marine Ecosystems approach.

In this work it is presented a comprehensive approach to the analysis of the coastal ocean off central Portugal up to longitude -12° W. By relying on Earth Observation data from the MODIS sensor, a long time series (2003-2014) of Sea Surface Temperature (SST) and Chlorophyll-a Concentration was created. The daily SST images were used to extract the location of the thermal fronts (and Chlorophyll-a fronts for selected dates) at a spatial resolution of 1 km.

A novel and comprehensive set of frontal probability maps were thus generated for a number of temporal resolutions including weekly, monthly, seasonal, and yearly versions. These maps were analyzed to support the identification of trends on the location, direction, and connection to other oceanographic and geographic features.

Fronts off Portugal tend to be strongly influenced by the sub-meridionally aligned coastline, with features such as Cape Carvoeiro also contributing to the formation of important frontal regions. The shelf edge is also a relevant area of frontal activity, similarly to finding reported by other studies in comparable regions, such as California. The Western Iberia Buoyant Plume is also a determinant driver of frontogenesis during the winter, and the impact of the Western Iberia Winter Front is also visible in the frontal map.

The SST and Chlorophyll-a Concentration datasets, were used in an attempt to create a novel bioregionalization scheme based on the use of Dynamic Time Warping (DTW) as a means to analyze the dissimilarity of monthly time series with a 4 km spatial resolution. The dissimilarity matrices were used to feed a clustering attempt which created a set of nested regions that segmented the coastal ocean off Portugal. The maps thus created highlighted the processes at

play in the different sectors of the study area, which were separated by regions of heightened frontal activity that could occur throughout the year or in specific seasons.

The frontal maps and the biogeographic classification relied on the analysis of a massive dataset using innovative techniques. In fact, the DTW-based clustering approach was used for oceanographic purposes for the first time, to the best of our knowledge. This approach promises significant advances in the field partly because it may be easily replicated and escalated to other, broader regions.

The analysis of frontal and bioregionalization maps made it clear that the processes taking place in adjacent (or somewhat distant) watersheds were determinant. The formation of river discharge plumes (and associated fronts) as well changes to coastal ocean color were determined to a large extent by rainfall and land cover processes inland.

As such, a land cover mapping initiative was initiated, focusing on soil imperviousness. A complete Impervious Surface Area (ISA) map was created using Landsat-8 imagery and machine-learning techniques (Regression Tree Model) for the entire Mondego river watershed in Portugal. The resulting map provides 30-meter spatial resolution ISA estimates (0-100%) with a Mean Average Error (MAE) of 1.6% and Root Mean Square Error (RMSE) of 5.5%, along with strategies for its regular update. This is currently the best product available in the watershed and amongst the most accurate in Europe.

The product was used to support an assessment on the connection between the anthropization of the watershed and water quality (and color). In fact, Earth Observation data directly imaged the occurrence of eutrophication events in dams of the region. Furthermore, Landsat data were also used to demonstrate the impact of human activities on ocean color, potentially affecting the retrieval of information and hampering the development of high resolution frontal maps and bioregionalization attempts near-shore.

Rainfall, an important element connecting land and ocean through rivers was also analyzed. In particular, the possibility of employing Satellite Rainfall Estimates (SRE) for watershed studies was evaluated. The limited freely available rainfall datasets from *in situ* gauges reinforced the need for alternative (or redundant) and survivable data sources. SREs showed moderate correlation with *in situ* sources at a daily scale, but improved over monthly intervals ($R = 0.80$). SREs are affected by several variables described in the thesis, including terrain complexity.

The information generated in the project, for the different domains, is relevant to a number of stakeholders. The relevance of disseminating the information cannot be overstated and therefore a stepwise program dedicated to the development of information sharing systems was initiated, resulting in the launch of the global TRMM Explorer and SmartBasins portals.

This work answered several important questions and opened new lines of research with important implications for scientific and operational applications in several different but interconnected domains.

Keywords: Ocean fronts; Ocean Mesoscale; Remote Sensing; Biogeography; Natural resources management; Time series

Resumo

As Frentes termohalinas oceânicas constituem fronteiras naturais entre massas de água que apresentam características distintas, incluindo temperatura, salinidade, direcção e velocidade da corrente e produtividade. As frentes podem ser descritas como descontinuidades e ecossistemas, promovendo o desenvolvimento de comunidades biológicas complexas e vibrantes que incluem uma ampla gama de organismo do fitoplâncton microscópico aos grandes mamíferos marinhos ou aves.

As frentes são também as fronteiras naturais e mensuráveis que delineiam regiões evidenciadas pela análise de variáveis biofísicas a diferentes escalas.

As unidades biogeográficas, e as frentes que separam estas regiões naturais, são influenciadas por um conjunto complexo de variáveis tanto locais como remotas. Fenómenos como o afloramento costeiro, a interacção entre diferentes correntes (e.g. Iberian Poleward Current e a Portugal Current) ou as plumas geradas pela descarga dos rios, podem no seu conjunto gerar frentes por vezes fortes e sazonais. Assim, a compreensão das massas de água e frentes, o objecto de muitos estudos biogeográficos, exige um forte conhecimento das dinâmicas oceânicas e dos processos que se desenrolam nas bacias hidrográficas, como sugerido pela abordagem seguida na definição dos Large Marine Ecosystems.

No presente trabalho é apresentada uma abordagem exaustiva à análise do oceano costeiro ao largo do centro de Portugal até a uma longitude de -12° W. Recorrendo de forma extensiva a dados de Observação da Terra provenientes do sensor MODIS, foi construída uma série temporal longa (2003-2014) de Temperatura da Superfície do Oceano (SST) e Concentração de Clorofila-a. As imagens SST diárias foram utilizadas na extracção da localização das frentes térmicas com uma resolução espacial de 1 km.

Mapas de probabilidade de ocorrência de frentes inteiramente novos foram gerados com diferentes resoluções temporais incluindo semanais, mensais, sazonais e anuais. Estes mapas foram analisados de forma a identificar tendências na localização, direcção e correlação com outros fenómenos e traços oceanográficos e geográficos.

As frentes ao largo de Portugal são fortemente influenciadas pela costa alinhada aproximadamente ao longo de um plano meridional, com acidentes geográficos como o Cabo Carvoeiro a contribuir para a formação de importantes regiões de actividade frontal. O talude continental constitui também uma área relevante para a formação de frentes, como sugerido em estudos anteriores efectuados em regiões comparáveis, como a Califórnia. A Western Iberia Buoyant Plume é também determinante no processo de frontogénese durante o inverno e o impacto da Western Iberia Winter Front é também visível nos mapas gerados.

Os dados de SST e Concentração de Clorofila-a foram usados na tentativa de criação de um novo esquema de bioregionalização baseado na utilização de Dynamic Time Warping (DTW) como método de análise de diferença das séries temporais. As matrizes de diferença foram depois utilizadas na definição de *nested clusters* que segmentaram o oceano costeiro ao largo do centro de Portugal. Os mapas assim criados evidenciaram os processos que se desenrolam nos diferentes sectores da área de estudo e que são separados por regiões de elevada actividade frontal, que pode ocorrer ao longo do ano ou em períodos específicos.

Os mapas de frentes e a classificação biogeográfica assentam na análise de um conjunto de dados extenso usando técnicas inovadoras. De facto, a utilização do método baseado em DTW foi empregue, tanto quanto sabemos, pela primeira vez numa aplicação oceanográfica. Esta abordagem promete avanços significativos na disciplina em parte por ser facilmente replicada e escalável a outras regiões mais vastas.

Da análise destes dados tornou-se claro que os processos que ocorrem nas bacias hidrográficas adjacentes (ou mesmo distantes) são determinantes para a dinâmica costeira. A formação de plumas de descarga fluvial (e frentes associadas) e a influência sobre a cor do oceano são determinadas em grande medida pela precipitação e processos associados à cobertura do solo em terra.

Assim, uma iniciativa de mapeamento da cobertura do solo foi iniciada centrada na impermeabilidade dos solos. Um mapa de Impervious Surface Area (ISA) foi criado a partir de dados Landsat-8 e técnicas de *machine-learning* (Modelos de Regressão em Árvore) para a bacia hidrográfica do Mondego, em Portugal. O mapa resultante fornece estimativas de ISA (0-100%) com uma resolução espacial de 30 metros e um Erro Médio Absoluto de 1.6% e Raiz Quadrada do Erro Quadrático Médio de 5.5%. Este é actualmente o melhor produto disponível para a bacia e encontra-se entre os mais precisos da Europa.

O produto foi utilizado como auxiliar na determinação de relações entre a antropização da bacia e a qualidade da água. Correlações significativas foram encontradas que podem afectar potencialmente a cor da água (e qualidade) de uma forma mensurável. Na realidade, dados de Observação de Terra sublinharam a ocorrência de episódios de eutrofização em barragens da região. Ademais, os dados Landsat permitiram demonstrar o efeito das actividades humanas na cor do oceano, potencialmente afectando a obtenção de informação e limitando o desenvolvimento de mapas de frentes de alta resolução e tentativas de bioregionalização em sectores próximos da costa.

A precipitação, elemento importante que liga terra e mar através dos rios foi também analisada. Em particular, avaliou-se a possibilidade de empregar Estimativas de Precipitação por Satélite (SER) foi avaliada. As limitações identificadas na disponibilidade de dados de

precipitação *in situ* reforçaram a necessidade de procurar uma alternativa viável e duradoura. As estimativas por satélite apresentaram uma correlação moderada com os dados *in situ* à escala diária, melhorando com agregação mensal (R = 0.80). As estimativas são afectadas por diversas variáveis incluindo a complexidade do terreno.

A informação gerada neste projecto, nos seus diferentes domínios, é relevante para diversos stakeholders. A relevância associada à disseminação da informação deve ser sublinhada e por essa razão um programa faseado dedicado ao desenvolvimento de sistemas que sirvam esse propósito foi desenvolvido, resultando no lançamento dos portais globais TRMM Explorer e SmartBasins.

Este trabalho procurou assim responder a diferentes questões relevantes e abrir novas linhas de investigação com implicações científicas e operacionais em domínios distintos mas relacionados.

Palavras-chave: Frentes oceânicas; Mesoescala oceânica; Detecção Remota; Biogeografia; gestão de recursos naturais; Séries temporais.

PART I – Introduction

Chapter 1. Introduction

Coastal oceans harbor some of the most important but also vulnerable ecosystems in the world (Costanza et al., 1997, Longurst, 2007). The natural resources available in coastal oceans and land areas, together with the importance to industry and trade, create the conditions for a strong anthropization of these regions. As a result, more than half of the world's population is concentrated along a narrow strip of land within 100 km of the coastline. As a consequence (or a cause perhaps) these areas assume a disproportional role in the global economy (Brown et al., 2013) and along with it, critical levels of pressure are exerted over natural resources and ecosystems.

Still, data needs are still significantly underserved and much research is required in order to guarantee an efficient management of coastal resources and an effective protection of critical ecosystems. In a context of climate change, emerging trends affecting critical variables can have profound impacts on the coastal ecosystems and livelihood of coastal communities. Significant warming trends have been measured in a sobering 71% of the planet's coastlines (Lima et al., 2012), including off Western Iberia (Relvas et al., 2009). This is even more disturbing if we consider that the ocean's temperature explains to a great extent the current ecosystem structure along the world's coastlines (Belanger et al., 2012). This is in line with the findings of Roff et al. (2012) who suggested that physical data (most notably Sea Surface Temperature (SST)) can explain the distribution of biota. Emerging changes in temperature patterns may therefore disrupt current patterns and structure in perhaps unpredictable ways.

The combined effects of these changes create the conditions for an emerging climate-related set of national security risks (CSAG, 2016). This challenging environment can only be addressed adequately if reliable scenarios are built. These must be capable of describing future physical, ecological, and societal conditions found in coastal areas worldwide, and therefore require the development of comprehensive datasets describing current and past conditions. In this context, understanding the distribution and structure of water masses is an important stepping-stone towards the development of comprehensive ecological models and forecasts for the costal environment.

However, the coastal ocean is highly dynamic and heterogeneous, challenging attempts to create arbitrary boundaries based on partial datasets created from traditional observation techniques, administrative and historical reasons, or outright convenience. Understanding the physical structure of these heterogeneous water masses is thus both a momentous and central task currently driving significant advances in oceanography and biogeography (Longhurst, 2007).

However, and because marine biota is much harder to sample and observe (a significant portion is microscopic) than land-based organisms, defining regions and locating (dynamic) boundaries is not as straightforward as in terrestrial studies (IOCCG, 2009). As such, alternative methods are often sought, creating a plethora of methodologies combining physical and biological variables (UNEP-WCMS, 2006).

Off Portugal, current biogeographic systems (e.g. Longhurst, 2007) fail to provide the detailed perspective required for management purposes or the regional segmentation of the ocean for scientific applications. Despite previous efforts (e.g. Relvas et al., 2007, Peliz et al., 2009, Peliz et al., 2005) the boundaries that divide such regions – thermohaline fronts – are also still poorly understood and described.

In the best tradition of the Large Marine Ecosystems (Sherman et al., 2004), we shall connect oceanic biogreography with adjoining watersheds in this study, thus creating a true continuum that acknowledges the connectivity and driving forces influencing both domains collectively.

The complexity of such an endeavor is best explained by a workflow capable of detailing the different processes and features that occur in the land and coastal domains that are intimately connected. Figure 1.1.highlights such processes and illustrates how these were analyzed in this study, through the development of a complex set of products based on a massive amount of Earth Observation and *in situ* data.

As such, in the second part of this thesis, after a presentation of the study area and problems we aim at addressing, we present a comprehensive new method based on Earth Observation data to segment the ocean. It is applied to the coastal ocean off central Portugal and discussed in some detail, in hopes of constituting a precursor for a similar global application.

The regions identified are only relevant if divided by real, measurable boundaries. Such boundaries albeit elusive, exist in form of the aforementioned thermohaline fronts. Oceanic fronts are interfaces highlighted by enhanced multivariate gradients between different water masses that constitute both a natural boundary and an ecosystem in themselves (Acha et al., 2004, Belkin et al., 2002, Longhurst, 2007). With a ubiquitous presence across the world's oceans, thermohaline fronts are extremely relevant for marine life at all levels, influencing the entire food web and the very physical structure of the ecosystems (Otero et al., 2009). These are narrow regions where the physical conditions lead to an enhancement of primary productivity and generate optimum conditions for the survival of marine life (Santos et al., 2011, Acha et al., 2004, Olson et al., 1985). These traits contribute to the build-up of entire food chains that rely on such systems (Castelao et al. 2008). Fish, marine mammals, and even birds are commonly observed in the vicinity of frontal regions, attesting to their relevance as a key element of the ecosystems (Polovina et al., 2001, Nevitt, 1999, Olson, 2002).

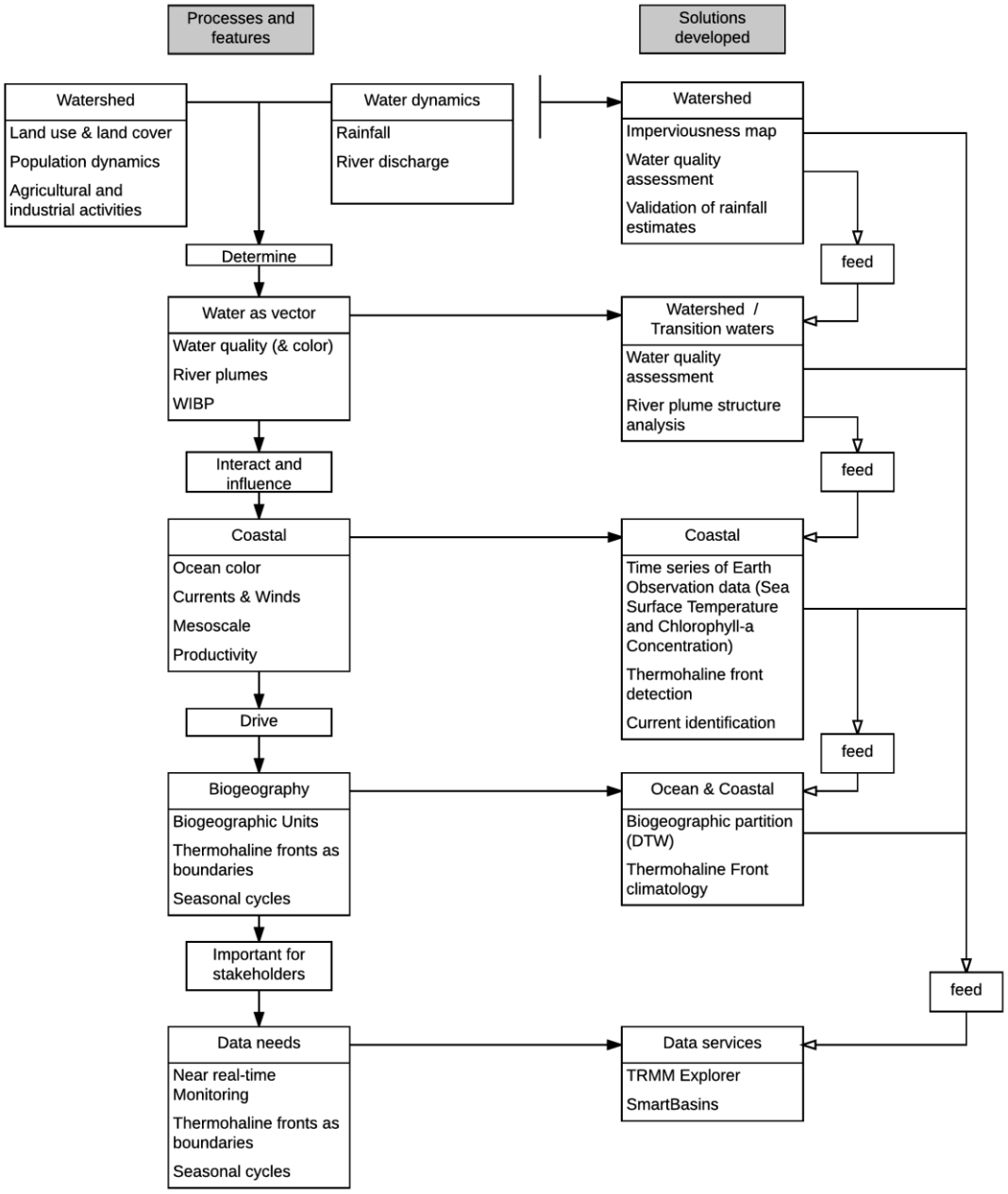


Figure 1.1. Generalized workflow of the thesis, showing the different processes and features mapped and analyzed with the ultimate goal of studying the natural boundaries of the ocean – thermohaline fronts. On the left hand side, it is depicted the processes and features of interest and how they are correlated, from watershed to the coastal environment. To the right, the products and services created in the course of this study.

Multiple processes drive front formation, from upwelling to river plumes, and understanding the distribution of these features is paramount to a range of scientific and operational applications.

Off the Portuguese coast, fronts are widespread and connected to several processes and geographic features (Mason et al., 2006) not unlike those observed in comparable regions (e.g. Castelao et al., 2008). However, our study area still lacks an in depth analysis of thermohaline front distribution and spatial patterns, even though these are important structural elements of the ecosystem. A time series was thus built (2003-2014) using Earth Observation data to detect and characterize fronts and then compare their distribution with oceanographic, topographic, and atmospheric drivers.

However, a significant part of the aforementioned pressures affecting coastal areas originate inland, in the countless watersheds that transport water, nutrients, and pollutants into the ocean (Lima et al., 2012). Water connects land and oceans in what constitutes a true continuum that must be widely recognized as such. Furthermore, several coastal features, like buoyant plumes, are generated by watershed-related processes, which must therefore be studied in detail to guarantee an accurate understanding of the dynamics at hand.

This continuum explains the third part of the thesis, which addresses the influence of land-based activities over water quality and color. The anthropogenic impact over the imperviousness of watersheds - an important variable affecting water quality - and our ability to monitor it is described. Rainfall, which binds the different components of the system, plays a central role in the establishment of the watershed-coastal system, while also constituting an important resource to the population. Monitoring precipitation from space opens new perspectives on the development of innovative critical real-time applications (Mantas et al., 2015), while contributing to the study of coastal features such as river discharge plumes and associated fronts.

All the aforementioned information is potentially important for a wide range of applications and to an extended set of stakeholders, internationally. Many research projects fail to connect the data generated with those who may benefit of the knowledge created in the process. We attempted to tackle this immense challenge through the development of a stepwise program dedicated to the development of data processing and distribution platforms. These were aimed at providing cost-effective tools to visualize and interpret critical data seamlessly, with the focus on an international audience composed of very diverse set of stakeholders.

These are the principles guiding the structure of this thesis, which aims at providing new, data-driven perspectives on different variables critical for ecological management and the operational (and sustainable) exploitation of the ocean. We expect our work to be capable of filling important gaps in our understanding of the ocean structure off central Portugal, while contributing to a growing body of information concerning land cover dynamics and atmospheric monitoring. The development of data portals was an ancillary challenge, which

we hope to promote as an important part of research projects that generate new datasets that can be employed in subsequent studies and applications.

The activities described in this thesis were heavily reliant on Earth Observation data and geospatial technologies, with the development of several new methodologies to generate and analyze information, some of which of significant value to stakeholders. Earth Observation systems are indeed bound to change our knowledge and perceptions on coastal processes and revolutionize regional oceanography and biogeography alike.

Researchers can now observe and measure critical oceanic variables with increasing accuracy at unprecedented rates. If this disruption promises groundbreaking advances it also calls for robust big data mining tools and approaches to be fully fulfilled. This project was designed under this assumption and at a time of paradigm shift, when new missions are routinely launched and new ways of extracting the information contained in the imagery are presented to the scientific community at an astounding pace. As such, we hope this study may offer relevant contributions to both the understanding of coastal processes and features as well as to the technologies employed in their research.

The thesis includes a combination of sections presented in the traditional form and as peer-reviewed papers published in scientific journals, duly identified as such.

Chapter 2. Ocean color missions and data

October 24, 1946 marks the dawn of space-based remote sensing, as a V-2 rocket was launched from the White Sands Missile Range in the United States of America and captured an image of Earth from an altitude exceeding 100 km (Fraser, 1985).

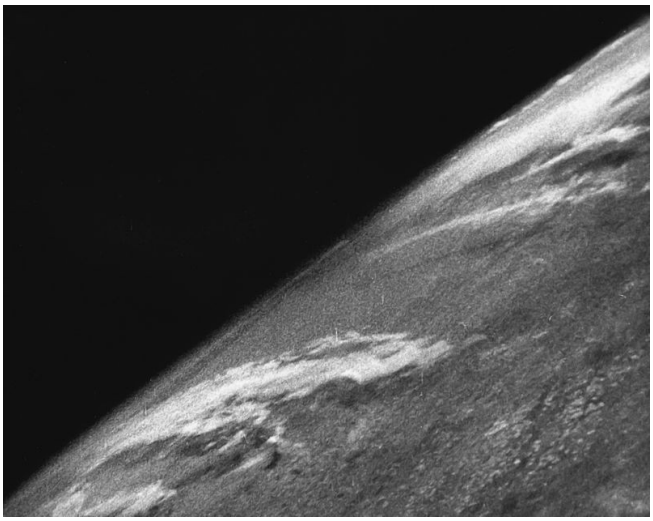


Figure 2.1. Photograph of Earth acquired by a 35 mm camera onboard a V-2 rocket launched from the United States of America on October 24, 1946.

The crude image depicted in figure 2.1 is a reminder of the simple beginnings of a now thriving and expanding space industry. A constellation of approximately 1419 satellites currently orbits the planet (UCS, 2016). This figure is hardly static with new satellites being added every month by multiple operators, both private and public.

Many of these satellites are destined to provide Earth Observation (EO) capabilities, monitoring the planet from different perspectives. With varying spatial, spectral, and temporal resolutions, researchers and operational users alike, have access to a plethora of data, which could only be imagined before that historical day in 1946.

The emergence of consumer-grade electronics and miniaturized sensors gave birth, in the past 5 years, to the development of yet another industry that promises to revolutionize remote sensing as well. Today, researchers can buy cost-effective remotely piloted or even autonomous platforms for low-altitude imagery acquisition for the same cost of a Landsat image prior to the opening of the archives. This revolution, if it is already transforming land sciences is still capable of only very modest contributions to oceanic monitoring and research due to limited range and imagining capabilities.

As such, in this work we focus on the use of satellite-based remote sensing technologies to explore the ocean off central Portugal. In this chapter we shall describe, very succinctly, the history and principles underpinning the use of EO sensors to detect, map, and interpret oceanic features at different spatial and temporal scales.

2.1. Missions

The field of ocean radiometry is based in the assumption that ocean color is determined by the pigments and particulates present in the water in a proportional way to its abundance (McClain, 2009). As such, we shall address primarily optical sensors throughout this chapter. Other important sensors including altimeters and Synthetic Aperture Radar (SAR), but because they were not used in this study, are not included in this summarized description of present and historical missions.

For a comprehensive review of the history of ocean remote sensing we suggest the recent and extensive work by Acker (2015).

The Coastal Zone Color Scanner (CZCS) was a pioneering mission launched in October 24, 1978 (curiously 32 years after the V-2 photograph) aboard the Nimbus-7 satellite.

The sensor was able to image both multispectral and thermal channels, at a spatial resolution of 825 m (Nadir) (Hovis et al., 1980).

Table 2.1.1. Bands of the Coastal Zone Color Scanner, launched in 1978.

Band	Central wavelength (nm)
1	443
2	520
3	550
4	670
5	750
6	1150

The CZCS outlived its operational design life and acquired thousands of images that helped revolutionize oceanography and opened a new window on processes that were up until then poorly understood or entirely unknown. For the first time, biotic and abiotic features could be detected, monitored, and the spatial arrangement accurately represented.

The mission left an important and lasting legacy, laying the very foundations of subsequent ocean radiometry missions (Antoine et al., 2003).

Despite the seminal role of the CZCS, other satellites were equipped with sensors capable of imaging the ocean, albeit not being designed exclusively (or at all) for that purpose. Signal-to-noise ratios and radiometric resolution are often barriers that preclude the exploitation of all EO missions for oceanographic purposes. Nonetheless, the last years of the 1970s were prolific and exciting in terms of ocean radiometry.

Starting with TIROS-M, NOAA satellites were equipped with the Very High Resolution Radiometer (VHRR) and later with the Advanced Very High Resolution Radiometer (AVHRR). As we shall discuss latter in this work, this sensor was paramount to the early study of the mesoscale, including of thermohaline fronts (Bernstein et al., 1977, Legeckis et al., 1978). In fact the seminal work of Cayula and Cornillon (1992) on front detection relies on AVHRR data from the NOAA-7 satellite.

The AVHRR was developed iteratively in a stepwise manner, as advances were tested and validated in proof-of-concept missions that pushed the technology ahead with each variant of the sensor. In its latest incarnation, the AVHRR/3 is a 4 channel radiometer, capable of imaging the planet at a spatial resolution of 1 or 4 km depending on the acquisition mode. This sensor builds on the experience and capabilities of earlier versions, starting with the 1978 4-channel sensor onboard TIROS-N.

The Landsat program, offers interesting examples of a set of satellites that although not designed for ocean radiometry were explored from its early days as potential (and later confirmed) data sources on the marine environment.

The first in the series was launched in 1972 and was equipped with an early version of the Multispectral Scanner, which included 4 bands with a spatial resolution of 80 meters. After Landsat-3 (launched in 1978), the sensor included a thermal band. However, it was with Landsat-4 and Landsat-5 (1982 and 1984) that the sensors started to resemble more closely those equipping the latest version (Landsat-8).

The Landsat program, due to its longevity and relative continuity of sensors offers important insights into the evolution of coastal regions. It is important to emphasize the role of Landsat in multiple aspects of environmental monitoring from watersheds to coastal areas and waters (Wulder et al., 2012, Williamson, 1997). The program spurred countless applications, including in the study area (e.g. Mantas et al., 2013). This was partly due to a breakthrough policy that enabled the free access to the archive maintained by the United States Geological Survey (Wulder et al., 2012b). In fact the open archives and the increased use of Landsat scenes, created the competitive environment required for other agencies worldwide to follow suit with similar policies.

Table 2.1.2. Simplified description of the bands available in the sensors (OLI and TIRS) equipping the Landsat-8 satellite.

Band	Wavelength (nm)	Spatial resolution (m)
1 (OLI)	433 – 453	30
2 (OLI)	450 – 515	30
3 (OLI)	525 – 600	30
4 (OLI)	630 – 680	30
5 (OLI)	845 – 885	30
6 (OLI)	1560 – 1660	30
7 (OLI)	2100 – 2300	30
8 (OLI)	500 – 680	15
9 (OLI)	1360 - 1390	30
10 (TIRS)	10600 – 11190	100 (30)
11 (TIRS)	11500 - 12510	100 (30)

Marine applications of Landsat data range from coastline mapping, turbidity assessment, coral reef monitoring, and sea ice detection (e.g. Thomas et al., 2002, Almonacid-Caballer et al., 2016, Hedley et al., 2016). These are but a subset of a growing body of potential uses given to coastal and marine scenes.

The applications emerge from a greater focus on the development of technical requirements that meet the necessarily stringent needs of the ocean color community (Hu et al., 2012). These include improved signal-to-noise ratios, the quality of thermal bands, and the introduction of a new band (433-453 nm) dedicated primarily to coastal studies (Vanhellemont et al., 2015, Irons et al., 2012). In this study, we shall use Landsat data when high resolution information, namely close to the shore is useful to detect or interpret specific features. The revisit period of Landsat is not sufficient to capture the inherent variability of the ocean, particularly in the coastal zone. Yet, the use of multi-mission data from 'virtual constellations' may provide the necessary coverage for at least some applications (Vanhellemont et al., 2016). Such a synergy can be found in the combined use of Landsat and Sentinel-2 data, which can provide higher frequency multispectral scenes with comparable spatial resolution.

Yet, modern ocean radiometry can be considered to have started with the launch of the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) onboard GeoEye's OrbView-2 satellite in 1997.

The sensor was pivotal to initiate continuous and accurate ocean color measurements with a focus on phytoplankton quantification and characterization. The sensor was thus able to

collect data in 8 spectral bands at variable spatial resolution (1.1 and 4.5 km). Revisit time (1 day) was also an important aspect of the system, which is an important limiting factor of several other missions (e.g. Landsat).

Because the sensor was developed and was onboard a private satellite, access to information was restricted and often provided upon a fee. Therefore, despite the technical capabilities offered by the system, its widespread use was somewhat limited by financial aspects.

Shortly after, a set of satellites from NASA and the European Space Agency (ESA) continued the legacy of SeaWiFS, which continued to collect data until 2010. These newer satellites included Envisat (ESA), TERRA (NASA), and AQUA (NASA). The two later satellites are equipped with multiple sensors, of which MODIS (Moderate Resolution Imaging Spectroradiometer) is the most pertinent to the field addressed by this study. With TERRA being launched on December 18, 1999, and Aqua on May 4, 2002, two MODIS sensors were operational and ensuring daily global coverage for the past 14 years.

If SeaWiFS had initiated a new age in the observation of the oceans, MODIS guaranteed its continuity and greatly expanded the resources available to researchers and operational marine data users. The open access policy of raw and higher-level products (including data at least up to level 3), favored the massive consumption of MODIS imagery worldwide. This imagery was made available at different websites maintained by NASA (e.g. at the Ocean Biology Processing Group) and other organizations.

MODIS is a complex and resilient instrument capable of passively acquiring information in 36 different bands, including multispectral and thermal data. Different bands were designed for different purposes, with Signal-to-Noise ratios reflecting such disparate design objectives and requirements. Band 9 (438 - 448 nm) of MODIS has a design SNR of 838, whereas the somewhat equivalent band 1 of Landsat 8 (433-453 nm) has a design SNR of 232 (Vanhellemont et al., 2015). This simple comparison is indicative of the advantages of MODIS over other EO sensors in the field of oceanography, where small radiometric differences in the targets can be significant. Still, the limited spatial resolution of the oceanic bands (1 km) can hamper the use of MODIS in coastal areas or transition waters.

Nonetheless, MODIS has been used extensively in oceanography to this date, including in the study area (Mantas et al., 2013). Marine applications range from the study of currents (Yuan et al., 2008), analysis of turbidity and coastal plumes (Shi et al., 2010, Petus et al., 2010, Shang et al., 2016), and research on Sea Surface Temperature (Kilpatrick et al., 2015), to name a few examples.

A variety of products has been developed for or adapted to MODIS since launch, including Chlorophyll-a concentration (O'Reilly et al., 2000), Remote Sensing Reflectance (Rrs_nnn)

(Carder et al., 2003), Normalized Water-leaving Radiance (nLw_nnn) (Gordon et al., 1994), and the aforementioned Sea Surface Temperature, SST (Brown et al., 1999).

Despite the similarities, the MODIS sensors equipping TERRA and AQUA are not entirely equal, and calibration issues plagued MODIS-TERRA, often leading researchers to prefer MODIS-AQUA for marine applications (Franz et al., 2007).

Sea Surface Temperature estimated from satellite sensors is paramount to a synoptic observation of numerous critical climate applications (Kilpatrick et al., 2015).

The retrieval of SST data relies on the correlation between emitted infrared radiation and that variable (Robinson, 2010). Several aspects must be addressed prior to SST estimations are made. This is a process not unlike other products in which the development and processing requires accurate atmospheric correction steps. Over the ocean that is particularly important – and difficult.

The definition of Sea Surface Temperature is also relevant, if it is to be used as an essential Climate Data Record. Although it may sound trivial, SST can be different, when measured by different instruments (including if estimated by infrared or microwave spaceborne sensors). The measurements are affected by the heat fluxes occurring in the upper layers caused by numerous forces including insolation, wind and wave action amongst others. These vertical thermoclines lead to differences in the first meter of water that can reach up to 1 K (or higher) (Robinson, 2010) and must therefore be acknowledged and estimated.

Instruments like buoys or vessels tend to collect data lower levels, often called bulk temperature measurements. The difference between skin and bulk measurements is a problem known to remote sensing experts for several decades (e.g. Schluessel et al., 1990) and has therefore been addressed in algorithm development efforts (e.g. May et al., 1998).

The diurnal thermocline is also a source of differences between the upper (~0.1 m) and bulk temperatures (~5 m). The SST algorithm applied to MODIS data and used in this study (Brown et al., 1999) mentions both problems and how these were addressed. Still, it is important to keep in mind that differences shall persist to a certain extent, which must be taken into account when complex interpretations are made on temperature gradients, front detection, and time series trends.

In this study, we employed the dual-band long-wave Sea Surface Temperature algorithm, which is considered to be accurate within 0.4 K. The algorithm is used in the software (l2gen) of the Ocean Biology Processing Group, as the standard SST processor.

The algorithm employs a dual-band approach, including Band 31 (10.780–11.280 μm) and 32(11.770–12.270 μm), with temperatures obtained by inversion through a correlation of

radiance and blackbody temperatures (Brown et al., 1999). In addition to the bands, the algorithm requires additional information including a baseline SST (Reynolds et al., 2002).

Because MODIS SST products are widely and successfully used by the scientific and operational community, it is widely accepted as a standard data source on ocean temperature.

On the other hand, the Chlorophyll-a Concentration algorithm, resorts to bands in the visible portion of the spectrum to estimate the near surface fields. In fact, obtaining data on the concentration of bio-optical pigments is only possible in this range (IOCCG, 2008). Chlorophyll-a Concentration is used as a proxy on phytoplankton biomass, thus offering users important information on the foundations of the food web (IOCCG, 2008)

The profound impact of ocean color is undeniable (Barber et al., 2000) and highlighted by the growing number of scientific papers published every year on this topic.

In this study we adopted the OC3M algorithm, which is also used as the default option for MODIS processing at the Ocean Biology Processing Group. The OC3M algorithm resorts to 3 different bands (443, 488, and 547 nm) and a number of coefficients, which empirically correlate *in situ* measurements and remotely sensed reflectance values.

Morel and Prieur (1977) raised important concerns regarding the validity of bio-optical variables in complex coastal waters. In this seminal study, the authors suggest a classification into two domains: Case-1 and Case-2. In Case-1 waters, phytoplankton determines the overall ocean color given its relative dominance. This is not the case in Case-2 waters, where other pigments do not vary linearly with phytoplankton and thus inorganic materials determine ocean color. This classification, albeit simplistic and sometimes used carelessly (Mobley et al., 2004), still offers an efficient way to classify waters and determine the overall reliability of the estimates.

In the future, we expect ocean radiometry to provide more than the basic variables currently held as climate data records (e.g. SST and Chlorophyll-a Concentration) but in order for that to be achieved it is pivotal to advance the characterization of water constituents.

Future sensors namely the Sentinels of the European Space Agency or the Joint Polar-orbiting Satellite System of the National Oceanic and Atmospheric Administration, along with currently active ones, shall constitute true virtual constellations able to provide different critical measurements across the spectrum.

Furthermore, the installation of sensors onboard non-traditional platforms such as the International Space Station (e.g. HICO) or fleets of small CubeSats as those launched by the private company PlanetLabs, open new perspectives to ocean observation from space . Complementing these data with that autonomous aerial vehicles and *in situ* assets of ever increasing accuracy and lower cost, will provide the data required to match the growing

computing power of cloud-based numerical models. This triad will enable nowcasts and forecasts to become increasingly powerful and able to provide definitive answers to growing societal challenges, including climate change.

In the next chapters we shall describe in greater detail the study area, the foundations, algorithms, and results of the use of ocean radiometry to determine the location and characteristics of ocean fronts off central Portugal and attempt a bioregionalization of these seas.

Chapter 3. Mesoscale phenomena and the Partitioning of the Ocean

3.1. Oceanic fronts: a mesoscale expression of the ocean

The mesoscale is sometimes described as the 'weather of the ocean', the inherent variability of water masses that affects regional patterns of key variables (e.g. Sea Surface Temperature) in relatively small areas (distances of 10-100 km) and short periods of time. The ocean mesoscale thus constitutes the combined response to variability at small temporal (sub-seasonal) and spatial (sub-basin) scales (Relvas et al., 2007).

Much of the variability at the mesoscale is created by the interaction of different forces, which translate into flows with high Reynolds number and thus heightened turbulence (Robinson, 2010). However this turbulence is not devoid of structure, but instead promotes the development of structures that work towards the reestablishment of the geostrophic equilibrium.

In the context of an oceanic basin, mesoscale features have a seemingly small size and temporal span. Nonetheless, the cumulative effect of these features may play an important role in the regional and global climate and contribute decisively to structuring biological communities across the globe (Xu et al., 2015, Otero et al., 2009).

The study of mesoscale variability has been aided by the advent of space-based marine observations, which provided much of the missing information required to understand the patterns that were incompletely described by *in situ* assets (Robinson, 2010). This variability provides the much sought-after structure of the open ocean and hence its relevance to a variety of fields such as biogeography and resource management (Nieblas et al., 2014).

Amongst the different mesoscale manifestations of the dynamic ocean, fronts are perhaps one of the most ubiquitous. Fronts have pervasive roles across multiple domains, from ecology (Bakun, 2006) to the dynamics of carbon flux and climate (Liu et al., 2016, Xu et al., 2015, Graham et al., 2013), but our body of knowledge on these features still needs to be significantly enhanced. The common perception of a continuous, almost unstructured ocean is still prevalent amongst important swaths of the population. Even government agencies worldwide tend to establish geometric administrative and management units that therefore have no adherence to the reality.

Le Fèvre (1986) points out that albeit the common perception, the ocean is hardly a domain characterized by 'progressive changes and smooth gradients' of key variables. Decades of oceanic observations, including by invaluable satellite sensors, highlighted the sharp

discontinuities that separate different water masses. These discontinuities – known as oceanic fronts – are central to our understanding of the dynamic ocean circulation patterns. These narrow zones where water masses with different characteristics come into contact are a true 'meeting of waters' in the words of Acha et al. (2004). Thermohaline fronts display maximum horizontal gradients of density, salinity or temperature (to name a few physical variables) and may also have a vertical expression (Belkin, 2002).

The aforementioned *meeting of waters* promotes the development of convergence areas, along a relatively narrow axis at the surface, of significant physical and ecological relevance, which is partly due to the separation of different water masses and the vertical motion promoted along the feature (Owen, 1981, Acha et al., 2004). Therefore, the contact between water masses often occurs as a sharp discontinuity, and not as a progressive mixing zone, despite some cross-front exchange in the vicinity of the boundary.

Fronts are ubiquitous in the global ocean and coastal seas and quite often detectable from shipboard observations made by scientists and mariners alike with nothing but the naked eye (Breaker et al., 2005). Such observations are reported historically, especially because of the ecological significance of these mesoscale features, which in turn affect economic activities such as fisheries. An example of the historical importance of fronts comes from Japan, where these features were named "Siome" and were known to fishermen, who could detect them visually in the sea surface (Uda, 1938, Yanagi, 1987). In the following section, we describe the ecological significance of fronts in greater detail, in light of the unique traits that characterizes them.

Figure 3.1.1. Left: A front separating the waters from Winyah Bay and the coastal Atlantic Ocean, as seen from an airplane. Notice the accumulation of debris or foam along the frontal axis. Image credit: NASA, GES DISC. Right: A front separates coastal (light blue) and ocean waters at the Gulf of Alaska. Image credit: Ken Bruland/USGS.

Fronts are thus, in essence, sub-vertical boundaries that separate water masses of different density and are visible at certain wavelengths (which may include the visible spectrum) from a bird's-eye view as they transect the ocean surface (figure 3.1.2) (Acha et al., 2004). Yet, fronts can be described in terms of the discontinuities they represent (Breaker et al., 2005) or of the flows associated to their axis (Chapman, 2014). These coexisting traits are defining elements of these features, which jointly contribute to the dual nature of fronts as boundaries (Margalef, 1997) and ecosystems (Longhurst, 2007).

If fronts can become barriers to marine life (and other materials) (Santos et al., 2004), they may also contribute to its dissemination and support in a variety of situations (Woodson et al., 2012). The broad spectrum of characteristics and possible impacts over the coastal oceanic structure and ecosystems supports a complex view of fronts, far from the one-dimensional character that sometimes was championed in early reports.

The ecological relevance and complexity arises from the 3-dimensional structure of these pervasive structures. Contrasting water masses are separated by a wedge that can reach depths in excess of 200 m (Armstrong et al., 1987), which crosscuts the surface where horizontal flows develop driven by the Coriolis force, along the frontal axis, in an attempt to establish a dynamic equilibrium (figure 3.1.2). These currents occur orthogonally to the gradient-driven convergent flows that power the accumulation of materials at the surface, along the frontal boundaries (Acha et al., 2004). The convergence and shear instabilities may play an important role in the enhancement of primary productivity (Longhurst, 2007), retention mechanisms (Santos et al., 2004), and in the establishment of important visual cues for oceanic organisms both above (Nevitt, 1999) and below surface (Belkin et al., 2002).

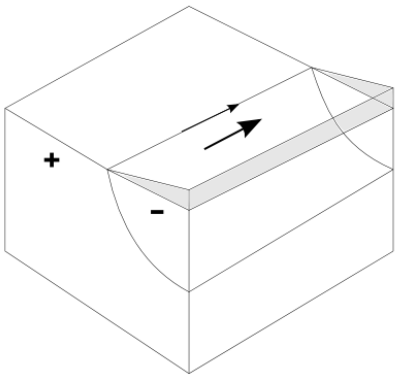


Figure 3.1.2. A typical and simplified depiction of the structure of a thermohaline front, with isopycnal separating low density (-) and high density (+) water masses. A flow is generated along the frontal axis and the low density water is forced vertically. The shadowed area represents the vertical displacement of the water mass.

At the temporal and spatial scale of this study, the majority of fronts to be detected will result of gradients created by differences in horizontal and vertical thermal advection as in the similar Californian system (Castelao et al., 2008). But different water masses are characterized by differences in several variables beyond temperature. In fact oceanic fronts are recurrently named thermohaline fronts after the combination of major gradients in thermal and salinity fields. Gradients of other variables capable of influencing ocean color may – and generally do - also occur over the same regions, but not necessarily in the same magnitude or superimposed to the thermal frontal axis (Belkin et al., 2009).

Discussing fronts is in fact complex due to the variety of shapes, sizes, and magnitudes these features can take. As Belkin et al. (2009) mentioned, fronts may occur over short or very long periods of time and even display a seasonal behavior. Similarly, small fronts may be formed by local and ephemeral phenomena or, in a strong contrast, extremely long and nearly continuous fronts can extend over entire basins and govern important processes of global significance.

Despite the differences, all fronts are relevant and can play a role in the structure of the ocean and ecosystems and must therefore be studied using appropriate data and methodologies to depict the features in different spatial and temporal scales. The challenges and complexity inherent to the study of these heterogeneous features should not be underestimated. Datasets adequate to study the Kuroshio Front in the Pacific Ocean must necessarily differ from those employed to research the fronts created by the Mondego river plume, in Portugal. Differences in the size and magnitude of both fronts calls for customized or flexible solutions, adequate at the horizontal scales that characterize each feature.

The diversity of traits and views that characterize fronts are, at least partly, due to the different processes driving frontogenesis, which include mechanisms of variable complexity. Some can be explained by rather straightforward drivers, such as river plumes, coastal topography, and bathymetry, while others constitute a response to major ocean circulation patterns and winds. In fact, major Polar, Subtropical and Equatorial fronts are amongst the most relevant of these features for the biogeography of the ocean (Longhurst, 2007).

In this context, it is possible to classify fronts into at least three major groups. These include 1) Open Ocean fronts; 2) Shelf-edge and upwelling fronts; and 3) Tidal fronts and river plume-driven fronts at shelf seas.

In reality the three categories often coexist, and fronts generated by each mechanism – or by a conjunction of several – occur in the coastal and shelf-edge regions, where they can play a significant role in the regional ecology and economy, as it will be analyzed in subsequent sections.

Off the coast of California, where the homonym current can be found, the coexistence of different types of fronts is representative of the complexity behind frontal patterns in regions dominated by Eastern Boundary Currents (as in the area of study of the present work, influenced by the Portugal Current) (Breaker et al., 2005).

If, near the coast, the influence of river discharge and coastline features create small-scale fronts that can have a variable expression throughout the year, bottom topography in the neritic zone can lead to semi-permanent features. There, the interaction of water masses with local geographic features such as canyons and seamounts creates complex patterns of frontal activity, which can be recognized in satellite imagery. The shelf-edge is, naturally and in this context, a major driver motivating the formation of significant fronts that often occur over the 200 m isobaths (Castelao et al., 2008).

Upwelling is also an important mechanism responsible for the formation of seasonal fronts, which may be recurrently observed in similar regions (Castelao et al., 2008). This phenomenon can be caused by persistent upwelling-favorable winds and wind stress curl. In the case of the Portugal and California Currents, these winds are northerlies, which persist through several months (Fiúza et al., 1982) and may even motivate off-season events during winter (Santos et al., 2004).

Offshore forcing caused by Ekman transport during upwelling events, leads to the seaward motion of water masses, which are replaced by cold, nutrient-rich waters, separated horizontally from oceanic waters by a prograde front (Mooers et al., 1978, Longhurst, 2007). The upwelled waters will first generate fronts close to the shore, which migrates progressively seaward as the winds intensify. When the equilibrium is reached a somewhat stationary front may be formed, constituting the aforementioned semi-recurrent upwelling front characteristic of several coastal regions worldwide (Breaker et al., 2005, Mann et al., 1991).

The seasonal character of upwelling fronts is shared with that of other features, such as meanders and filaments, which also play an important role in the formation of local, highly variable fronts (Castelao et al., 2008). It is also important to emphasize that, as mentioned before, front generation and distribution are the result of the complex interplay of variables. Even in the case of upwelling fronts (and upwelling itself), coastline geometry and bathymetric features will also influence the distribution, persistence, and magnitude of said fronts.

Frontal evolution upon formation also depends on a variety of factors, some of which inherently connected to the processes leading to their onset. The stability of the fronts is only apparent, as Coriolis force will lead to progressive meandering and eddy shedding, the later evolving as separate entities (Robinson, 2010) and may create so-called “eddy-streets” that propagate over long distances (Longhurst, 2007). The simple motion of water masses driven by

local or distant forces, will lead to a displacement of fronts, leading to significant inter-annual variability, especially in maps covering broader areas.

It is the significance and complex nature of fronts that spur a growing interest over their development cycle, distribution, and impacts at the regional and global level. In subsequent sections and chapters we shall analyze the structure, spatial patterns, and functions of fronts in greater detail, with a strong emphasis in the area of study, located off central Portugal. Unlike the customary basin-wide studies on major regions and discontinuities, we focus on the regional mesoscale variability capable of influencing numerous processes of physical, ecological, and economic significance.

3.2. Ecological and economic significance of fronts

The relevance of fronts can hardly be overstated. Not only do they separate different water masses, each with specific properties and influencing (or supporting) ecosystems differently, but fronts are in themselves relevant pieces of the ecological puzzle.

The ocean as a fluid system is characterized by the ubiquitous presence of fronts, which are an inherent component of the ecosystem (Acha et al., 2004). First and foremost, fronts can be seen as permeable ecotones (Margalef, 1997), dividing water masses and being responsible for the retention of fish larvae, phytoplankton, and nutrients, which will experience limited cross-frontal transport. But, considering that fronts are also areas where there is a measurable increase in biological activity it is possible to consider these features as ecosystems as well. Furthermore, under specific conditions, fronts may actually act as conveyors, distributing marine organisms from distant areas into their adult habitats (Woodson et al., 2012).

But, as Longhurst (2007) pointed out, fronts are simultaneously habitats (in themselves) and a transition between different regions. In fact fronts constitute one of three different main ecotones found at sea, several of which related to mesoscale activity. These include convergent and divergent fronts, mesoscale eddies and filaments, and finally, ecotones that are driven by semidiurnal tides.

This complex and perhaps ambiguous or duplicitous nature of fronts (perhaps motivated by an incomplete or biased perspective on marine systems) becomes apparent upon a close inspection of the processes that take place in these regions.

Paradoxically, fronts are meeting places of excellence. Mixing occurs horizontally and vertically, which promotes the enhancement of primary productivity that has been reported for several decades (Olson and Backus, 1985). Understanding these processes using traditional observation techniques is difficult, but thanks to Earth Observation sensors, this enhancement

can be objectively and routinely recorded for both scientific and operational purposes (e.g. Acha et al., 2004).

This is particularly important for a range of applications, from ecosystem assessment to the deployment of information systems mapping potential fishing zones (PFZ) as available, for instance, in India (ESSO, 2016) and the United States of America (Roffer's, 2016).

PFZ advisories based on satellite data are feasible because ocean color and mesoscale features, such as fronts, are not limited to the surface but have a subsurface expression, sometimes significant. Fronts can thus have a physical impact in the redistribution, and blocking, of marine organisms at depths of as much as 200 m, due to cross- and along-front flows created by the strong gradients (Breaker et al., 2005)

Furthermore, floating organisms are drawn to fronts because the flow is often convergent in the cross-front direction (Castelao et al., 2008). The convergence leads to an accumulation of an ever greater amount of both organisms and detritus while retaining new production and thus influencing all levels of the trophic chain. The convergent flow is, in fact, the foundation over which frontal ecosystems develop and flourish, more or less permanently, depending on the persistence of regional current patterns across seasons.

Acha et al. (2004) summarizes some of the plethora of significant impacts frontal systems have over marine biological communities worldwide. Amongst the different roles played by fronts, the authors specify the aforementioned enhancement of primary production, the establishment of sheltered areas and of feeding grounds for nektonic and benthic species or larvae. Such widespread effects generally have an equally transversal impact over communities. In the case of frontal systems this assumption holds, as animals in the higher trophic levels are attracted to regions with higher gradients.

That attraction is caused by the enhanced biological activity generated by the unique physical conditions made possible by the ocean's frontal systems, which create valuable feeding grounds that can be detected using visual (and other) cues. When fronts occur at upwelling areas, their role becomes even more relevant and may contribute to the development (or enhancement) of relevant ecosystems, even in the oceanward side of the system due to cross-front transport (e.g. Mooers et al., 1978). On the other hand, in the shoreward side of the upwelling front, the enhancement of primary productivity creates improved conditions for pelagic fish to feed (Garrido et al., 2006).

But fronts can not only attract organisms, but separate different species assemblages. In California, salmon and albacore tuna are found in different sides of thermohaline fronts, with each preferring waters with different characteristics (Breaker et al., 2005). Catches by trolling of albacore tuna in the Pacific are known to be closely correlated with frontal systems, in

particular of the Transition Zone Chlorophyll Front (and associated thermohaline boundary) (Polovina et al., 2001).

Understanding these relations of biota and fronts (and water masses they separate) is thus essential to promote a sustainable and efficient management of marine living resources.

The motility (or lack thereof) of different biota may also contribute to the distribution of life across frontal regions (Longhurst, 2007). Floating organisms are likely to find a barrier in fronts, precluding their further expansion offshore, thus creating an effective biological retention barrier. Such barriers are actually important to guarantee the successful development of eggs and larvae and prevent their offshore spreading into unfavorable grounds (Santos et al., 2004).

Off-season upwelling events lead to the development - and oceanward migration - of a frontal system with a negative impact on small pelagic fish recruitment off Portugal (Santos et al., 2001). This comes to show that frontal regions play an important role in ecosystems, and disturbances in the delicate seasonal balance may have long-lasting impacts across the food chain.

Satellite technology helped researchers understand frontal dynamics, and map the inherent variability of these features with unprecedented detail, opening new perspectives on the subject. The detection of fronts is detailed in a latter section, but the importance of accurate spatial and temporal data on the mesoscale and frontal systems is acute. Considering their relevance, and the resources now available to detect the features, it is then logical to include ocean fronts in ecological and habitat maps, which are important not only from a scientific perspective but from an operational one as well (Breaker et al., 2005).

An increasing body of information supported by such cartographic advances, suggests how marine organisms at different trophic levels and from very diverse regions, rely on fronts for a number of activities throughout their life-cycle. The distribution of bird breeding grounds, for instance, appears to be correlated with fronts, as these animals forage preferentially along these boundaries (Acha et al., 2004). This is true for a number of coastal birds, for which such relations were already determined.

Different species may rely on either visual or olfactory cues – or both – and the concentration of predators, plankton, and fish is unlikely to go unnoticed when organisms scour the ocean in search for food (e.g. Nevitt, 1999). The semi or fully persistent nature of some frontal systems favors a continued exploitation of the resources found therein.

It is noteworthy that a wide range of complex marine organisms, including fishes and whales can and do actively pursue the detection of fronts (Olson, 2002). Different mechanisms can be employed depending on the nature of the organism, but all methods leverage on the set of unique characteristics displayed by fronts.

The identification of fronts by some organisms may result of the direct detection of preys or predators alike (Nevitt, 1999), producing a 'snowball' effect along the favorable grounds. On the other hand, Bakun et al. (2006, 1994) suggested that the concentration of floating materials along frontal axis may provide important cues to mobile organisms. As aforementioned, fronts are convergent areas, where materials of different origins tend to accumulate, creating 'lines' of foam, debris, and biogenic products. These can be detected by humans (the Japanese 'Siome') and marine life alike, with birds for instance, being in a privileged position to do so while in flight. Bird communities are actually profoundly influenced by the distribution of marine fronts at different levels and throughout their life-cycle (Veit, 1995, Acha et al., 2004).

Other organisms, such as turtles and whales, are known to travel and feed along major thermal fronts (Polovina et al., 2001). In the case of whales, the effect of fronts over underwater propagation of sound waves may provide a useful mechanism for detection of the features and subsequent navigation in an environment that often lacks significant positional aids (Belkin et al., 2002).

Frontal regions are also known to influence the concentration of sharks over seasonal and near real-time scales (Miller et al., 2015). In fact, not only do sharks tend to follow these boundaries and rely on them as important feeding grounds, but longline fishing vessels tend to follow similar spatial patterns with significant overlap (Queiroz et al., 2016).

The aforementioned examples offer but a small sample of a growing body of information that describes fronts as preferred habitats and ecotones for a multitude of marine organisms. In Portugal, we are still lagging behind other analog regions in our knowledge of these correlations and on the development of systems capable of exploiting these findings. The development of such systems is impossible until a comprehensive survey of frontal systems is performed, as suggested in this study, and near real-time methodologies implemented to guarantee a continuous, cost-effective monitoring of the ocean.

Frontal regions (thermohaline and Chlorophyll a) are the natural boundaries of an ever changing fluid medium. These boundaries, some persistent and others ephemeral, segment shelf seas and the open ocean in ways that can be traced using reproducible methodologies based on the interpretation of EO data. In the following section we address the dynamic field of Biogeography and how mesoscale features, fronts in particular, can play a pivotal role in the definition of representative partitions of the sea and in their subsequent interpretation.

3.3. Biogeographic Classification of the Ocean

Unlike terrestrial biomes, easily discernable even to the layman, the ocean is often regarded as a mostly continuous and uniform water mass. This is a common, albeit erroneous, perception shared by a significant part of the population. Still, sailors, fishermen, or scholars have long been aware of the inherent complexity and dynamic nature of the ocean. Spatial and temporal variability, at different scales, is an observable reality, which although more subtle than the stark contrasts offered by deserts and alpine ecosystems is nonetheless responsible for the extreme diversity of oceanic life found globally.

Ancient sources give testimony to early observations of coastal and ocean dynamics and of the interplay of different processes. These sources often recount the observations made by seafarers, who employed their somewhat rudimentary knowledge of the ocean and coastal systems in support of their trade. Herodotus for instance explains how “(...) if from a vessel bound to Egypt, the lead be thrown, at a distance of a day's sailing from the shore, it will come up at the depth of eleven fathoms covered with mud, plainly indicating that it was brought by water." (Book 2 Chapter 5 Herodotus). This seemingly simple sentence entails a surprising understanding of the basic mechanisms driving the transport of sediments towards the ocean and their maximum extent in the seafloor from the river Nile delta. Such empirical notions highlight the relevance of understanding the heterogeneity of the ocean, at all depths, to address important practical questions that, in so many instances, are still current (so much of this classical knowledge being forgotten towards the dramatic civilization collapse experienced in late Antiquity).

However it was only at the dawn of the Age of Discoveries and at the pinnacle of the so called Age of Sail that a comprehensive knowledge of the currents, winds, and biology of the world’s oceans emerged. The pioneering works of early researchers such as Mary Sommerville (1780-1872), recognized the discontinuous distribution of living organisms and the connection of this heterogeneity to environmental variables, such as water temperature.

Recently, the deployment of autonomous sensors, the multiplication of scientific cruises and the advent of Earth Observation satellites, opened new windows into the formidable dynamics of the ocean – coastal and deep (IOCCG, 2009) (Figure 3.3.1).

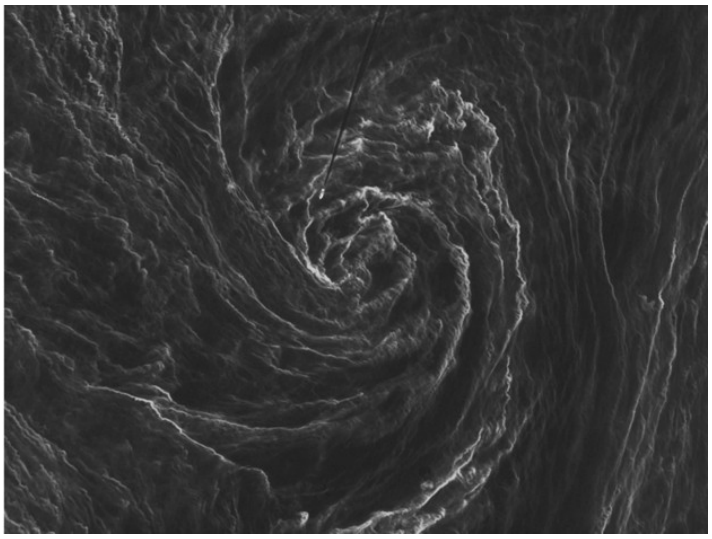


Figure 3.3.1. Sentinel-2A image of an oceanic eddy highlighted by an algal bloom (light pixels) in the Baltic Sea. The eddy is also traversed by a ship, which leaves an imprint in the ocean color along its wake. Image credit: ESA.

Considering the variability of the marine system, as already alluded to in the previous chapter, partitioning the ocean into physical and ecologically coherent units seems not only desirable but feasible, notwithstanding the herculean dimension of the task. This effort of regionalization aims at segmenting a broader area into several homogeneous zones, using a methodology capable of dealing with the dynamic nature of the ocean. This effort must equally be applied at a pre-defined resolution, in order to deliver results that are compatible with the requirements for which it is developed (Callejas-Jimenez et al., 2012).

Longhurst (2007) in his seminal work defines three requirements that should be met in a biogeography of the oceans. Such a partition should describe the distribution of species, explain how these form ecosystems regionally, and should identify the regions where these conditions are met. However, the limited amount of information and the unequal wealth of such data across regions greatly limited the chances of success in addressing this topic globally, in the last 150 years.

Furthermore, a significant amount of the oceanic biota is microscopic and commonly "expatriated", raising significant difficulties to the partitioning effort. This fact, together with the vastness of the (3-dimensional) ocean and the common apparent absence of *a priori* barriers to the distribution of life hinders the deployment of comprehensive sampling strategies that can lead up to the basis of a detailed biogeography of the oceans.

Spaceborne ocean color sensors take center stage in this context by providing synoptic views of the world's oceans at adequate temporal and spatial resolutions (IOCCG, 2009, Longhurst, 2007).

If it is clear that Biogeographic Classifications (BC) partition the ocean into ecological units from a set of biological and physical data, the different approaches adopted by researchers and agencies worldwide contributed to the current lack of consensus on which system to adopt (UNEP-WCMC, 2006). In fact, a lack of consensus emerges in the very selection of datasets to be used in the partitioning effort, ranging from schemes leveraging a diversity of biological data to others combining it with physical variables such as Sea Surface Temperature (SST) or bathymetry (Rice et al., 2011).

This ongoing debate is symptomatic of a dynamic field, as it attempts to find an identity shared across different disciplines and objectives - a task not easy to achieve. So far, and in response to the aforementioned challenges, several regional systems were devised to address regional needs (Spalding et al., 2007). Such solutions are sometimes contradictory, but the experience gained from each iteration contributes to a growing body of information on strategies to efficiently tackle the challenges of BC.

However, the development of objective (and scalable) regionalization strategies is much needed. The lack of consensus and of objective, reproducible systems is particularly problematic, considering the relevance of reliable regionalization approaches in support of the management of the ocean ecosystems, not only from a global perspective but also locally (Nieblas et al., 2014).

Biogeographic Units (BU) provide the necessary geographical framework from which different applications targeting a sustainable exploitation of resources can be designed. Rice et al. (2011) highlighted a number of potential applications for BC. These included the development of a scheme under which the status, trends and threats can be assessed, for an ecosystem-based and proactive approach to the management of resources and activities in the marine domain.

It is doubtful, however, that local managers and stakeholders can leverage existing broad, global BC into daily operations in limited territories, perhaps several hundred kilometers wide. Such systems fail to represent spatially the intrinsic dynamics of the local ocean, including the seasonal and mesoscale variability that can have profound impacts on a wide range of variables and activities. The promise of “Big Data” mining solutions may address both the global and local needs, building on previously unimaginable volumes of data to classify the world’s oceans at very high spatial and temporal resolutions.

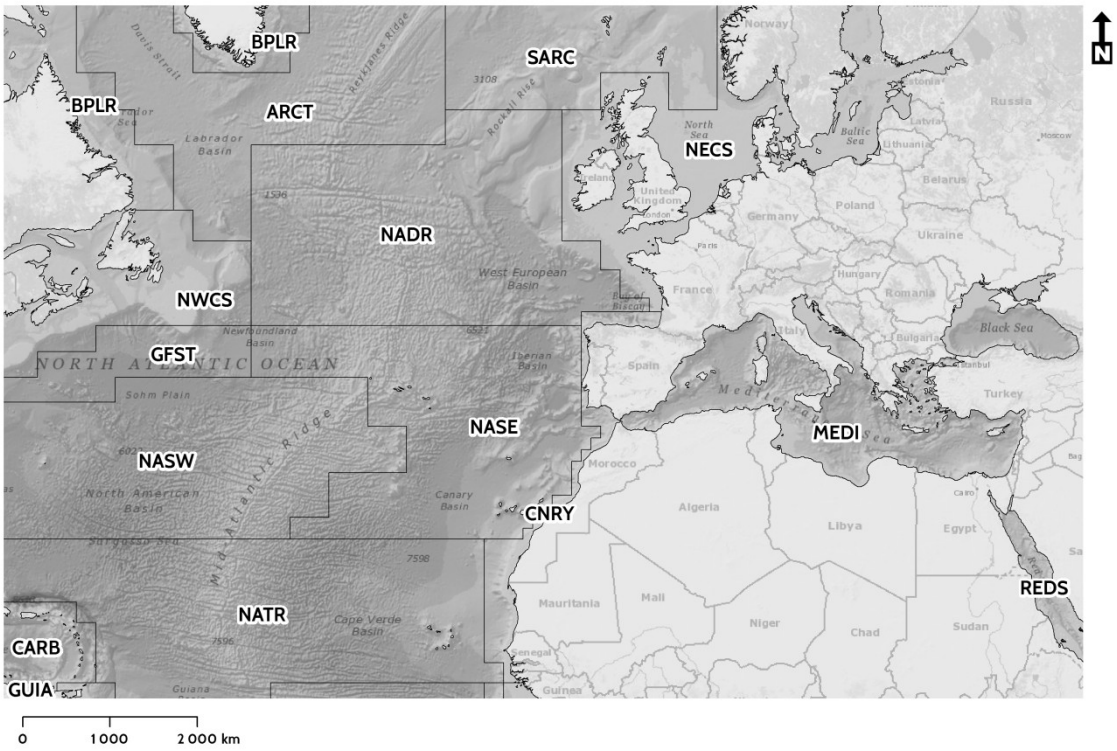


Figure 3.3.2. Ocean provinces according to the system by Longhurst (2007). Province codes included in the map for reference purposes.

On the other hand, existing partitions based on political and administrative boundaries, prove to be very inefficient tools for the management of the ocean and coastal systems, considering that only occasionally do these “virtual” man-made boundaries adhere to the geographic and ecological reality (Spalding et al., 2012). National and regional boundaries at sea are also limited to the coastal areas, leaving significant areas of the open ocean unclaimed, and therefore unclassified under the political order.

Political instruments can nonetheless incorporate the concepts driving Biogeographic Classifications and even spur a renewed interest in the field. This is the case of the EU Marine Strategy Framework Directive, which calls for the development of regions, through close partnerships between member-states, based on "hydrological, oceanographic and biogeographic features" (EC, 2007). European seas are indeed a good example of the need to implement trans-national management policies based on non-political biogeography, considering the patchwork of national jurisdictions in the continent, oblivious to the ecological units of the territory.

But regionalization efforts go beyond the (important) identification of ecosystems and provide more than blueprints for the implementation of management efforts. In particular, calculating certain variables is pivotal to improve our understanding of fundamental processes occurring

in the ocean. The carbon cycle, for instance, can only be properly understood if accurate estimates of primary productivity are available.

The parameterization of primary productivity algorithms requires an understanding of several variables, including the assimilation number (P^B_M). This variable describes the maximum rate of phytoplanktonic carbon change as a result of primary production, normalized to chlorophyll concentration. It is dependent on light, temperature and nutrients and may, through proxies, be estimated from satellite data (Sathyendranath et al., 2009). However, work leading to such estimates often stressed the specificity of this variable to regional autotrophic assemblages (Platt et al., 1988) and the domain-dependence of some of the primary productivity parameters (Devred et al., 2007, IOCCG, 2009).

These facts led to several efforts to identify optimum biogeographic classifications (often based on sea surface temperature) and characteristic P^B_M values found within them (Picart et al., 2014). Key ocean parameters might thus be organized in a form of "piecewise continuity", or at least simplified in such a manner compatible with practical modeling purposes (IOCCG, 2009, Longhurst et al., 1995).

The "aqueous media" leads to an unavoidable dynamic nature and heterogeneity of the units or regions of the ocean (Hooker et al., 2000). Therefore, any biogeographic classification system that aims at maintaining its applicability over time, must address explicitly - or implicitly – the ocean's dynamic nature.

Longhurst (2007) recognized that one of the main limitations of his approach to BC lies in the static nature of provinces and of their boundaries. Devred et al. (2007) recognized this problem and adopted an alternate methodology capable of incorporating changes in the location of water masses. But before addressing the nature of boundaries, it is paramount to determine the coherence of the partitions.

In order to do so, it is necessary to assess whether the natural distribution of species (and key variables) is discontinuous and can be predicted from a spatial and temporal perspective (Longhurst, 2007). Ecoregions must thus constitute areas that can be clearly differentiated from neighboring ones, driven by disparate forces, which may include upwelling, geographical barriers or dominant currents, amongst many other variables (Spalding et al., 2007). To accomplish this task, a good understanding on how environmental factors influence biogeography is required, or the field will be limited to the "simple" description of species distribution without a real discussion of the subjacent causes driving the spatialization (Belanger et al., 2012).

In the context of a changing climate, with significant modifications to the normal regional values of key ocean variables (IPCC, 2013), our ability to understand and forecast future biogeographic patterns is of utmost importance.

It is also relevant to understand the intrinsic variability of key variables (e.g. SST) at sampling sites, considering that these may be located at the boundaries of different biogeographic units (Devred et al., 2007). This raises, once again, the question of adopting static or dynamic boundaries in BC systems. If, on the one hand, static boundaries simplify the classification process, on the other, they can introduce ambiguities and result of sporadic conditions at the time of the study, irrelevant to the long-term distribution and recruitment of marine organisms. Static boundaries would also fail to accommodate climate change (or long term cycles), as BUs defined under a set of conditions would become obsolete over time.

Boundaries in the static versions of BC must be understood as "zones of change", which are not truly static but "blurred and variable over time" (Spalding et al., 2012).

However, both visions (static and dynamic) may not be entirely opposite and attempts to reconcile them are perhaps desirable, if the best of both systems can be preserved in a unique scheme. The balance between both approaches is sometimes achieved through the adoption of established partitioning systems and the integration of new statistical methodologies capable of defining – dynamically - the boundaries of the regions. An example of such an attempt is offered by Devred et al. (2007) as aforementioned. This study offers an interesting contribution to this purpose. In it, researchers use SST and Chlorophyll a data to identify the position of dynamic boundaries of the provinces previously defined by Longhurst (1995). The method resorted to clustering the datasets using *a priori* knowledge on the values to be set as reference for each region, an approach that proved to be useful at the province-level.

This is actually one of several studies demonstrating that static boundaries fail to represent the dynamic variability of the ocean. This variability is routinely identified in *in situ* datasets (Nieblas et al., 2014) and should thus be incorporated into biogeographical classification systems.

In fact variability is not just a nuisance in the cartographic process but a real concept of biological significance. The Transition Zone Chlorophyll Front (TZCF), for instance, is a basin-scale variable feature found in the North Pacific Ocean that is found between the subtropical gyres and the subartic gyres, each with different average Chlorophyll a values. In satellite imagery, the front is depicted as a strong gradient in the Chlorophyll a concentration.

This broad region, found between the Subarctic and Subtropical Frontal Zones, or roughly between 32° N and 42° N (Roden, 1991) was subject to the "only major change from the

arrangement in the first edition" of Longhurst's book (2007), and classified as the North Pacific Subtropical and Polar Front Provinces (NPST and NPPF).

The relevance and consequences of the transition (frontal) zone and its intrinsic variability are wide-reaching. Not only has it posed significant challenges to the biogeography of the Pacific Ocean but it is known to influence marine life extensively. Polovina et al., (2001) described how loggerhead turtles (*Caretta caretta*) followed changes to the 1997-1998 location of the transition zone, suggesting the relevance of this feature to the species. Furthermore, catches of albacore tuna seem to be correlated as well with the location of the frontal zone, which migrates seasonally over 1000 km. However, interdecadal variability of the climate over the regions affecting the transition zone (Chai et al., 2003) adds further challenges to a system that must be designed to accommodate both short and long-term change.

Considering the importance of the location of the transition zone to marine life, and its variability, it is legitimate to set in the requirements of any new classification system (especially pelagic schemes), the ability to accommodate the variability, which must be traced in a simple and cost-effective manner.

The aforesaid example pertaining the TZCF is but an extreme example of the widespread variability found along the contact lines – or regional boundaries - of different water masses, at different scales. Even from an operational perspective of exploitation of marine resources, the static definition of boundaries, albeit easier does not respond to the true needs of users (Gregr et al., 2012). Static boundaries are certainly useful to meet the requirements of long-term management policies but perhaps less effective for everyday management purposes (Spalding et al., 2012).

However, is this static nature of management units still applicable in today's increasingly connected and sampled world? Can technological advances contribute to remove the essentially operational constraints that impose the adoption of static boundaries?

Marine Protected Areas (MPA) are generally set in legal documents as static units subject to special management policies aiming at the conservation and sustainable use of natural resources worldwide (Wood et al., 2007). The static nature of reserves obeys to administrative and procedural practices, which often limit the protected area to regions of undeniable ecological and cultural significance or defined by simple practicality.

However, a growing body of information tends to support the establishment of dynamical MPAs in response to real ecological and cultural needs. Game et al. (2009) argue that communities may accept dynamic MPAs more easily, if they understand the rationale motivating temporary restrictions instead of fixed static restricted areas within a mutating environment.

This is a particularly appealing concept when the objective of the MPA is to promote a sustainable fisheries sector. MPAs are known to have a very positive impact in the regulation of catch and biomass levels (e.g. Rodwell et al., 2004). Still inter-annual variability of key variables leads to different recruitment and spatial distribution of stocks, such as in the aforementioned case of albacore tuna in the TZCF, or as a consequence of the complex interplay of variables in the definition of sardine stocks off Portugal (Santos et al., 2004). As such, it does seem reasonable to adjust the extent of protected areas in accordance to the range of the variability that effectively affects targeted ecosystems and species or, alternatively, adjust the extents of the MPA dynamically in response to inter-annual changes in the bio-physical variables.

The biogeography of many marine organisms, namely of ecologically restrictive ones, offers important insights into the relevance of boundaries (as barriers between water masses) in structuring ecosystems and species assemblages (Bauman et al., 2005). These relations are relevant both for present and paleontological analysis, despite ongoing debate over ecophenotypic variability. These limitations can only be addressed by a growing body of information regarding key marine organisms, which may in turn support a better understanding of the relevance of dynamic boundaries in the biogeography of the ocean.

Furthermore, BUs may represent essentially, “sets of species”, more than mandatory associations (Belanger et al., 2012). These species are then capable of adapting to a set of variables in certain, more or less successful and shared ways.

This leads us to yet another question, in the context of the emerging patterns in key variables driven by global climate change, can static (and traditional) partitioning systems hold?

Longhurst (2007) writes about a recent “corruption” of biogeography as a consequence of the pressure to which some ecosystems are subject. Such pressures lead inexorably to changes in species distribution and stocks caused by changes in temperature, salinity, and other variables (Longhurst, 2007, Belanger et al., 2012). It is “virtually certain” that the upper ocean has warmed in the period between 1971 and 2010 and “likely” that warming also occurred from the 1870s to 1971. The warming rate in recent decades reached alarming values ((0.11 (0.09 to 0.13)° C per decade in the upper 75 m with increased thermal stratification and growing gradients of salinity found between regions (IPCC, 2013). Such trends may even be more expressive in some regions of the world, hampering the application of static classification systems in the near future, at least if they are not updated.

In our changing world, it thus seems logical to adopt dynamic systems, which can adapt to new conditions and accurately reflect the properties of the water masses. The aforementioned corruption of biogeography (and its indicators) also furthers the case for physical-based BC.

Such systems offer unbiased information on bio-physical variables, which may contribute to a deeper understanding of the impact of change over biological communities. Classifications would no longer incorporate the reduction of stock or extirpation, but rather be used as the bio-physical baseline for studies downstream. Physical variables can also be collected routinely from Earth Observation sensors and other automated systems, enabling a rate of information (and classification updates) that would be impossible using traditional techniques.

For the reasons presented, in this study, we aim at addressing the biogeography challenge through the design and implementation of a dynamic classification system based on Earth Observation data. This system shall be compared against the distribution and persistence of mesoscale features, which are natural barriers (or connectors) to the biological communities.

But, if it is widely recognized that insufficiencies in the understanding of species distribution persist and motivate the development of classification systems using alternate data sources, it is equally true that we are far from having a complete depiction of the variability of key environmental variables like Sea Surface Temperature of Chlorophyll a concentration. We also lack, for the most part, the values that were characteristic to each region before the current and generalized changes. In fact, 71% of the coastal areas are warming as mentioned before, at varying temporal rates and in complex spatial patterns (Lima et al., 2012).

Satellite data, which constitutes the most efficient method from which global data can be mined, is only available for the most recent decades. Simultaneously, typical errors associated to instantaneous SST measurements by EO sensors is now roughly equivalent to the average warming observed since 1971. Continuous series of bio-optical variables, for instance, became available only since the launch of SeaWiFS in 1997.

This creates insurmountable obstacles to the construction of historical time series depicting baseline conditions for the world's marine ecosystems. It is therefore necessary to consider the hypothesis that current time series are already incorporating the effects of global changes, namely to the climate.

Coastal ecosystems are actually the most exposed to the aforementioned changes, highlighting both the challenges and urgency of adequate and holistic monitoring strategies addressing these areas (Halpern et al., 2008). It is important to emphasize that coastal waters suffer the cumulative effect of direct anthropogenic pressure (i.e. fisheries, pollution) as well as indirect changes driven by impacts to the adjacent terrestrial ecosystems (Lima et al., 2012).

Considering the unequal distribution of data on living organisms and the poor distribution of *in situ* sources for physical data, even in developed countries, satellite data seems to be the obvious source of information to attempt a comprehensive and updatable methodology for global BC.

Satellite sensors offer an increasing amount of information, and researchers worldwide suggest new products based on numerous algorithms designed to retrieve biological and physical variables from coastal and open ocean waters.

The plethora of remote sensing products, with particular emphasis on SST and Chlorophyll a concentration, has been levied to devise new strategies to address BC, at both regional and global scales.

Oceanic thermal fronts (OTF) are located in the contact areas between different water masses (Breaker et al., 2005). Therefore, OTF are intrinsically connected to the main surface oceanic currents and provide an unambiguous representation of the boundaries found between biogeographic regions (Longhurst, 2007). These features can become an invaluable aid to the identification of homogeneous biogeochemical provinces, separated by natural barriers, unhindered by the caveats of strategies built on *in situ* data (Nieblas et al., 2014). Oceanic fronts are discontinuities, as aforementioned in the previous chapter, and thus seem logical candidates to a dynamic partition of the ocean, despite the semi-stationary nature of many of these features.

The role of fronts as partitioners of the marine space is reinforced by observations suggesting the differential distribution of biota across different sides of fronts. Breaker et al. (2005) mention an example off the coast of California, where salmon and albacore tuna are found in opposite sides of the local major upwelling fronts. This reinforces the role of mesoscale phenomena, and fronts in particular as ecotones, creating (somewhat permeable) barriers in an otherwise continuum system (Margalef, 1997).

Thermal fronts, as the name clearly suggests, represent more or less sharp gradients in the temperature fields (and other variables). This leads us to consider that the differences in the biota found across fronts are probably the result of that gradient, which separates regions with diverse horizontal and vertical temperature profiles, further aided by physical isolation mechanisms. Under this assumption, biogeographic partitioning should look into the thermal properties of coastal and ocean waters worldwide, in order to identify objective patterns. Such patterns may co-occur with specific biotic assemblages, while driving temperature-dependant physiological processes (e.g. assimilation number).

The relation between abiotic variables and biodiversity are long known (Spalding et al., 2007). Sea Surface Temperature in particular, has been reported to play a major role in the overall biogeography of the ocean (Tittensor et al., 2010). Temperature has been reported to predict between 53 and 99% of today's coastline benthic biogeographic structure (Belanger et al., 2012). In fact, these authors state that "a few readily acquired oceanographic parameters are sufficient to predict biogeographic patterns". This is of the utmost importance to guarantee

that a dynamic and objective biogeographic partition of the ocean does indeed become feasible at an adequate spatial resolution and cost, while relying on assumptions and requirements contributing to model parsimony.

Remembering the rapidly increasing ocean temperatures is never redundant. In the Eastern Atlantic, where this work's area of study is located, the number of extreme hot days is increasing at an alarming rate (8.4 +/- 6.6 additional hot days per decade in average) (Lima et al., 2012). This poses new questions and stresses the urgency of developing efficient methods not only to monitor SST values (Belanger et al., 2012), but to spatialize the seasonal and inter-annual thermal patterns into coherent units, separated by the ubiquitous fronts.

The first-order link between biotic and abiotic variables supports the assumption under which marine biological communities are structured according to the patterns of said variables, including SST (Belanger et al., 2012). This is a particularly meaningful connection considering the complexity of marine life surveys when compared to the monitoring of variables retrievable from remote sensing imagery. Sea Suface Temperature is routinely measured from space for several decades, opening the door to continuous monitoring and dynamic spatialization of the sea on time scales previously impossible. Considering the aforementioned lack of biological data, especially in coherent resolutions and homogeneous quality, it is necessary to rely on alternate data sources, used as proxies, to establish biogeographic classifications (Roff et al., 2000, 2003).

In our changing world, our ability to relate physical variables – and their discontinuities – with the distribution and characteristics of life forms and ecosystems, opens the door to unparalleled models of past and future biogeographic partitions of the world's oceans (Belanger et al., 2012).

It must be emphasized that many BC studies, including those relying on physical variables to partition the ocean, lack a validation of the assumptions considered when the experiments are designed (Gregr et al., 2012). Yet, and although supported by a growing number of studies, which support the use of such relations, as aforementioned, it is challenging to conceive a feasible methodology to validate BC systems based on physical variables, when the biological data is clearly insufficient, especially as the distance from the littoral increases.

On the other hand, and as aforementioned, the dynamic nature of biogeographic boundaries must be considered. Inter-annual variability, which may be easily traced using remote sensing data, is harder to monitor when the data sources for such classifications come from biological surveys. Yet, under the match-mismatch hypothesis (Durant et al., 2013, Cushing, 1990) even the short term (e.g. annual) climate variability may have profound effects on effective recruitment. This "climate-mediated mechanism influencing populations" (Durant et al., 2013)

influences species differently (Costello et al., 2006) depending on complex relations of trophic synchrony and top-down/bottom-up controls. Therefore, caution is needed when formulating hypothesis connecting the variability of variables including SST and Chlorophyll a concentration with annual recruitment or abundance figures.

In the Western Iberia Upwelling Ecosystem (WIUE), it is known that a decline in the productivity of small pelagic fish (SPF) can be attributed, at least partially, to changes in the upwelling regime, as well as other environmental changes (Santos et al., 2007). The increased frequency of winter upwelling events, leads to a dispersal of eggs and larvae with immediate consequences to their survival. The thermal signature of upwelling events can be captured by remote sensing sensors (SST), as aforementioned in previous sections. This makes it conceivable that such signatures would have an impact on dynamic biogeographic classification systems based on, or at least including SST data as part of the inputs. It is also likely that a continuous monitoring and classification system would be able to detect the anomalies and spatialize the variability even before the ecological consequences were recognizable to researchers and operational users.

It is thus important to reconcile our understanding of the ecological responses to physical parameters with the spatial and temporal scales of the data available to monitor those variables. This shall constitute the next step towards a comprehensive and dynamic biogeography of the oceans, with applications at the global as well as at the local scale. Sea Surface Temperature plays a pivotal role in this effort, as a firmly established Earth Observation product, retrieved at moderate and high spatial resolutions for the global ocean at high temporal frequencies. Sea Surface Temperature fronts follow this trend, representing natural boundaries between distinct water masses that act frequently as barriers to the dispersal of organisms.

A wide diversity of methods and source datasets has been employed in the process of creating biogeographic classification systems in recent years. The different approaches reflect different visions, data availability, and even the field of expertise of the authors who created them.

We shall now present a non-exhaustive list of existing classification systems, both global and regional, with an emphasis on schemes based on reproducible statistical methods relying on bio-physical data. We aim at providing an overview of different options available and the methodologies used to produce them, focusing on the advantages and disadvantages of each approach.

The increasing amount of information available, data mining software as well as the growing power of off-the-shelf commercial computers opened complex clustering and classification operations to a broader range of users. Satellite data, in particular, along with expanding *in*

situ observation networks made possible by advances in autonomous sampling and communication systems, created new opportunities for biogeographic studies.

It is important to emphasize, nonetheless, that purely taxonomic systems of reference persist and these occupy a rightful and useful place in the context of BCs.

An example of such a taxonomic-based system is offered by Spalding et al. (2007). This work of reference presents the Marine Ecoregions of the World (MEOW), a nested system comprising a set of realms, provinces, and ecoregions covering the world's coastal and shelf seas. In this system, parsimony is often mentioned as a relevant attribute of the scheme, which "sought to minimize further divergence from existing systems, yet still to obtain a truly global classification System".

Data sources include biogeographic assessments in peer-reviewed literature, ecoregional assessments conducted by NGOs, classification systems put forth by public agencies as well as inputs from several authors.

The authors of that study had the declared intention of designing a system capable of addressing real management needs, and therefore with higher spatial resolution than some of its counterparts. The provinces, in this work, are considered as the result of "evolutionary isolation", yet they recognize the limitations of a static system meant to classify an ever-changing marine environment. The boundaries of the regions are to be used as indicative guidelines and never as a rigid and constant boundary.

Most of the Iberian Peninsula and the Bay of Biscay are integrated into a single Ecoregion (27, Realm: Temperate Northern Atlantic; Province: Lusitanian; Ecoregion: South European Atlantic Shelf). This segmentation level can hardly be used within the study area or even for national management goals, considering that it does not differentiate ecosystems occurring in Portuguese (continental) shelf seas.

The system suggested in the seminal work by Longhurst (1995, 2007) divides the world's oceans into a set of biomes (4) and provinces (54). This is a static system that relies very heavily on remote sensing data and a manual interpretation of the biogeography of the ocean.

The author emphasized that the boundaries of the provinces are not fixed, but instead constitute the average position across seasons and muti-year variability. In the study area off Western Iberia, two different biomes, and provinces, are recognized. The offshore area is included in the Westerlies biome, North Atlantic Subtropical Gyral Province (East) (NAST-E). On the other hand, closer to the shore, the area of interest in included in the Coastal Biome as part of the Canary Current Coastal Province. The two regions are separated by a rectilinear line in the along-shore direction over the entire western façade of Portugal (Figure 3.3.2).

At several locations, the CNRY province does not cover the entirety of the shelf in the Portuguese (continental) sector or the estimated width of the Poleward Current (roughly 40 km). This leads to important limitations to the regional applicability of this system and even the use of Longhurt's system for the automated or semi-automated retrieval of oceanic variables for intercomparison purposes.

Spalding et al. (2012) highlights how Longhurst's approach, despite some strong similarities with the systems adopted by taxonomists, takes on a different path due to the extensive use of EO data. Such divergences, when detailed comparisons are made, lead to important differences. These preclude a universal adoption of the system, motivating ongoing efforts to create new or adapted schemes.

It must be borne in mind that Longhurst's provinces constitute a global partitioning effort, based on a method that is hardly automated and therefore offers a best-effort (and exhaustive) approach to a BC. Despite the divergences and critics, it is amongst the most widely used systems worldwide.

In fact, Devred et al. (2007), expands the work by Longhurst, by addressing the problem of variable boundaries using objective statistical methods applied to EO data.

In this study SST and phytoplankton biomass (Chlorophyll a concentration values) were acquired from MODIS AQUA data. Conductivity, Temperature, and Depth (CTD) data were also acquired at a cruise in order to support the validation of the method. K-means clustering applied to the SST and Chlorophyll a (log-transformed and scaled) data using R, was the method adopted by the authors to assign individual pixels to the different provinces, following Longhurst's scheme. 7 different clusters were created, using 6 centers, or variables, including SST, Chlorophyll a, water depth, province coefficient, latitude and longitude. Province coefficients were defined by the authors and the method is debatable, although the results seem to support the chosen method. Overall, the methodology enables the dynamic mapping of ocean provinces at a high resolution, especially when compared with traditional biogeographic maps of wide areas as is the case (see their Figure 3). The seasonal variation of the position of different water masses is thus represented seamlessly, providing EO data is available (which might be a problem at certain latitudes).

The method suggested by Devred et al. (2007) has the advantage of leveraging a previous classification system, but expanding it to move from a static to a dynamic approach. Validation of the method in terms of its ecological significance is lacking but such efforts contribute to the overall goal of reducing BC diversity while improving available cartography. Nonetheless, the use of non bio-physical variables in the clustering process, some of which only indirectly

related to relevant ecological processes may hinder the applicability of the process and its long-term reliability.

A different study, by Nieblas et al. (2014) used a combination of different datasets to create a comprehensive classification of the Mediterranean and Black Seas. Once again, as in Devred et al. (2007), these authors aimed at defining subprovinces, as subdivisions of the ones suggested by Longurst.

Data included SST from AVHRR, Chlorophyll a data from MODIS, Ssalto/Duacs sea level anomalies and geostrophic velocity anomalies (u,v) from Aviso, as well as bathymetry data from NOAA. These sets are available at very different spatial resolutions, complicating the clustering process. The data itself were used to derive even further variables, with an emphasis on mesoscale features, such as frontal intensity. In fact, the authors rely on a combination of "classical and mesoscale" features to generate the classification maps.

Instead of using the original data (except for bathymetry), synthetic variables were generated after the application of Principal Component Analysis (PCA). PCA is often used as reliable tool to reduce the redundancy of datasets and reduce the number of inputs prior to clustering or other statistical operations. 5 subprovinces were thus created, which are in line with previous studies. The results obtained in this study emphasize the importance of mesoscale features to bioregionalization efforts. In fact, borders of the subprovinces defined using mesoscale features are more stable over time than those using the full array of variables. The use of PCA raised several interesting questions, pertaining seasonal and interannual variability, but the authors failed to deliver comprehensive explanations to such changes.

However, this small subset of works raises the important question of the different terminologies used in BC studies. Different authors frequently define new segmentation levels, despite the common aspects shared by all schemes. How to differentiate provinces from subprovinces and which criteria are to be used, remain important, but open questions.

Several studies rely entirely on remote sensing data to create objective and reproducible schemes that can often be replicated worldwide and throughout long periods using multi-mission data. For this to be possible, a common set of variables must be retrieved by spaceborne sensors, which must be cross-calibrated to guarantee true data continuity.

Among the studies leveraging remote sensing data, we emphasize that of Callejas-Jimenez et al. (2012). In this work, the authors generate a dynamic classification (regionalization) of the Gulf of Mexico using Normalized Water-leaving radiances (nLw) at a spatial resolution of 1 km. nLw data were retrieved from daily MODIS-AQUA imagery and nLw 412 nm and nLw 488 nm were selected as the input for the classification after an analysis of the different variables using PCA. This method demonstrated the importance of both products in describing variability, thus

limiting the redundancy of using other, similar variables. The regionalization was then pursued through the use of Standard Empirical Orthogonal Function (SEOF) (combining the two aforementioned products). The SEOF yields unitless fields, which can be used to generate the partitions, or as in the nomenclature used by the authors, dynamic biogeographic regions.

The classification system generated two provinces (ocean and coastal), which included several sub-provinces and regions (11). Interestingly, the authors successfully accounted for seasonal differences, including hurricane and non-hurricane conditions. Intra and inter-annual variability is an important factor that is often neglected in works detailing new BC.

The provinces and regions identified in the study account not only for the overall circulation patterns and dichotomy between coastal and ocean environments, but successfully incorporate mesoscale features and seasonal events such as river runoff into the regionalization.

The aforementioned example, offers an interesting approach to address regional biogeography, using a set of replicable and well defined variables. These are variables measured from satellite sensors, and therefore, unaffected by dissimilar surveying efforts across the region, supporting the homogeneity of the data coverage and complying with the aforementioned requirement of reproducibility.

Optical properties often reflect the regional hydrodynamics and are an important aspect governing the distribution of marine life, namely during critical stages of the life cycle. Furthermore, in Callejas-Jimenez et al. (2012), PCA highlighted the redundancy of several variables and the primacy of the nLw set over Chlorophyll a or SST products. However, and not unlike systems that rely on SST data, optical products are only effective up to the penetration depth (or first optical depth) (Callejas-Jimenez et al., 2012). As it was also aforementioned, the surface signature is often an expression of the complex interplay of variables including currents and frontal surfaces, which often have a sub-surface expression. In mixed waters this assumption is even more accurate, considering the limited vertical variability.

However, and despite the relevance of the presented methodology – and of many other similar studies relying on non-SST EO products – some questions remain.

In particular, it remains to be demonstrated in several cases if the statistical variability, as described by PCA for instance, matches the biological variability. In other words, are the selected products capable of separating regions with indeed different conditions to support life? The use of physical data has been proven to explain the distribution of taxa (Belanger et al., 2012, Roff et al., 2000), especially when temperature is employed (Tittensor et al., 2010). However, optical variables such as nLw have been added to the arsenal of biogeographers

much more recently and therefore lack the extensive validation already conducted on other products.

Data availability is often debated in biogeography works, but no less attention should be paid to the definition of suitable criteria for the establishment of reliable ecological regions. The work by Gregr et al. (2012) follows a different approach to those describe thus far. The author identified 16 different classification systems that were considered as potentially suitable for delineation of Canadian Ecologically and Biologically Significant Areas (EBSA). The criteria for the establishment of the areas include biological and physical aspects, most of which are mentioned in previous works.

The classification systems used include quantitative (multiple methodologies) and expert approaches, with data sources including physical, biological, ecological, and even political sets.

The authors conclude that the evaluated systems, despite individual strengths and weaknesses, are still "disconnected from all but the lowest trophic levels and the most sedentary creatures".

The focus on dynamic boundaries is highlighted by the authors (as in several other studies), given the need for efficient management of the EBSAs. The relevance of clustering methods is also supported, although the authors mention that these are only partially ecological given their reliance on physical data. Still, the feasibility and reproducibility criteria are met and thus partially fulfill the requirements for the definition of ESBAs. Nonetheless, the presented ESBA approach, merging physiographic (physical data) and zoogeographic (e.g. species distribution) offers an interesting approach. Despite the clues offered to future integrative works that bring together the different sets of data and approaches into novel, and integrated systems, it is difficult to ascertain the feasibility of such an effort, for large areas or when high frequency updates are deemed necessary. In fact the dynamic boundaries, the authors explicitly support, seem difficult to be supported in a system that requires extensive expert judgment and biological data.

The paper demonstrates how, despite the numerous systems proposed, reconciling them with specific conservation or management criteria is difficult, at best. It seems more likely that in the near future, managers will continue to rely in systems put forth by the scientific community, despite a persistence lack of consensus. It would be desirable, nonetheless, the establishment of joint, comprehensive programs aiming at the development of truly integrated BC designed from scratch to meet the requirements of the different end-user communities.

Another relevant classification system is set by the Large Marine Ecosystems (LME), which partitions the coastal ocean into 64 regions for the development of ecosystem-based assessment and management activities. The Global Environmental Facility (GEF) addressed the

increasing pressure to coastal areas through the development of the concept of an LME-based management approach (Sherman et al., 2009).

Interestingly these are focused not only on coastal areas, which the system partitions, but it is also focused on watersheds adjacent to these regions. The system, in itself, is based in a combination of four key ecological criteria including bathymetry, hydrography, productivity, and trophic relationships. As such, the transboundary approach to regionalization offers an effective tool for resource management and interjurisdictional cooperation. The LME approach, championed by several government agencies worldwide (e.g. NOAA) is also based on a set of five "modules", which contribute to evaluate the state of the LMEs as objectively as possible, further supporting inter-comparisons and long-term analyses and forecasts. The five modules include a number of indicators arranged within the key topics, which include productivity and oceanography, fish and fisheries, pollution and ecosystem health, socioeconomics, and governance.

LMEs and their boundaries were established using a set of criteria that included Earth Observation data, in this case Primary Productivity retrieved from SeaWiFS. The complexity of creating a set of global biogeographic regions is well exemplified by the LME. Despite the use of EO data, as aforementioned, other data sets were included in the analysis. Some of these include extensive works analyzing the physiographic traits of each region that called for extensive review from experts worldwide (NOAA, 2004). As such, this is a system that incorporates both objective (and quantitative) elements but also expert judgment and validation.

The Western Iberian Peninsula (WIP) is included within LME 25 (Iberian Coastal). This large area also covers the Algarve coast and, to the North, the Spanish waters of the Bay of Biscay. Clearly, the level of detail is also reduced seaward, with a simple static boundary established beyond the continental shelf.

This is clearly a system designed to address management concerns at a regional level, within a common global framework. The progressive delineation of the LMEs, led by k. Sherman ever since 1990 (Sherman et al., 1990) is representative of the massive effort behind this system. In line with such an effort is the subsequent effort by the Global Environmental Facility, to sponsor numerous projects addressing pressing research and governance topics. These projects are developed within the spirit of the LME approach, addressing both coastal and watershed issues worldwide, effectively "bridging the gap between land-based activities and LME waters" (Sherman et al., 2009). The LMEs are also on the basis of extensive research on a wide range of ecological and governance studies, including fisheries (e.g. Frank et al., 2011). It

is also indicative of the commitment towards this system by several organizations and researchers globally.

The perspective adopted in the delineation of LMEs is different of the approach followed in the current study, which is in part due to the global coverage of the system. Nonetheless, the underlying concepts and perspectives driving the choice of criteria are perhaps amongst the most sensible and adequate for a truly integrated watershed-coastal management of the natural resources and ecosystems. As such, we will strive to integrate the watershed/coastal perspective, considering the true continuum that exists in an otherwise apparent dichotomous reality.

The summary presented thus far is not exhaustive, nor does it aim to be, since so many systems have been put forth by authors and agencies worldwide. Often, the systems seek to address specific concerns or meet the objectives of various stakeholders, or simply incorporate new data sources previously unavailable to the scientific community.

The different systems rely on different metrics and criteria, rarely agreeing in the boundaries or segmentation levels, despite an overall agreement in broader lines. A conclusion that can be drawn from the analysis of the listed systems is that hierarchical systems (e.g. Longhurst, 2007), based on statistical criteria, provide an objective way to adjust the segmentation to different scales, as needed.

Nested systems provide the much-needed flexibility required to apply the BC to different settings and communities of end-users seamlessly. The level of detail and the sheer complexity brought by the number of classes can be tailored to cater for the objectives of stakeholders of the classified region. If, for non-specialist management purposes, a complex system with numerous regions may be unwieldy, it may be adequate for MPA zoning or other scientific applications requiring detailed data sources.

It must also be concluded that the development of new BC must indeed address specific needs or attempt to introduce novel approaches that can truly improve the partition of the ocean. Yet, reconciling new systems with existing ones is also relevant and a stepping stone towards an integrated approach to BC. It is beyond doubt that a unanimous system is still elusive, and this shall continue to be the case for as long as data insufficiencies or evolving data mining technologies continue to offer new insights on the structure and dynamics of coastal and ocean ecosystems. As such, it is unlikely that the convergence into a single system will ever be possible, and efforts should instead be directed at improving available regionalization systems and simultaneously attempt an integration of the new partitions into established umbrella global schemes.

3.4. Earth Observation technologies for the study of fronts

The ubiquity of thermal (and ocean color) fronts in the coastal environment coupled with the spatial and temporal transiency that sometimes characterizes them, introduces significant challenges to their monitoring and mapping. The reduced cross-front length further difficult the detection of fronts, which require active surveys (e.g. cruises) to efficiently measure the gradients in the optimum bearing. The cost and infra-structure required to accomplish such task is overwhelming and precludes both long term research and the development of applications resorting to regular observations.

Earth Observation technologies take center stage as a viable alternative data source, capable of support monitoring efforts of wide swaths of the ocean with different missions in a wide range of spectral and spatial resolution. Satellites and other space- or airborne platforms are thus at the heart of this analysis, which, nonetheless, considers the supplementary use of *in situ* assets.

Satellite data provide the synoptic view of the ocean that supports not only the identification of the fronts but also of the phenomena that generate them in an “unprecedented” way (Longhurst, 2007, Benazzouz et al., 2014, Otero et al., 2009). Data from spaceborne sensors is also particularly suitable to image the mesoscale variability of the ocean and track the rapidly evolving SST fields at such scales (Relvas et al., 2009).

The use of EO data (SST) for front detection is actually – and perhaps surprisingly, considering the persisting knowledge gaps – employed for several decades (Szekielda, 1976, Bernstein et al., 1977, Legeckis, 1978). These original works relied on the primitive sensors that equipped early scientific EO satellites. A summary of the history of ocean color sensors was provided earlier and as such we shall not dwell for long in these considerations.

Nonetheless, the work of Legeckis (1978) is particularly meaningful, considering the breakthrough use of a combination of geostationary and polar-orbiting satellites. The author combined data from the Very High Resolution Radiometer (VHRR) from NOAA polar-orbiting satellites and the Visible and Infrared Spin-scan Radiometer (VISSR) sensor onboard the Geostationary Operational Environmental Satellite (GOES). The large number of daily acquisitions by GOES’s sensors provides users with the chance of imaging the ocean’s surface more frequently and fill the data gaps caused by cloud cover and other obstacles (Breaker et al., 2005). This happens at a cost though, with the acquired data being available at a lower spatial resolution, leading researchers to supplement it with data from polar orbiting satellites.

In fact, studies as soon as the '70s and '80's of the 20th century leveraged the use of higher resolution data from sensors such as Landsat's Multispectral Scanner System (MSS) (e.g. Fèvre et al., 1983). Févre's study (1983) not only identifies a thermal front but characterizes the enhanced biological activity found therein, in one of the earliest examples of such correlations using EO data at this resolution.

But, if in the early days of satellite-based frontal studies, much of the analysis was conducted manually, the advent of widespread robust computing technologies introduced automated or semi-automated front detection strategies. These strategies are often divided into two categories including histogram-based or gradient-based tools, depending on the method employed to identify the gradients in the imagery.

Some of these tools were developed as part of efforts towards the implementation of computer vision systems or at least benefited of advances in that field (e.g. Canny, 1986). Still, some of the most important works in the field of front detection date several decades and include the papers by Canny (1986) or the seminal works by Cayula and Cornillon (1992, 1995). The later is indeed such a reference that many research studies and applications still rely on the original methodology, adapted to modern sensors and software. We too, in this work, adopted Cayula and Cornillon's (1992, 1995) approach and for this reason, we shall describe it in greater detail shortly.

Considering all the challenges and data gaps often mentioned in the literature, an appropriate methodology to depict fronts in a cartographic projection was deemed necessary. The unequal sampling of oceanic areas by optical sensors called for a correction of front detection fields thus leading to the development of frontal probability maps.

These maps are mentioned by different names but are based on a common methodology, which calls for the calculation of the ratio between the number of detections and the number of observations (e.g. Breaker et al., 2005, Ullman et al., 1999). These frontal probability maps can be composited for different time scales, in order to highlight the regions where activity is recurrent (Breaker et al., 2005). Relations between the measured activity and other variables, including currents, bathymetry, coastline geometry, can then be determined with greater security.

The recurrent lack of data, which may obscure some patterns of mesoscale activity (Relvas et al., 2009), sometimes is addressed by compositing several images over time spans as long as one week. However, this approach is not advisable, as the mutating SST fields will lead to a blurring of the features and overall misrepresentation of the frontal fields (Miller, 2004, Otero et al., 2009).

In this study we address the detection of fronts by applying the algorithm to single images without the application of SST compositing strategies. Frontal probability is then calculated over variable periods including monthly, seasonal, and yearly intervals.

Front detection follows Cayula and Cornillon's (1992, 1995) approach, as aforementioned. This approach was first thoroughly presented in a paper (Cayula et al., 1992), which we shall now summarize to provide the essentials on the front detection methodology employed in this study. A latter section provides additional details, namely of the threshold values adopted and overall processing workflow.

The first steps in front detection include pre-processing (into atmospherically corrected SST) and cloud screening. This is an integral part of the methodology because clouds over the ocean will create a sharp temperature gradient, not unlike that of thermohaline fronts we aim at detecting. In fact, prior experience in the analysis of Level-2 MODIS imagery suggests that SST imagery is often erroneously cloud-masked over high magnitude oceanic thermal fronts.

The algorithm of Cayula and Cornillon operate at the window and local levels. This means that a broad range of fronts can be detected, by varying the scale of the analysis. This allows for a more efficient search of both sharp and smooth fronts, either well represented in the histogram or not.

Firstly, the image is segmented into multiple, overlapping windows, which span a pre-defined size (16x16, 32x32, 64x64,…). The data retrieved from the pixels within the windows is then subject to statistical analysis to search for bimodal distributions, which would indicate two populations are present, each with different SST ranges. The two populations must be separated by a front, which is characterized by a given temperature threshold. Several artifacts may be generated and therefore a cohesion assessment of the output is required and applied afterwards. The last step at the window level is to determine the location of the edge pixels, but without creating the final front segment.

Once the window-level processing is complete, the algorithm moves to the local-level processing. In it, a contour-following algorithm attempts to identify continuous edge pixels, which may form a front. It evaluates not only the proximity of the edge pixels but also the direction, introducing several conditions to connect (or discard) identified pixels into seamless segments. The local-level processing is also effective in determining and correcting errors or misidentifications occurring in the earlier steps because it integrates the information from different, redundant windows.

Several thresholds are used throughout the algorithm, along with complex mathematical operations included in each task, which albeit seemingly simple, requires a complex and interconnected set of operations over high quality input data.

3.5. The Study area

The study area addressed by this work includes the Atlantic coastal and open ocean area off central Portugal located between land and meridian 12.0° W, bounded to the south by parallel 39.2° N and extending to the north up to 41.2° N (Figure 3.5.1). The study area also includes the adjacent Mondego river watershed. The Mondego river, the largest under exclusive Portuguese control meets the Atlantic Ocean at the central latitude of the study area. The ocean segment of the study area covers a total area of 58414 km^2, after exclusion of the intersected land areas. To that figure, the 6670 km^2 of the Mondego river watershed must be added. This is an area of significant size, which could only be studied with adequate sampling frequencies resorting to the already described remote sensing technologies.

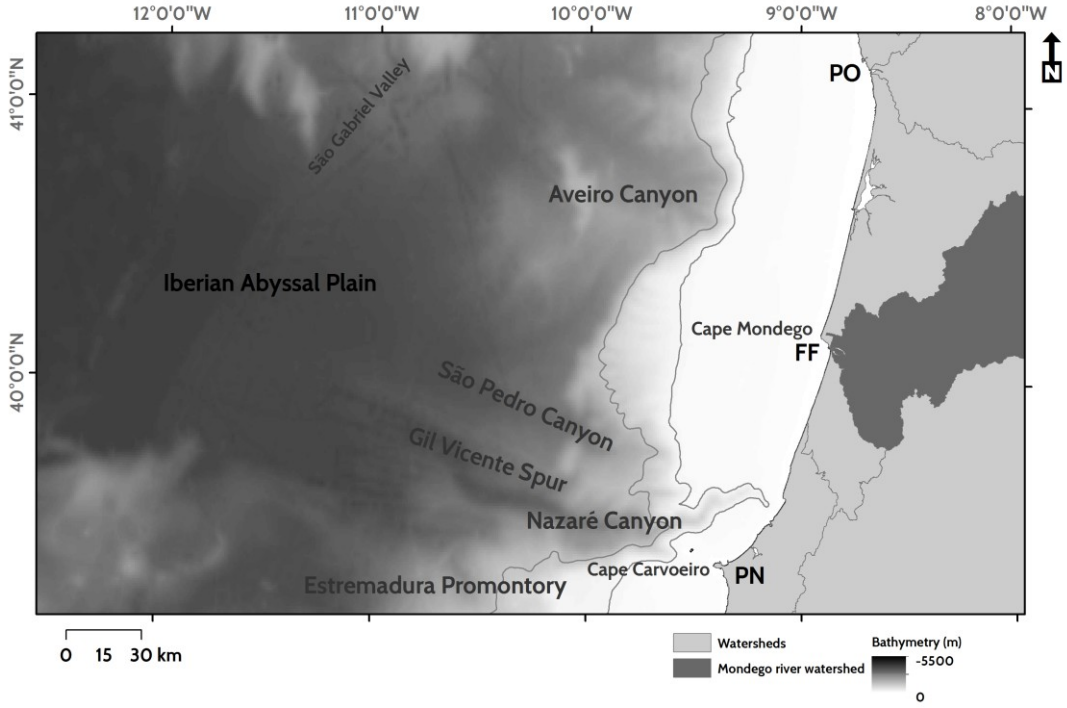


Figure 3.5.1. The Study Area including the name of major bathymetric features. The Mondego river watershed is also highlighted. PN: Peniche; FF: Figueira da Foz; PO: Oporto.

In this section we shall describe fundamentally the oceanic section of the study area, with the Mondego river watershed being described in a latter chapter.

The waters off the Western Iberian Peninsula (WIP) are located within the Eastern North Atlantic Boundary, sharing several traits with other similar areas, such as California.

The western Iberian coast is the northern limit of a well developed and studied upwelling system, sometimes called the Eastern North Atlantic Upwelling System. This system stretches from the northern end of the WIP to Africa (Nykjaer et al., 1994).

The coastline of the WIP is generally aligned meridionally, including in the area of study. Both to the North (Bay of Biscay) and to the South (Algarve and Gulf of Cadiz) the coastline is oriented in a zonal manner but these areas are beyond the scope of our analysis.

Locally, the coastline is cut by estuaries, capes, and "rías" (in Galicia) creating some variability in the local circulation patterns and in key physical and biological variables. Some of these coastal features are connected to several rivers that flow towards the Atlantic Ocean. Within the boundaries of the study area, the most relevant rivers are the Mondego, Vouga, Lis, and Douro. The Tagus and Sado rivers (to the South) are also relevant to the study area, namely during winter conditions, when a northward flow is present that transports lenses of freshwater into the area of interest.

The continental shelf is only moderately developed or even narrow in certain areas, extending to an average of 50-60 km from the coast. The shelf edge is characterized by steep slopes and cut by two major canyons (Nazaré and Aveiro) in the southern and northern extremes of the area of study. The cross-shelf axes of both canyons are roughly aligned along latitude lines and thus orthogonally to the coast.

The open ocean area immediately adjacent to these features is occupied by the Tejo Abyssal Plain and the Iberian Abyssal Plain, which together form the Iberian Basin (Peliz et al., 2005). Important seamounts to the North (Galicia Bank) and South (Gorringe) generate obstacles that are relevant to the regional circulation patterns and communication with the aforementioned zonally aligned areas.

The study areas was defined upon the inspection of the coast and bottom topography, in an attempt to restrict the analysis to a coherent, yet diverse area where multiple features could contribute to the observed pattern of mesoscale activity. As such, the study area is bounded to the South by Cape Carvoeiro and to the North by the Douro river mouth. The first creates a significant topographic shift in coastal alignment, in close proximity to the axis of the Nazaré Canyon. The Douro river, on the other hand, is the source of a large volume of freshwater during the rainy season, masking the signature of other water masses and currents, while contributing to the development of the Western Iberia Buoyant Plume (WIBP), which is described latter. To fully include both features entirely, the area of study was slightly expanded beyond both, as depicted in figure 3.5.1.

As mentioned earlier, the diversity of topographic features is a significant driver of variability, leading to both convergent and divergent phenomena, which include the development or

intensification of upwelling or shadow zones, filaments, and eddies (Peliz et al., 2002, Mason et al., 2006).

The so-called Joint Effect of Baroclinicity and Relief (JEBAR) (Huthnance, 1995), which is thought to play a role in the development of currents over the slope, is profoundly influenced by topography. The geography of the coastal areas, together with density gradients, generates poleward slope currents that cannot be solely explained by wind stress. These poleward currents, which are also recurrent off WIP, will also be described latter, but nonetheless offer an additional example of the relevance of the local geography in the context of the broader regional or even global circulation pattern.

The regional climate, as described by the Köppen-Geiger climate classification (Kottek et al., 2006, Rubel et al., 2010) is of the type Csa and Csb. This means the climate is essentially warm temperate with a dry summer (C). Temperatures in the summer may be hot (Csa) or warm (Csb), the latter found in the northern sector of the study area.

Annual precipitation varies across the study area, and most information is available from land-based weather stations. In the Mondego watershed, the highest precipitation levels are found in the headwater regions, to the East where they exceed 1500 m per year. In contrast, in the Mondego river estuary, annual precipitation is less than 900 mm. In terms of temperature, the mean annual values are, as expected, lower in the highlands (< 12° C) and higher in the vicinity of Coimbra (>20°C)(APA, 2012).

The Atlantic Multidecadal Oscillation (AMO) is responsible for cyclic variations in the SST fields (Enfield et al., 2011). These cycles must be accounted for when long time series are analyzed, considering the sensible differences in the measured temperature values. The variability is recurrent at periods of 35 to 80 years (Santos et al., 2011) and may thus be beyond the archive of satellite measurements, or even of *in situ* records for many regions.

The variability is likely to induce changes in the distribution or recruitment of several species, namely at the interfaces between regions, where it is reasonable to expect higher levels of change.

The area of study, as previously mentioned, lies in the northern end of an important upwelling system located at a region under the influence of a broad Eastern Boundary Current. This upwelling system is actually known by several designations, each failing to attain unanimity even in recent works. The choice of a name may seem trivial or secondary, but the lack of agreement is symptomatic of the challenges faced by researchers in describing the complex system. The Canary Upwelling Ecosystem (CUE) – one of the possible names and the one adopted in this work - indeed stretches from the Bay of Biscay (to the North) to Senegal (to the South). Other possible names of this system include the Western Iberia Upwelling Ecosystem

(for a subset of CUE) (Santos et al., 2007) or North Atlantic Upwelling Region (Relvas et al., 2007). The continuity of the system is broken by the Gulf of Cadiz, from where Mediterranean waters flow with a non-negligible impact on the characteristics of the waters off WIP.

The CUE is an important ecosystem, one of four major upwelling regions worldwide that sustain significant commercial fisheries (Benazzouz et al., 2014). The ecologic and economic relevance spurs from the natural conditions and the prevalence of upwelling-favorable winds along the equally suitable meridionally aligned coastlines. Similar areas can be found elsewhere including in California, Namibia, and Peru, where some of the most productive marine ecosystems can be found (Carr et al., 2003).

The study area is found at the northern boundary of the trade wind belt, with alternating equatorward (predominant in the summer) and poleward (predominant in the winter) winds (Mason et al., 2006). The location of the Intertropical Convergence Zone (ITCZ) along with the Azores High Pressure center oscillate throughout the year, which in turn creates favorable (summer) or unfavorable (winter) wind stress conditions for upwelling to occur in the CUE. The reversal of dominant winds in the region is thus an important process driving relevant changes to oceanic circulation and the occurrence of upwelling (of nutrient-rich intermediate water) and downwelling events. These events, although characteristic of specific seasons may actually occur year-round, depending on either the relaxation of upwelling-favorable winds in the summer or its opposite in the winter (e.g. Santos et al., 2004). Mesoscale activity, in the area of interest, can also be explained (over 70%) by the alongshore wind stress and the inherent variability, often in temporal scales of days (Alvarez-Salgado et al., 2003).

The upwelling events at WIP can be detected in satellite imagery, from the interpretation of SST anomalies across zonal gradients (Wooster et al., 1976). These gradients, as it will be described in a latter section, highlight the rise of cold, nutrient-rich waters along the coast, which are in stark contrast with the warmer off-shelf waters at the same latitude. The contrast is relevant not only from an ecological standpoint but also to the oceanographic interpretation of satellite imagery and the accurate detection of mesoscale features using gradient-based methodologies.

Nonetheless, and despite the importance of wind stress, the region is also influenced by the broader regional circulation patterns. Of these, the North Atlantic Current (NAC) and the Azores Current are probably the most relevant. A simplified map of the region is depicted in figure 3.5.2, which highlights several of the features described in this chapter and throughout the thesis.

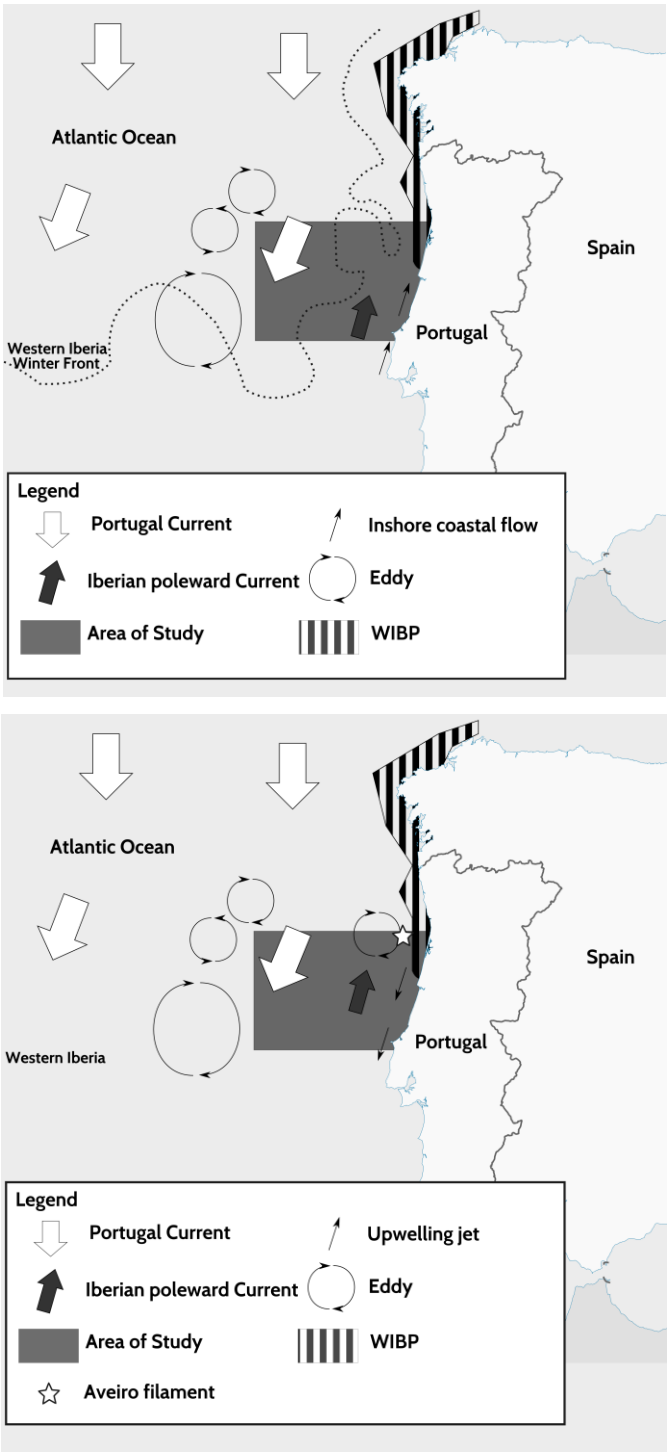


Figure 3.5.2. The main oceanic currents in the area of study. Top image: Winter, Bottom image: Summer. Adapted from Peliz et al. (2005), Mason et al. (2006), and information of the author.

The North Atlantic Current is found to the North of the Iberian Peninsula, and it is the northern limit of the Portugal Current (PoC). The PoC is a southward-flowing, slow current that can be found off WIP westward of 10° W up until 24° W (Perez et al., 2001). This system is capable of significant transport, reaching over 3 Sv in the upper levels (Stramma, 1984).

On the other hand, the Azores Current is found to the south of WIP and joins, at least partially, the Canary Current that flows southward along the African shelf. The Portugal Current is thus located in the inter-gyre region (Peliz et al., 2005), which is characterized by a complex interplay of currents and counter-currents that generate significant spatial and temporal variability. Some of the variability develops at the mesoscale as a result of the in-depth and surface interaction of different water masses, generating shear-related features, which then interact with the bottom and coastal topography.

Overall, the North Atlantic sub-tropical gyre is anticyclonic and the North Atlantic Current and Azores Current are bridged by the southward flowing Portugal Current (Perez et al., 2001). Stramma (2001) estimates that the North Atlantic Current is the most significant in terms of average transport (35 Sv). The Azores Current follows (12 Sv) while the Canary Current is limited to a transport of 5 Sv. As such, the Portugal Current is seemingly the least important in terms of transport, despite the regional relevance of the current to local ecosystems and as a connector of NAC and the Canary system.

However, the Portugal Current is, as aforementioned, one of several currents present in the study area off the WIP.

A poleward current, found usually in the subsurface, can have a surface expression during winter, as upwelling-favorable winds (northerlies) relax (da Silva, 1992). This current can be detected recurrently over the shelf and is often named the Iberian Poleward Current (IPC) (Mason et al., 2006). As in other aforementioned topics, the IPC is identified by different names such as the Portugal Coastal Counter Current (PCCC) (Ambar et al., 1994).

The IPCs – or poleward flows – are known for several decades, and have been the subject of studies, albeit more recently than other phenomena occurring off WIP such as the seasonal upwelling (Frouin et al., 1990, Haynes et al., 1990).

As in other Eastern Boundary regions, poleward flows like the IPC are located over the shelf and are generally under 100 km wide (Mason et al., 2006). Peliz et al. (2005) reported that the Iberian Poleward Current extends (at least in the subsurface) as far as 70 km westward of the shelf. Unhappily, surface current meters were not deployed as part of that study and so limited information is available on the surface expression of the current at such distances. The same study reported, nonetheless, that the current was active near the coast even in the summer, when upwelling conditions are more frequent, showing how the IPC persists year-round at least in depth. Despite the persistence, the changes to the surface expression are non-negligible, potentially influencing important ecological and physical processes in the shelf and therefore impacting economic activities alike. The IPC's seasonality is suggested by several authors, highlighting the winter-time activity from September to April, with peaks in

December-January, after which upwelling takes center stage (e.g. Cola, 2003, Relvas et al., 2007). During IPC peak months the prevailing winds off Portugal are southerlies, further supporting the formation of both the northward flow and coastward transport (Isemer et al., 1987).

But it is clear that the IPC does not cease entirely and instead becomes (possibly) restricted to deeper layers while it migrates offshore as suggested by vertical profiles describing the current's core (Peliz et al., 2005). Still according to the same author, the seasonal cycle is at least partly connected to the variability detected in the meridional gradient, which weakens in the summer. Yet, wind stress may influence the surface expression of the current, masked by the seasonal upwelling and Ekman transport offshore. The seasonality of the IPC is just another piece of the complex, highly seasonal traits of the CUE, which are also clearly measurable off central Portugal.

The velocity of the warm and salty IPC has been measured in several, non-exhaustive studies. The lack of systematic *in situ* data limits our ability to fully describe and understand the transport capabilities of this feature. It is actually noteworthy that one of the most important currents flowing off Portugal has been subject to such a limited exploration effort. Measurements made available in the literature suggest velocities ranging from 0.1 to 0.3 $m.s^{-1}$ (e.g. Haynes et al., 1990, Martins et al., 2002). Because the flow is driven by baroclinicity, a coastward flow develops (perhaps assisted by the aforementioned southerlies), that considerably weaker than the IPC proper (Mazé et al., 1997). The coastward (or eastward) flow is relevant, despite the weak transport capability, considering that it may generate a barrier to the dispersion of materials (including biogenic) oceanward during critical reproduction periods of small pelagic fish (Santos et al., 2004).

The IPC is conspicuous in satellite SST imagery because of the its thermal contrast, when compared with other currents and features in the vicinity. The thermal (1-1.5°C) and salinity (0.1-0.3 psu) gradients have been described in the literature (Relvas et al., 2007), despite the somewhat sparseness of data currently available regarding the current. The isopycnals associated to the IPC downwell near the coast by up to 200 m and as far as 40 km (Peliz et al., 2005). Again, the core of the current, even when there is a surface signature is centered at greater depths, sometimes around 100 m (Peliz et al., 2005).

The current can be seen in satellite imagery developing northwards off central Portugal, namely in winter-time imagery when the IPC is regaining strength and assuming a measurable and distinctive surface expression and signature. Peliz et al. (2003) describes the intrusion as a "narrow slope-trapped tongue-like structure", which propagates northward along the 200 m

isobath. It is clear in the satellite imagery that the IPC is very turbulent and linked to numerous mesoscale features, such as rich eddy fields (Peliz et al., 2003).

Poleward flows are known to occur at Eastern Boundary regions and are generated by meridional density gradients, as aforementioned (Peliz et al., 2003). The gradient is probably the most important variable at play and has been discussed for the area of study in several research papers (e.g. Van Aken, 2001). However, the effect of bottom topography and coastal alignment and geography should not be neglected. The Joint Effect of Baroclinicity and Relief (JEBAR) has been described as the driving force behind the poleward flows in Eastern Boundary Current areas, such as California (Huthnance, 1984, Gay et al., 2009).

This is thus a baroclinic current system, which interacts with equatorward currents either below or above (during summer). The current eventually interacts with the topography generating complex mesoscale features (such as eddy dipoles) that can persist over long periods (Peliz et al., 2003).

Different authors suggest the IPC develops following a number of recurrent stages – 3 or 4 depending on the authors – that describe a process that starts with the intensification of the meridional gradient. We will adopt the 3-stage development process in this description, while highlighting the differences found in the descriptions. We acknowledge the dynamic nature of the process, one that, as in so many cases is difficult to classify into distinct non-overlapping categories.

Furthermore, one must also consider the evolution of the system not only through time but also through space. The maturity of the IPC along the coast of Portugal may vary and display different characteristics in its spatial domain. In this work we rely on the descriptions of Peliz et al. (2003, 2005) and Relvas et al. (2007), to describe the development of the IPC off WIP.

The first stage of development is often referred to as the adjustment stage. In this stage, which takes place for 2 to 3 weeks, a light intrusion of water can be detected moving northward, clearly trapped to the bathymetry. The tongue is the result of baroclinic adjustment, which generates the northward flow. The flow is not yet turbulent but spans over 150 km (wide) and can be 400 m (deep). Simultaneously, and in response to the poleward flow, a weak counter-current can be detected. For an example of satellite images depicting the stage, Relvas refers the readers to Pingree and Le Cann (1990).

The next stage can be named the Eddy Development Stage. The second moment takes place in the 20 to 80 days after the onset of the flow. It is characterized by a decoupling of the current from bathymetry. This is partly due to the development of the turbulent character, with the formation of numerous eddies. These eddies tend to form at key topographic features, including the Aveiro Canyon (which is found within the study area) and the Estremadura

Promontory. Eddies (both cyclonic and anticyclonic) are also formed in the vicinity of other relevant subsurface topographic features but may not be anchored to them. These eddies are sometimes classified as SWODDIES (Slope Water Oceanic Eddies), which are usually anticyclones that can persist for many months in the region (Pingree et al., 1993).

The IPC is likely to encircle Galicia Bank and promote transport towards areas outside the slope (and the basin), progressively detaching from bathymetry as aforementioned, possibly within the study area, in the Aveiro Canyon region (Daniault et al., 1994, Mazé et al., 1997). This is in line with findings of a thinning of the flow towards the north (by 50% in width) (Peliz et al., 2005)

During this stage, the turbulent character becomes evident in satellite imagery not only in the form of eddies (and dipoles) but also in the countless meanders that are formed and transform the tongue into a complex feature that expands beyond (and behind) the slope. During this stage, a deep (> 500 m) equatorward flow is present, thus creating a vertical discontinuity.

There is no consensus regarding the third and final stage of the IPC. Although the divisions are somewhat artificial, considering the continuous nature of the phenomenon and the natural inter-annual variability, we believe the events surrounding the final stage – or stages - can indeed be consolidated into a single period.

This stage can be classified as the Decay Stage and, as the name foretells, describes the final moments of the IPC (as a surface current), taking place approximately 80 days after the onset of the current.

As the IPC evolves, the previously formed eddies begin to interact, namely to the north of Cape Carvoeiro. This interaction contributes to the formation of several dipoles and add complexity to the flow, which is simultaneously migrating north and westward. Westward migration is caused by advection, which is the result of dipolar interaction and the beta effect (which accounts for changes in meridional value of the Coriolis Parameter along a constant latitude). At this time, anticyclonic structures are prevalent and are separated by intrusions of colder waters.

As this stage evolves the turbulence grows and numerous structures can be detected in the satellite imagery. The structures are now omnipresent and found well beyond the slope (> 100 km). The detachment of the slope current tends to happen at specific areas, where bottom topography or even the coastal geography promotes such migration (Peliz et al., 2003).

Thermal fronts are created, which separate the IPC intrusion (warm and saltier) from the surrounding ocean (westward) or nearshore (eastward) waters (Peliz et al., 2002). In fact, the strong contrast seen in SST imagery can be exacerbated by the presence of cold waters

originating in the mainland through river runoff. These plumes of low salinity water are described next in greater detail considering its relevance in the context of the study area.

The IPC, highlighted by this stark contrast with surrounding flows, is therefore creating a transition zone, separating coastal and open ocean waters. In fact, and as aforementioned, the intrusion is a warmer, saltier water mass surrounded by waters that are colder and generally fresher (Peliz et al., 2002). This is an important aspect considering that its detection based on SST imagery is based on these assumptions. SST-based detection is considerably less complex and costly than traditional surveys, which require several current meters and multiparameter probes (e.g. CTD) to be deployed at multiple locations offshore.

But even when using SST data, it is useful to rely on temperature anomalies, rather than absolute values (Peliz et al., 2002). The anomalies can be calculated as the temperature difference coastal waters and offshore, open-ocean waters at longitudes beyond 10° W. It is useful to study long-term climatologies to understand the natural variability of the values and the westernmost extent of IPC (or other) currents and features. This approach is used latter in the study and is described in greater detail in the section dedicated to methodologies employed to extract and analyze the coastal environment using EO data sources.

Although poleward flows, like the IPC, are restricted to deeper layers for most of the year, and therefore have a limited impact over pelagic fish populations (Mason et al., 2006), these currents may have a surface signature during important moments of the life cycle of marine organisms.

The IPC is associated to the development of Harmful Algal Blooms (HAB) namely in the Rías Baixas in Galicia, which can have a detrimental effect on the local industry of mollusk production (Relvas et al., 2007). The toxicity is often associated to the “untimely” onset of the IPC, and to changes in the survivability and recruitment of planktonic organisms.

The IPC is responsible for the development of regions with a strong thermal gradient. This gradient – or thermal front – generates an effective barrier which separates the different water masses and the organisms or materials found therein. Such regions are also convergent, and may generate important areas affecting both phytoplankton, fish eggs, and larvae (Santos et al., 2004, Ribeiro et al., 2005).

The barriers have also been credited with blocking the flow of pollutants towards the coast (Álvarez-Salgado et al., 2006), which further suggests that the IPC may indeed form a separation between coastal and oceanic waters, with limited penetration of the flow nearshore.

The IPC is a complex current that interacts with several other features, inherently complex as well. As such it is easy to reckon the wide range of impacts the flow can induce in the

ecosystems off Portugal and other areas dominated by Eastern Boundary Currents. Such differences can be spatial, temporal or both and may disrupt the carefully choreographed life cycles of marine organisms inhabiting the CUE off the WIP.

An opposite flow (coastal counter-flow) to the IPC is recorded closer to the shore. This may result of the presence of a buoyant plume of fresher water that results of seasonal increased river runoff.

This feature is frequently called the Western Iberia Buoyant Plume (WIBP). The WIBP is a low salinity feature that occurs near the shore and is generated by river runoff (highly seasonal) of the western coast of Portugal and Spain.

The plume is often identified in the vicinity of the Mondego river estuary and extends northward under the influence of the Vouga, Douro, Lima and Minho rivers. Although not usually considered a part of the WIBP, the Tagus estuary is also the source of a significant plume of low salinity water that extends to significant distances, sometimes as far as the shelf edge.

Typical salinity values are often below 35.7 psu (Peliz et al., 2002) but this figure is highly dependent on the freshwater input and wind stress conditions.

The WIBP and the IPC are highly variable phenomena, not only from a seasonal standpoint but also in terms of inter-annual variability. This is caused by the cumulative effect of numerous variables such as watershed-wide rainfall levels in the adjacent river basins, just to name an example.

These features and their variability have an important impact over important processes, including transport patterns thus influencing the survival of marine organisms (Santos et al., 2007).

Because of the variability of both river runoff and wind stress conditions, a precise multi-year delimitation of the WIBP is impractical, if not impossible. Nonetheless, it is not unusual for the plume to extend well beyond the shelf edge to depths of over 1000 m (Peliz et al., 2002). The plume isn't, even so, vertically homogeneous across the shelf. The WIBP tends to be thicker close to the shore (~ 20 m), thinning westward, where it may be reduced to a small, under 10 m-thick layer (Peliz et al., 2002). The seaward side of the feature is generally more stratified considering that the WIBP waters become superimposed on the coastal waters (Santos et al., 2004). The thickness of the WIBP depends on multiple variables and processes including wind stress, other currents, and Ekman transport (Relvas et al., 2007), some of which we will address in greater detail.

The WIBP is indeed influenced by the dominant wind regime off western Portugal, which promotes convergence towards the coast. During winter, when south-westerly winds

dominate and the buoyant plume is at its maximum (due to seasonal rainfall levels), onshore Ekman transport develops. The convergence thus promoted, and the plume itself, creates a stratified environment with particular nutrient availability and optical climate (Otero et al., 2009).

However the convergence also results in the formation of inshore currents, albeit weak, but that can have an important impact over transport patterns in the region and persist even during summer (Peliz et al., 2002). Strong poleward currents also develop, and despite the small width of the flow (< 10 km) it is a hallmark of winter-time nearshore circulation patterns, capable of influencing transport of sediments and organisms. Marta-Almeida et al. (2002) reported a northward current with a velocity of 1 $m.s^{-1}$ in the vicinity of the Douro river, generated by the combined effect of river runoff and favorable winds.

Still, during winter upwelling events, a southward displacement of the WIBP is apparent in satellite imagery published in previous studies (e.g. Otero et al., 2009). Under upwelling conditions, motivated by northerly winds, temporary equatorward surface flows may develop, which reverse upon relaxation of the winds.

Another consequence of winter upwelling conditions is the offshore advection of the WIBP under enhanced Ekman transport westward. The transport can move water masses (and pigments) several dozen kilometers per day (Ribeiro et al., 2005), but more research and additional observations are required. The IPC, further offshore, stops the WIBP from extending further west, limiting simultaneously an even more significant thinning of the plume (Santos et al., 2004). At the contact between both features, significant mesoscale activity can be detected in the form of eddies, meanders, and other instabilities.

The WIBP is characterized by a subsurface salinity maximum (core), which is located close to the upwelling front in the summer, extending beyond the 1000 m isobath. The thickness of the WIBP in the summer is also variable, decreasing mid-shelf and increasing again westward, becoming an important contributor to a near-surface low salinity northward flow (Peliz et al., 2002).

Otero et al. (2009) suggests that geostationary satellites may provide a better view of the short-term variability of river plumes, which constitute a pivotal aspect of the WIBP and where the thermal and optical anomalies are stronger. The variability is often connected to the wind, which may generate changes in plume structure in short periods (i.e. hours). The sensors equipping geostationary satellites are, nonetheless, often coarser in terms of both spatial and spectral resolution. Unless improved sensors are deployed, the promise of rapid and nearly continuous monitoring of estuarine environments and coastal lenses will remain unfulfilled.

Alternatively, a large constellation of polar orbiting satellites could be equally efficient to monitor coastal ecosystems and hydrodynamics.

Either public or private (e.g. PlanetLabs's Doves), such fleets may provide high frequency acquisitions over day and night at high spatial and temporal resolution and often high signal-to-noise ratios. These are essential requirements to adequately image coastal areas considering not only the extreme variability at different time scales but also the impact of coastal weather. In the study area, and the WIP, frequent cloud cover in the winter (when the WIBP is at its maximum) and fog limits the acquisition of quality data over these regions very often, namely in the WIP.

The aforementioned features contribute to the formation of a thermal and salinity front (or set of fronts), separating the inner shelf of the coastal (and offshore) waters (Otero et al., 2009). It must be emphasized that during periods when the WIBP is at its maximum development the IPC is also active. Therefore, three different flows (and water masses) may coexist at the surface. Two poleward flows are present, the first associated to the WIBP nearshore and the second – the IPC – which is initially trapped to the slope. Further offshore, the broad Portugal Current slowly moves equatorward as part of the broader Atlantic circulation patterns. Further complexity is added by the vertical discontinuities and sub-surface expression of currents and at the surface, by the interaction of the flows with topography.

Salinity values at the Mixed Layer Depth front show only slight differences between studies, and were found to be within the 35.6-35.7 range (Otero et al., 2009, Peliz et al., 2002).

Vertically thin, the WIBP tends to cool faster than neighboring coastal waters. This lends the WIBP an even higher contrast when compared with other water masses (e.g. IPC), and even a lower temperature than that of the parent rivers and estuaries. This feature of the WIBP is caused by convection, which in this case, is restricted to a smaller vertical mass (Santos et al., 2004). Considering the different thickness of the layer across the inner-shelf, it is possibly, if not likely, that the cooling effect will manifest itself differently throughout the WIBP. On the other hand, when the conditions favor downwelling (coastal convergence) the WIBP waters tend to warm (Relvas et al., 2007), which may hinder a gradient-based detection strategy.

Detection of the plume can be accomplished upon interpretation of satellite imagery and in particular of SST data. Although the river plumes often display a distinctive color due to the elevated sediment load, the surface temperature of these features is also relatively different from that estimated from surrounding waters. This fact supports a detection strategy based on the analysis of thermal anomalies. The anomaly itself is variable throughout the year. In the winter, the WIBP waters are characterized by a negative anomaly, as shelf waters are warmer.

On the other hand, in the summer, river runoff (even though limited) generates a positive anomaly while upwelling pumps colder waters towards the surface (Torres et al., 2007). In summary, WIBP detection in (thermal) satellite imagery must account for seasonal specificities and intra-seasonal variability. Changes to the prevailing winds, for instance, may induce a relaxation of upwelling events in the summer, leading to changes in the SST anomaly profiles, which require a careful analysis (Otero et al., 2009).

Simultaneously, the heat exchange with the atmosphere may difficult the detection of the WIBP when the thin layer warms rapidly and efficient detection was limited to the first days after the onset of the feature (Ribeiro et al., 2005).

The development of convergence zones in response to the development of the WIBP can promote an increase in the amount of biological organisms near the front axis (Otero et al., 2009). The accumulation in the seaward side is coincident with a localized vertical growth of the WIBP, due to convergence and the contact with the IPC (Relvas et al., 2007). This accumulation may provide foraging grounds or enforce a range barrier to small planktonic or pelagic organisms.

The contact between the WIBP and the IPC promotes strong mixing, due to the turbulent nature of the contact between flows (similarly to what happens at the limits of the IPC, where opposing flows generate increased turbulence partly due to shear stress).

Furthermore, the WIBP is to known to offer adequate conditions for the development of phytoplankton along with larval fish (Santos et al., 2007). Variability in this feature may thus have an impact on the balance of the ecosystem, namely through changes to the recruitment of small pelagic fish.

The WIBP is thus an important and recurrent feature that must be adequately monitored. Such monitoring is not without challenges considering the extreme spatial heterogeneity of the coastal ecosystem off Portugal.

Some of the complexity is created by the geography of the area under the influence of the Portugal Current. The Gulf of Cadiz is also more than a simple obstacle separating the Portugal Current from the Canary Current. Through this gap, Mediterranean Water enters the Atlantic, flowing at depths of about 1000 m towards the North. This water mass is typically denser (higher salinity and temperature) and can be found along the Portuguese coast, namely in the area in the vicinity of Cape São Vicente (Mason et al., 2006).

These Mediterranean waters mix with Atlantic Intermediate Waters, creating a density plume which can acquire significant speeds in excess of 1 $m.s^{-1}$ (Relvas et al., 2006).

The interplay between the Mediterranean Water and bottom topography (amongst other variables, some of which poorly understood) lead to the development of the so-called

“meddies” – eddies of Mediterranean Water. These are predominantly generated near Cape São Vicente but their formation can also be detected in other important geographic accidents, some of which in the proximity of the study area (Estremadura Promontory) (Serra et al., 2002, Bower et al., 1997).

It is important to note that meddy propagation tends to be southward, after an initial northward displacement towards the Gorringe Bank (Relvas et al., 2007). Although surface signatures of meddies are possible (Carton et al., 2002) these may not always be obvious. Still, data from floats revealed that surface motion appears to be, at least occasionally correlated to these features (Relvas et al., 2007).

Meddies, despite the seemingly infrequent surface expression can be easily detected in vertical profiles of temperature and salinity (e.g. ARGO floats) and such results have been published by several authors (e.g. Peliz et al., 2005). The elusive character of the features, and the interaction with other water masses and currents, makes them still poorly understood. The limited impact these features have on the study area – according to the current body of knowledge – precludes a more detailed description of meddies in this work. Still we acknowledge that additional work would be relevant, namely to understand the northward movement of eddies across Gorringe Bank and determine the northern range of the features before dissipating.

The analysis above suggests that across a longitudinal transect, as far as longitude 24° W, the variability of the study area is significant. Different (and opposing) currents are found in close proximity, influencing the ecosystem in very dissimilar ways. The system is adapted to this variability and marine life depends on these differences to prosper in the waters off WIP. Nonetheless, and despite the inherent relevance of each system, upwelling events are arguably the most distinctive aspect of the region.

Upwelling is an important aspect of the ecosystem, even reflected in its name. It translates into significant ecological and economic traits typical of high productivity regions. But because upwelling occurs in the coastal area of WIP in a specific context, we believe it would be adequate to address it after a description of other features. The IPC, for instance, has a surface signature in the absence of upwelling conditions, highlighting the complex interplay of variables that command the occurrence of different, complementary phenomena.

However, upwelling is arguably one of the most structuring elements of the ecosystem, which will now be summarized. Previous authors addressed upwelling dynamics in great detail, and as such, we do not wish emulate such efforts, but instead present the key aspects of upwelling that may influence mesoscale activity off WIP. Summer circulation patterns off the WIP are better understood and described than the wintertime counterparts (Mason et al., 2006). This is

– at least partly – due to the importance of upwelling to the maintenance of the ecosystem (aptly named the Canary Upwelling Ecosystem).

The climate off Iberia favors the development of seasonal northerly winds, which initiate and support the development of upwelling cells along the coast (Fiúza et al., 1982). The process is initiated as the Azores high pressure cell moves north, highlighting the importance of this feature and of its variability over the WIP circulation and ecological patterns. The northerlies are further enhanced by the presence of a low pressure system (of thermal origin) over the Iberian Peninsula (Relvas et al., 2007).

The observation of upwelling phenomena in the Portuguese coast is reported for several decades (at least) (e.g. Fiúza, 1983) and it is thus reasonably well described. Overall, in these events, the aforementioned equatorward winds promote Ekman transport offshore, which causes an imbalance leading to a rise of cold, nutrient-rich water to the surface from depths in excess of 100 m.

Similarly to what was described regarding the evolution of the IPC, coastal upwelling is also subject to a step-wise development, with several identifiable, yet interconnected stages occurring in a sequential order. Still, there is also significant variability and both the upwelling onset date(s) and intensity varies with wind direction, magnitude, and persistence. As such it is inadequate to study or forecast upwelling from the mere analysis of climatologies (Alvarez et al., 2011), but instead the employment of tandem temperature (*in situ* or satellite SST) and wind measurements is more suitable to describe upwelling conditions for ecological and operational applications.

The upwelling season is active predominantly between July and September (Wooster et al, 1976), although upwelling-favorable winds can occur year-round. When favorable winds develop, the newly upwelled water (Eastern North Atlantic Central Water – ENACW) leads to a rise in the isopycnals near the coast (Tenore et al., 1995, Mason et al., 2006) and generates a signature that can be detected in satellite imagery, which consists initially of a nearshore, small and meridionally aligned mass of cold water. However, as upwelling strengthens and further develops, the initially restricted band expands westward, and prominent jets anchored at recurrent locations such as the Aveiro Canyon become noticeable in satellite imagery (Relvas et al., 2007). These filaments can transport a significant amount of upwelled water (and materials therein) offshore, sometimes to distances of over 250 km (Peliz et al., 2002).

The upwelling and thermal gradient between the cold and coastal waters generates a front (or more accurately, a series of fronts along a generic meridional frontal region), which define a barrier between the shelf waters and the offshore region. This is an unstable region subject to the short-wave frontal instabilities that propagate at high velocities along the flow, generating

a contorted region marked by jets, filaments, and eddies (Barth, 1994). This is not a static front but it evolves with the advancing upwelling season, moving further offshore and creating a distinctive region dominated by cold waters coastward of the discontinuity (Austin et al., 2002). The movement is partly due to Ekman transport offshore, which is also connected to the generation of a geostrophic southward current in the area of study (Mason et al., 2006).

Strong along-front flows with surface and subsurface expression are also generated by the strong thermal and salinity gradients found at upwelling fronts (Belkin et al., 2009, Armstrong et al., 1987). Such flows may play an important role in the local circulation and distribution of materials (including biogenic) (Woodson et al., 2012), adding an additional layer of complexity to the interpretation of coastal flows at the mesoscale or smaller spatial and temporal scales.

Frontal systems in upwelling regions are often mobile, moving offshore with the consolidation of the upwelling season, as aforementioned for the Portuguese case. In California, a similar trend was observed, with upwelling fronts moving hundreds of kilometers offshore, as the phenomenon develops (Breaker et al., 2005).

Upwelling off the coast of Portugal is then characterized by the spatial and temporal variability of the structures that are associated to this phenomenon. The phenomenon is not limited to a uniform narrow band located along the meridional Portuguese coast. Instead, this is a dynamic phenomenon with focal cells from where the upwelled waters emerge and spread. Simultaneously, shadow zones, where the upwelling signal is weaker are also present, creating strong contrasts not only cross-slope but also along-shore. The interaction with other features and currents also contributes to the heterogeneity with the development of significant mesoscale activity which includes the formation of eddies. Finally, filaments develop at certain locations (as in other upwelling systems) and reinforce the heterogeneity of the coastal waters off Portugal. The variability is discernable in medium and low resolution satellite imagery (e.g. MODIS and Landsat), which highlight the richness of features rendered by the gradients in the SST and Chlorophyll a concentration fields.

Upwelling shadow zones are characterized by warmer waters, due to a combination of factors that include a weaker upwelling, higher stratification, and particular circulation patterns (Mason et al., 2006). These shadow zones are created by the interaction of the prevailing winds and currents with the coastal geography.

Examples of such shadow zones, where upwelling signatures are less prominent and can therefore be detected in satellite imagery, can be found southward of Cabo da Roca (Moita et al., 2003).

The filaments are, nonetheless, amongst the most interesting and less studied upwelling-related features. These are narrow but long structures characterized by colder waters that

detach from the coast and propagate eastward for more than 150 km. These features can be seen and tracked using satellite imagery and are known to move at velocities of up to 0.28 $m.s^{-1}$ (Smyth et al., 2001). Recent studies by Cordeiro et al., 2015, attempted to describe and characterize these features off WIP resorting to satellite and model data.

Haynes et al. (1993) attributed the development of the filaments to the presence of capes, which force the movement of the structure offshore. Still, other forces may apply, namely in the absence of complex coastline, with wind stress (Batteen et al., 1992), frontal instabilities taking on the main role in generating the filaments (Peliz et al., 2002). Shear stress at the interface between the equatorward and shelf-edge poleward flows can also be an important contribution to the phenomenon (Peliz et al., 2002, Cordeiro et al., 2015). On the other hand, Roed et al. (1999) go as far as saying that the filaments are not created through the action of topography. Once again, more research is needed and consensus is still distant as it happens with so many other oceanographic processes.

Still, the paper by Cordeiro et al. (2015) suggests that filaments are actually correlated with wind stress, and featureless coastlines are less likely to spur filaments, in a complex interplay of drivers. Considering the exhaustive analysis of long time series and the matching results between satellite and model data, the paper by the aforementioned authors is likely to provide a balanced view of the genesis of these features.

Nontheless, the absence of consensus is illustrative of the state of the art concerning the features and the need for further research to shed light over the formation mechanisms, which are most likely to include a set of contributing forces rather than a single driver.

Amongst the filaments commonly observed off the WIP, the Aveiro Canyon filament is the most notable and widely mentioned in the bibliography. This filament is recurrently detected in satellite and *in situ* data, extending over hundreds of kilometers offshore at the approximate latitude of the Aveiro Canyon. The filament can be traced in SST imagery as a cold anomaly that protrudes into the warmer coastal waters and becomes increasingly convoluted as it moves eastward and interacts with other flows.

In the vicinity of Aveiro, a return flow equatorward of the filament generates a warmer signature that contrasts with the cold, nutrient rich upwelled waters found in the nearshore environment (Strub et al., 1991). Nonetheless, Peliz et al. (2002) alerts to the fact that despite the occurrence of eddies and dipoles in the vicinity of the filament, this is not recirculated by action of the mesoscale features. Nonetheless, the topographic forcing of the Aveiro Canyon together with the action of the IPC promotes the development of the filaments and its permanence offshore.

These structures are evident transport mechanisms from the coastal environment towards the offshore domain. The high velocities and spatial scale of the feature suggests it must have a non-negligible effect over biogenic and other drifting materials. This is potentially valid at least in the upper layers (Relvas et al., 2007) and considering the recurrence of the feature, it should probably be studied as an inherent component of the ecosystem within the area of study.

The temporal variability of upwelling is also associated to the long term variability of the atmospheric and oceanic cycles, such as the Atlantic Multidecadal Oscillation. The cycles, of both cooling and warming, are measurable using long term SST analysis, from *in situ* and satellite sources, providing that consistent time series exist. The time series show that the average temperature near the coast is lower than that measured offshore (Santos et al., 2011). This is probably due to the effect of upwelling, which can occur throughout the year despite the prevalence of the phenomenon during warmer months. Upwelling is thus an important component of the overall SST profile of the coastal area and disruptions of the cycle can probably affect the thermal signature of the area of study.

However the analysis shows that the correlation between the cycles and AMO decreases with proximity to the coast (Santos et al., 2011). Gradual changes to SST in the study area (and worldwide) are hence important to guarantee that an appropriate monitoring of the responses of the system to the emerging trends created by global climate change.

In the study area, and in the broader region where it is located, general warming trends have been measured in recent decades. This trend is not exclusive to the WIP but has been recorded throughout the Eastern North Atlantic (Casey et al., 2001).

Emerging trends in the vicinity of the study area have been reported by different authors (e.g. Relvas et al., 2009), and are corroborated by different datasets, which reinforce the alarming hypothesis. Relvas et al. (2009) in a comprehensive study on different sections of the Portuguese coast suggested the warming trend may be as high as 0.037° C/year (1960-2005). For satellite-based measurements the estimates range from 0.022 to 0.033° C/year in recent data sequences starting in 1985 and ending in 2008.

The same author also suggests that the warming is occurring in a way that exacerbates the thermal contrast between the coastal and offshore waters due to a misalignment of the trends in the different sectors. Warming is apparently stronger off central Portugal, but with areas that lag behind the general tendency, such as the region to the south of Cape da Roca.

Considering the somewhat limited time series available and used in regional studies off WIP, the warming may be, at least partly, due to natural variability of the system.

It is known that waters off WIP are influenced by the North Atlantic Oscillation (NAO), which may introduce long-term variability in the measured temperature fields (Santos et al., 2005). It

has been suggested that warming periods off northern WIP are somehow correlated with low NAO cycles (Peliz et al., 2005). Therefore, the continued monitoring of the region is essential in order to expand existing time series and detail the spatial patterns emerging from the data.

The use of long time series based on satellite measurements is paramount in the measurement of SST (and other variables) over wide swaths of ocean, even though the data is limited to recent decades. Still, and despite this limitation, the growing archive of satellite data is useful to analyze emerging patterns in temperature and other variables, considering how impractical the deployment of massive float arrays is. Furthermore mesoscale activity, so often related to the interaction of flows with contrasting characteristics, is also likely to evolve into new patterns considering the strong forcing from thermal-related phenomena. Therefore, and at this scale (temporal and spatial), satellites offer the best solution for a continuous monitoring leading to the development of a complete image of the warming ecosystem (Relvas et al., 2009).

Thermal changes to the ecosystem, may introduce a plethora of stressors into existing communities. In fact, the detection of species from warmer waters has been reported already in the Northeast Atlantic (Brander et al., 2003). Simultaneously, the warming trends (and the apparent heterogeneity of said warming) are likely to modify (albeit slightly at first) flow patterns, which can displace natural barriers (frontal regions) or promote transport into new regions as the system attempts to find a new balance. The aforementioned modification of winter upwelling events, for instance, is known to play an important role in the recruitment of important fish species, such as the sardine (Santos et al., 2004).

Detecting upwelling plumes is possible from the analysis of satellite SST imagery given the stark contrast between the cold, mixed upwelled waters and the coastal warmer waters (Relvas et al., 2007). Once again, and as in the case of the IPC, this is a gradient-based detection, which takes advantage of the relative temperature differences to ensure an accurate detection.

One of the strategies employed to monitor upwelling with limited effort and efficient results is through the calculation of the Upwelling Index (UI).

The UI is the surface temperature difference between waters at coastal locations where upwelling is active and offshore regions at the same latitude (in meridionally-aligned coasts), where the impact of upwelling is inexistent or negligible (Alvarez et al., 2011, Benazzouz et al., 2014). The data used in this calculation can either be daily or temporal composites (e.g. 8-day fields), in order to reduce noise and maximize coverage. For the area of study, previous studies calculated the UI as the SST zonal mean between longitudes 11-12° W, at the latitude of the sampling points (Relvas et al., 2007).

When the UI is calculated from satellite imagery, it is often abbreviated as UI_{sst}. The UI_{sst} is extremely useful because it simplifies the development of long and coherent time series of SST anomalies. The analysis of such time series yields important data including the inter-annual variability (Alvarez et al., 2011).

The UI_{sst} tends to assume negative values during the primary upwelling season (spring and summer), due to the negative thermal anomaly that prevails in regions dominated by upwelled waters. This is particularly effective when the average UI_{sst} is considered, highlighting the reduced or absent upwelling activity from January to May across the study area and a peak of activity in August and September (Alvarez et al., 2011).

Moored or drifting systems (e.g. the MONICAN network in the Nazaré Canyon) can also provide important information on upwelling, especially when sea surface data is collected simultaneously with atmospheric variables. Nonetheless, *in situ* systems or networks have a limited capability (unless they are deployed in significant numbers) to depict upwelling events and capture the spatial heterogeneity that characterizes this phenomenon.

Variability in the upwelling magnitude, time, and persistence is known to have significant ecological consequences off the WIP. In particular, the population of small pelagic fish, such as the economically relevant sardine, can be affected by variability in the upwelling regime (Santos et al., 2007).

Several mechanisms can explain the influence of upwelling over biological communities, some of which are not dissimilar to others mentioned in other features found in the area of study. Upwelling contributes to enhance the productivity of waters that would otherwise have a significantly lower load of nutrients. But this only addresses part of the true impact. The frontal area formed by the upwelled waters creates a “biologic retention zone” in the inner shelf, potentially limiting the zonal movement of organisms or other biogenic materials and creating dissimilar cross-front species assemblages (Breaker et al., 2005, Mason et al., 2006).

Furthermore, the flows (cross and along-slope) formed throughout the upwelling season (and episodes) contributes to the movement of materials either promoting or negating access to favorable development grounds. Upwelling also generates strong thermal gradients between the loci where it is occurring and other areas. The SST gradient may reach over 3° C along the Portuguese coast (Relvas et al., 2007), creating unique physical conditions on which diverse biological communities may thrive. There are thus spatial and temporal forcing mechanisms that strongly influence the stability and life story of organisms across the ecosystem. We must thus emphasize once again the need to monitor the variable patterns continuously, using methods enabling a holistic perspective of the (simultaneous) phenomena occurring in the region. Still, the region is undersampled and understudied, when compared with equivalent

systems elsewhere (e.g. California) and a significant investment is required to further advance our knowledge on the ecosystem.

But from the different features described so far, none is as elusive and difficult to monitor as Internal Waves (IW). IWs can be succinctly described as waves traversing the interior of a fluid, often in the interface between two layers with different properties (Apel, 1987).

These waves are often slow and have large amplitudes being difficult to detect at the surface. Although a detailed description of these features is beyond the scope of this work, some notes are called for, considering the relevance of IWs to the spatial organization of relevant features and the detection of fronts using satellite data.

IWs are found in diverse regions, whenever the natural conditions foster the stratification of the ocean or powerful tides. Bottom topography also plays an important role in the genesis of these features. Two types of waves can be found in the study area, including Internal Tides (IT) and the shorter and phase-locked Internal Solitary Waves (ISW) (Muacho et al., 2013).

Internal tides are found quite ubiquitously in shelf edge regions, including the WIP (Jeans and Sherwin, 2001). In the WIP, tides are predominantly controlled by the semi-diurnal M2 and S2 components (as well as K1) and propagate as a result of the combination of Kelvin-like waves and bottom topography (Relvas et al., 2007).

Two traits make IWs particularly relevant. Firstly, they can travel for thousands of kilometers, and secondly they induce vertical motion of up to tens of meters, displacing the seasonal thermocline and particles found in the water column, including those of biogenic nature.

As such, IWs can contribute to the movement as well as the mixing of nutrients and organisms, playing an important part in the coastal ecosystem.

Detecting IWs is best done with SAR data, which returns a series of alternating dark and bright bands (Alpers, 1985). In the study area, studies based on SAR and *in situ* data revealed the ubiquity of IWs, which are formed along the west coast of Portugal with focal points centered in the proximity of prominent bathymetric features such as seamounts, promontories, and canyons (Jackson, 2004, Muacho et al., 2013, da Silva et al., 1998).

The solitons retrieved from SAR data in previous studies, were separated by short distances of up to 1.5 km, while traveling at velocities of 0.45 $m.s^{-1}$ with amplitudes of nearly 250 meters in some cases (Small, 2002, Jackson, 2004). The wave packets these features often form suggest these are essentially non-linear but dispersive trains, but our understanding of the role of these features in the ecosystem is still limited (Relvas et al., 2007).

Considering the lack of abundant historical SAR data and the inherent difficulty of deploying thermistor chains offshore, much research is needed to uncover the persistence, seasonality, and origin of the different IWs. Still, Muacho et al. (2013) reported that the action of IWs was

capable of influencing the distribution of chlorophyll concentration and primary productivity in the region of the Nazaré Canyon. Internal Tides were found by these authors to be correlated with higher concentration of Chlorophyll, because they exposed the autotrophs to higher light intensity.

Such relation had been reported earlier, and attempts have been made to quantify the impact of IWs in the enhancement (Lande et al., 2008). The hypothesized process is rather simple, describing how the passive autotrophs may gain access to stronger light intensities as they are transported vertically by the IWs. As such, photosynthesis is enhanced and primary productivity is positively affected in the disturbed areas.

This is an important fact, which underlines the importance of integrating different data sources for the interpretation of ocean datasets. The enhancement and patchiness created by IWs is reflected in Ocean Color data. Albeit no information is provided regarding the formation of Chlorophyll a fronts, these are likely formed, even though SST fronts may not always be present in this case.

The ocean off WIP is thus a region of high dynamism, characterized by multiple currents and features within a short distance from the shelf edge. As in the Californian system, a Coastal Transition Zone (CTZ) is therefore present off the WIP. The CTZ is a region of varying width that marks the transition between the coastal and oceanic areas (Brink et al., 1991). Both can be distinguished from neighboring regions by their energy, which is higher in the coastal area.

This is also a highly variable area and where the presence of mesoscale phenomena and frontal boundaries is more likely to occur. The aforementioned phenomena and features contribute to the development of the CTZ, which is highly variable in space and intensity of the components.

The CTZ in California is a highly productive area characterized by cold, nutrient-rich waters and the presence of long upwelling filaments, which develop predominantly in certain periods (as in Portugal) and extend over 300 km offshore (Kosro et al., 1991, Moisan, 1994). In fact the CTZ of both the California and Portugal systems share several similarities, as it was the case with other phenomena and features including the broad circulation patterns, coastal counter-currents, and upwelling ecosystems.

In the case of Portugal, the CTZ is often associated with the slope area, where the development of abundant mesoscale features including eddies (and dipoles), along with the presence of the poleward flow, create a dynamic boundary between regions (Peliz et al., 2003).

The abundance of mesoscale features, as can be hypothesized from the aforementioned characteristics of fronts or eddies, is likely to play a significant role in the modulation of both oceanic and atmospheric processes. Fisheries and a significant proportion of commercial ship

traffic occur at or near the CTZ. It is thus both an ecologically relevant and vulnerable region that changes with the seasons and inter-annual variability.

The analysis of the described phenomena yields an important conclusion. Thermal and Chlorophyll a fronts are ubiquitous and are often important markers of true horizontal and vertical discontinuities in the oceanic continuum. The convergent nature of fronts fosters the development of processes that can hardly take place within the body of the water masses they divide. As such, they are not mere contacts between regions but rightfully an ecosystem, making of the mapping and characterization of such features an important oceanographic task (Belkin et al., 2009).

Off Portugal, and within the study area, the presence of fronts has been recognized in previous studies (e.g. Relvas et al., 2007, Miller, 2009, Peliz et al., 2005), but these never presented a systematic, and long term analysis of frontal systems using medium resolution satellite data. A significant amount of studies addressing fronts in the study area, are not actually focused on these features. Instead they address other aspects of the regional oceanography or ecology, using fronts as an additional variable (e.g. Peliz et al., 2005). Because of this, front analysis is often restricted to specific seasons or based on coarse data sets.

Furthermore, frontal probability maps reported by different authors are often contradictory either due to different approaches, data sources, or vision. It is therefore urgent to create a solid, high resolution (spatial and temporal) dataset describing thermal fronts in the Portuguese coast, namely in the vicinity of important fishery chubs such as Peniche, Figueira da Foz, and Aveiro. Such studies often identify hotspots for the occurrence of seasonal or permanent fronts, namely at the southern section of the Nazaré Canyon (Marta-Almeida and Dubert, 2006). Still, much research (and data) is still necessary to create a proper climatology of fronts. Additionally, the dataset now presented should be updatable and distributable, in order to provide a real contribution to the scientific and operational use of frontal maps into the future.

The next chapter describes the effort to map and characterize thermal fronts, and the water masses they divide, in an effort to recognize the distribution, persistence, and correlation with other phenomena. To this end, the use of high-frequency and long-term data acquisitions by Earth Observation data were pivotal, providing a holistic and new perspective on the ocean off central Portugal.

Part II. Biogeographic analysis and frontal climatology

Chapter 4. Biogeography and Frontal Climatology from optical data

4.1. Remote sensing data and processing

As mentioned in a previous chapter, at the core of the project is a comprehensive set of Moderate Resolution Imaging Spectroradiometer (MODIS) data acquired from the Ocean Biology Processing Group Website (OBPG) of the Goddard Space Flight Center at NASA. The MODIS sensor is found in two NASA satellites, AQUA and TERRA. Despite the longer operational life of TERRA (launched in 2000, unlike AQUA, which was launched in 2002), known calibration issues (Franz et al., 2007) motivated the exclusive use of imagery acquired by MODIS AQUA. This approach supports the development of a consistent single-mission time series, unaffected by inter-sensor discrepancies (Donlon et al., 2007). Considering that the AQUA satellite was launched May 4, 2002, the remainder of the year was excluded from the series, which starts in January 1, 2003, and ends in December 31, 2014. Avoiding the integration of data from the available months in 2002, contributes to the consistency of monthly and annual fields.

In order to create a complete time series of Sea Surface Temperature and Chlorophyll a concentration, a total of 3450 images were thus acquired. The wide swath of the MODIS sensor enables, very frequently, for a single scene to image the entire study area. However, when necessary, adjacent images were mosaicked to create a single daily product.

The data were downloaded as HDF files in Level 0 and processed with the SeaDAS 7 software package (Baith et al., 2001) into Level 2. Several products were generated including Sea Surface Temperature (°C), using the Long-Wave Sea Surface Temperature product (Brown and Minnett, 1999), Chlorophyll a concentration ($mg.m^{-3}$) (O'Reilly et al., 2000), Remote Sensing Reflectance (Rrs_nnn) (Carder et al., 2003), and Normalized Water-leaving Radiance (nLw_nnn) (Gordon and Wang, 1994). Furthermore, 'true-color' images were generated and exported into GeoTiff format for visual analysis and support interpretations.

Level 2 products, with a spatial resolution of 1 km, are generated by the l2gen library from Level 1 versions with full geometric, radiometric and atmospheric correction, and integrate numerous flags for quality control. The flags adopted in the study, to mask low-quality pixels included atmospheric correction failure, Land pixel detection, Sun glint and others. The full list includes: ATMFAIL, LAND, HILT, HISATZEN, STRAYLIGHT, CLDICE, COCCOLITH, LOWLW, CHLWARN, CHLFAIL, NAVWARN, MAXAERITER, ATMWARN, HISOLZEN, NAVFAIL, FILTER,

HIGLINT, SSTFAIL, SSTWARN, each described in the SeaDAS software documentation and OBPG website (OBPG, 2016).

The daily SST and Chlorophyll a products were subsequently exported to GeoTiff format and archived using ArcGIS 10 (various versions from 10.0 to 10.2) for later processing and analysis.

To minimize the chances of error and optimize the production of the datasets, an automated batch script was created, taking advantage of SeaDAS command-line processing capabilities. The Graph Processing Framework's Graph Processing Tool (GPT) functionality was leveraged to create the script, which was designed to handle the different operations required in the processing chain. The script was fully tested and outputs were compared against data from different periods processed manually, to guarantee that all requirements were being met, including pixel masking. The resulting products were also submitted to a rigorous post-production quality control process, which included the random selection of Level 2 products and comparison with manually processed versions. Pixel values and masking were evaluated and in the event of discrepancies being found, the script would be corrected and the entire collection reprocessed.

Because the number of observations is important for subsequent analysis, the total, annual, and monthly pixel count was calculated for a regular grid at a spatial resolution of 1 km.

The daily Level-2 data were also mapped into monthly Level 3 gridded products with a spatial resolution of 4 km. In the processing, four layers are generated including mean SST or Chlorophyll a concentration, Standard deviation, number of passes (input files), and number of observations (number of values used to generate the Level 3 pixel). The monthly Level 3 files were then used to generate monthly climatological files (2003-2014) containing the average and standard deviation for SST and Chl-a in each pixel of the study area. The Level 3 products were also exported into GeoTiff format and archived using ArcGIS.

SST and Chlorophyll a front detection protocols are described in an ensuing section. Nonetheless, several derivative products were also generated from the Level 2 and Level 3 SST and Chlorophyll a data. One of such products was the Upwelling Index (UI_{SST}). The UI_{SST} is calculated as the difference between the SST field in each pixel and the mean of the values retrieved for pixels at the same latitude but located between longitude 11.5° W and 12.0° W. This approach enables estimating the surface temperature difference between the open ocean and the coastal areas, expressed in °C (Nykjær et al., 1991).

Data from the Landsat program were also acquired for coastal analysis, but in this case, from the United States Geological Survey Earth Explorer website (EarthExplorer, 2016) maintained by the Earth Resources Observation and Science (EROS) Data Center. The imagery was used in selected areas, in support of interpretation of the lower resolution MODIS data, whenever

deemed necessary. Additionally, the Landsat data were used to retrieve the location of river plumes generated by the Mondego river at Figueira da Foz.

The scenes from Landsat -5/-7/-8 were acquired at Level 1 and processed to correct for atmospheric effects, prior to production of Surface Temperature and multispectral products. The correction was applied to the scenes using an implementation of MODTRAN4 radiative transfer model (Berk et al., 1998) available at ENVI, through the FLAASH module (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes). Considering the primary application of these data to coastal areas, in the vicinity of land, the aerosol type was set to "Maritime", which was considered adequate given the assumption that maritime wind is prevailing.

The post-processing tasks applied to the thus corrected Landsat data, are described in greater detail in subsequent sections.

In support of the demanding data archival effort associated with the large volume of information processed (original and derivative), all files were accompanied by descriptive metadata and a data handling log was created. The metadata file contained useful information regarding the file contents, production date, and version, for rapid retrieval of processing details from the log. In turn, the data handling log included the details of the processing run, including potential errors or warnings raised by the scripts in the process.

4.2. *In situ* data sources

Despite the wealth of data provided by remote sensing data sources, *in situ* measurements are still invaluable to calibrate, validate, and supplement those products (Smale et al., 2009, Mantas et al., 2015).

Several *in situ* data sources are available for the global ocean and many others to regional seas. Several repositories were designed to store either *in situ* information, such as in the case of SeaBASS, maintained by NASA (Werdell et al., 2003) or a combination of field, remote sensing and modeling products and services as available at the European Copernicus Marine Environment Monitoring Service (formerly known as "myOcean") (Nardelli et al., 2013, CMEMS, 2016).

In the context of this project, two systems were used to retrieve oceanic *in situ* data, to support the validation of the remote sensing data sets and provide additional information, unavailable from remote sensors at the current state of the art. To this end, three data sources were acquired for the study area and period from the CORIOLIS portal (Villéon et al., 2003). This portal offers seamless access to several relevant oceanographic data sources, searchable through a simple map interface. In particular, it was deemed necessary to retrieve information

that could provide an insight into 1) the vertical structure of the ocean in selected areas, 2) provide surface Sea Surface Temperature and Salinity measurement match-ups.

Data from the ARGO float network (Gould et al., 2004) was used to retrieve Temperature and Salinity along vertical profiles. The ARGO floats are designed to operate in continuous park and profile mission cycles, with an operational design of approximately 150 such cycles. Although several models are used in the network, they share common traits and an overall mission cycle. These cycles, have a duration of approximately 10 days and include a drift stage at a depth of 1000 m and a profiling ascent, from an even greater depth. As the float ascends it collects over 200 temperature and salinity measurements, which are then transmitted via satellite.

As of April 21, 2016, an array of 3918 floats was deployed and actively delivering information. The information broadcasted via satellite from the ARGO floats is received in three Global Data Assembly Centers (GDACs), from which it is disseminated in near-real time. Additional information on the network is available at the ARGO network official website (http://www.argo.ucsd.edu/).

A total of 482 vertical profiles were downloaded and processed. Processing included the conversion of pressure values to depth, following UNESCO's recommendations (Fofonoff et al., 1983). A script in python was created to recursively search the profiles from the archive folders, segment the different profiling dives into separate files, and finally apply the necessary conversion equations to each profile.

Pressure to Depth conversion is calculated in two steps. Firstly it is determined the value of the gravity (g) variation with latitude and pressure:

$$g\ (m.s^{-2}) = 9.780318\ .\ [\ 1.0 + (\ 5.2788 \times 10^{-3} + 2.36 \times 10^{-5}\ .\ x)\ .\ x\ ] + 1.092 \times 10^{-6}\ .\ p$$

where,

$$x = [\sin (\text{latitude} / 57.29578)\ ]^{2}$$

p = pressure (dec)

After g is calculated, the depth stored in the ARGO profile is converted to pressure as:

$$\text{depth (m)} = [(((-1.82 \times 10^{-15}\ .\ p + 2.279 \times 10^{-10})\ .\ p - 2.2512 \times 10^{-5})\ .\ p + 9.72659)\ .\ p]\ .\ g^{-1}$$

The script, as well as its documentation, is available online at the project website (www.smartbasins.com/argo). The profiles were also mapped in ArcGIS, to simplify the comparison of the *in situ* data with the satellite and ancillary data.

Additionally, temperature data from drifters were also retrieved. The data is collected by passive lagrangian floats equipped with temperature and salinity meters. Data from the drifters were used to provide a set of *in situ* SST measurements and current information.

Data from scientific cruises, including shipboard thermosalinograph were also acquired. These data were particularly important to validate the SST front detection methodologies. To this end data match-ups were identified and frontal maps compared against the profiles generated by the *in situ* assets.

4.3. Ancillary data

Regional bathymetry was downloaded from the website of the Instituto Hidrográfico (Portuguese Navy). The West Iberian Bathmetry Model (WIBM2009) was selected (Quaresma et al., 2013). This is a 1 arc-second digital elevation model compiled from different bathymetry databases and referenced to the Portuguese hydrographic zero. The data were converted from ESRI shapefile to GeoTiff in support of the development of derivative products including slope and aspect rasters at ArcGIS.

Furthermore, river flow and water quality data were acquired from the Portuguese Hydrologic Information System (SNIRH) online portal (SNIRH, 2016) for the period between January 1, 2003 and December 31, 2014. Two sets were retrieved from the database. The first includes instantaneous discharge values (2 stations) and Stage level (9 stations), from the Mondego, Vouga, Douro, and Lis watersheds. When available (n = 5) rating curves were used to estimate river discharge from Stage data made available at SNIRH. These data were used to determine the optimum dates for river plume analysis and to compare the dimensions of the WIBP with river outputs.

The second set retrieved from SNIRH includes water quality variables from 19 water quality stations within the Mondego river watershed. Table 5.1.3.1 summarizes the location and values calculated for the period between January 1, 2011 and December 31, 2013. For those measurements below the limit of detection (LOD), the value was replaced by half the LOD whenever the total number of replacements does not exceed 15% (EPA, 2000). For each site (n=19), the associated sub-basin was delineated using the ArcHydro extensions for ArcGIS.

The medians of each station (2011-2013) were calculated and log-transformed to mitigate seasonality (Mehaffey et al., 2005) before further analysis.

4.4. Time series analysis

Ocean ecosystems are spatially organized, with contiguous provinces separated by objective boundaries limiting water masses with distinct biogeophysical characteristics. However, the

vastness of the ocean and the dynamic nature of the boundaries limit the usefulness of traditional *in situ* techniques to accurately map the distribution of these regions.

As aforementioned it is thus necessary to partition the ocean to monitor it efficiently using long time series capable of capturing the intricate variability found at different spatial and temporal scales.

Remote sensing data offers an unprecedented and still unmatched source of information that can be used by the scientific community to extract important insights into the dynamic of physical and biological variables and processes.

In this study, we tackle the processing and interpretation of a large set of Earth Observation data, mostly acquired by the MODIS-AQUA sensor, to determine the location and persistence of thermal fronts, characterize coastal and ocean waters in terms of long time series, and organize such waters into coherent units using dissimilarity metrics and clustering methods.

The aforesaid data processing methodologies were followed by the application of extensive data organization and mining techniques, which are detailed in the following sub-sections. Furthermore, the connection of the frontal regions with the seasonal and annual SST and Chlorophyll a cycles will be detailed, as we put forward a new regionalization model. This model, which can be integrated with existing systems, has the advantage of being reproducible, scalable, and objective.

This section (4.4) focuses on Sea Surface Temperature and Chlorophyll a dynamics, relying on standard and emerging techniques for data mining and interpretation.

4.4.1. Sea Surface Temperature and Chlorophyll a concentration

The analysis of Sea Surface Temperature (SST) in the area of interest between January 2003 and December 2014 relied on the outputs of a processing chain described in earlier sections. Nonetheless, there were significant post-processing steps required prior to data classification and analysis. In this section we shall describe the overall processing of the SST and Chlorophyll a data prior to implementation of regionalization schemes and front detection protocols. The two variables shall also be succinctly described within the framework of the present study.

We do not aim at an exhaustive characterization of SST dynamics but instead highlight the differences between the main drivers of spatial heterogeneity within the study area. In particular, we emphasize the role of upwelling, WIBP, and currents on the development of complex and highly seasonal spatial patterns. Understanding the overall design and timing of the patterns, and how they are depicted in SST imagery, is paramount to design and validate successful regionalization scheme. Interpreting the distribution of frontal boundaries requires

the cumulative understanding of the aforementioned spatial patterns and a comprehension of the regionalization. We thus adopted a stepwise approach to the challenge of mapping and classifying the coastal region off central Portugal, which is mimicked in the organization of this and latter sections of this thesis.

The first step was to evaluate whether the number of valid SST pixels was sufficient to perform the tasks required to accomplish the objectives of this work. To that end, all valid pixels were identified and counted using a simple arcpy script. The script iterated through data directories and assigned a value (0: no valid data; 1: valid data) to a temporary file, all of which were subsequently summed to yield monthly, annual, and total counts.

A valid SST pixel is any cell where the SST measurement retrieved from the satellite imagery (Level 2) is non-null. Null values were identified with flags, each representing a different issue that precluded a safe estimate. The problems range from proximity to land to the presence of clouds or sun glint. This is therefore not a proxy for cloud cover, but an overall assessment of data retrievability.

The resulting map, which displays the number of valid SST pixels as a continuous grayscale is depicted in figure 4.4.1.1.

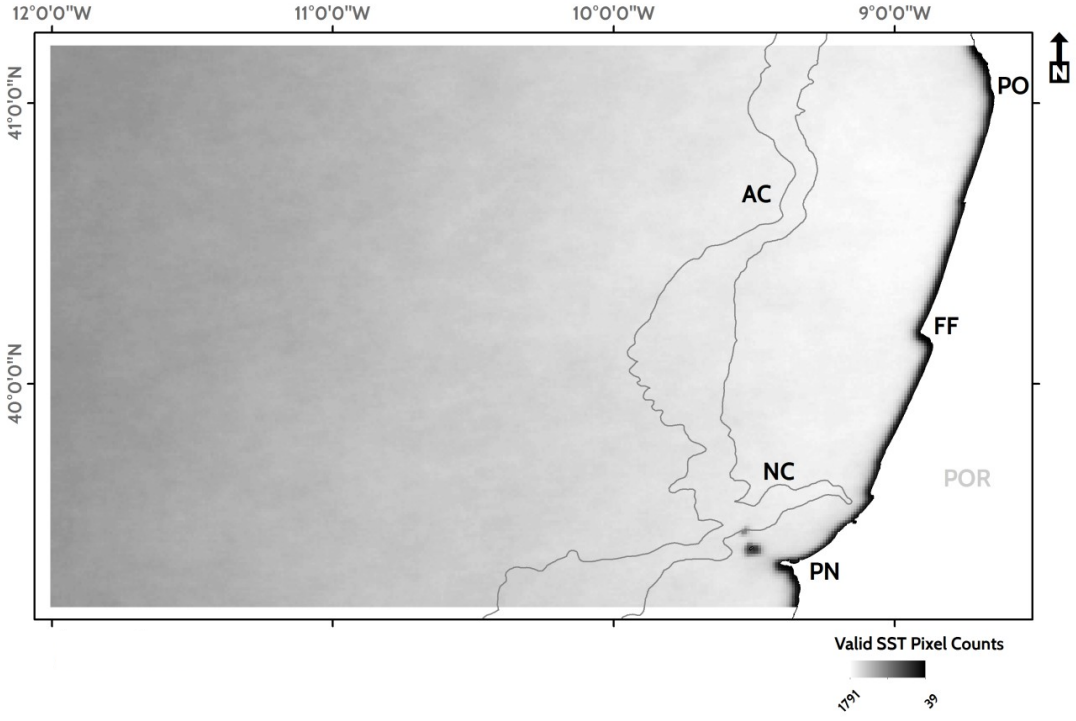


Figure 4.4.1.1. Total count of valid Sea Surface Temperature measurements per pixel, for the entire period of study (01/2003 – 12/2014). PN: Peniche; FF: Figueira da Foz; PO: Oporto; NC: Nazaré Canyon; AC: Aveiro Canyon; POR: Portugal. The 200 and 1600 m isobaths are represented as dark lines.

A total of 3450 images were used to build the SST time series (processed to Level 2), but the pixel with the highest valid SST count returned 1791 observations (*ca*. 52% of the potential maximum). Other pixels, especially near the coast (or the Berlengas Islands) reached values as low as 39. These are pixels that are located at the interface between land and ocean, including estuarine waters, bays or sectors between islands and capes, or that for other technical reasons fail to meet the quality control requirements.

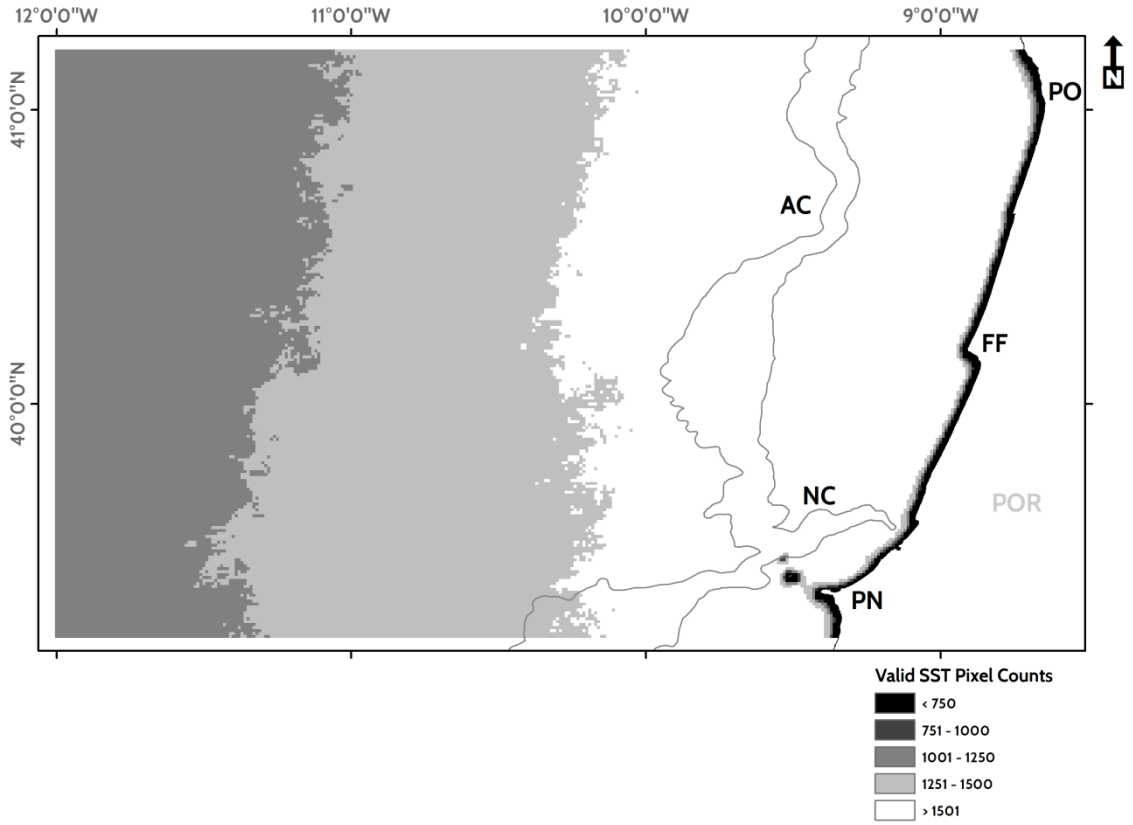


Figure 4.4.1.2. Classified total count of valid Sea Surface Temperature measurements per pixel, for the entire period of study (01/2003 – 12/2014). PN: Peniche; FF: Figueira da Foz; PO: Oporto; NC: Nazaré Canyon; AC: Aveiro Canyon; POR: Portugal.

The classified map of the total valid SST pixels (Figure 4.4.1.2) showed very clearly that these are distributed in a non-random fashion, with meridional bands of cells with similar retrieval probability. Near the shore, there is a band less than 5 km wide, where few or no data were successfully retrieved from MODIS imagery.

The analysis of the classified map provides an interesting depiction of data availability, showing how the pixels within the shelf (and beyond) fall within the 1500-1800 interval, thus guaranteeing that frontal boundaries there may be detected more efficiently than in other,

westward regions. The shelf and shelf edge regions are indeed the most relevant for the study considering the processes that occur there and economic relevance of the area.

The data gap affecting the littoral area is not negligible, especially in the vicinity of dynamic areas such as estuaries and capes. This gap may preclude the successful retrieval of SST values for the construction of time series and the detection of mesoscale features including thermal fronts.

The algorithm employed to detect fronts calls for the existence of contiguous SST pixels with a certain gradient. Therefore, isolated pixels are useless to that end. Still, they may be used to create time series, if the retrieved monthly fields are sufficient.

Overall, the pixel count data provides a useful tool to prepare the creation of a regular point grid that can be used to extract the values from raster cells. This is particularly important because, for several applications that will be described shortly, the information must be stored in tables or vector files.

The grid was designed to match the center of the 4-km pixels (Level 3) processed from the Level 2 SST files, and created at ArcGIS. This reduces the volume of data to be analyzed while guaranteeing each cell contains uninterrupted data that can be explored as a continuous time series. The Level 3 pixels are a spatial and temporal synthesis of the lower level data, as aforementioned, providing monthly (or lower frequency) SST fields. As such, data losses are mitigated through temporal and spatial averaging, although it must be kept in mind that the pixels in the vicinity of the coast and other underrepresented areas are incorporated into the grid.

Post-processing of the data tables enables the removal of inadequate cells, namely when the percentage of blanks is such that may compromise the accuracy of the final averaged estimate. Figure 4.4.1.3 displays the location of the grid cells used to extract the data from the different Level 3 products generated. Some cells appear to be located over land but are in fact only partially on land, and include data from the lower level (level 2) files (ocean-only), which are produced at a higher spatial (and temporal) resolution.

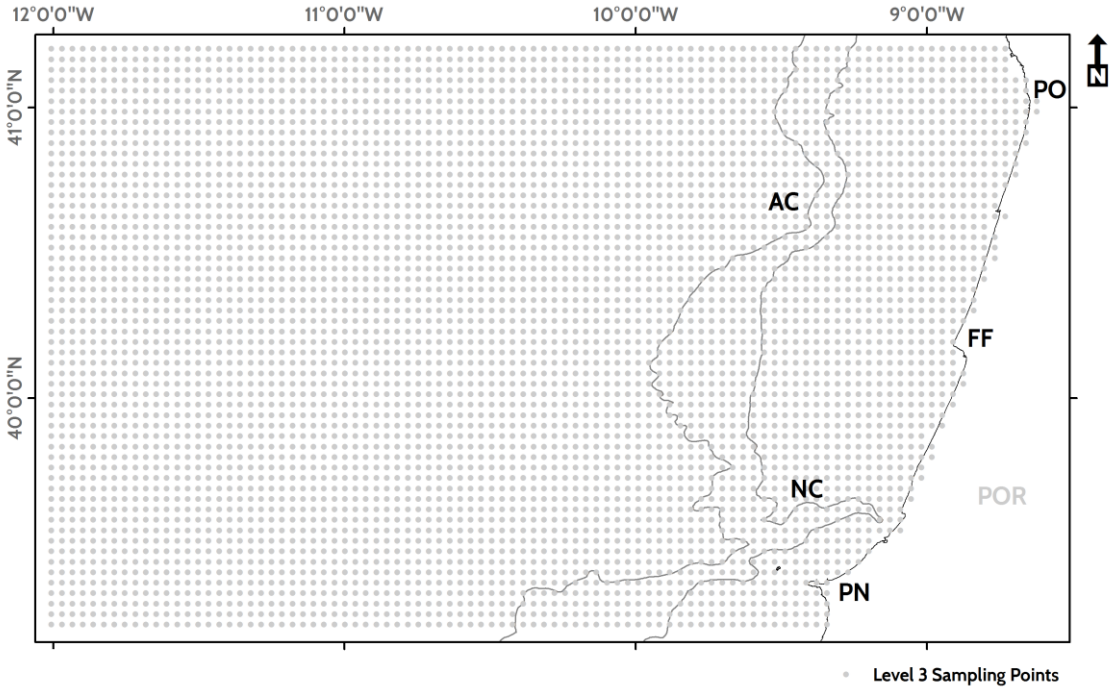


Figure 4.4.1.3. Regular grid used to extract data from the Level 3 files, including SST and Chlorophyll a concentration. Some pixels are located near the shore but all include data from the higher spatial resolution Level 2 files. PN: Peniche; FF: Figueira da Foz; PO: Oporto; NC: Nazaré Canyon; AC: Aveiro Canyon; POR: Portugal.

The regular grid contained 4871 points centered in the Level 3 pixels, each identified by a unique code.

Upon extraction of the SST and Chlorophyll a values, a database was created to store the information in preparation of subsequent analysis. Data tables were created in Microsoft Excel and archived according to the unique identifier assigned to each point. It was at this point of the processing chain that a first quality control was performed on the series and those failing to meet the requirements excluded from the analysis. The quality requirements included the successful retrieval of at least 90% of the time series without the absence of two sequential values. The evaluation was conducted semi-automatically using a python script that iterated through the data directories, counted the number of valid points and assigned a quality score to each. Missing values were assigned an identifier “NA” and, for valid pixels, these were interpolated using the software R (R Core Team, 2013) and the package ‘Zoo’ (Zeileis et al., 2005). Using the *na.approx* command, it is possible to iteratively interpolate all missing values, identified as “NA” from the data frame. In this case, the values were interpolated using the average of the neighboring values.

Zonal transects were designed to retrieve data from relevant regions where SST fields may differ as a consequence of different currents and processes at play at the local level along a longitudinal profile. The data were expected to detect the zonal variability in the temperature fields, as reported in previous studies (e.g. Peliz et al., 2002).

Figure 4.1.1.4 depicts the location of the zonal transect, which were designed to sample regions with different characteristics or in the vicinity of major bathymetric features (e.g. Nazaré and Aveiro Canyons).

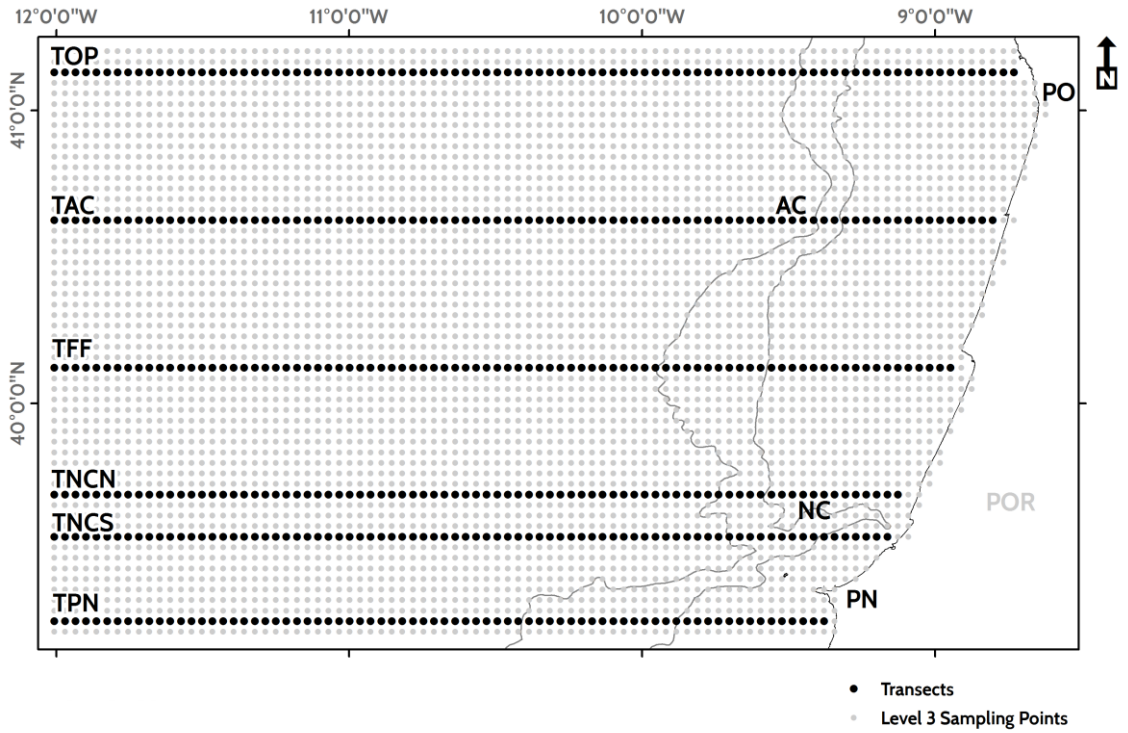


Figure 4.4.1.4. Zonal transects used to subset the regular grid used to extract data from the Level 3 files, including SST and Chlorophyll a concentration. TPN: Transect at the latitude of Peniche; TNCS: Transect southward of the axis of the Nazaré Canyon; TNCN: Transect northward of the axis of the Nazaré Canyon; TFF: transect at the latitude of Figueira da Foz; TAC: Transect at the latitude of the Aveiro Canyon; TOP: Transect at the latitude of Oporto. Please see text for more information. PN: Peniche; FF: Figueira da Foz; PO: Oporto; NC: Nazaré Canyon; AC: Aveiro Canyon; POR: Portugal.

The zonal variability can be seen in figure 4.4.4.5, which represents the time series of SST values extracted at the Easternmost and Westernmost points of the Figueira da Foz transect (TFF).

To assess the linearity of correlations between the coastal and offshore series, Pearson correlation coefficient (*r*) was calculated as follows:

$$r_{xy} = \frac{Cov(XY)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \quad \text{(Eq. 4.4.1.1)}$$

where X and Y are the variables under study (coastal and offshore SST).

For points P2453 and P2536, the value of *r* was of 0.88, significant for a $p < 0.01$. Still, and despite the linear correlation, which shows a similar and matched seasonal warming and cooling trend, the mean values are quite different. The overall mean SST for the offshore point is of 17.1 (2.54) ° C and of 15.8 (1.88) ° in the coastal pixel. One-way ANOVA revealed that, for $p<0.05$, the two set of values were significantly different. The difference is also significant when only the winter months are considered, despite the smaller difference in the values found for each point.

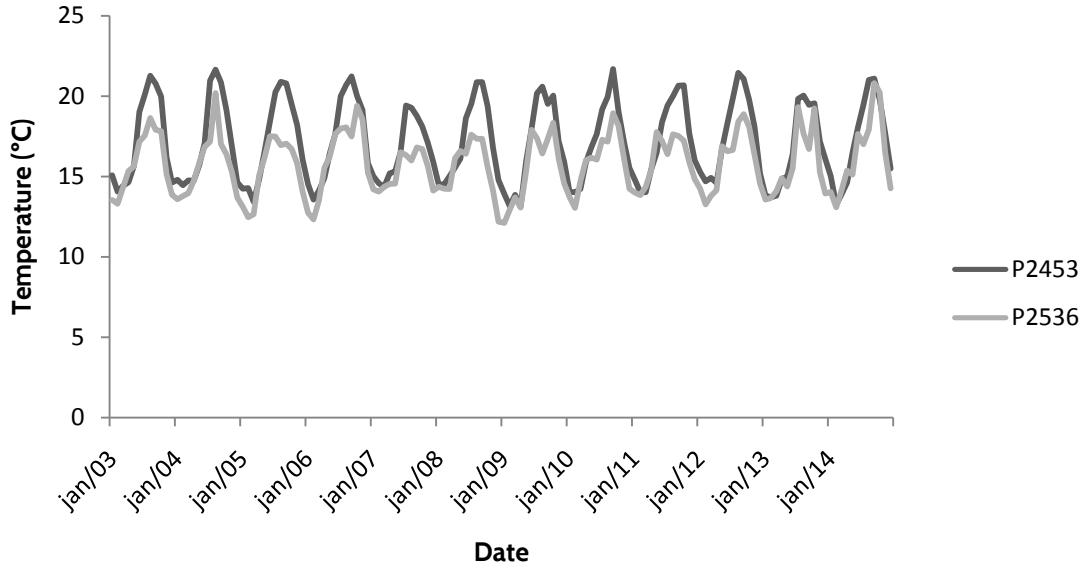


Figure 4.4.1.5. Plot depicting the monthly SST time series (01/2003-12/2014) of the westernmost (P2453) and Easternmost (P2536) points in the Figueira da Foz (TFF) transect. Notice the higher temperatures offshore during the summer.

The Upwelling Index, mentioned in an earlier chapter, is based on the temperature differences found between coastal and offshore points, as those depicted in the figure above. During the prime upwelling season (warmer months), the cold and nutrient rich waters form an initially narrow band near the coast. This upwelling signature is clearly visible in the imagery, and translates into a unique time series characterized by colder waters in the warm months (when compared with offshore pixels). On the other hand, the effect of the WIBP is less recognizable,

considering that the temperature value between coastal and offshore water is not sharp (despite statistically not similar, after One-way ANOVA, significant for a $p<0.05$).

The SST data offers important clues on the processes at play in the region. The onset of currents such as the IPC, the development of upwelling filaments or the interaction of currents and the formation of mesoscale features are all discernible in satellite imagery. Yet, and as aforementioned, these phenomena have strong subsurface signatures and are also vertically bounded by other currents and water masses with diverging traits. We shall now describe, very succinctly, the differences between the surface thermal signatures and the subsurface horizontal and vertical variability. This is a realm beyond the reach of satellite-borne optical sensors, but that can be at least partly studied by emerging technologies, such as satellite altimetry.

Figure 4.4.1.6 shows the location of a subset of 3 measurements, acquired in different seasons, by the same ARGO float. These measurements were acquired at the latitude of the Figueira da Foz Transect (TFF) albeit some were retrieved from locations at a short distance to the West of the boundary of the study area (12° W).

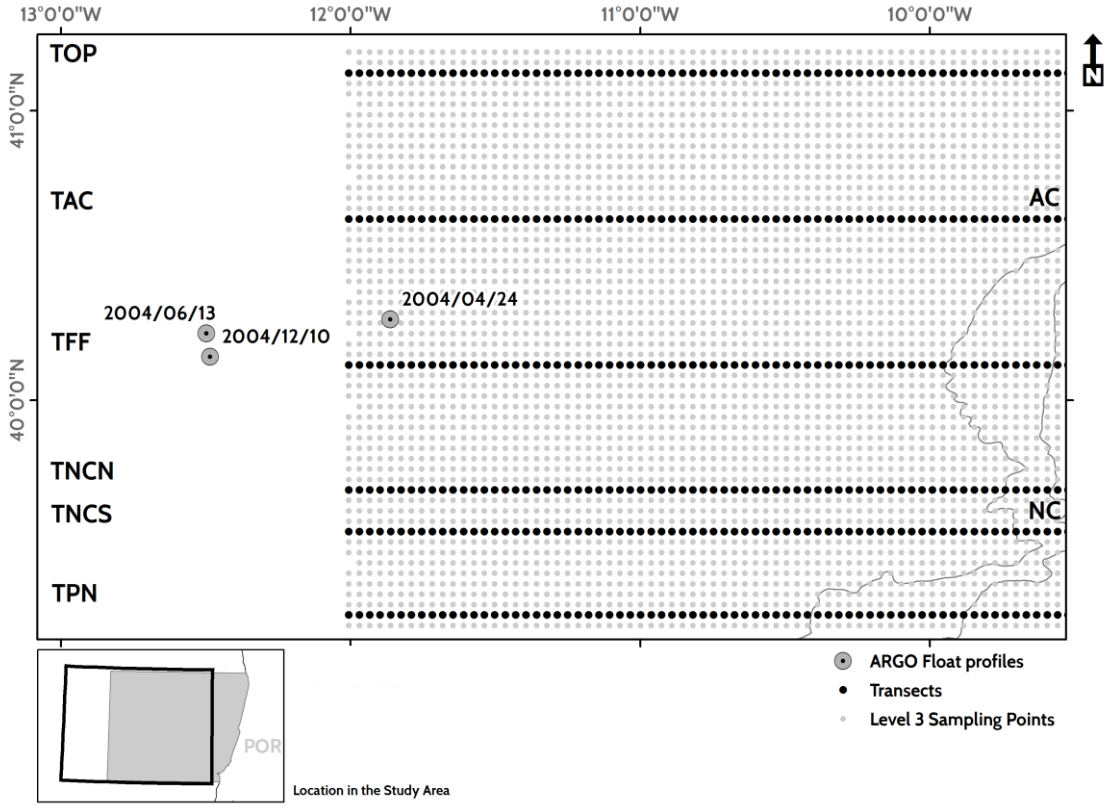


Figure 4.4.1.6. Location of three ARGO vertical profiles acquired in different seasons at approximately the latitude of the Figueira da Foz Transect (TFF). NC: Nazaré Canyon; AC: Aveiro Canyon.

The three vertical profiles are depicted in figure 4.4.1.7 and show the typical variability of the seasonal thermocline. The June 13th profile depicts a sharp decline in temperature over depth in the first tens of meters, characteristic of warmer months. On the other hand, the April profile displays the distinctive pattern of a deeper mixed layer (and thermocline).

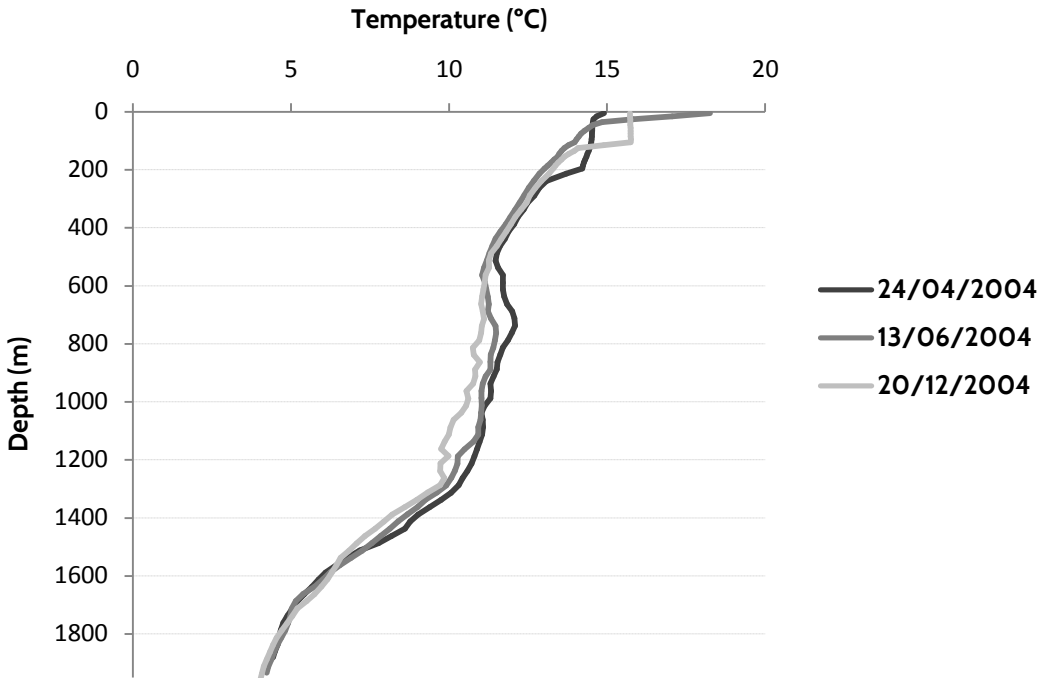


Figure 4.4.1.7. Plot depicting the three ARGO temperature profiles acquired in 2004 and projected in figure 4.4.1.6. The profiles are representative of the variability in the depth of the seasonal thermocline, and the highlight the need to consider the differences between SST and subsurface temperatures.

The depth and surface temperature of the mixed layer is of the utmost importance in satellite-based studies and its variability is what plots such as that of figure 4.4.1.7 depict (over a background given by currents and mesoscale features). It is also clearly noticeable in the profiles, the influence of Mediterranean Outflow Water (Bower et al., 2002) at the intermediate depths, between 600 and 1200 m, recognizable by the warmer temperatures recorded there.

Again, it is noteworthy to recommend the joint use of SST and *in situ* assets to understand the physical and ecological processes at play in the ocean. Otherwise, researchers are bound by a limited view of the 3D dynamic space that the ocean represents. The presence of Mediterranean Outflow Water can only be detected if vertical profilers are employed. MEDDIES, which also have a Mediterranean connection and are strongly influenced by

topography, can be detected by satellite altimeters, but that is one of several manifestations of the much broader subsurface movement of waters into the Atlantic (Serra et al., 2002).

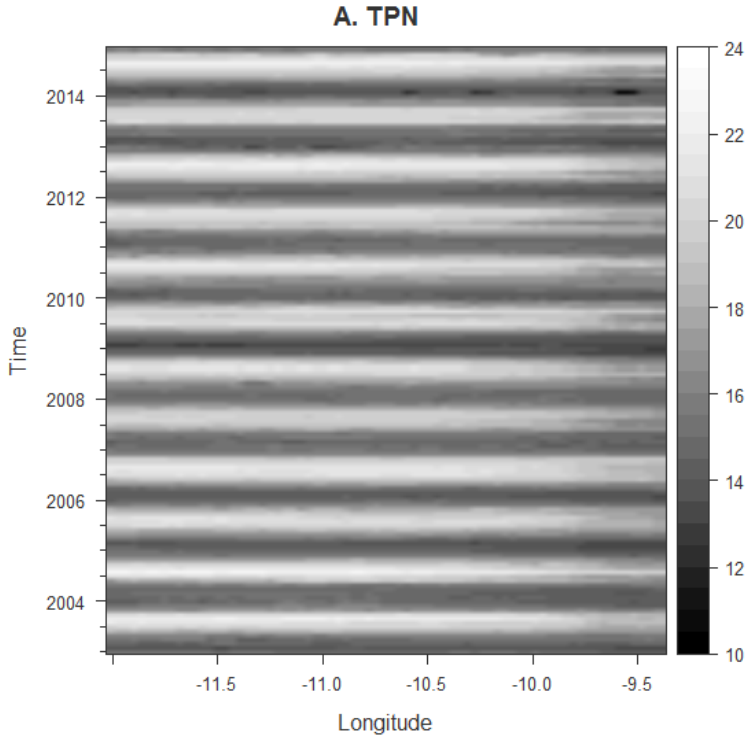
SST data can't reproduce the vertical complexity of the ocean, but the processes that drive this variability are often detectable through the surface expression, which is translated in specific patterns that can be interpreted from the imagery. SST time series are of particular interest considering that not only becomes possible to fill in the (sometimes numerous gaps) in the coverage but also reconstruct currents and the evolution of relevant features. Still, and considering the objectives of this work, which aims at characterizing thermal fronts retrieved from surface data, it is beyond the scope of the study to address these vertical profiles in greater detail. In fact, the ARGO profiles were used only when doubts rose regarding the interpretation of specific features. ARGO data were also used to provide additional information to supplement trajectory and temperature data from surface drifters. It is nonetheless relevant to emphasize the importance of the vertical structure of the oceans, to which optical sensors are mostly blind.

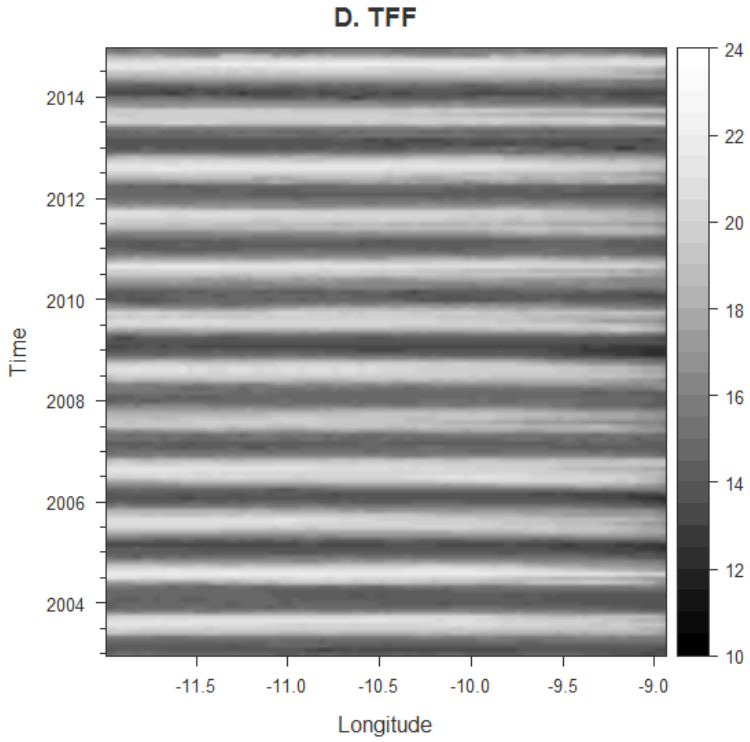
The very nature of the ARGO floats precludes an extensive – and systematic – use of the array within much of the study area. Depth limitations that bind ARGO floats to the open ocean, removes this set of data from the set of admissible sources to study the shelf and shelf edge region. Still, and considering previous studies (e.g. Relvas et al., 2007), it is exactly where data coverage is in deficit where fronts are predominant and our study is therefore focused.

Because the study area is characterized by a high degree of variability, the transects were analyzed in greater detail in order to search for: i) seasonal variability; ii) spatial (zonal and meridional) variability, and iii) trends in the time series.

To that end, the data contained in the pixels constituting each transect were extracted and processed using R. The libraries "rasterVis" (Lamigueiro et al., 2014) and "Zoo" (Zeileis et al., 2005) were used to fulfill the data analysis requirements. In particular, the data contained in the geodatabase were exported into csv files prior to being imported into R. Hövmoller plots were then generated for each transect. These plots offer a simple yet robust tool to visualize long time series when one of the coordinates (x or y) is constant. The zonally aligned transects provided an opportunity for the use of these plots, which we will now describe.

Figure 4.4.1.8 depicts the series of Hövmoller plots generated for all the transects.


A. TPN
Time
2014
2012
2010
2008
2006
2004
-11.5
-11.0
-10.5
-10.0
-9.5
Longitude
24
22
20
18
16
14
12
10


D. TFF
Time
2014
2012
2010
2008
2006
2004
-11.5
-11.0
-10.5
-10.0
-9.5
-9.0
Longitude
24
22
20
18
16
14
12
10

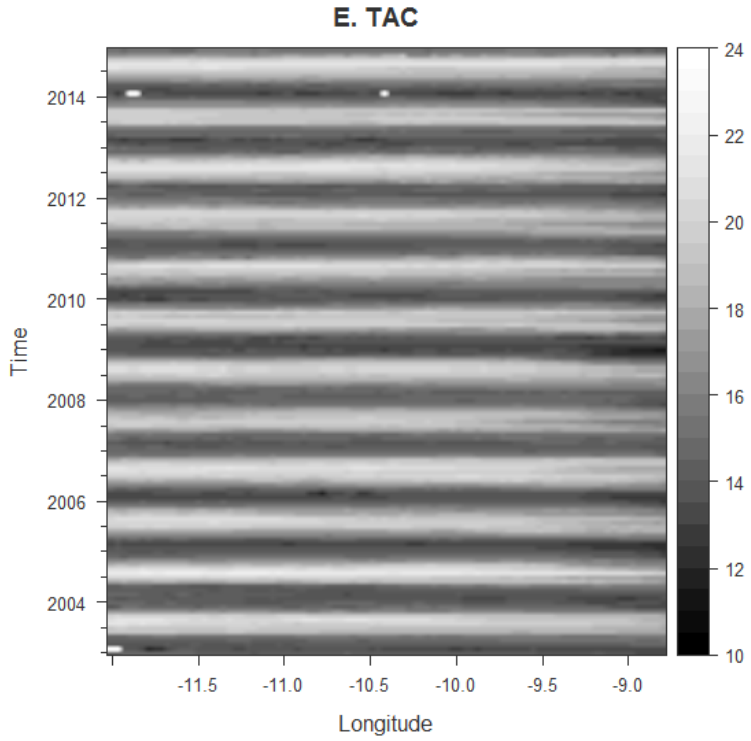


Figure 4.4.1.8. Selected SST Hövmoller plots of several transects depicted in figure 4.4.1.4. A: TPN; D: TFF; E: TAC.

The SST Hövmoller plots highlight how the processes at play off Portugal vary with latitude and longitude. These plots highlight the continuous nature of the changes operated over the SST fields in both time and space. In particular, it is evident a coastward displacement of higher SST values during the warmer months. This trend progresses northward but it is interrupted in the Figueira da Foz Transect (TFF). In this region, an inversion of the trend is detectable, with colder waters expanding westward, on par with the lower temperatures characteristic of higher latitudes.

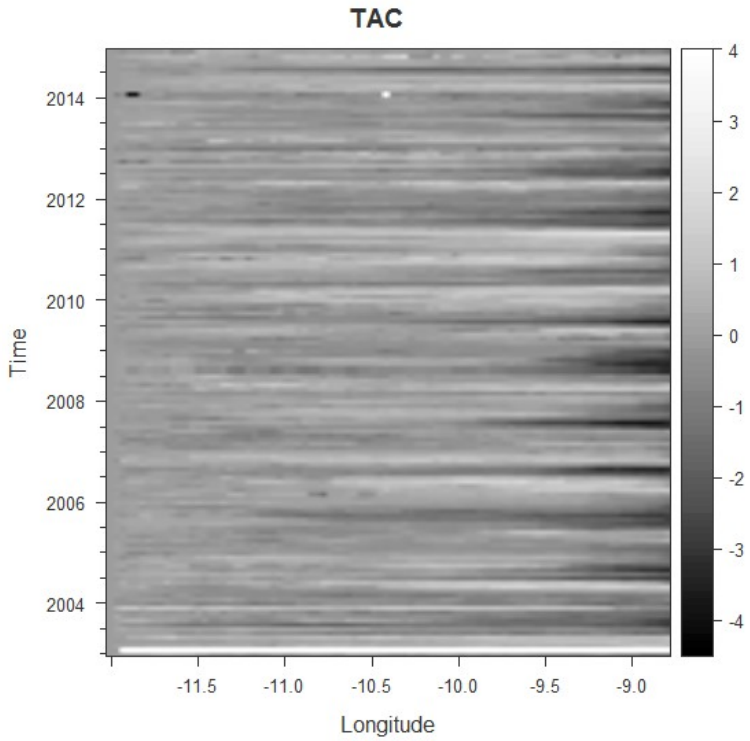
Considering the angle at which the coast is aligned, parts of the transects located at the same longitude may actually be affected by different processes considering the different distances at which they lay from the coast. The residual presence of the WIBP in the warm months may also help explaining some of the differences observed in the transects, which are likely to be more significant northward of the Mondego river. Still, the development of upwelling cells in the Portuguese coast is heterogeneous. As highlighted by Relvas et al. (2007), in the vicinity of Cape Carvoeiro (especially to the South of this feature) a prominent upwelling cell is present, a conclusion drawn from the analysis of SST anomaly and frontal probability maps. The SST

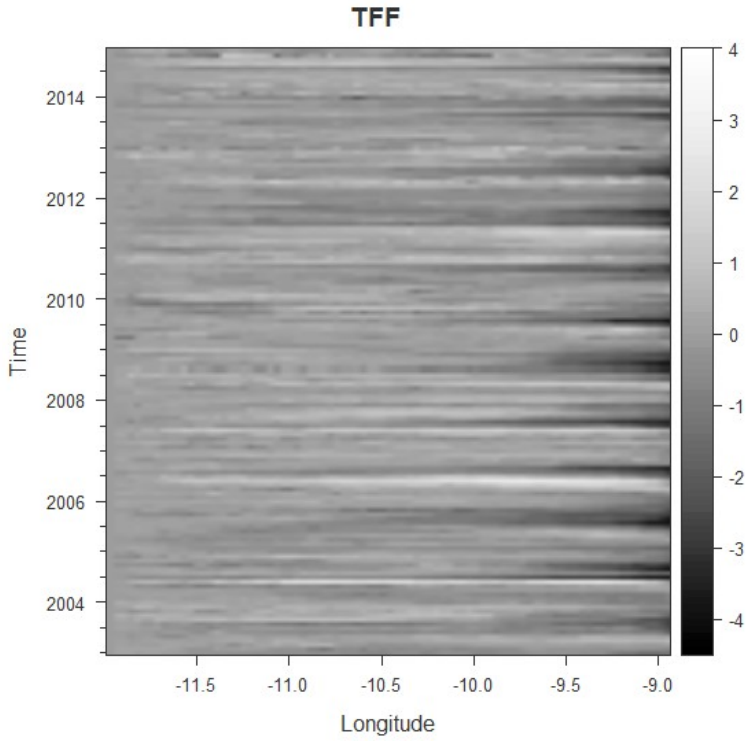
Hövmoller plot (A: TPN) supports this hypothesis, showing a westward distribution of lower SST values to the South of the Cape. Transect TNCS, just north of the geographic feature, emphasizes the distribution of warmer waters closer to the shore, suggesting a weaker upwelling pattern.

Further North, at the latitude of Aveiro (transect TAC), the influence of the recurrent upwelling filaments is also expressed. Albeit less prominent than in the satellite imagery, where the spatial structure of the feature is clearly discernable, the SST fields here are consistent with strong upwelling activity. This activity is no longer confined to a narrow strip near the shore but expressed as a strong advection of colder waters several hundreds of kilometers towards the ocean.

To understand whether the apparent differences are statistically meaningful, the three points closest to the shore in the TPN, TNCS and TFF transects were averaged to generate monthly fields (12). Using one-way ANOVA, no significant differences were found for this set, for a $p < 0.01$. The intra-annual variability and small variability between the absolute SST values may be responsible for this apparent lack of differences.

Even when the UI_{SST} is calculated, significant differences ($p < 0.01$) are not found for the aforementioned transects. Figure 4.4.1.9 depicts the Hövmoller plot of the UI_{SST} calculated for transect TFF. The duration and westward expansion of the upwelling band (and filaments) seems to vary, as expected, despite the overall cyclic nature of the season. The negative values of the UI_{SST} in the cold months may be due to the effect of the WIBP. The plume in the winter generates a thin lens of waters that are colder than those at the shelf edge. On the other hand, in subsequent months, including in early spring, the effect of the WIBP, which is still pronounced in rainy years, introduces a positive anomaly that lasts throughout the season and is only attenuated by the onset of upwelling. This effect becomes more pronounced in the northern transects (TFF, TAC, and, especially, TOP).


TAC
Time
2014
2012
2010
2008
2006
2004
-11.5
-11.0
-10.5
-10.0
-9.5
-9.0
Longitude
4
3
2
1
0
-1
-2
-3
-4


TFF
Time
2014
2012
2010
2008
2006
2004
-11.5
-11.0
-10.5
-10.0
-9.5
-9.0
Longitude
4
3
2
1
0
-1
-2
-3
-4

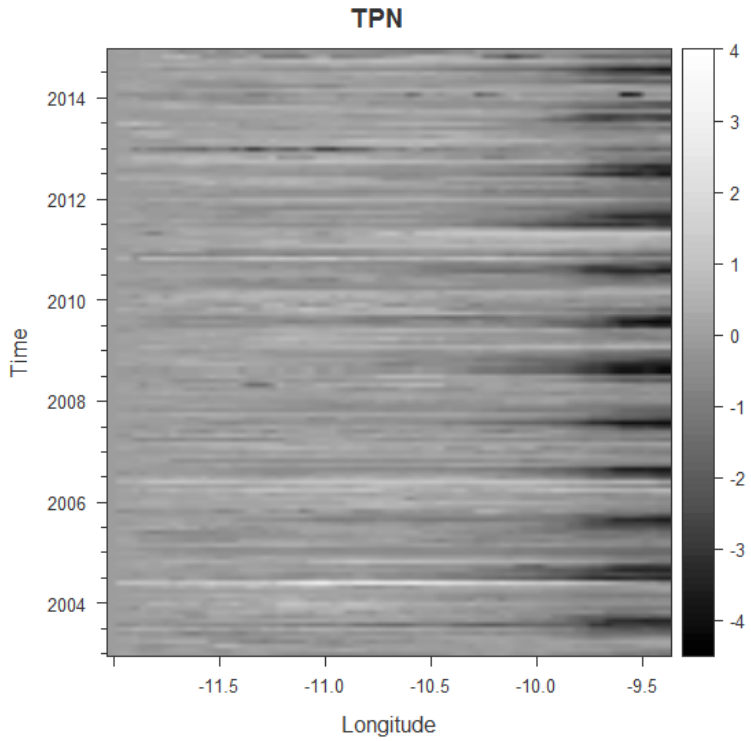


Figure 4.4.1.9. Hövmoller plot of the Upwelling Index for the TAC, TFF, and TPN transects. Notice the positive pulses just before the onset of the upwelling season.

In fact Figure 4.4.1.9 shows how the positive pulse before the upwelling season is much more prominent in the TFF transect than what was measured in southern transects (e.g. TPN). If the WIBP was not at play, the pulse was expected to be more subtle, considering the overall latitudinal cooling of the waters in the study area. However, this potential connection cannot be explained by the SST and UI_{SST} alone.

When the UI_{SST} is compared against monthly river discharge at Coimbra's "Açude" dam, no statistically significant correlation was found either with the full set (all 144 months of data) or the subsets Jan/Feb/Mar and Apr/May. The river discharge at the dam was used as a proxy for river outflow at the mouth, because no better data is available at this scale. As such, resorting to river discharge information is not sufficient to forecast SST-based gradients outside the primary upwelling season, as expected. Integrating additional data, including wind, air temperature, currents, and bottom topography is likely to be necessary to develop a complete image of the variability of UI_{SST} in the nearshore environment.

In the present study, no attempts were made to generate a model of the WIBP or other features, based on the aforementioned data. However, the basis for a future system capable of delivering data to feed such models was laid, as we will describe in subsequent sections.

The presence of the WIBP, and its complexity, can be easily appreciated in figure 4.4.1.10, which compares the plume in 2003 and 2008 using Landsat-7 ETM+ and MODIS AQUA data.

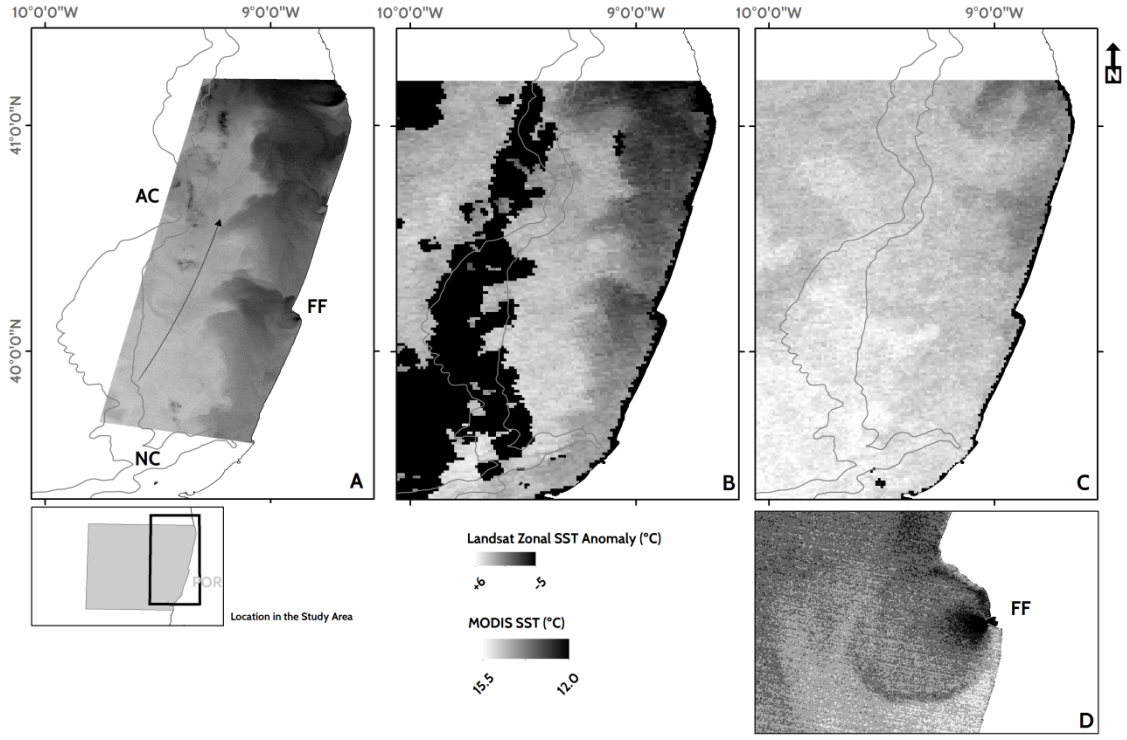


Figure 4.4.1.10. The Western Iberia Buoyant Plume as seen in satellite SST imagery. A: Landsat Zonal SST Anomaly in January 24, 2003, calculated as the temperature difference between the coastal and a central reference area off Figueira da Foz (Path 204 Row 032). The black arrow indicates the direction of the flow. B: MODIS AQUA SST image acquired in January 24, 2003, for reference purposes. Notice the lack of detail when compared with the Landsat-7 match-up. Black pixels were flagged (clouds, land contamination, etc…) C: MODIS AQUA SST image of January 27, 2008. Notice the lack of a well developed WIBP. D: Subset of image A, showing the river discharge plume near the Mondego river estuary. Please see text for more information.

Landsat-7 ETM+ (and subsequently Landst-8 OLI) imagery can provide a detailed view of the mesoscale and sub-mesoscale features occurring in shelf waters. The WIBP was successfully identified in the imagery, through a gradient method in which a pseudo-zonal anomaly was calculated. Considering the narrow oceanic band imaged by the ETM+ sensor, a significant portion of the WIBP off Oporto is found beyond the western edge of the scene. This would preclude the calculation of an accurate true zonal anomaly. Instead a reference area off Figueira da Foz was used. The remaining pixels were then compared against the average SST value for the reference box. An alternative would be to retrieve the reference values from the

MODIS match-up, but that is not always possible given the frequent cloud cover further offshore, as it was the case in this image.

In the Landsat scene, two aspects are worth mentioning. Firstly, the WIBP is developed along the entire section of the coast imaged by the sensor. The westward development of the plume varies, reaching less than 30 km at the latitude of Figueira da Foz while being advected more than 60 km off Oporto. Although no *in situ* data is available, the comparison of successive MODIS AQUA imagery suggested the Plume was generally being displaced northward and especially further offshore. Advection eastward progressed at high velocity with zonal (and meridional) a displacement of approximately 0.1 $m.s^{-1}$. Interestingly, the velocity is similar to those reported for the IPC (Huthnance et al., 2002). Future research may seek to integrate drifters with satellite data for plume velocity and direction characterization.

The matchup of MODIS-like sensors with Landsat or equivalent systems is of the utmost importance. The lack of spatial resolution on MODIS products is offset by Landsat imagery, which, nonetheless, lacks the frequency provided by wide swath sensors. The WIBP is a good example of an oceanic feature that is best studied using such combination of data. The rapid changes to which is subject calls for high frequency acquisitions, but the intricate patterns generated within the plume and especially at the interface with other water masses (e.g. IPC) require higher resolution sensors.

Figure 4.4.1.10.D shows one such complex pattern, at the mouth of the Mondego river estuary. A river discharge plume is often observable in satellite imagery, including SST and multispectral data. The plume is extremely relevant for a number of ecological, hydrodynamic, and ecological applications (Cunha et al., 2002). In the context of our study, the Mondego river plume provides a source of thermally contrasting waters that contribute to feeding the WIBP. Throughout the study period, the expansion of the northern breakwater protecting the Figueira da Foz harbor, contributed to a significant change in the plume's shape and direction. The new asymmetrical nature of the breakwater pair generates a southward facing plume front and modifies the recirculation patterns as well. This is beyond the scope of the project, but it has the potential to change frontal dynamics and average spatial distribution in the area.

Figure 4.4.1.11 depicts the January 24, 2003 tidal plume, which was generated in a period of unstable winds and strong river flow.

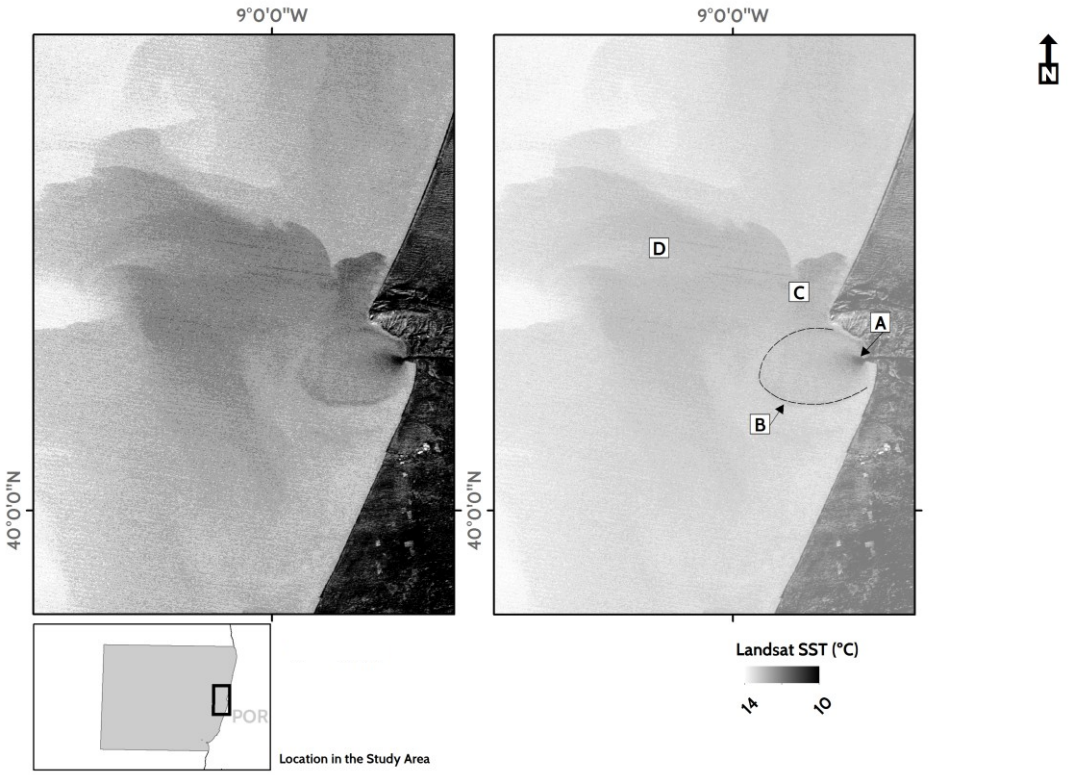


Figure 4.4.1.11.Mondego river plume and the WIBP. Left panel: the river plume as seen in Landsat-8 ETM+ SST image of January 24, 2003. Right panel: Annotated SST image, showing the (A) source, (B) tidal front, (C) recirculation area, and (D) the far-field. Please see text for explanation. The SST scale refers to the left panel. The image in the right panel was subdued to simplify the superposition of markers.

The tidal plume generates a sharp elliptical thermal front (akin to a termination shock) where the along-front temperature is lower than that measured within the halo created by the outflow. The plume is restricted to the upper layers and subject to intense shearing. The frontal area's hypothesized increased thickness is likely due to convergence phenomena (in the zonal plane) that characterizes typical frontogenesis processes in these settings.

The extension of the northern pier at the Figueira da Foz harbor led to changes in the shape and area of direct of influence of the plume. Still the description and analysis that will follow is applicable to both pre- and post-extension scenarios, despite specific differences beyond the scope of the study.

The bar-built estuary (Fairbridge, 1980) of the Mondego river is divided into two arms (north and south), each with different hydrodynamic and ecological conditions (Marques et al., 2002). The estuary is subject to a mesotidal regime, with maximum levels exceeding 3 m (Cunha et al., 2002). Data acquired from SNIRH (2016), demonstrated that the river flow possesses a

strong seasonality but also inter-annual variability, not even river control policies and infrastructures were able to eliminate entirely.

In this context, the Mondego river plume fits into the model described by Horner-Devine et al. (2009). This model, itself an evolution of earlier work by Garvine (1982), provides an interesting and replicable framework to classify river discharge plumes.

The plume is generated by the variable flow that results of the combined action of upstream dam discharges and river runoff. Upon reaching the estuary, the proximal source of the plume, the feature is modulated by tidal action. The plume then interacts with coastal waters, which may either belong to the cold freshwater buoyant lens created by the cumulative outflow of WIP rivers or, further West, warmer and more saline shelf waters. Several intermediate stages are recognizable in satellite imagery and, considering the thermal gradients produced, may play a role in structuring the distribution of frontal boundaries (thermal, salinity, and ocean color) in the region.

On ebb tide, the source water propagates westward, generating a sharp and roughly elliptical front characterized by colder waters, which can be located as far as 10 km away from the river mouth. The data available precludes us from calculating the westward velocity, or establish the exact sequence of steps leading to frontal development, which only a high frequency sensor or *in situ* assets could provide.

Still according to the conceptual model and other works by Horner-Devine et al. (2009, 2006), the water from the tidal front then recirculates (Figure 4.1.1.11.C) under the influence of Earth's rotation and river discharge, with a movement proportional to the inertial radius. Both masses are separated by the tidal front, with the recent discharge probably overlapping, at least partially, the 'older' more mixed (and saline) water. This is a dynamic process in which the strength of the tide and the river flow work together to influence the salt-wedge and buoyant freshwater plume generated in the estuary (Cunha et al., 2002).

As the plume continues to interact with coastal waters, it becomes subject to Earth's rotation, currents and wind, and seemingly, river momentum is no longer relevant. At this point, the plume is considered to have entered the far-field. Nonetheless, and as shown by figure 4.1.1.10, whenever river discharge is less prominent, the river plume – and the WIBP – become much less conspicuous, and restricted to the pixels nearest to the shore.

The relevance of studying river plumes, and the fronts generated by these features cannot be overstated. River plumes and the frontal boundaries they generate are sought after by marine life as a recurrent feeding ground (Zamon et al., 2014). This is the consequence of a productivity enhancement and overall increase in the availability of prey (Morgan et al., 2005). Satellite data can provide a synoptic view of river plumes off the Portuguese coast. Still, further

research is required to determine the exact structure and variability of the river plume, in its different stages off the Mondego estuary, along with the real impact of the features over biological communities.

We conclude by emphasizing the fact that the second highest river flow of study period was recorded at the Coimbra “Açude” dam in January 2003. The images are thus particularly clear and representative of the plume dynamics and frontal boundaries generated.

Once more, the relevance of integrating data from different EO and *in situ* systems is highlighted, especially during the development of models that can then be used to interpret satellite imagery.

Other relevant features in the study area are the upwelling filaments anchored to the Aveiro Canyon (herein named the Aveiro Filament), which can have a significant expression oceanward.

The literature mentions the Aveiro filament recurrently, and the sequence presented in figure 4.4.1.12 demonstrates that it is indeed a frequent feature in the region, likely to have a significant impact in local transport and frontogenesis.

The analysis of SST imagery throughout the study period revealed that the large filament occurs predominantly between July and September, but the size and exact location have a certain degree of inter-annual variability, and may occur whenever northerly winds promote upwelling events of sufficient duration and intensity. Previous studies related the wind maxima and SST minima, despite the 1-day lag verified (Oliveira et al., 2009). This lag is, nonetheless, meaningless at the temporal scales under analysis in this work.

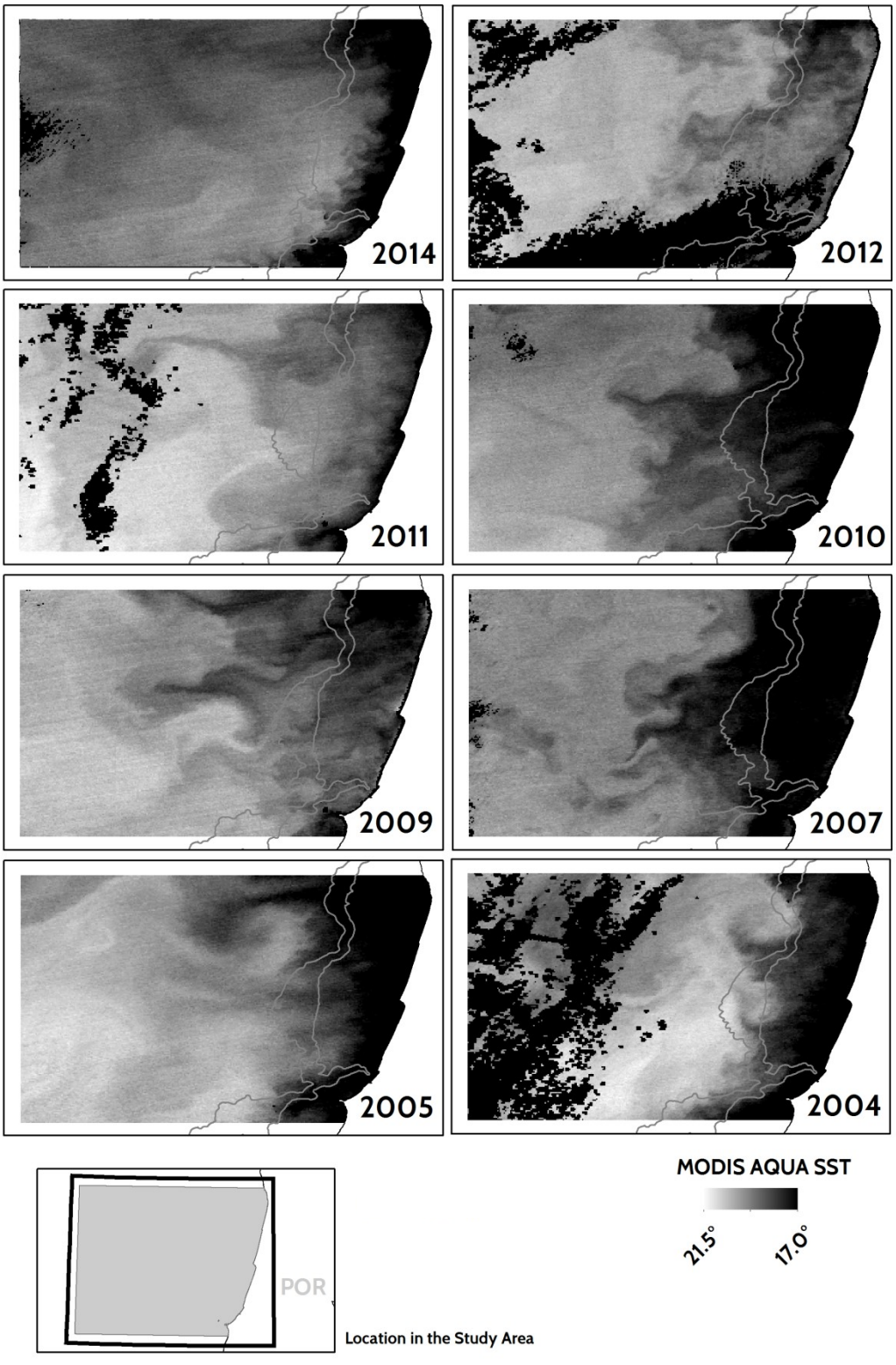


Figure 4.4.1.12. (A)

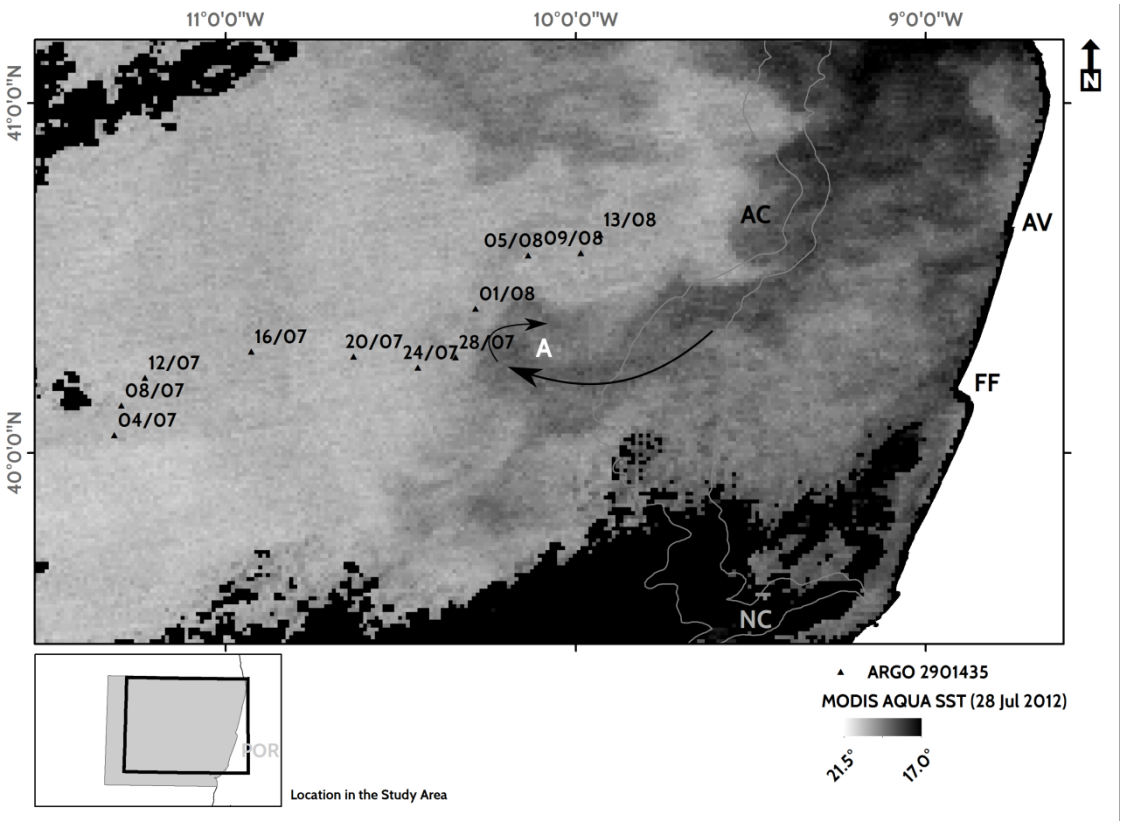


Figure 4.4.1.12. (B)

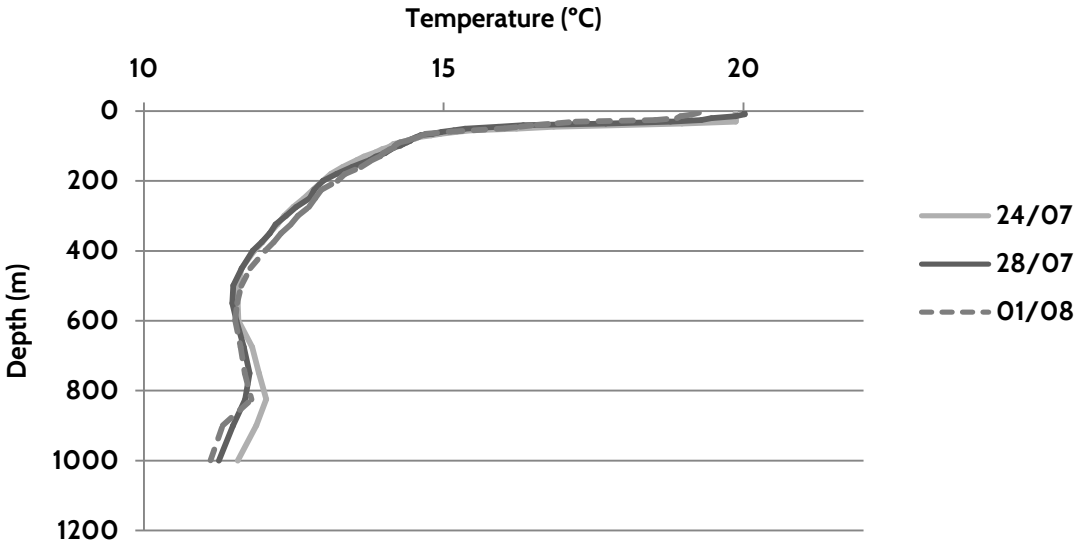
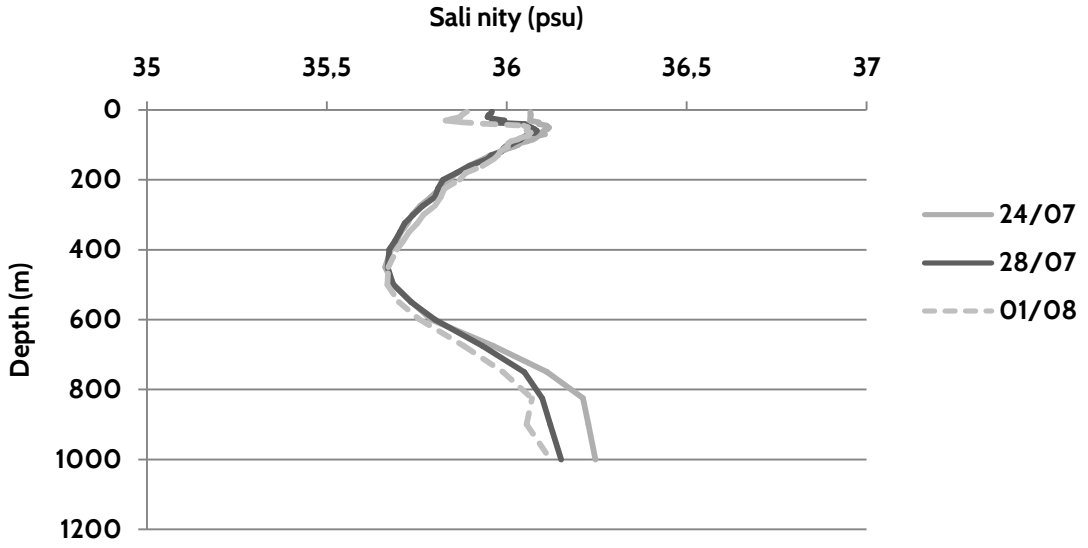


Figure 4.4.1.12. (C)



(D)

Figure 4.4.1.12. Upwelling filaments off central Portugal. A: Panel depicting the recurrent large filament detected in the vicinity of the Aveiro Canyon (2014/07/11; 2012/07/28; 2011/08/27; 2010/08/15; 2009/08/30; 2007/08/30; 2005/08/21; 2004/09/16). B: Detailed view of a large filament (Aveiro Canyon) detected in imagery of July 2012. A coincident ARGO track and profiles are available and depicted. This image is further discussed in latter chapters. Black arrows show flow direction reconstructed from time series. A: Anticyclonic eddy; AC: Aveiro Canyon; NC: Nazaré Canyon; FF: Figueira da Foz; AV: Aveiro. C & D: ARGO temperature and salinity profiles (platform ID: 2901435, July and August 2012), showing differences between three points in the vicinity of the filament. Notice the smaller influence of Mediterranean Water in the later profile and a lower temperature and salinity in the upper layers.

In fact the Aveiro Filament is not always anchored to the Aveiro Canyon but it can occur to the North, roughly at the latitude of Oporto. Both filaments may actually occur simultaneously, as seen in the SST image of 2009.

The filaments are generally associated to bathimetric and topographic features (Haynes et al., 1993) and, therefore, anchoring points are rather recurrent (Cordeiro et al., 2015). Still, wind forcing plays an essential role in determining not only the location but also the dimensions of the structure (Cordeiro et al., 2015). Because Cordeito et al. (2015) conducted an exhaustive survey of filaments, which included the study area, we shall not attempt to determine the source of the filaments. Instead, we focused in identifying the features and determine the

correlation between the filaments and thermal fronts. This effort also included the smaller, but pervasive meanders, as seen throughout the coast in the image of 2004 in figure 4.1.1.12.

Despite our purpose, the analysis of the SST time series still yields interesting data that may be explored in the future. Highlights include the similar shape and evolution (despite the different ordinal date) of the filaments (e.g. 2007/2010/2012) or the dimensions of the eddy fields (cyclonic and anticyclonic) generated at the thermal front between upwelled and coastal waters (e.g. 2005).

Figure 4.1.1.12.B shows an interesting example of the interaction between the upwelling filament and coastal waters. An ARGO float (platform 2901435) drifts Eastward towards a filament that developed in July 2012. The ARGO trajectory is subsurface, emerging only at the end of the cycle (which lasts 10 days) for data broadcasting. The float's trajectory was diverted by the westward-moving filament. ARGO floats submerge towards a target depth of approximately 1000 m and drift until diving further to reach the profiling depth. This strongly suggests that the flow associated to the filament has a subsurface expression, capable of diverting the float at some point. This may be in apparent contradiction to the conclusions by Relvas et al. (2007), which suggest that the filaments are restricted to a thin surface layer no more than 100 m deep.

The Temperature-Salinity profiles, on the other hand, provide interesting information considering that – at the surface – the float seemed to do little more than graze the advancing front of the filament. If at the surface, the sharp thermocline is likely due to surface warming, a deeper thermohaline front is suggested. This deeper front, located between the isopycnal 26 and 27 separates the less saline and colder waters presumably under the influence of the filament from the coastal warmer and saltier waters (Figure 4.4.1.13, plotted with the OCE package for R). This is in line with findings by Flament et al. (2000) in the California system, where similar jets occur during the upwelling season. However, caution must be applied in the interpretation of these data, considering that the front is probably strongly divergent and therefore the ARGO float is unlikely to have penetrated entirely into the filament's core. Additional data, from CTD for instance, would prove invaluable to further characterize this important upwelling-related structure.

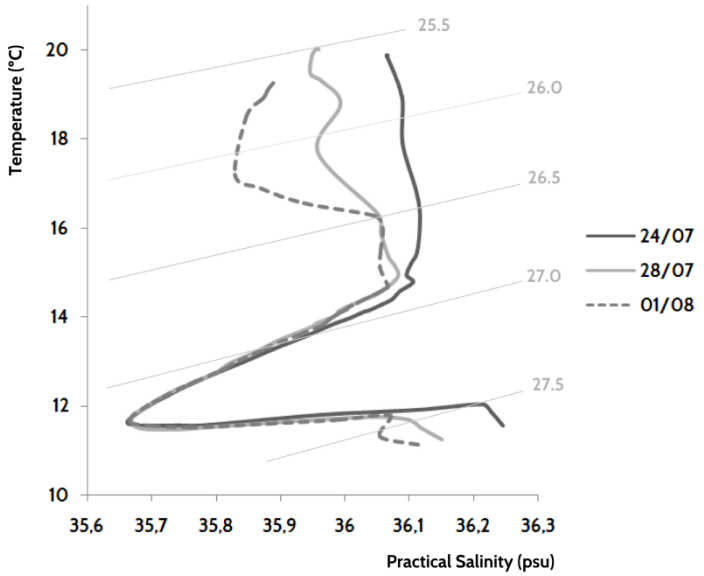


Figure 4.4.1.13. Profiles acquired by an ARGO float (platform 2901435) and represented as a T-S diagram with overlaid isopycnals (grey solid lines).

The profiles are extremely relevant, as they provide information unavailable to satellite-borne sensors and future studies should attempt to acquire profiles both inside and outside the filament's front, considering the variable thickness of the coastal features (Flament et al., 2000). Furthermore, the subsurface expression of fronts is of the utmost importance for research and applications, from an ecological and economic perspective alike.

The seasonal variability can, even so, be characterized by the integration of SST imagery and data for the development of a synoptic view of the costal dynamics.

The different zonal transects, along with the imagery provide a wealth of information, which greatly exceeds the scope of this work. The differences in the UI_{SST}, for instance, are of great relevance to ecological studies and important source of information on the dominant flows and intra- and inter-annual variability of upwelling. Inherent variability in the features, together with emerging patterns caused by global climate changes (Casey et al., 2001, Relvas et al., 2009), can disrupt important processes and influence the recruitment of species or their spatial distribution (Santos et al., 2004).

Figure 4.4.1.14 compares the average monthly UI_{SST} values throughout the year, highlighting the influence of different oceanographic and meteorological phenomena. The plethora of different currents, water masses, and transient phenomena such as river discharge pulses, influence the SST values estimated by the sensors. Nonetheless, a recurrent pattern was discernable in the imagery as depicted below, which shows the seasonal variability. This variability can be described as an alternation between an upwelling regime (negative UI_{SST}

values) and a poleward flow to which important riverine inputs are added nearshore. The simple model is then disrupted by off-season events, such as the somewhat frequent upwelling episodes occurring during winter (Santos et al., 2001).

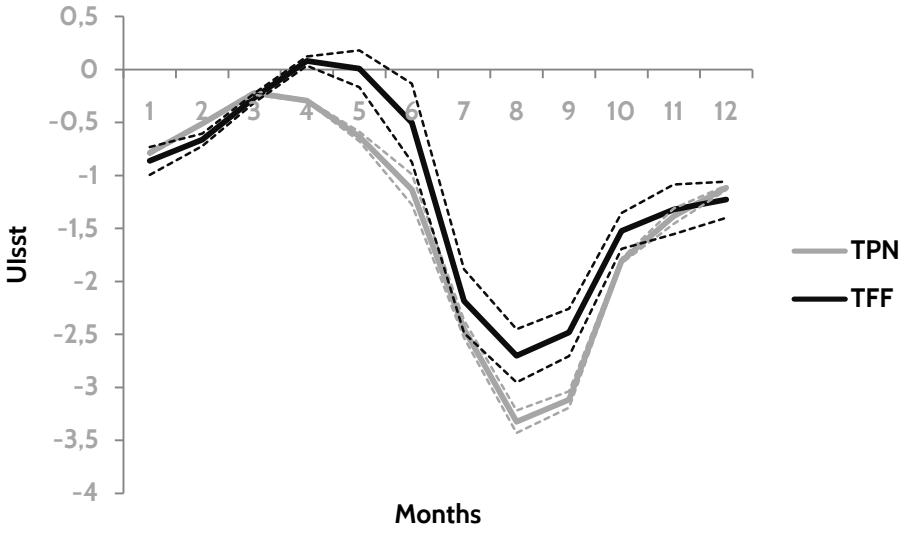


Figure 4.4.1.14. Comparison of the time-averaged Upwelling Index for the three points closest to the shore in the TPN and TFF transects. The minimum (maximum upwelling) value is common for both areas and occurs in August. The highest value occurs in March for TPN and April in the case of TFF.

Despite the size of the study area, the data presented thus far highlights the spatial heterogeneity that characterizes it. The heterogeneity may in fact represent a temporal mismatch of the processes in different areas. It is noteworthy that the time of occurrence of minimum UI_{SST} values are common to both transects, yet, the timing of maximum values differ. If for the Figueira da Foz (TFF) transect the maximum occurs in April/May, at TPN (Peniche), the highest values were in March. These transects were selected due to their representativeness and location. TPN, clearly different from the remainder and the TFF, located in a central region within the study area and known to be influenced by a multitude of features we have described in this document (e.g. WIBP, IPC, coastal counter-currents, upwelling).

In fact, a meridional transect can highlight such differences. Figure 4.4.1.15 depicts the location of the transect, which includes all pixels located along a line at 9.45° W. The points, because they are located at different distances from the coast, are influenced by different features. While a point nearshore may be influenced by the WIBP, another may return warmer values because it is located within the warm tongue of the IPC. Albeit the caution such a

transect calls for, it is still illustrative of the barrier located in the region between the axis of the Nazaré Canyon and Cape Carvoeiro.

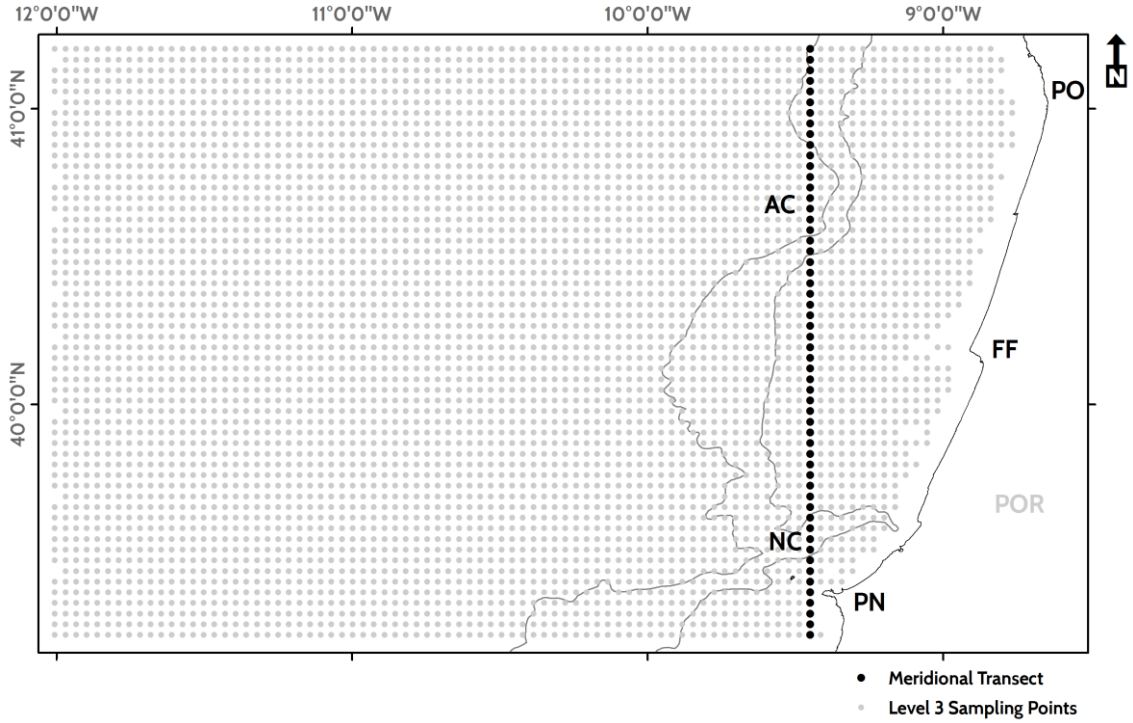


Figure 4.4.1.15. Map depicting the location of the meridional transect, located along longitude -9.45°. PN: Peniche; FF: Figueira da Foz; PO: Oporto; NC: Nazaré Canyon; AC: Aveiro Canyon; POR: Portugal.

The SST Hövmoller plot depicted in figure 4.4.1.16 represents the zonal variability along the meridional transect. In the plot, the SST values for each latitude (constant longitude) are represented over time. In it, the time series values abruptly change southward of latitude 39.4° N, particularly in the warm months. Although the topography could suggest a shadow-zone, where upwelling would be curtailed by slower winds blocked by the Peniche Peninsula, the data suggests quite the opposite.

When the information from the different transects (zonal and meridional) are combined, it is clear that the area southward of the Peniche-Berlengas line is the most dynamic area within the study region in terms of upwelling or alternatively where its signature is clearest. The abrupt differences cannot be solely explained by the increasing distance of the points to the coast, because the zonal transects suggest a similar scenario. The proximity to the Nazaré Canyon, together with the topographic constraints imposed by the Berlengas Islands,

motivates a strong and complex flow. This leads to the development of mesoscale eddies leeward of the Cape and an impressive cold-water signature typical of a strong upwelling cell.

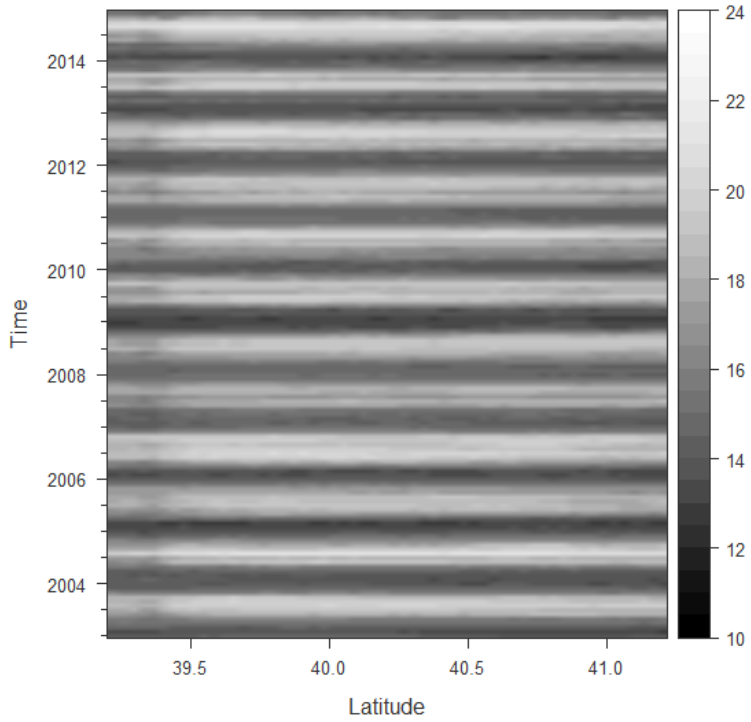


Figure 4.4.1.16. SST (° C) Hövmoller plot of the meridional transect depicted in figure 4.4.1.15. Notice the abrupt change of temperature profiles in the southern area.

When the transect is analyzed in greater detail, the same pattern described before reemerges. In figure 4.4.1.17, a temporal subset of the transect series was plotted for January 2003 and August 2003. These are representative of the cold and warm season, and highlight the differences found along the transect.

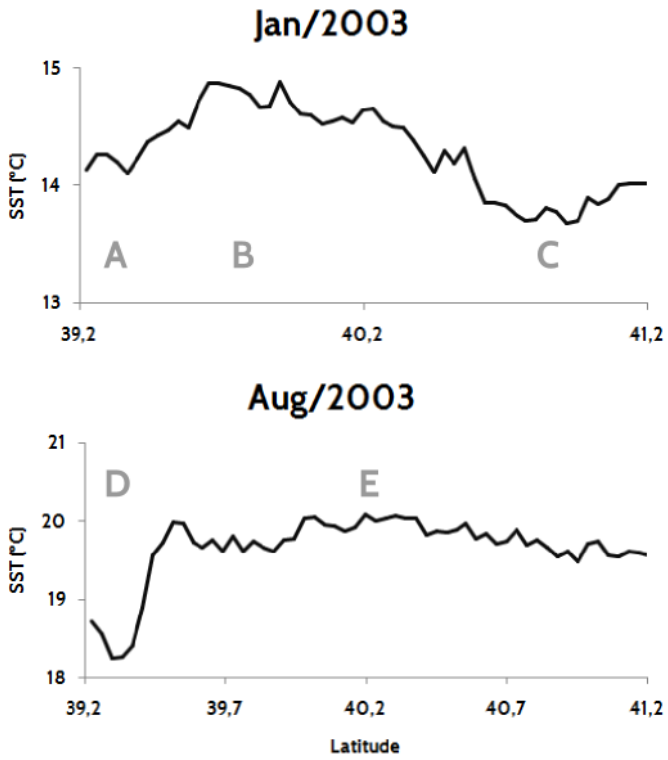


Figure 4.4.1.17. SST (° C) plots of the meridional transect for January 2003 and August 2003. See text for more information.

In January 2003, three different regions are traversed by the profile. A) represents a water mass with intermediate temperatures between B) and C). These points are located to the south of the Peniche-Berlengas line and are therefore part of the special region already emphasized. Interestingly, the temperatures at B) rise sharply only to fall at latitudes approaching 41°. We hypothesize that the plot is representing a section of the IPC, with its warmer waters, which are probably progressing northwards along a topographically controlled tongue. To the North, at C), the WIBP assumes an important role as a consequence of the significant river discharge of the rivers to the north of the Mondego, particularly the Douro River. Was the transect designed to match a constant distance from the coast and these features would be less conspicuous or even absent of the plots (e.g. IPC).

In the August 2003 plot, which is also similar to plots of other years, two regions can be distinguished. Leeward of Cape Carvoeiro (*cf*. A), a region of low SST values (D) is in stark contrast with neighboring areas to the North (E). In fact the temperature difference exceeds 1.5° C in the most extreme case. This is the signal created by the strong upwelling cell we described earlier and that could be connected to the Nazaré Canyon as well. Windward of Cape Carvoeiro - when the northerlies are blowing and promoting upwelling - the coastal area coastward of the shelf edge is more homogeneous due to the westward expansion of the

upwelling band. If the plot depicted an earlier month, when upwelling is not fully developed, it would be possible to detect differences along the transect akin to those of the January graphic. In August, the presence of the WIBP is only vestigial and restricted to a narrow band near the river mouths. As such, the influence of this feature over the plotted SST values is minimal. The Nazaré Canyon on the other hand, seems to play a role, albeit discreet, in the thermal profile. Centered at approximately 39.7° N, a region of lower SST values is observable, and may reflect the water transport along the axis of the Canyon towards the coast.

SST and thermal fronts generally occur simultaneously with boundaries and gradients of other variables (Belkin et al., 2009). Simultaneously, it has been demonstrated several decades ago that the optical environment is also an important driver of SST itself, with turbidity influencing (positively) SST (Charlock, 1982). In light of these facts, and also considering that several spaceborne sensors lack the ability to retrieve thermal data, understanding Chlorophyll a dynamics in the study area becomes relevant from multiple perspectives.

In particular, we are interested in understanding how SST and Chlorophyll a correlate and how the later is influenced by the thermal properties of the water. It is known that Chlorophyll a concentration is strongly influenced by abiotic variables, including SST (Alvarez et al., 2012).

It is also known that the Chlorophyll a concentration is highly variable, with seasonal peaks, especially in the warm months (Alvarez et al., 2011). These peaks can occur, in the western coast of the Iberian Peninsula in February, April/May, and July/September (Alvarez et al., 2012). Nonetheless, it is important to emphasize that many studies on Chlorophyll a dynamics are based on remote sensing technologies, applied to the coastal ocean.

The use of EO data in the coastal domain is particularly problematic, as we have mentioned before, due to the complexity of the optical environment.

The seminal work by Morel et al. (1977) and Gordon et al. (1983) led to the division of ocean waters into so-called “Case-1” and “Case-2”. Coastal waters are generally classified as Case-2, considering that water color is determined by the presence of a complex mix of phytoplankton and other constituents such as ‘yellow substance’ and suspended particles. These are not linearly correlated with phytoplankton, rendering its retrieval problematic.

River runoff, for instance, or wave-induced turbidity, can lead to the development of a seasonal optically complex environment, thus limiting our ability to rely entirely on standard algorithms to build Chlorophyll a time series. Although we lacked *in situ* data to validate the Chlorophyll a fields retrieved from MODIS, we assumed the limitations while acknowledging the importance of the dataset as a cost-effective source of information.

Chlorophyll a concentration is strongly controlled by the recurrent upwelling events, including in the (less frequent) winter episodes (Ribeiro et al., 2005).

During winter, Chlorophyll a concentration reaches 0.3 $mg.m^{-3}$ off WIP, a figure substantially lower than the values measured in the warmer months (Peliz et al., 2004).

Figure 4.1.1.18 depicts the time series of Chlorophyll a concentration, retrieved from the Level-3 MODIS AQUA data, in approximately the same region as the SST plot of figure 4.1.1.5.

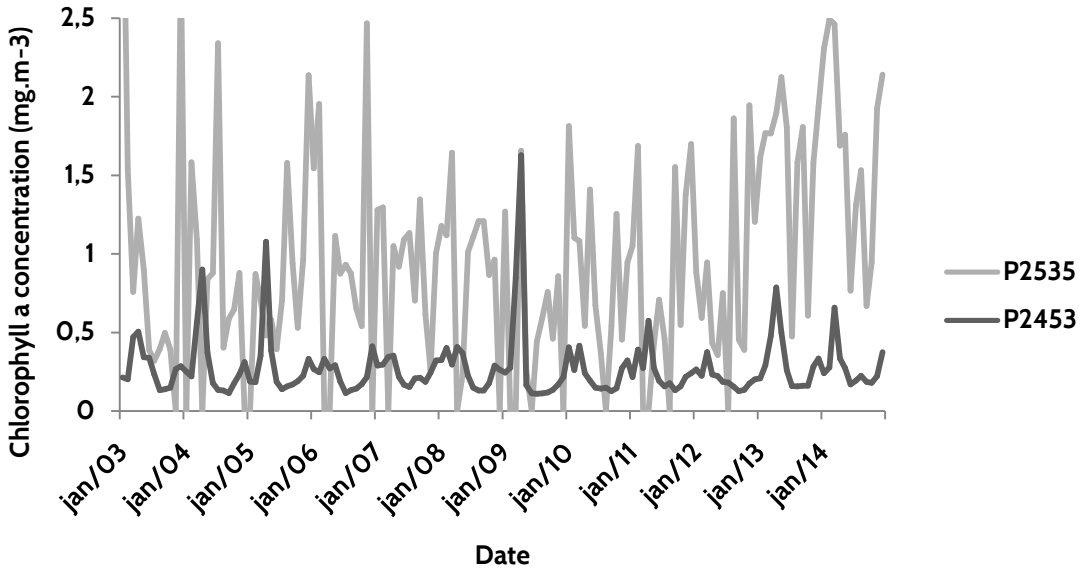


Figure 4.1.1.18. Plot depicting the monthly Chlorophyll a concentration (MODIS AQUA) time series (01/2003-12/2014) of the westernmost (P2453) and Easternmost (P2535, P2536 did not contain sufficient valid information) points in the Figueira da Foz (TFF) transect. Notice the higher values offshore occurring in April (Spring bloom) and more predictable cycles.

Notice how the western pixel (P2453) displays a much more cyclic and clean time series. On the other hand the coastal site depicts generally higher values and a less defined seasonality. We hypothesize that the 'noisy' time series of the coastal pixel is, at least partially, due to the fact these are Case-2 waters, with significant influence of river discharge from the Mondego river (and WIBP) and coastal turbidity levels.

The data from the coastal site shows some of the peaks and troughs in magnitude mentioned by Alvarez et al. (2012). However, point P2453 is characterized by a main annual peak centered in April and the lowest values occurring in the summer. The spring bloom is likely to play an important role in this cycle, which is less influenced at this longitude by coastal upwelling events.

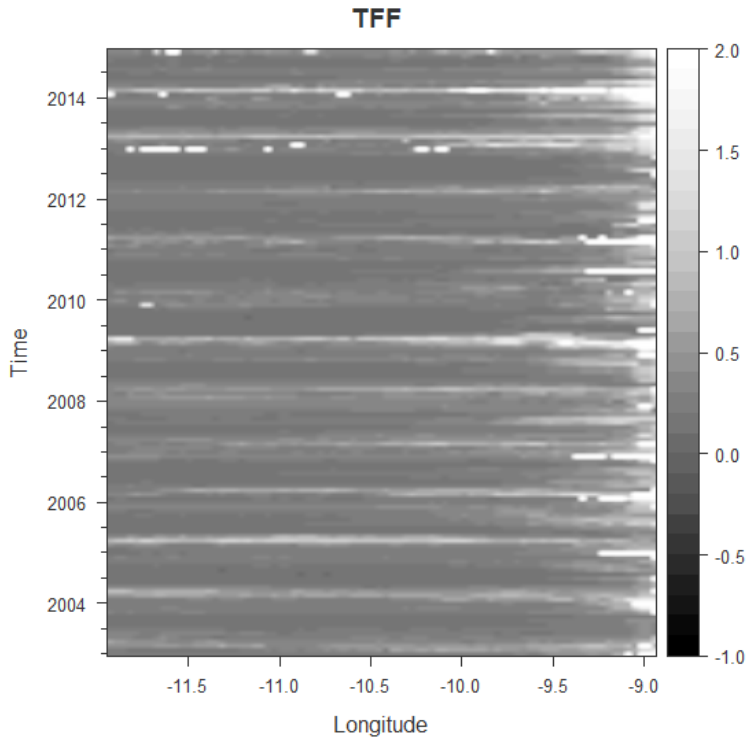


Figure 4.1.1.19. Chlorophyll a concentration (MODIS AQUA) Hövmoller plot of transect TFF depicted in figure 4.4.1.4.

Figure 4.4.1.19 shows a Hövmoller plot of Chlorophyll a concentration (MODIS AQUA) of the TFF transect (at the latitude of Figueira da Foz). The recurrent high pulse in April is shared by most points across the zonal transect. However, significant complexity from the shore to longitude -9.5° W complicates the analysis at the coastal level.

The Chlorophyll a concentration estimated by the sensors depends on a number of variables, including the depth of the mixed layer and the hydrodynamic conditions in the upwelling centers (Moita et al., 2003, Oliveira et al., 2009). Ekman transport offshore will further contribute to the movement of biogenic materials westward, leading to a complex pattern, which is, as SST, described by multiple variables, flows, and features (including at the mesoscale).

In fact, not only Chlorophyll a concentration values differ spatially (sometimes associated to mesoscale or topographic features or the simple location within a plume), but the profile of the communities can also vary, sometimes with important implications for the ecosystem. This is the case when specific conditions foster the development of toxin-producing organisms (e.g. *Gymnodinium catenatum*), at specific regions within the broader upwelling areas (Moita et al., 2003). The same author also emphasizes the heterogeneity of concentration values within a cell, with the highest estimates found in the coastward region.

The timing and duration of the phenomena can also have important implications in the structuring of phytoplankton communities (Silva et al., 2008) as do the general patterns of mesoscale activity (Marchesiello and Estrade, 2007).

As such, it is plausible that, depending on the region being considered, the correlation between SST and Chlorophyll a concentration differs.

Even more subtle features, such as internal waves, can affect the distribution and concentration of Chlorophyll a. As previously reported, the enhancement is promoted by a temporary rising of the Deep Chlorophyll Maximum by action of the IWs in the study area (Silva et al., 2002). Considering the limited cartography produced for these features, it becomes clear that many variables are still poorly understood or spatialized.

To understand the correlation between SST and Chlorophyll a, the Pearson correlation coefficient was calculated for the time series of SST/OC pairs in the study area. Only Level 3 pixels with sufficient data, as previously addressed were used in the analysis. The test aims at assessing the linearity of the correlations between both variables.

The results are depicted in figure 4.1.11.20, with the results being represented per pixel. Notice that many nearshore pixels were not used in the analysis, which is the result of the difficulties in retrieving reliable data in that environment. The processing chain developed to this end, masks coastal pixels frequently to avoid introducing more error into the time series, than that inherent to Case-2 waters.

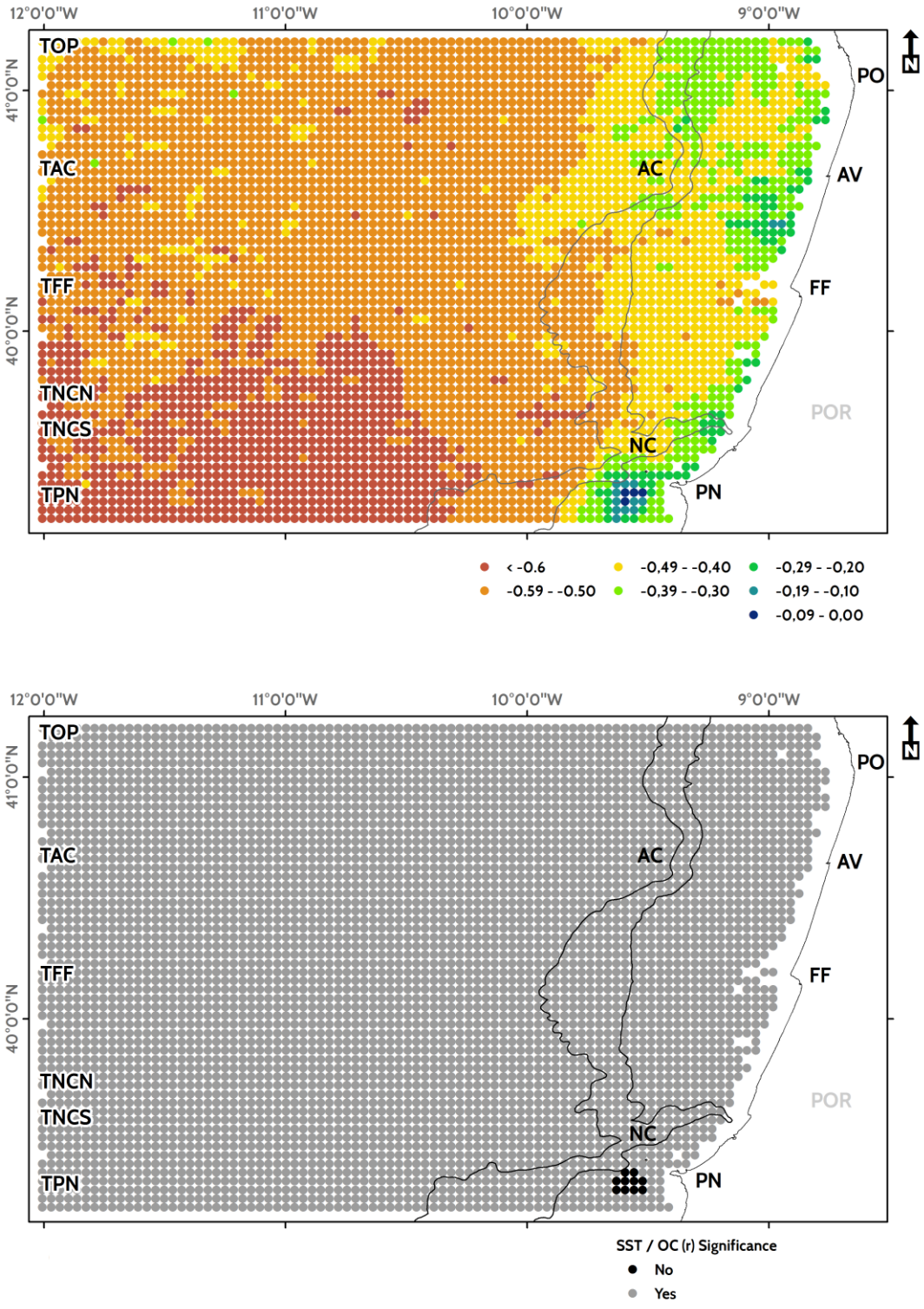


Figure 4.1.1.20. Map depicting the (top) Pearson correlation coefficient and (bottom) the significance of the correlation for SST and Chlorophyll a (OC) values for $p < 0.01$. The location of the zonal transects is included. The 200 and 1600 m isobaths are represented as thin dark lines. PN: Peniche; FF: Figueira da Foz; PO: Oporto; AV: Aveiro; NC: Nazaré Canyon; AC: Aveiro Canyon; POR: Portugal.

The figure highlights several aspects worth mentioning. Firstly, it becomes clear that the correlation between SST and Chlorophyll a obeys to a spatial structure. The distribution of correlation values (*r*) is not random, but instead spatialized. This is also true for the significance of said correlation, with only a small pocket of neighboring pixels southward of Berlengas fail to display a significant correlation. We hypothesize that the lack of a significant correlation there is the consequence of the much debated re-circulation leeward of prominent capes, of which Cape Carvoeiro is an important representative (Oliveira et al., 2009, Mason et al., 2006). The Berlengas island wake may also play a role, as we shall discuss latter in this work.

The Aveiro Filament is also represented in the figure, as a lower correlation area that protrudes from the coast at approximately the latitude of Aveiro. Comparison of the figure with figure 4.1.1.12 emphasizes the connection, which is certainly not coincidental.

It is also noteworthy the general, narrow band of low *r* values along the coastal area. This band is likely caused by a combination of factors. Southward of the Cape Carvoeiro-Berlengas line, it is likely due to the aforesaid patterns of re-circulation and mesoscale activity leeward of the Cape. To the north, and especially from the Mondego river towards Oporto, the band is probably a consequence of the combined effect of the WIBP, upwelling filaments, and nearshore circulation patterns. Notice how the band expands eastward with latitude, leading to an oceanward displacement of the nearshore domain. The displacement is coherently aligned with both the development of the topography-anchored upwelling filaments and the development of sizeable freshwater buoyant plumes, which are wider in the vicinity of the Douro river. Still, areas such as the Nazaré Canyon also display complex patterns and localized variability in the correlation. Overall, the patterns show a strong topographic connection, suggesting that both hydrodynamic and atmospheric forcing mechanisms are at play, with the shelf edge marking an important (albeit generalized) barrier between the coastal and oceanic domain.

The development of phytoplankton communities is often a dynamic process in which the organisms respond to multiple variables (e.g. SST, wind force and direction, etc…) (Breitbarth et al., 2007, Carpenter et al., 2004). Still, the map also suggests very clearly that the processes have a varying weight depending on the region, and the correlation coefficient offers an interesting tool to determine the levels of complexity. Straightforward regions, such as those in the southwest corner of figure 4.1.1.20 show a high correlation (< 0.60, significant for $p < 0.01$) between SST and Chlorophyll a, serving baseline against which other regions can be compared. Again, caution must be exercised when analyzing the correlation between SST and Chlorophyll a in coastal areas, considering the difficulties inherent to the retrieval of bio-

optical variables in Case-2 waters. At least some of the low correlation values yielded nearshore may be the consequence of such hindrances.

In this context of complex spatial and temporal variability that characterizes the region of study, which is due to a plethora of drivers, it becomes important to question whether the individual time series may be compared objectively. To this end, metrics capable of ingesting the large volume of information available and detect patterns and similarities in the data are called for.

The similarities must be quantified in order to establish connections between points or regions. This is the focus of the next section, which addresses the use of Dynamic Time Warping (DTW) as an objective method to deal with large sets of data and organize them into a clusterable distance matrix. This effort will constitute the basis of a new, scalable regionalization model applied for the first time in the study area.

4.4.2. Dynamic Time Warping and Clustering of SST and Chlorophyll a concentration

In the previous section, it was recognized that SST and Chlorophyll a concentration varied across the study region and that the relation between both variables was apparently spatially structured.

The spatial structure and variability of the series and trends, which are accurately captured by the satellite sensors, create an interesting opportunity to attempt a regionalization scheme. Such a scheme would have to rely on the similarity of time series in order to cluster the pixels into coherent partitions. The “unique and homogeneous” units could then be compared against natural borders, including thermal or ocean color fronts (Nieblas et al., 2014).

Using a dynamic methodology to partition the coastal ocean would offer several benefits, including the objectivity of the metrics employed, the semi-automated definition of the optimum number of regions, and the dynamic nature of the boundaries delimiting each BU.

These were the set of requirements targeted in this study, as a new approach to biogeographic classifications was proposed. Ultimately, the objective was to compare the resulting regions with the existing body of knowledge on the area of study and identify the level of coherence between BU boundaries and frontal regions.

It is widely known that phytoplankton communities require specific SST conditions in order to bloom (e.g. Carpenter et al., 2004, Moutin et al., 2008). These conditions are connected to physiological processes that occur at an optimum rate at specific temperatures (and Photosynthetically Available Radiation) (Breitbarth et al., 2007). Moreover, the time series of SST and Chlorophyll a can display significant variability at different scales (namely interannual),

which clearly hinder the simple usage of climatological averages for ecosystem or biogeographic studies and applications (Thomas et al., 2003).

Using objective metrics in order to identify common traits between different grid cells is thus of the utmost importance in the development of a robust classification framework. However, the classification of time series (especially in long sets) is particularly difficult considering the dynamic nature of data, which is still connected through a specific (and meaningful) order (Łuczak, 2016).

Any BC scheme should attempt to be based on objectivity, scalability, and reproducibility. When the BC is based on the use of time series, the complexity of detecting patterns in the data must be considered. This is in fact a field in rapid progress considering the persisting challenges to the detection of dynamic ranges or anomalies, for instance (Łuczak, 2016). In fact, several environmental classification systems, such as the Köppen-Geiger system (Kottek et al., 2006) are developed without taking into account the similarity between local variables (Netzel et al., 2016). Indexing time series lies at the heart of the effort pursued in this study, since we aim at clustering EO data into coherent units based in their dissimilarity. Several metrics can be used to determine the differences between the shapes of the series, including the widely known Euclidean Distance and the least known Dynamic Time Warping (DTW) (Łuczak, 2016).

However, and because Euclidean distance has been proven to be less reliable, DTW is the method of choice in this study because it is capable of detecting similarities in shapes out of phase (Keogh et al., 2004). This method also enables the implementation of a clustering effort based on the unsupervised identification of similarity patterns (Netzel et al., 2016).

The non-linear Levenshtein distance-based Dynamic Time Warping (DTW) meets these requirements since the method is capable of measuring the distance between points in time series, can be applied to large datasets, and has been thoroughly documented, despite a lack of applications in the field of oceanography. DTW was first introduced in the 1970's as a data mining tool designed to find similarities in time series (Sakoe et al., 1978, Giorgino, 2009). This methodology has been used in several applications, most notably in speech recognition tasks, although Earth science datasets have also been analyzed with DTW in the past (Petitjean et al., 2012, McIntosh et al., 2005). DTW is a very flexible method, with variations of the original approach enabling the development of applications relying on these adaptations to entirely new fields (Giorgino, 2009).

Mantas et al. (2015) employed this technique successfully to analyze long time series and generate dissimilarity matrices for rainfall data over the Peruvian Andes. In this example, the methodology successfully recognized similarities in the rainfall patterns, which were then

compared against land cover dynamics and general atmospheric patterns in the region. Still, it has been difficult to find applications of this technique to oceanographic data, in particular to SST or Chlorophyll a concentration. The lack of such examples is hard to explain considering the advantages introduced by the technique, but the computational cost associated to the method may, at least partially, be to blame for the situation.

Still, the seminal work by Berndt et al. (1994) already foresaw the need to use robust data mining solutions (including DTW) to analyze the massive and "inherently temporal" datasets generated daily by NASA (and other agencies). As such, we recognized in this study an opportunity to apply DTW to a large dataset as a precursor to future BC efforts based on this technique, while offering an objective and unprecedented map of the coastal regions off central Portugal.

DTW enables the comparison of "one-to-many points" in an elastic manner, providing a method to compare time series even if it is required to "stretch or compress" said sets (Ding et al., 2008). The temporal axis can thus be compressed and stretched and a good fit found between the series (Berndt et al., 1994). This is, naturally, the foundation of the method's ability to determine the similarity of shapes, despite of phase mismatching, as aforementioned.

In summary, two time series are analyzed, a query ($X = (x_1,\ldots,x_N)$ and a reference set ($Y = (y_1, \ldots, y_N)$), through the creation of a matrix, in which each element (*i*, *j*) matches the 'alignment' between points (Giorgino, 2009).

Afterwards, a 'warping path' (W) (figure 4.4.2.1) is created to link the elements of both series for a minimum possible distance (Berndt et al., 1994).

Dynamic Time Warping values can then be calculated from potential paths, selecting the ones that minimize the distance between the elements of the time series. DTW can thus be calculated as:

$$DTW(X,Y) = \ minW\ [\sum_{h=1}^{p} \delta(\omega_k)]$$

Where, δ is the distance between two elements of the grid. The output of DTW is thus particularly interesting as it provides an objective evaluation of the true differences separating the time series independently of phase-differences (Giorgino, 2009).

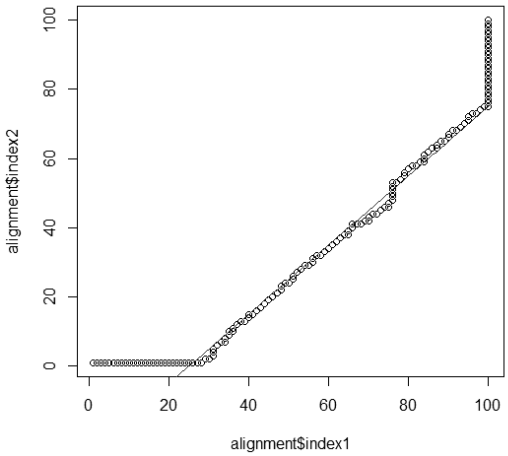


Figure 4.4.2.1. An example provided by R's 'dtw' package (Giorgino, 2009), showing a warping path between a query and a reference time series.

The time series analyzed under DTW can be single or multivariate (Giorgino, 2009), and that is another important attribute of the methodology. In this study, DTW shall be applied to SST, Chlorophyll a, and a combination of SST and Chlorophyll a.

The full dataset (January 2003- December 2014) was indexed for all datasets. A total of 701 424 values were ingested for the analysis of SST, 674 496 for Chlorophyll a, and 1 348 992 for the multivariate analysis. Furthermore, subsets of the time period were analyzed as well. These included yearly runs of the processing chain and two 3 year-blocks of SST data (2004-2006 and 2007-2009). The objective of the subsets was to evaluate whether the results obtained for the long time series (computationally expensive) were equivalent to shorter intervals.

A complex workflow was designed and implemented in R, which included a pre-processing stage, the calculation of the distance matrix and subsequent clustering. The latter is achieved through the employment of Hierarchical Clustering Analysis, applied to the distance matrixes generated by the DTW. The agglomerative Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Maechler et al.,2014) was the selected clustering methodology, due to the simplicity of calculation and computational effectiveness.

The first step in the processing chain included the replacement of missing values with an approximation based on the average of the immediate neighbors. Sampling points with substantial data losses were not considered. This explains the difference between the number of samples to be classified of SST (4871) and Chlorophyll a (4684). For multivariate analysis, only matching points were used, meaning that the analysis was restricted to the Level-3 pixels that contained both SST and Chlorophyll a pairs.

Using the "dtw" library available at R (Giorgino, 2009), the distance matrix is calculated for the time series using the Dynamic Time Warping methodology.

The resulting matrix is then used as the input of the clustering effort. In order to determine the optimum number of clusters generated by UPGMA, we approached this problem in two different ways. Firstly, we calculated the Within-cluster Sum of Squares (WCSS). WCSS is calculated as follows:

$$WCSS(k) = \sum_{j=1}^{k} \sum_{xi \in cluster\ \ j} \left\| x_i - \bar{x}_j \right\|^2$$

Where, k is the number of clusters and $\bar{x}_j$ is the mean of the sample in the cluster. The WCSS can then be compared for various values of k and the optimum number of clusters being determined through the heuristic, so-called 'elbow method'.

In a second approach, various values of k were tested and mapped to determine (visually) the correlation with known features in the study area. In particular, we were interested in identifying the correlation between the cluster areas and the spatial extends of the WIBP, IPC, upwelling cells, and other features.

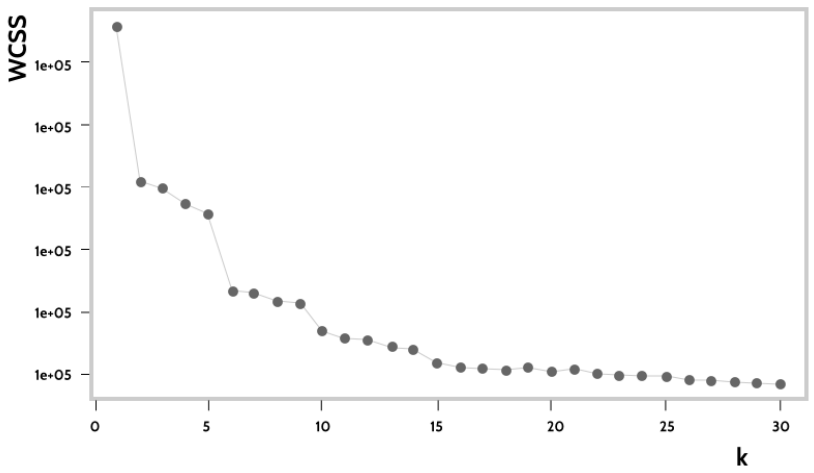


Figure 4.4.2.2. Plot of the WCSS values for different options of k for the SST time series (2003-2014). Notice the multiple breakpoints (or 'elbows') at k=2, k=6, k=10, and k=15.

Considering that SST is the primary dataset of our study, the number of SST-based clusters is the most relevant to subsequent analysis. Furthermore, the problems raised by Case-2 waters create further obstacles to the analysis of Chlorophyll a-driven clusters.

Figure 4.4.2.2 suggests some level of hierarchic organization of the SST pixels is possible. Several breakpoints are clearly discernable in the plot, including at k=2, k=6, k=10, and k=15. The step 'elbows' may indicate the clustered time series can be organized into several levels, depending on the scale of the analysis. A high-level Ocean/Coastal distinction is likely possible if k=6 is adopted. In this case, a number of clusters group coastal sites, namely in the vicinity of areas under the influence of river discharge (including smaller rivers like the Lis), or close to prominent geographic features (e.g. Cape Carvoeiro). A map representing the clusters using a k=6 is depicted in figure 4.4.2.3.

Westward, the contact between coastal and oceanic waters is approximately marked by the shelf edge, although this boundary is displaced towards the West to the north of Aveiro. This displacement is likely the consequence of the combined effect of the cold upwelling filaments anchored to the Aveiro Canyon, the WIBP, which is particularly developed to the North of Cape Mondego, and possibly the interference of the edge-trapped poleward flow further south.

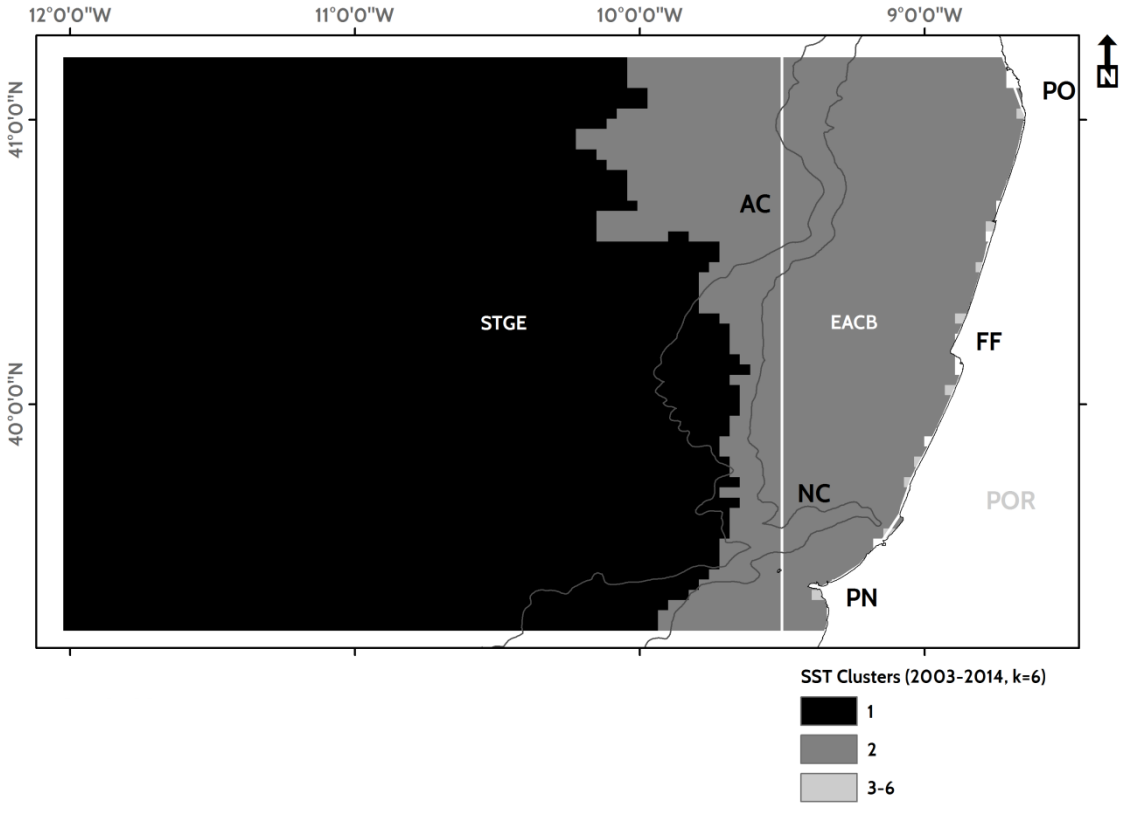


Figure 4.4.2.3. Simplified map depicting the clusters (k=6) of MODIS AQUA SST time series (2003-2014). The boundaries of Longhurst's provinces are depicted as a thick white line. PN: Peniche; FF: Figueira da Foz; PO: Oporto; NC: Nazaré Canyon; AC: Aveiro Canyon; EACB: Canary Coastal Province; STGE: North Atlantic Subtropical Gyral Province.

The SST clusters are not entirely aligned to Longhurst's provinces (Longhurst, 2007), which follow a generalized boundary, somewhat parallel to the shelf edge. Nonetheless, the coastal/ocean dichotomy seems to be expressed in the clustered data. The dynamic delineation of boundaries, following the work of Devred et al. (2007), seems to be made possible using the analysis of long time series through DTW.

The Canary Coastal Province and the North Atlantic Subtropical Gyral are described in detail in Longhurst's (2007) seminal work.

Still, the EACB is a large but generally narrow province which includes part of the study area, as it extends from Galicia to Senegal. Upwelling dominates primary production cycles and magnitudes in the province. Highly productive regions, including in the study area, are constrained to the aforementioned band, which does not usually exceed the width of the shelf and is commonly less than 40 km wide. Still, and according to the same author, it is not unlikely to measure Chlorophyll a concentration values in excess of 2.5 $mg.m^{-3}$, which is in line with observations made in the course of this study.

On the other hand, the North Atlantic Subtropical Gyral Province, which bounds the EACB offshore, offers some interesting, but contrasting features. Being much wider, it covers most of the Portuguese Economic Exclusive Zone, including the Azores, Madeira, the sea off continental Portugal, and also the Canary Islands.

Chlorophyll a concentration in the province is lower than that of the EACB, and it has been difficult to determine the exact cycles to which it obeys. Still, and despite the possibility of high concentration values throughout the year, some seasonality was observed in the portion of the province that is located within our study area. In fact, the Pearson Correlation Coefficient between SST and Chlorophyll a concentration was stronger within the area encompassed by the province within our region.

In order to advance beyond the separation of coastal and oceanic provinces suggested by Longhurst, further subdivisions were tested, considering the breakpoints depicted in the WCSS plot. To this end, new stratified maps were generated for k=10 and k=15. These clusters will be analyzed jointly, because they provide important insights not only into the methodology's performance but also on the circulation patterns of the region.

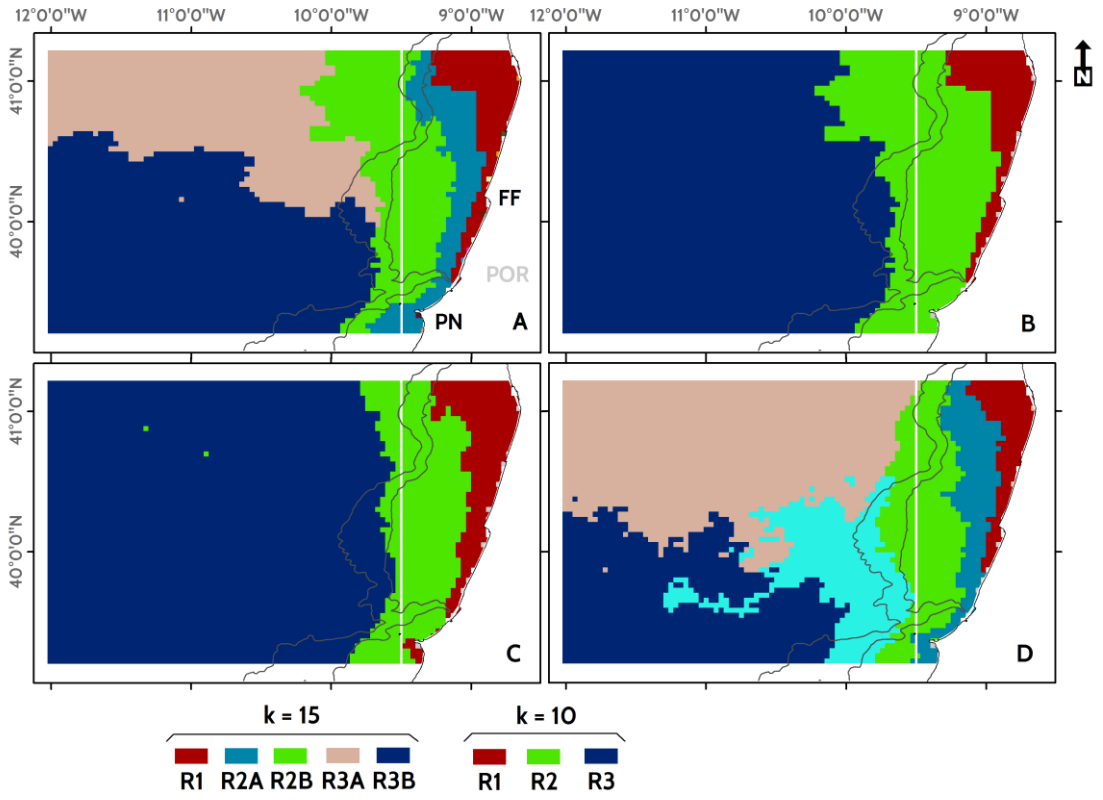


Figure 4.4.2.4. Comparison of several clustering options for the MODIS AQUA (Level 3) SST data. A: k = 15 applied to the 2003-2014 time series. B: k = 10 applied t the 2003-2014 time series. C: k = 10 applied to a subset of the time series (January 2007- December 2009). The colors and nomenclature of the regions follow the k = 10 standard. D: k = 10 applied to a subset of the time series (January 2007 – December 2007). The colors and nomenclature of the regions follow the k=15 standard. Longhurst's provinces are represented as white lines and the 200 and 1600 isobaths are depicted as dark grey lines. FF: Figueira da Foz; PN: Peniche; POR: Portugal.

Figure 4.4.2.4 highlights the complex patterns that organize the ocean off central Portugal. Increasing the value of k, leads to additional segmentation of preexisting regions. For a k = 10, three regions (except those in the vicinity of the shore, as aforementioned) can be identified. We adopted a code to identify the regions based on the hierarchical nature of the methodology. The baseline regions are set by the segmentation found using a k = 10. The regions generated by subsequent segmentation, using higher values of k, are denoted by a string added to the integer used to identify the larger region. Region 2 (R2), for instance, can be subdivided into R2A and R2B when a k = 15 is used. Similarly Region 3 (R3) can be segmented into regions R3A and R3B.

From all regions (k = 15), only R1 persists unaltered in both maps. The location and zonal development of the R1 region is consistent with the area occupied by the WIBP. The presence of a buoyant lens of freshwater from neighboring rivers is well documented. However, the satellite data, as described earlier, demonstrates that the plume is often located southward of commonly reported positions (e.g. Peliz et al., 2005). We hypothesize that the R1 region corresponds to the average position of the WIBP, which is, nonetheless, variable throughout the year. The comparison of SST time series data and animations as well as of analysis of the Hövmoller plots of zonal and meridional transects supports this interpretation. Upwelling activity in the area may also contribute to the good definition of the region. The temperature contrast between summer and winter months probably contribute significantly to create this well-defined and persistent region. Interestingly the axis of the Nazaré Canyon marks the southernmost boundary of the R1 segment.

On the other hand, Region 3, which is divided into R3A (North) and R3B (South), clearly represents oceanic waters, away from the immediate influence of the coast. The region is found beyond the 200 m isobaths, but at a varying distance from the shore. In fact the 1600 m isobath is not aligned with the boundary of R3, which is much closer to shore at the southern end of the study area. The Easternmost boundary of R3 is located at a distance of 33 km of the shore at the latitude of Peniche, but at a distance of 121 km of Oporto. At Figueira da Foz, a central location within the study area, the distance between the coast and R3 is of 68 km.

Between R1 and R3, R2 stretches from the inner shelf towards the vicinity of the shelf edge. Its boundary is irregular and only somewhat topographically constrained, with R2A (Eastern segment) following the general shape of the coastline (and R1), with greater development off Aveiro-Oporto. This area represents the core of the upwelling region, as depicted by several authors using *in situ* and satellite data (e.g. Relvas et al., 2007).The R2B region represents the outer limits of the upwelling region, which is characterized by mesoscale features created by the interaction of the upwelling jet and the offshore currents. The R2B region is thus dominated by the presence of seasonal eddies and filaments, with the presence of a shelf edge-trapped poleward current influencing SST time series in the winter.

We can then assume R2B to constitute the boundary of a Coastal Transition Zone (CTZ) as that suggested for the California system by Brink et al. (1991).

This is a region where contrasting flows, with different energy and often direction meet. Filaments anchored to prominent topographic features are also found there, especially in the northern half of the study area, where R2B is also more developed.

The existence of a CTZ in the study area has been suggested before (Peliz et al., 2003), but the DTW-based methodology offers, for the first time, a well bounded map showing its extents over a long time series containing 11 years of SST measurements.

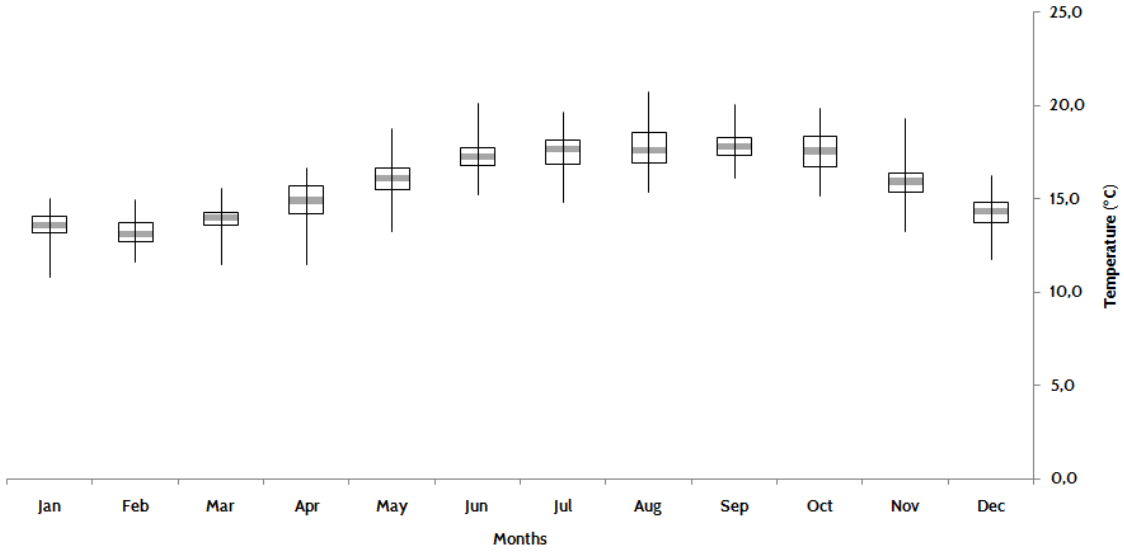


Figure 4.4.2.5. Boxplot of the SST time series for region R1.

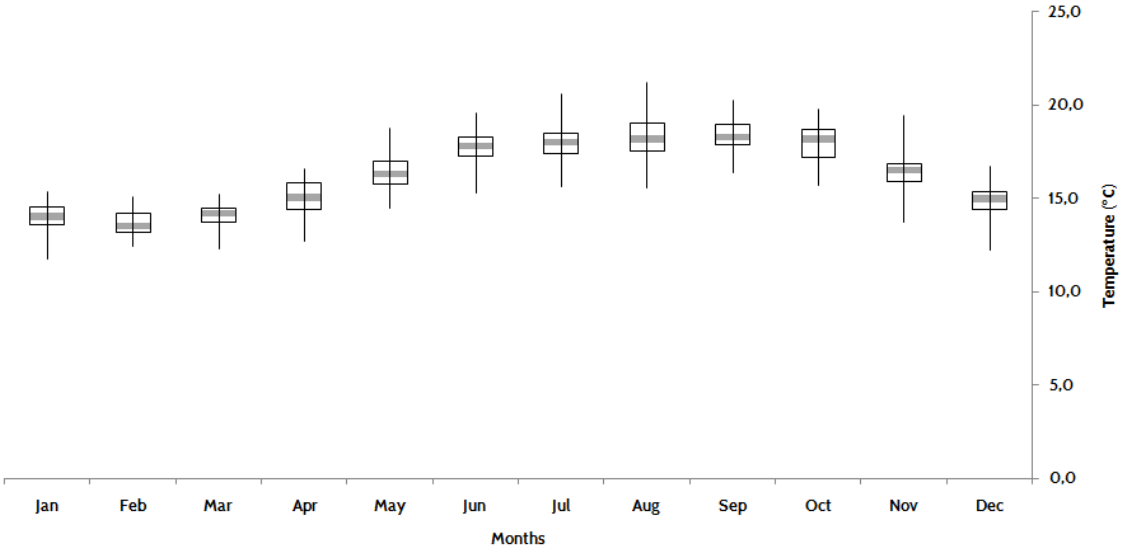


Figure 4.4.2.6. Boxplot of the SST time series for region R2A.

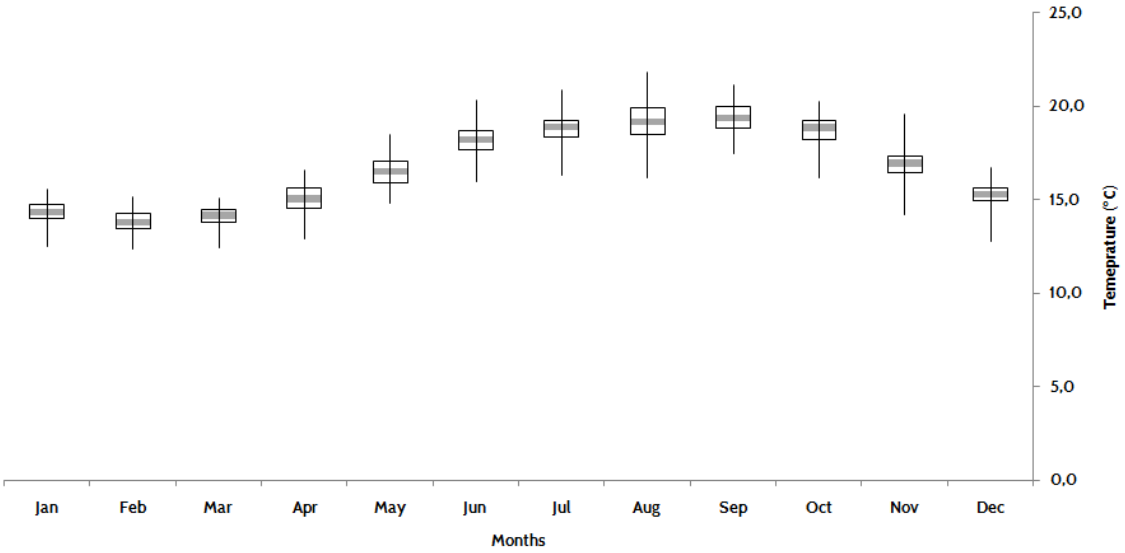


Figure 4.4.2.7. Boxplot of the SST time series for region R2B.

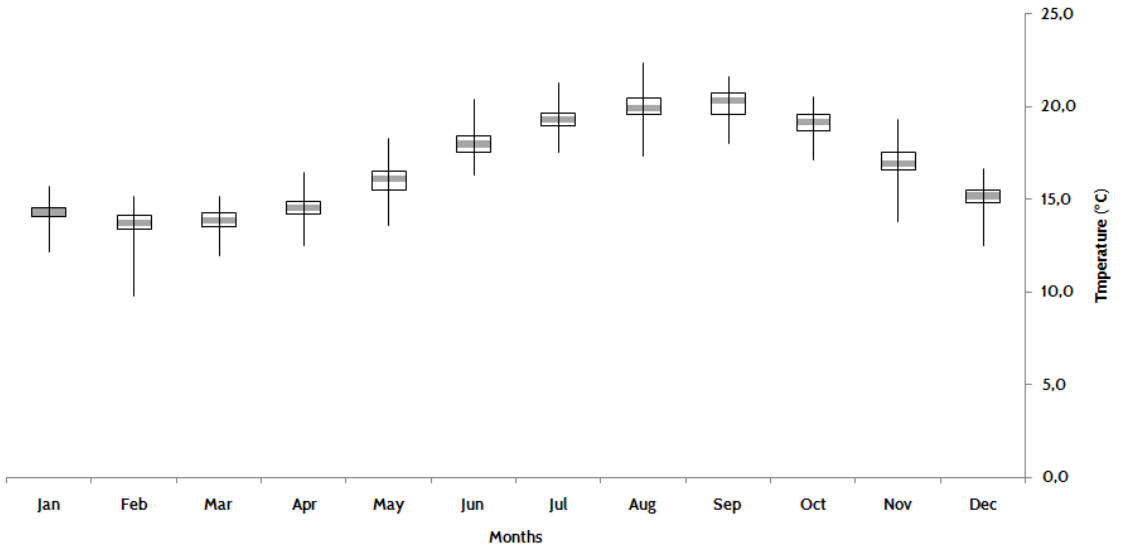


Figure 4.4.2.8. Boxplot of the SST time series for region R3A.

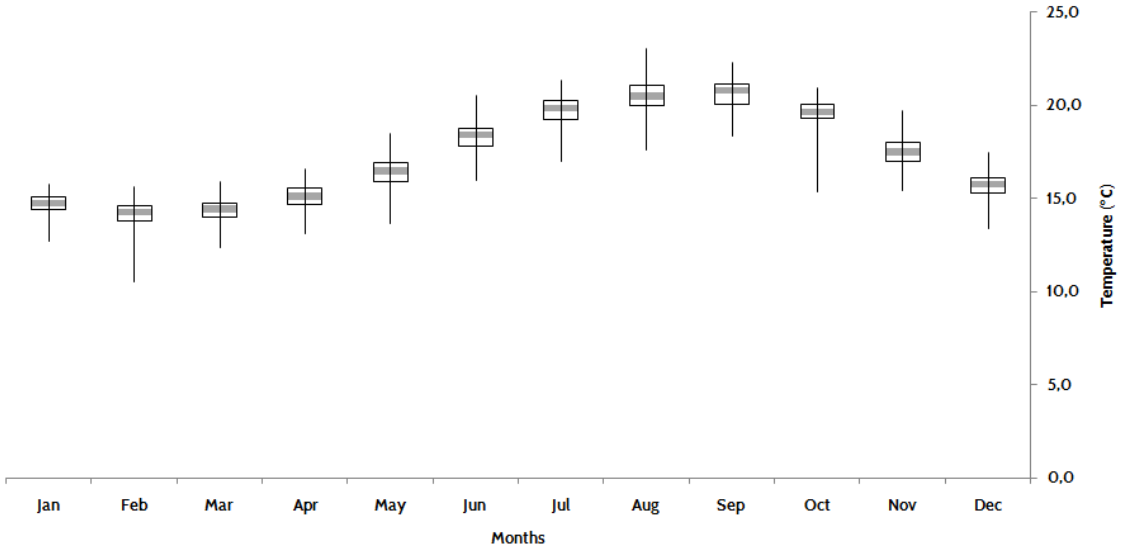


Figure 4.4.2.9. Boxplot of the SST time series for region R3B.

Figures 4.4.2.5-9 depict the maximum, minimum, Q1, Q3, and median SST values for each region (k = 15). The different regions show some level of internal variability, but all depict a common seasonal cycle of warming and cooling.

The regions are predominantly organized along meridional bands. This leads to differences in the absolute SST values recorded for each unit. This is in fact the basis for indices such as the UI_{SST}, which rely on zonal temperature differences between coastal and offshore points to determine the strength of upwelling events (e.g. Alvarez et al., 2011).

Figure 4.4.1.5 depicted the SST differences between a coastal pixel (located at R1) and an offshore pixel at the same latitude (at R3B). That image shows, in a simplified manner, how the temperature cycle differs in both regions. The temperature differences between both points (UI_{SST}) cannot be calculated for the regions in which both points are located. Considering the latitudinal differences in water temperature, such calculation would be meaningless. However, the differences in the mean SST values calculated for each region may provide interesting insights into the dynamics driving the distance matrices and subsequent clustering.

The standard deviation of the mean SST values calculated for the regions suggests the variability is negatively correlated to the upwelling index ($R= 0.98$, significant for a $p< 0.05$) (figure 4.4.2.10).

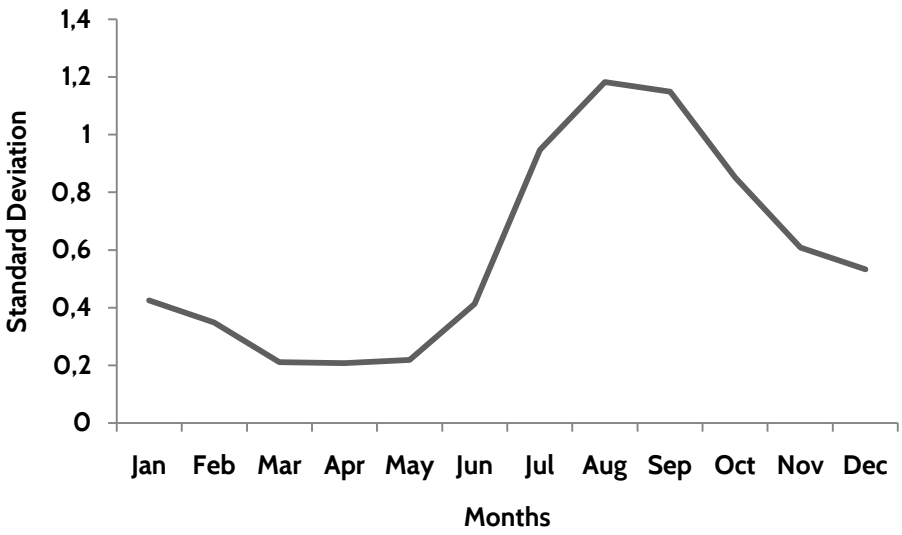


Figure 4.4.2.10. Standard deviation of the mean SST (°C) values per month for the regions identified in the DTW-based method (k = 15).

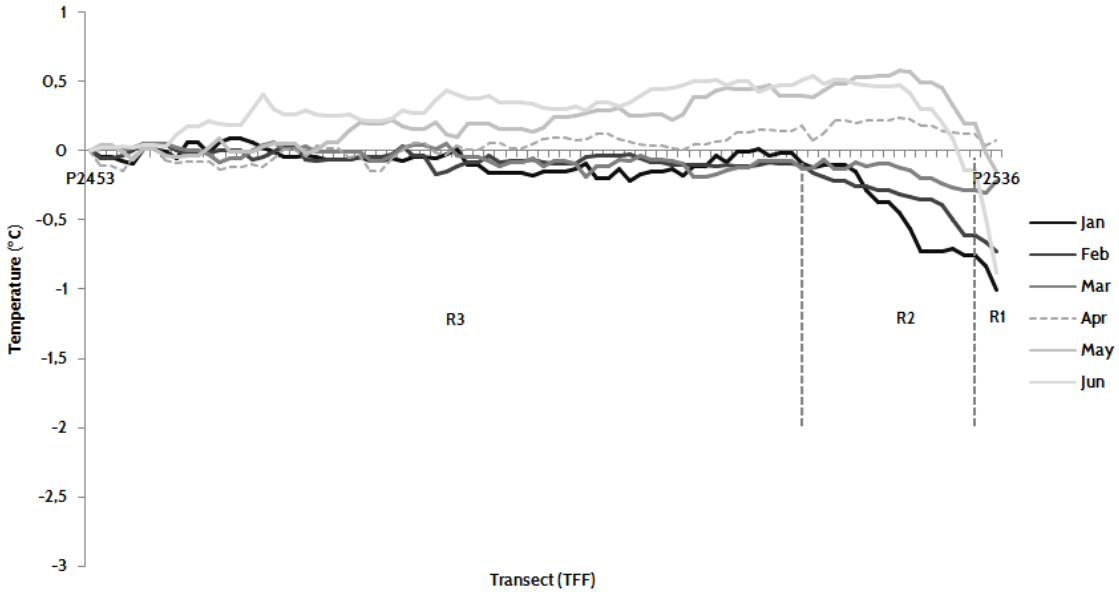
As expected, the standard deviation is lower in April and higher in August. This suggests the differences between the coastal and oceanic regions are higher in the upwelling months, when SST values decrease eastward but increase offshore. If the temperature differences in January, at the TFF transect, are lower than 1 °C, in August, the difference exceeds 2.5 °C. In April, interestingly, the mean SST difference between nearshore pixels and offshore ones is of less than 0.5 °C. Figure 4.4.2.11 shows how the temperature differences vary across the TFF transect in different months, using the temperature at the pixel farthest from the coast as the reference. This is thus a 'moving UI_{SST}', which may be useful to determine differences in the water masses inside the sampled area, namely in the shelf.

The results are in line with previous studies, although, interestingly, the UI_{SST} is often calculated for single pairs (coastal vs. offshore) and not for zonal transects, despite the information it may provide on the wealth of water masses that populate the Portuguese shelf and open ocean. Figure 4.4.2.11 provides interesting visual cues that support the interpretation of the biogeographic classification attempted in this study. The magnitude (and signal) of the UI_{SST} seems to be connected to the regions generated by our method. As such, some considerations on the seasonal dynamics are called for.

The data from figure 4.4.2.11 highlights the similarity of upwelling magnitude in the summer months (July, August, and September). Unlike the results reported by Picado et al. (2013), the strongest UI_{SST} values were found consistently in August (mean monthly UI_{SST} for the period Jan/2003-Dec/2014). The values exceed -2.5° C as reported by these authors, who also

highlight the differences found when the index is calculated from SST data or from Ekman Transport models. In the second case, the upwelling maximum occurs in July.

The literature itself present differing views and results regarding the upwelling maxima, which is nonetheless likely to occur in August-September with some likely inter-annual variability (Alvarez et al., 2011). It must also be considered the potential impact of a hypothesis several authors did put forth (e.g. Santos et al., 2004, Borges et al., 2003), which suggests the upwelling cycles are changing off the Portuguese coast, with unpredictable consequences to biological communities. Differences in the magnitude or frequency of either winter or summer events may disrupt long-standing seasonal patterns, considering the relevance of upwelling (and fronts) to phytoplankton biomass and the entire food chain (Peliz et al., 2004, Reese et al., 2011). For this reason, monitoring SST and Chlorophyll a concentration from space is paramount to ensure both a near real time capability and long term perspective of the dynamic system.



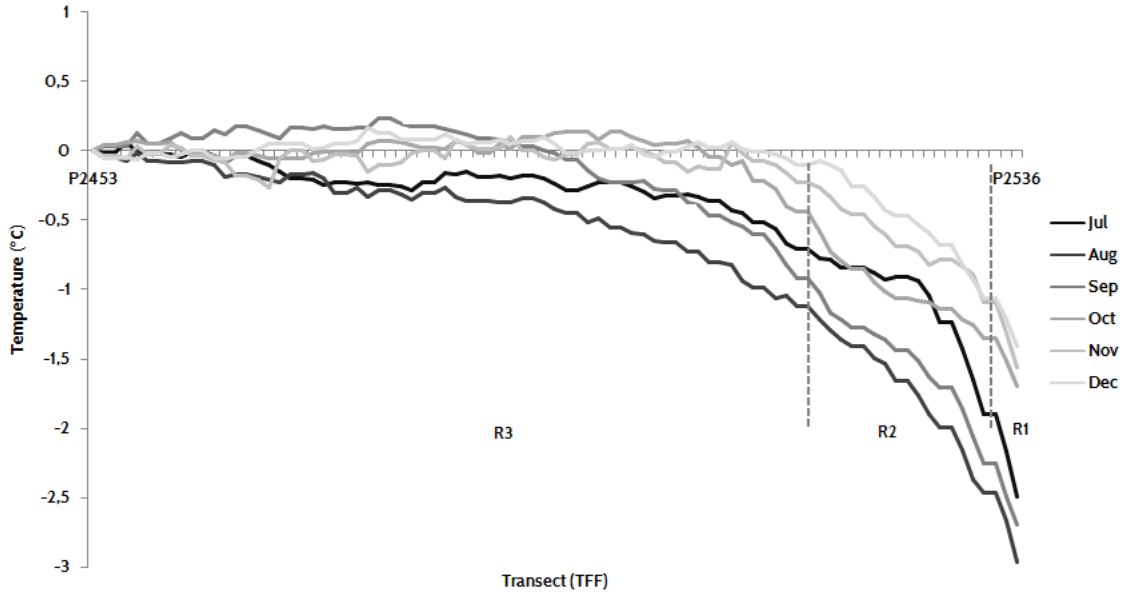


Figure 4.4.2.11. Mean SST difference in relation to a reference point (P2453, located at longitude -12.0°) calculated for the 2003-2014 time series (Top: Jan-Jun, Bottom: Jul-Dec). The dataset for April is represented as a dashed line for greater clarity. The boundary separating DTW-based regions (k = 10, R1, R2, R3) are represented as vertical dashed lines. Notice that the coastal pixels (R1) in April display a positive value, unlike data from all other months.

Albeit the discordance on the upwelling maximum, the April 'minimum' was found to be in agreement with data reported by Picado et al. (2013). The influence of river runoff and the relative temperature differences between the WIBP and offshore waters should be further investigated, as these are likely to influence significantly the values estimated for UI_{SST}.

A Landsat-8 Thermal Infrared Sensor (TIRS) image of April 25, 2016, albeit outside the study period, offers an interesting view of the Portuguese coast, off Figueira da Foz (figure 4.4.2.12). In it, a plume produced by river Mondego is clearly visible, displaying similar features to those described in earlier sections and by Horner-Devine et al. (2009). However, the temperature gradient is quite subtle. The waters entering the Atlantic Ocean no longer generate the cold lens that can be seen in the winter months. Instead, the warmer riverine waters enter an ocean that is, itself, warmer. The seasonal warming trend, the river water input at a higher temperature, and the absence of a developed upwelling cell, leads to an equilibrium that translates into a positive (or close to zero) UI_{SST}.

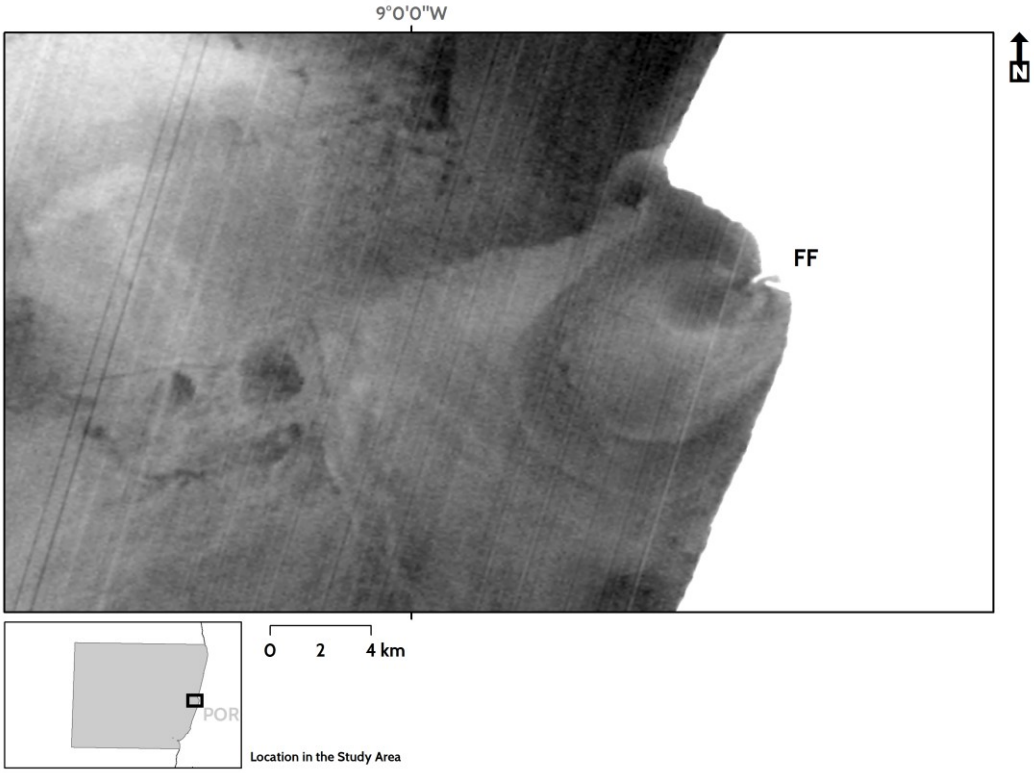


Figure 4.4.2.12. The Mondego river plume in April 25, 2016 as imaged by Landsat-8 Thermal Infrared Sensor (band 10). Warm temperatures represented in lighter tones. Diagonal lines are image artifacts.

Furthermore, a narrow band of warm water, extending 40 km westward, is detectable in the imagery, displaying an apparent poleward motion. Ship wakes seen in the image display a cold signature, suggesting that the warm water mass imaged by the satellite is restricted to a thin surface layer. As the ships traverse the layer, they induce the mixing with deeper, colder waters (figure 4.4.2.13). This observation calls for caution in the interpretation of EO data, as satellite sensors often image a surface, warmer skin, which may not always reflect the temperature in deeper (e.g. 1 m) layers. Albeit this potential limitation, skin temperature, and its variability is also relevant, considering that it reflects a plethora of drivers from mixing to daily warming trends as well as the underlying currents driving the overall motion of upper layers.

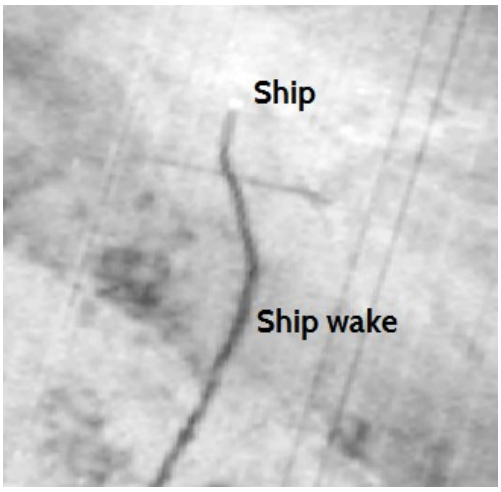


Figure 4.4.2.13. A ship heading North is seen in a Landsat-8 TIRS SST (°C) image (band 10) of April 25, 2016 at 40.93°; -8.89°. The ship's wake is depicted as a dark line. Warm temperatures represented in lighter tones. Diagonal straight lines are image artifacts.

It is equally noteworthy a seeming spatial correlation between the regions and the SST/Chlorophyll a Pearson correlation coefficient. The distribution of *r* (figure 4.1.1.20) seems to follow a trend, not only in the coastal areas – where Case-2 waters and upwelling dominate – but also in the open ocean. At R3B the values of *r* suggest a stronger negative correlation (<-0.6), with strength decreasing northward into R3A. We hypothesize that R3B is more regularly under the influence of the pronounced warm tongues of the IPC, despite the strong correlation (r =0.99, significant for a $p < 0.05$) between the mean monthly SST values measured for the two areas. It is, nonetheless, clear that there is a latitudinal correlation gradient between the two variables, which only further research can shed light on.

Possibly, the heightened mesoscale activity associated with the interaction between the IPC and other currents, which can be seen in satellite imagery throughout the study period, may also contribute to the co-variability of the datasets. This has been suggested as a hypothesis to describe the occurrence of variable correlation patterns off Baja California (Herrera-Cervantes et al., 2013).

Easier to explain is the dichotomy – and spatial patterns – characterizing coastal and oceanic co-variability of SST and Chlorophyll a. Overall, in coastal waters Chlorophyll a time series are affected by water constituents that do not correlate linearly with phytoplankton (i.e. the definition of Case 2 waters). As mentioned earlier, the Chlorophyll a time series are difficult to interpret and correlate because, at least to a certain extent, it is not Chlorophyll a that it is being measured. The closer the pixel is to the shore, especially in regions with significant river outputs or favorable seabed, the more complicated it is to retrieve accurate measurements of

Chlorophyll a concentration. Moreover, the nutrient inputs from adjacent watersheds will artificially fertilize the coastal ocean leading to larger, more persistent blooms. These reasons all contribute to a distorted annual cycle of Chlorophyll a concentration as measured by satellites (and *in situ* assets, to some extent).

The weakest correlation was found within the R1 and R2A regions, namely to the South of Cape Carvoeiro. There, a pocket of non-significant pixels suggests Sea Surface Temperature and Chlorophyll a concentration are not correlated whereas the spatial pattern suggests this was a non-arbitrary outcome. This is often considered an upwelling shadow zone (Mason et al., 2006) and it is possible that recirculation leeward of the northerly winds that propel upwelling is driving the pattern.

Analysis of the Chlorophyll a time series in that region returns a problematic vision of highly irregular patterns, likely caused by several drivers, some connected to the aforementioned Case-2 problem. Mesoscale activity in that area probably enhances the problem, through the development of complex cyclonic fields capable of retaining biogenic materials, and producing a complex optical environment. Still, the detection of this area, within the R2A region, despite difficulties in the interpretation of the satellite data, may herald an interesting opportunity for the mapping of shadowed areas.

Finally, the expression of upwelling filaments is also embedded in the development of the R2B region, which extends westward at the latitude of Aveiro Canyon. Once again the correlation between SST and Chlorophyll a concentration mimics the spatial distribution of the upwelling filaments, especially in the southern side of the axis of the Aveiro Canyon.

The analysis of the remainder maps of figure 4.4.2.4 shows that only 3 years of SST data (2007-2009) are seemingly sufficient to generate a map that, at the k = 10 resolution, resembles the one created with the completed time series. This is an important finding that offers the basis for the development of the routine application of the methodology to rolling archives. The application of the methodology to smaller time series generates considerable computational savings, which are relevant in an operational context.

On the other hand, despite general similarities with the prime regionalization map, the example presented in figure 4.4.2.4 of a classification based on one year worth of data (2007), shows important differences. All five regions (k = 15) are represented, approximately within their long-term spatial boundaries, but a new region is mapped at the triple point created by the contact of R2B, R3A, and R3B.

The stability of yearly classifications was evaluated for other years as well, with similar outputs. The use of one year of data generates less predictable outcomes, limiting the use of the methodology to short-term mapping applications. As in the map of 2007, new clusters are

often created that lack any apparent connection to other regions identified in other years or in the longer term efforts.

As such, it seems reasonable to assume that at least 3 years of high quality SST data are required in order to successfully deploy the methodology, off central Portugal.

The classification of Chlorophyll a time series was also attempted, using the DTW-based methodology.

The same methodology was employed to determine the optimum number of clusters, but very different results were obtained for the Chlorophyll a-based segmentation. In this case, a k = 3 was deemed sufficient, as pointed out in the WCSS plot depicted in figure 4.4.2.14.

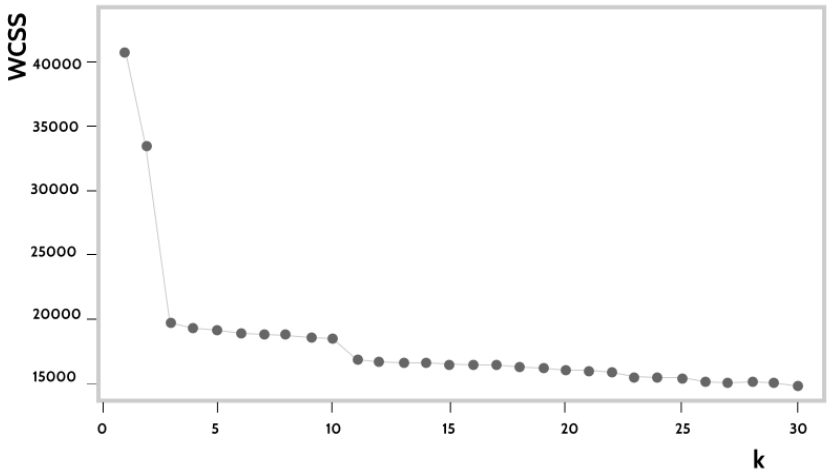


Figure 4.4.2.14. Plot of the WCSS values for different options of k for the Chlorophyll a time series (2003-2014). Notice the breakpoints at k=3, k=11.

Using the information retrieved from the plot, a map was generated, which represents the three main regions (figure 4.4.2.15).

The regions (C1, C2, and C3) obey to a logical distribution of the pixels according to the distance from the shore and thus of the influence of upwelling and other coastal processes (e.g. Alvarez et al., 2012, Moita et al., 2003).

As such, the interpretation of the map is fairly simple, with simple nuances within the inner shelf, caused by the proximity to major river estuaries. This proximity generates a different cluster (C1), where the Chlorophyll a concentration values throughout the time series are modulated by the joint action of oceanic (e.g. SST, mixing, etc…) atmospheric (e.g. FPAR), and coastal (e.g. river runoff and the WIBP) processes. Conversely, the impact of Case-2 water-related anomalies in the retrieval of Chlorophyll a concentration must be considered as well.

Ultimately, the Chlorophyll a clusters offer a dynamic view of the distribution of the coastal/oceanic boundary. Interestingly region C2 appears to be controlled by the coastline more than by bottom topography. The semi-constant distance of the C2/C3 boundary to the coastline is somewhat reminiscent of the R2A region generated by the k = 15 clustering of SST time series. Nontheless, the C2 region is less developed off Oporto, not showing the westward prominence present in the abovementioned SST-based map. The Nazaré Canyon, on the other hand, seems to play an important role in the development and orientation of the C2/C3 boundary. The northern side of the axis of the canyon marks an inflection of the boundary, which assumes a sub-meridional direction, in opposition to the roughly 45° angle it follows to the South (while enveloping the Berlengas Islands).

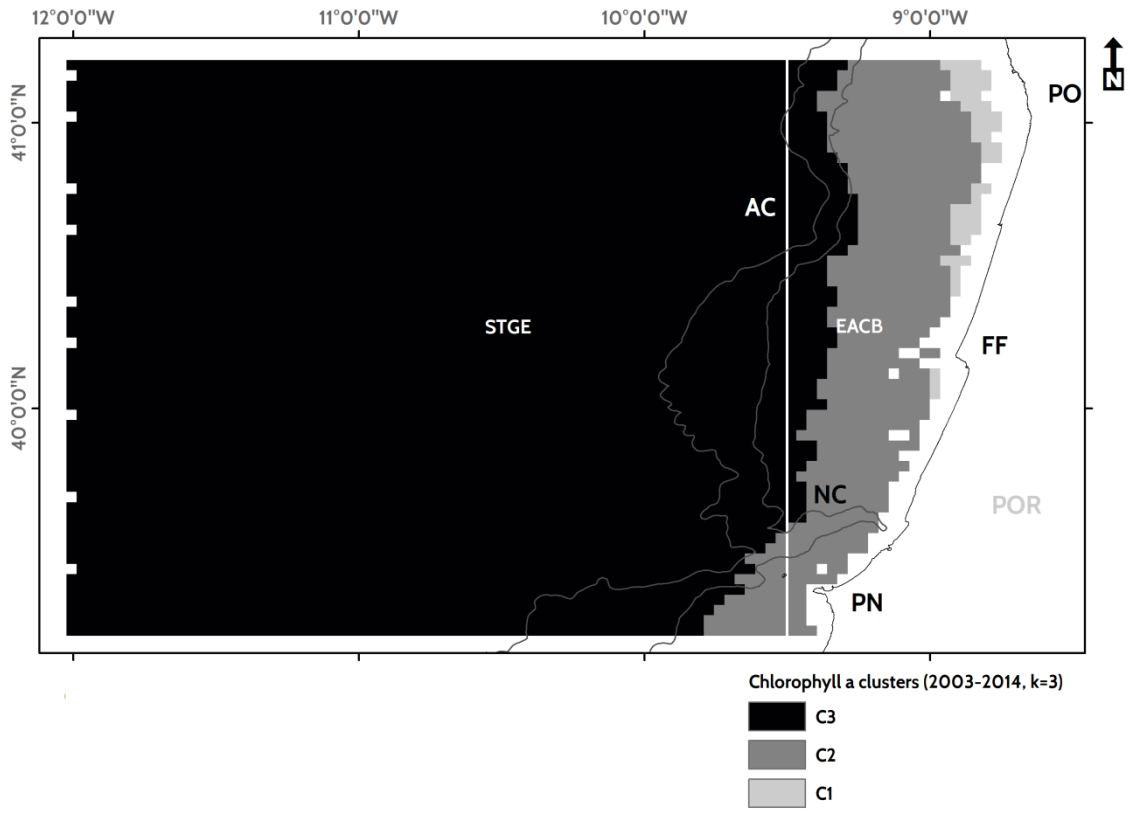


Figure 4.4.2.15. Map depicting the clusters (k=3) of MODIS AQUA Chlorophyll a concentration time series (2003-2014). The boundaries of Longhurst's provinces are depicted as a thick white line. PN: Peniche; FF: Figueira da Foz; PO: Oporto; NC: Nazaré Canyon; AC: Aveiro Canyon; EACB: Canary Coastal Province; STGE: North Atlantic Subtropical Gyral Province.

Considering the importance of both SST and Chlorophyll a, as defining variables of the marine ecosystem, it seemed logical to test the application of multivariate DTW. To this end, SST and

Chlorophyll a datasets were compared and points with high quality information of both variables used in the process.

Interestingly, the complexity of the SST-based DTW is not reproduced in the new clustering, but instead, Chlorophyll a appears to drive the regionalization (at least in terms of optimum number of clusters). Three regions seem to describe adequately the entire study area (figure 4.4.2.16), not unlike what was determined for the Chlorophyll a concentration-based clustering.

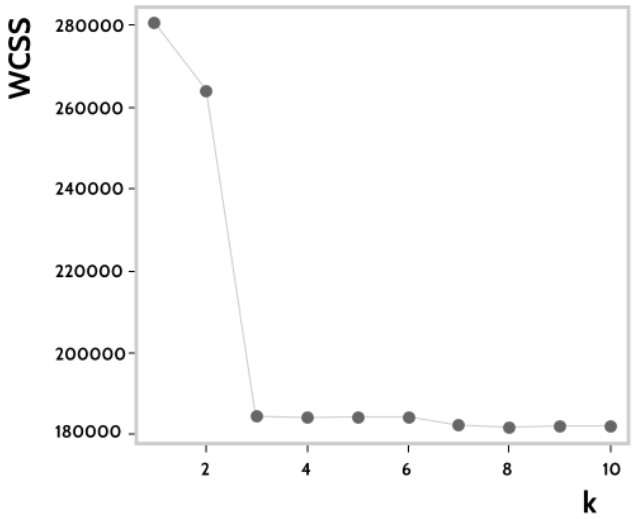


Figure 4.4.2.16. Plot of the WCSS values for different options of k for the multivariate SST-Chlorophyll a time series (2003-2014).

When the clusters are mapped (figure 4.4.2.17), the three clusters provide, once again, an interesting division into two major realms (oceanic and coastal). However, the coastal regions M1 and M2 differ from the previously described C1 and C2. The multivariate regions also differ from the original SST (k = 15) clusters.

In the combined analysis, the coastal clusters resemble the R2A and R2B created from the thermal data. Despite the similarities, the western region (C2) is less developed westwards, although the prominence off Oporto is featured, unlike what was observed with the Chlorophyll a concentration clusters. Furthermore the WIBP-related R1 is absent in this new classification, suggesting that other processes take center stage when both variables are analyzed together. The topographic constraints introduced by the Nazaré Canyon are omnipresent in all schemes, as is the strong control exerted by the orientation of the coastline. Still, the multivariate analysis seems to present an intermediate solution, which is truly a combination of the spatial patterns of physical and biological parameters.

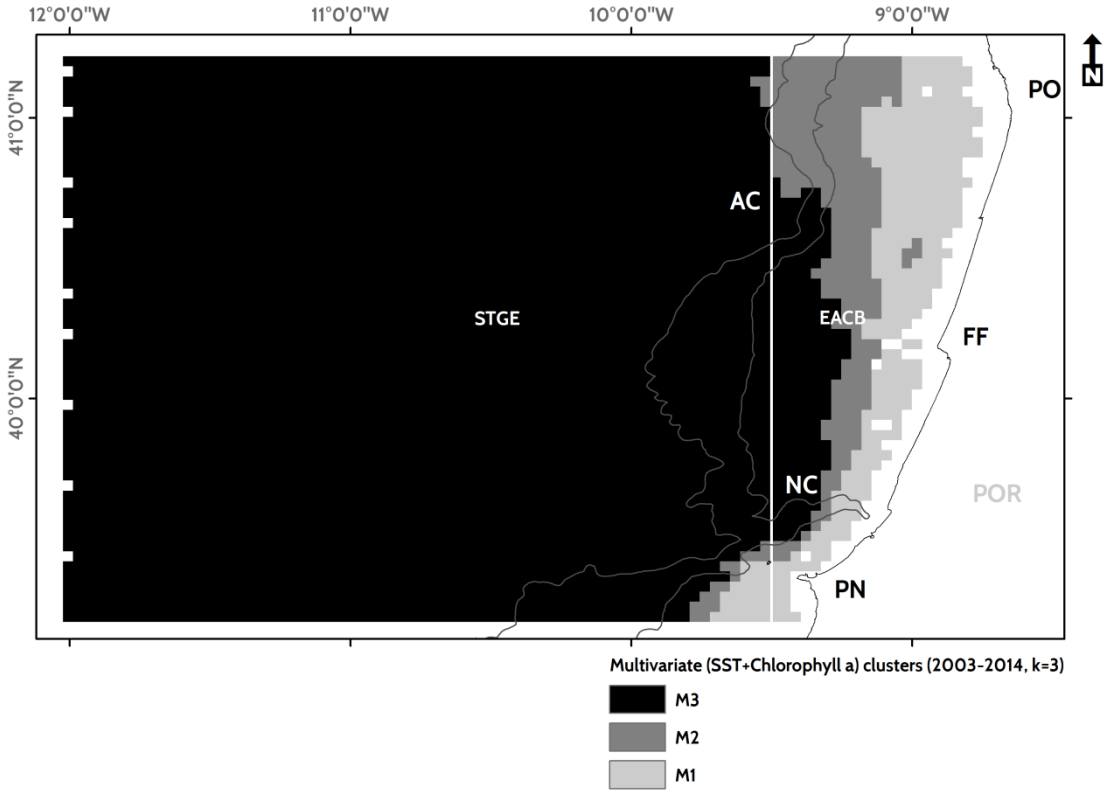


Figure 4.4.2.17. Map depicting the clusters (k=3) created from a multivariate (MODIS AQUA SST and Chlorophyll a concentration) time series (2003-2014). The boundaries of Longhurst's provinces are depicted as a thick white line. PN: Peniche; FF: Figueira da Foz; PO: Oporto; NC: Nazaré Canyon; AC: Aveiro Canyon; EACB: Canary Coastal Province; STGE: North Atlantic Subtropical Gyral Province.

Overall, the different schemes provide different levels of detail. If SST data can be segmented into multiple regions (at least 5) without adding significant amount of noise into the map, the clusters of the other variables may not. Both the Chlorophyll a and multivariate schemes are limited to three coherent regions within the study area. These are essentially coastal (2 regions each) and one major oceanic province.

The classification output reflects the differences between the coastal and offshore domains, as suggested in the classification by Longhurst (2007). The simple visual comparison of this system with either attempt presented in this study, demonstrates the advantages of using long time series and matching methodologies to design the ecoregions. The next step would be to

validate the segmentation using ecological data. This would allow the validation of the tool, not only for physical oceanography and ecological applications alike.

But in the end, it is also important to determine how to identify the boundaries of ecoregions in the context of everyday monitoring activities. How can the research and operational community identify the moving limits of ecoregions in a dynamic medium?

As we have reported, yearly data still offers limited insights into the spatial patterns of fronts, generating fairly erratic clusters, which are not always simple to interpret. The moving application of DTW to longer series, namely 3-year blocks, provides a useful and dynamic tool to determine the current distribution and spatial organization of well described regions. Still, at shorter time scales, these regions are inherently dynamic and the processes that drive the energetic discourse of the provinces must be tracked.

Oceanic fronts, as regions of heightened gradients in the fringes of water masses (Belkin, 2002) may offer a viable tool to determine the most current position of relevant regions. In fact, bioregionalization calls for the study of fronts, as these are the natural and "observable frontiers" of the ocean (Nieblas et al., 2014).

The framework we put forth in this study to classify the ocean off central Portugal, can only be completed by the - perhaps even more significant - development of a true long term climatology of oceanic fronts using a large archive of satellite data. This is the subject of the next sections, in an effort to tackle an existing problem in the scientific literature. Knowledge on oceanic fronts in Portugal is still limited, with few papers (e.g. Relvas et al., 2007, Peliz et al., 2005) addressing this pressing issue, with far-reaching implications for scientific and commercial applications.

4.5. SST and Chlorophyll a Front detection protocol

Earlier sections characterized the processing steps applied to MODIS AQUA data, in order to generate the Level 2 and Level 3 products used in the analysis of SST and Chlorophyll a patterns. The processing of SST and Chlorophyll a concentration, or the clustering of these variables constitute only a part of the broader effort to characterized biogeographic units and the boundaries separating them. To detect SST fronts, and develop a long term record of the spatial distribution of the features, additional processing steps are required.

These steps were applied to Level 2 imagery processed earlier, since applying front detection algorithms to composite images (e.g. 8 days) is inefficient (Otero et al., 2009).

The methodology of Cayula and Cornillon (1992) was applied, with adaptations being introduced to match the requirements of MODIS datasets and image formats.

The first step was to convert the values in the Level 2 imagery into integers. To this end the SST and Chlorophyll a concentration values were multiplied by 100 and 1000 respectively and the raster output stored into GeoTIFF format, at ArcGIS. All subsequent steps were implemented at ArcGIS as well, leveraging the Marine Geospatial Ecology Tools to implement an automated chain, driven by a customized arcpy (python) script.

The detection of the SST fronts followed the overall rules put forth by Cayula and Cornillon (1992), with the particularity that the temperature threshold for front detection was set to 40 (0.4 K) to take into account the detection limit SST using MODIS data and algorithms (Brown, et al., 1999). The window size was selected to encompass 16 pixels, with a stride of 1. This leads to a more computing-intensive process, which in return provides more robust detection capability. Using Cayula and Cornillon's (1992) methodology, it was calculated a minimum single population cohesion of 0.8675 (determines the spatial cohesion of a bimodal distribution) and a minimum global population cohesion of 0.8875 (determines the cohesion for the entire dataset).

For Chlorophyll a fronts, the parameters were adjusted, to meet the different value ranges of phytoplankton gradients. Specifically, the threshold for front detection was set to 40 (or 0.04 $mg.m^{-3}$), to take into account, in a conservative manner, the lower range of calibration and validation matchups used in the development of the algorithms (e.g. NASA, 2015). The remainder of the key parameters was kept unchanged. The window size and stride were maintained at 16 and 1, respectively. This leads to a minimum single population cohesion of 0.8675 and a minimum global population cohesion of 0.8875.

The data thus produced creates daily front maps (figure 4.5.1), which are frequently incomplete due to cloud cover, sun glint and other problems. Therefore, monthly composites were constructed, using the combined daily data. The latter were coregistered and mosaicked as follows. Front magnitude data were not preserved in the construction of the monthly composites. The daily fronts maps were converted into binary maps with pixels being assigned a value according to the presence or absence of a thermal or Chlorophyll a front (0= no front; 1: front). The rasters thus created are then summed, with the final integer assigned to the pixel representing the total detection count of the calendar month.

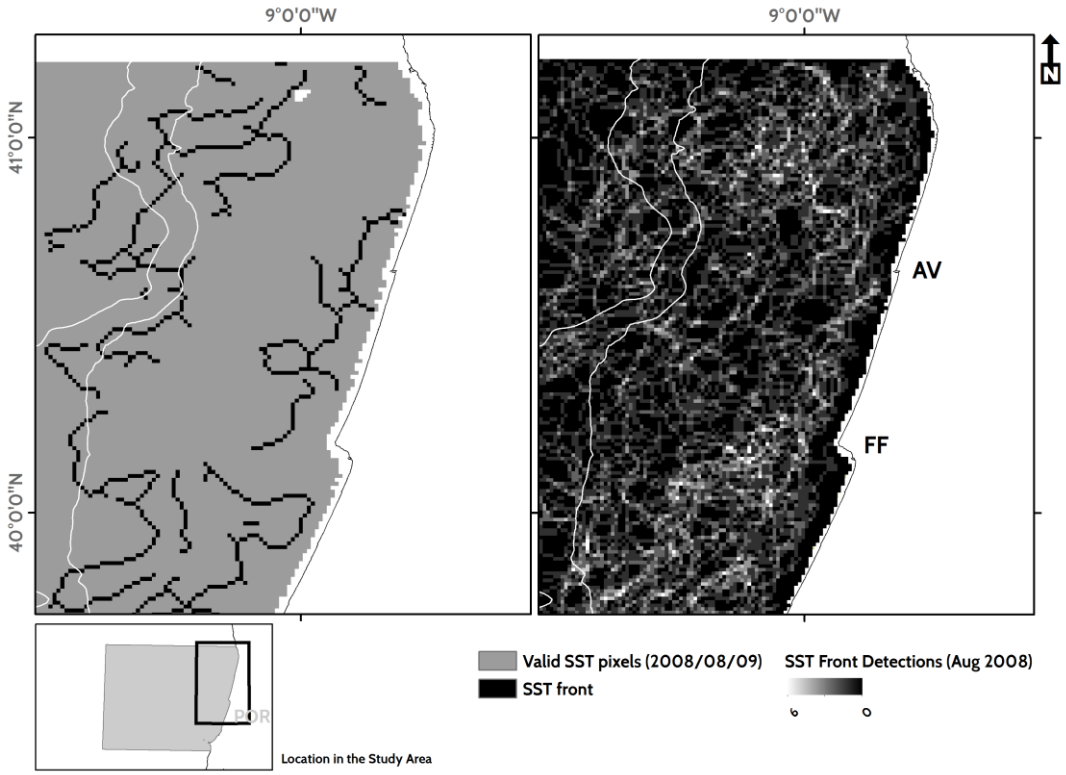


Figure 4.5.1. Left panel: Daily front detection map of August 09, 2008 with 200 and 1600 isobaths represented as well (white lines). Right panel: monthly composite of absolute front detections from August 2008 (right panel). Notice the widespread distribution of fronts in the shelf, but with higher values in the inner shelf and at the edge.

Because the daily maps often do not cover the entire study area as a result of the gaps introduced by clouds and other issues, as reported earlier, a correction method was introduced. The total number of valid SST (and Chlorophyll a) observations was calculated for each month. Afterwards, the total number of front detections per pixel is divided by the valid observations count, yielding a measure of detection probability. This Probability of Front Detection (POD) was used as the reference for frontal analysis in the study period (Ullman et al., 1999, Mavor et al., 2001). Monthly, Seasonal (Quarterly), and Yearly maps were generated using this methodology.

The proven nature and robustness of the method dismissed the need for a thorough validation. Such validation would also require extensive field work, including in distant regions of the study area, calling for a logistical support unavailable in the context of this work.

The use of surface drifters or ARGO floats was precluded by the characteristics of the fronts, namely their convergent nature and along-front flows. As highlighted in an earlier section (*cf.* figure 4.4.1.12), drifting assets may not cross the front and enter a different water mass, but instead be driven away from it, as both flows interact. As such, combing the archives for front-

crossing drifters would prove a cost-ineffective task, likely to be successfully completed only through the automation of the process, in an effort that would greatly exceed the advantages it could generate.

As an alternative, the CORIOLIS (CORIOLIS, 2015) archive was searched to retrieve temperature and salinity data acquired during scientific cruises. The data were collected through the deployment of thermosalinographs (TSG), and valid datasets were downloaded and stored into a geodatabase. The data were then compared with SST and SST front maps acquired in the same day, as depicted in figure 4.5.2. It was determined that, despite the mostly empirical assessment, frontal detections are in agreement with the predicted robustness of the methodology. In the depicted example, the frontal boundary was successfully imaged by MODIS AQUA and detected using the methodology already described. In similar match-ups the comparison yielded comparable results. The deployment of new *in situ* observation networks, including zonal moored arrays, as available in other regions of the world, may provide further information to validate and observe the vertical structure of thermal fronts.

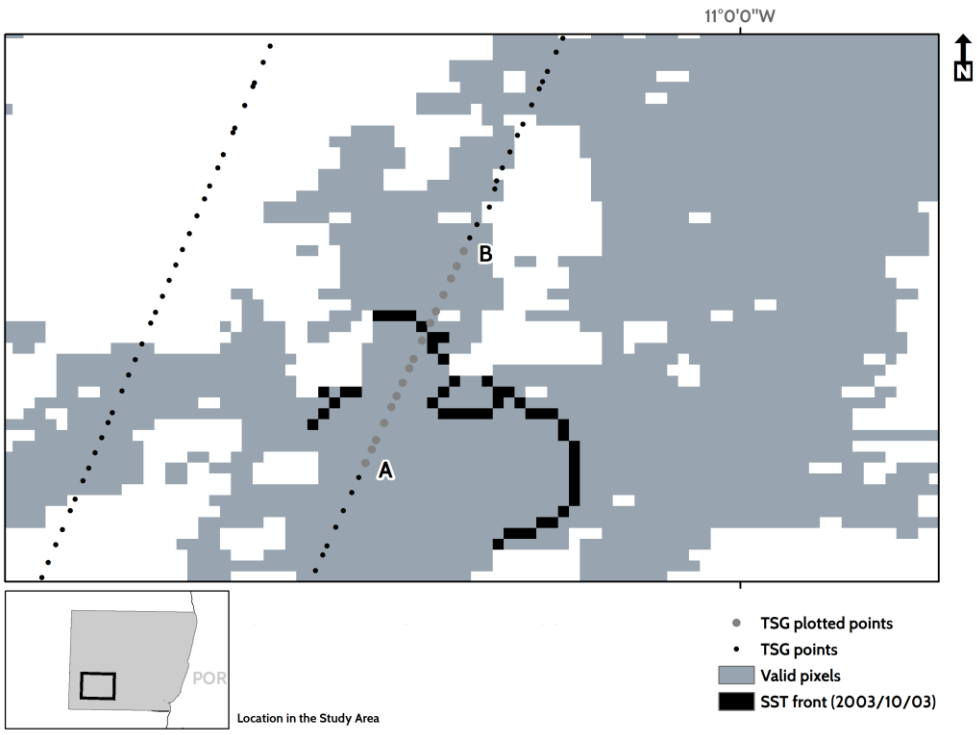


Figure 4.5.2 (Top)

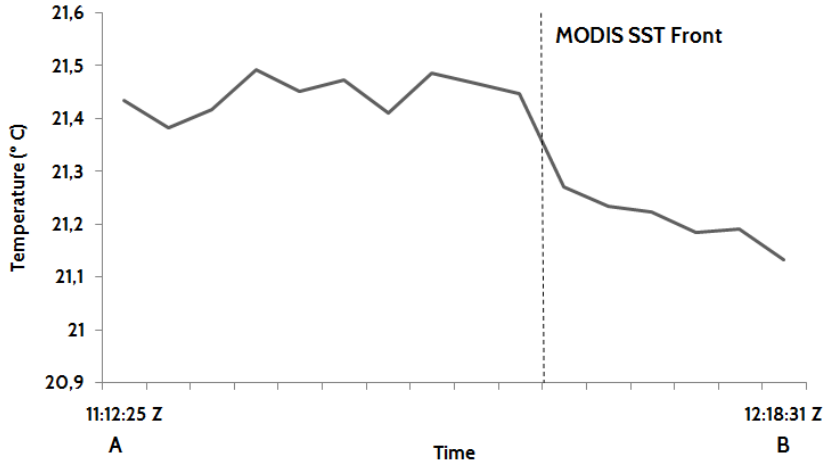


Figure 4.5.2. Top: MODIS SST front map overlaid by the location of the TSG data acquired by an *in situ* asset on October 03, 2003. Lower image: TSG data acquired from the French IFREMER's ship THALASSA with a temporal interval of *ca.* 4 minutes. The data were acquired October 03, 2003, offering a valuable match-up to the remotely sensed image. Notice how the thermal front was accurately detected in the MODIS AQUA SST image.

Figure 4.5.2 depicts a temperature gradient along an approximately meridional transect, with a temperature difference of almost 0.5° C. The interpretation of a time series, consisting of one week of SST imagery, centered in the cruise date, suggests the front is associated with the western border of a tongue of warm water progressing northward. The poleward flow was at that point evolving rapidly and interacting with neighboring water masses, creating increasingly developed mesoscale features in the region. The SST front is not particularly developed, with a moderate gradient, which, nonetheless, spans a short distance. The gradient exceeding the threshold, and the steepness of the slope contribute to the successful detection of the feature, which is somewhat convoluted and clearly represents a dynamic border, which is still evolving as the flows interact.

For Chlorophyll a fronts, it was impossible to accomplish a validation of the methodology. To surpass this limitation, the generated products were visually inspected, by overlaying the frontal boundaries to the original Chlorophyll a fields.

Overcoming the difficulties in validating Chlorophyll a fronts will require extensive research and the deployment of sizable sampling campaigns. That was clearly outside the scope of this work. Because Case-2 waters offer additional challenges, which were not yet overcome by existing algorithms, the quality of retrievals is still limited, when compared with the maturity of SST data. For these reasons, and considering the objective of the study, Chlorophyll a fronts were in fact processed only for a subset of the period of study, in order to assess the

collocation of these features with the thermal gradients and the development of a preliminary spatial distribution map, during the peak of the upwelling season (August). Future research may address the full spatial distribution of Chlorophyll a fronts, but only after the effect of non-linearly correlated water constituents over the algorithms is reduced.

For the August subset, the detection of simultaneous SST and Chlorophyll a fronts required an additional step. Since the area covered by valid pixels in the SST and Chlorophyll a image is different due to the specifications of each algorithm, a common retrieval area had to be defined.

To this end, an arcpy script was designed for ArcGIS, that iteratively analyzed the SST/Chlorophyll a image pairs and created a mask describing the common valid area. The mask was then used to extract the valid pixels in both images. These subsets were then submitted to the front detection procedure already described for each variable and the output converted to a binary (0= no front, 1= front detected). The SST mask image was then submitted to a 3x3 moving window filter, which attributed a value of "1" to the pixels neighboring the fronts. These intermediate step was important to ensure the detection of closely located Chlorophyll a fronts that were notwithstanding, perfectly aligned with the SST boundary. This is particularly important, considering the scale of work, and the fact that Chlorophyll a fronts may often be present in the immediate vicinity of the thermal counterpart, but not in the same pixel.

Both images were then evaluated by a conditional tool, which assessed the simultaneous presence of a Chlorophyll a front over an (filtered) SST boundary. A raster was then generated with a numeric code, where '0' would be assigned to front-free pixels, '1' to Chlorophyll a front-only, '2' to SST front-only, and '3' to simultaneous fronts.

This process enabled the development of a somewhat simple, yet efficient mapping process to detect fronts detected simultaneously in both images.

With the processing steps complete, the large dataset was stored and analyzed, with the most significant findings and cartography produced described in the following section.

4.6. SST front climatology

In this section we present a climatology of Sea Surface Temperature fronts, created from the analysis of daily MODIS AQUA imagery acquired between January 2003 and December 2014. The data shall be presented and discussed in a sequential manner, from the broader spatial patterns in the time series to the effect of specific features in monthly maps.

The relevance of a front climatology cannot be overstated. With extensive operation and scientific ramifications, thermal fronts are not just boundaries but also important regions for marine life (Reese et al., 2011). But the near real-time study of fronts using *in situ* assets is both difficult and time consuming. The scarcity of long term, high resolution frontal probability maps, precludes the development of solutions that would require such information. Simultaneously, the experience gained in the processing of a massive dataset, comprising thousands of daily satellite images, created the required conditions to develop near real-time and automated processing chains. Such chains are paramount in the planning and execution of scientific cruises or operational data distribution, allowing the timely identification of relevant features in the target areas.

This section describes the results obtained after the construction of an 11-year time series of daily frontal activity off central Portugal. The frontal probability maps helped characterize events and oceanographic features, offering a novel view of an important area of the Portuguese sea.

Figure 4.6.1 depicts the complete set of fronts detected in the time series (January 2003 to December 2014). The fronts were mapped in terms of Probability of Detection (POD), as described in the methodological section.

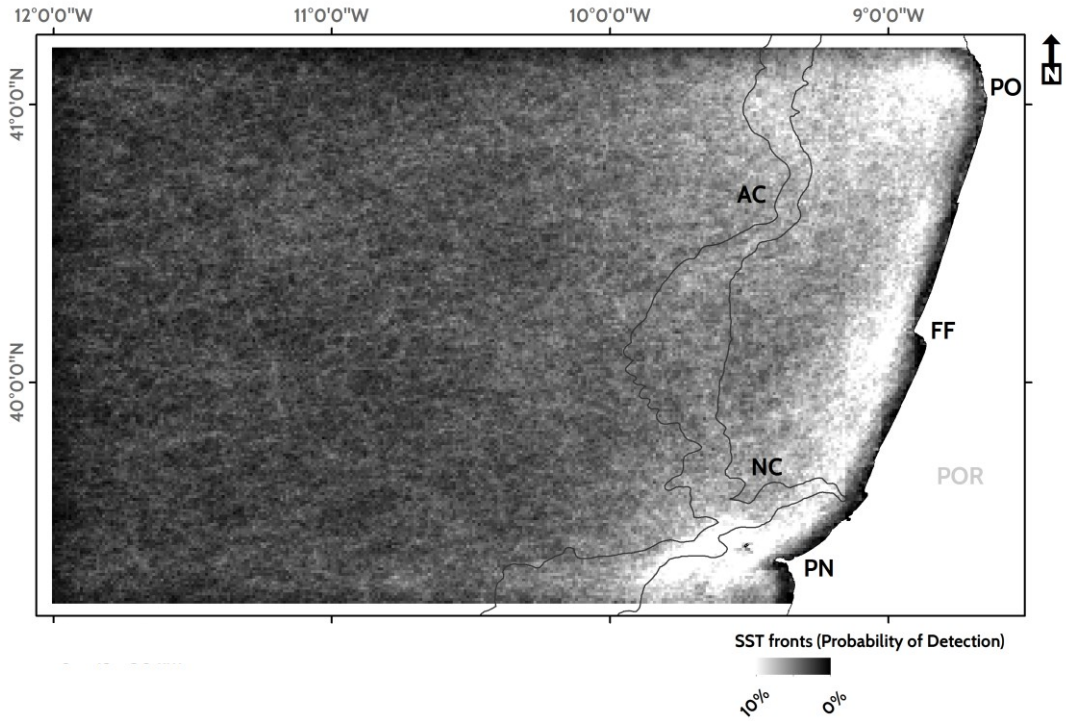


Figure 4.6.1. The Front Probability Map for the complete time series (January 2003 – December 2014). AC: Aveiro Canyon; NC: Nazaré Canyon; PN: Peniche; FF: Figueira da Foz; PO: Oporto; POR: Portugal. 200 and 1600 m isobaths represented as dark lines.

The long term analysis depicted in figure 4.6.1 can only provide a broad view of the occurrence of fronts, highlighting the regions where frontogenesis is more likely to occur. The map is insensitive, nonetheless, to seasonal variability and inter-annual differences in frontal distribution.

Overall, the SST frontal map shows maximum POD values reaching 10%, with most frequent detections occurring in a narrow band found along the coast. The high frontal POD in that band is generally limited to the first 25 km west of the shore line.

Along the coast, four hotspots can be identified by the generally higher values of POD. These regions are highlighted in figure 4.6.2.

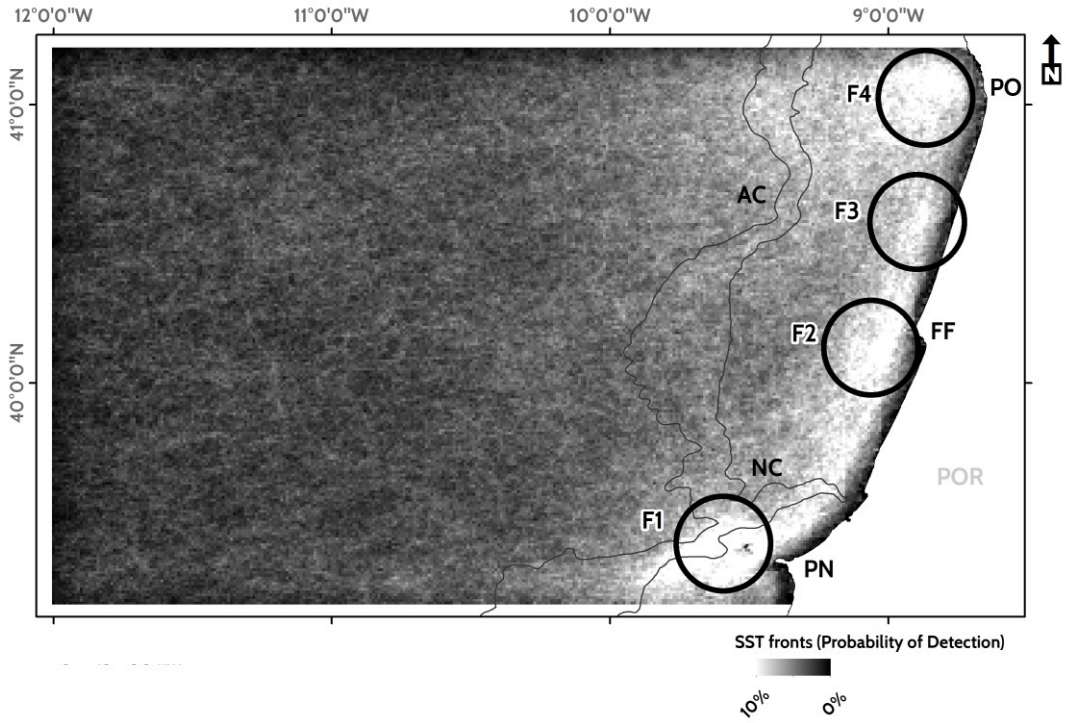


Figure 4.6.2. The Front Probability Map for the complete time series (January 2003 – December 2014) with centers of high POD marked (F1-F4). AC: Aveiro Canyon; NC: Nazaré Canyon; PN: Peniche; FF: Figueira da Foz; PO: Oporto; POR: Portugal. 200 and 1600 m isobaths represented as dark lines.

Three of the 'hotspots' (F2, F3, and F4) are located in the vicinity of major river mouths (Mondego, Vouga, and Douro). These are probably generated by the interaction of river output with coastal waters. In previous sections we described the development of river plumes (and fronts) that are part of the broader WIBP system. We hypothesize this signal was created by such system.

The other high POD spot (F1) is located in the vicinity of Cape Carvoeiro-Berlengas Islands and this had been identified previously by Relvas et al. (2007) and Peliz et al. (2005). These authors justify the recurrent presence of fronts through a combination of topographic control, tides, and meridional gradients.

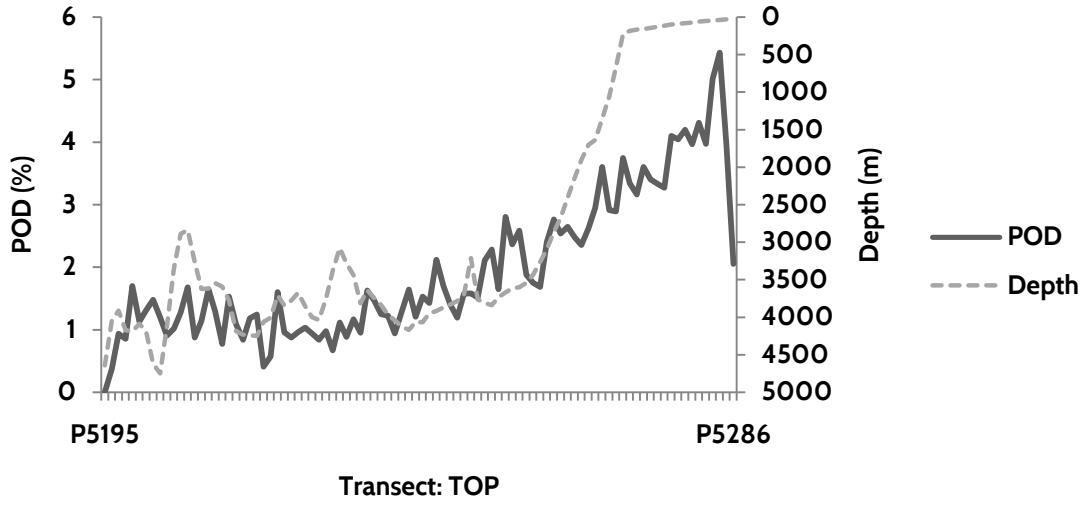
We emphasize the similarity in the distribution of the fronts in the F1 region and the spatial patterns of the clusters generated after DTW analysis of SST, Chlorophyll a, and multivariate data.

The POD values offshore tend to decrease rapidly, although a subtle increase in the vicinity of the shelf edge is observable. As Relvas et al. (2007) observed, the decrease in the values strongly suggests that several processes, including filament development are subject to much greater variability, preventing the development of regions of higher POD. Nonetheless, a comparison of our generalized map with that of Relvas et al. (2007) shows significant differences, some of which arise from a greater wealth of data, and higher spatial resolution used in our approach.

The influence of bottom topography over the recurrence of thermal fronts was assessed through the comparison of POD values against the West-Iberian bathymetry digital terrain model (WIBM2009) produced by the *Insituto Hidrográfico* of the Portuguese Navy (Quaresma, 2009).

The data were resampled and collocated, and subsequently extracted using the pixel centers of the Level 3 grid cells employed to create the transects mentioned in previous sections.

The results are displayed for three representative zonal transects (TPN: Peniche; TFF: Figueira da Foz; TOP: Oporto) in figure 4.6.3.



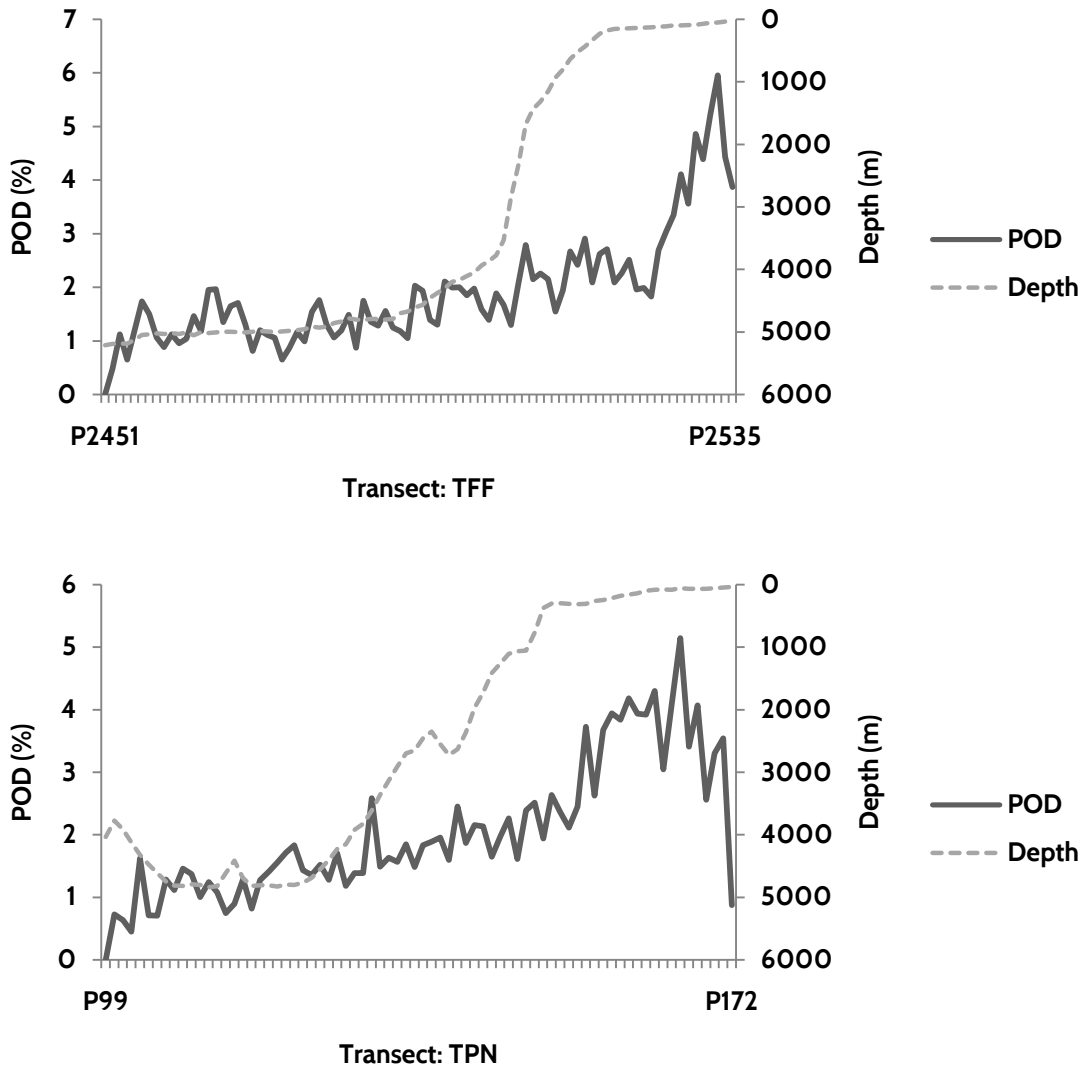


Figure 4.6.3. Plot depicting SST frontal POD and bathymetry in three transects (TPN, TFF, and TOP). Notice the higher POD values in the inner shelf.

The comparison confirmed previous findings that suggested a higher frequency of frontal activity in the shelf (Relvas et al., 2007, Castelao et al., 2008). It must be emphasized that the POD of frontal activity is perhaps the least studied aspect of all addressed in this study. This is especially true in the study area, where little information is available in the literature. As such, it is difficult to compare our results with previous studies, and will, therefore, attempt a comparison with equivalent systems, better explored than Portuguese waters.

The analysis of the full set can only provide a broad understanding of frontal dynamics, with specific features masked by the sheer number of detected fronts. The transects reflect that reality and little information can be extracted. It is noteworthy that no shelf edge front is

discernable. Most frontal activity is restricted to the inner shelf, with POD values decreasing before the edge. Despite localized increases in the vicinity of the slope, it is not possible to clearly identify such edge-related front, described in other studies (e.g. Breaker et al., 2005).

Across the study area, the maximum POD values vary, as previously mentioned, but the meridional variability is less pronounced than the zonal gradients. All three transects displayed the maximum values close to the coastline and range from 5 to 6% of observations.

It is also interesting to notice the variability of POD values across the transects, even in the open ocean. The POD curve is not smooth but punctuated by sharp, localized increases and troughs, which seem to suggest the heterogeneity of the spatial distribution of the processes governing the frontogenesis

But considering the likely spatial and temporal variability of fronts in the study area, it was deemed necessary to analyze these dynamics and adequately represent them in different formats. To this end we employed the Empirical Orthogonal Function (EOF) (or Principal Component Analysis, PCA), considering the robustness of the method and the ability to compare results with other regions. This is only possible because EOF has been adopted in the study of thermal fronts elsewhere, including California (Castelao et al., 2008).

The EOF was first published by Pearson (1901) and henceforth used abundantly in various fields, including remote sensing and oceanography. In summary, EOF generates a set of Modes (or Principal Components), after the application of an orthogonal transformation. This transformation creates a new set of linearly uncorrelated variables, which spur from a set of previously correlated data.

EOF offers several advantages, namely the simplification of complex datasets (matrices) into smaller, more manageable and unmixed outputs. We refer the readers to Shlens (2003), Wold et al. (2002), and Ringnér (2008) for detailed descriptions of EOF.

In this study, the matrices fed to the EOF, which was run at ArcGIS 10.2, were the rasters containing the values for POD of Thermal Fronts within the study area. The monthly mean fields, totaling 12 files, were thus analyzed and several outputs were generated for spatial and temporal analysis. Subsequently, the entire sequence of monthly POD maps (n=144) were also submitted to EOF analysis, for the creation of a time series with the fields calculated for each of the first three significant modes.

Following Castelao et al. (2008), we adopted the method devised by Overland and Preisendorfer (1982) to determine whether the Modes generated by EOF were significant (for $\alpha = 0.05$). To meet this requirement, each mode must describe at least 2.4% of the variance.

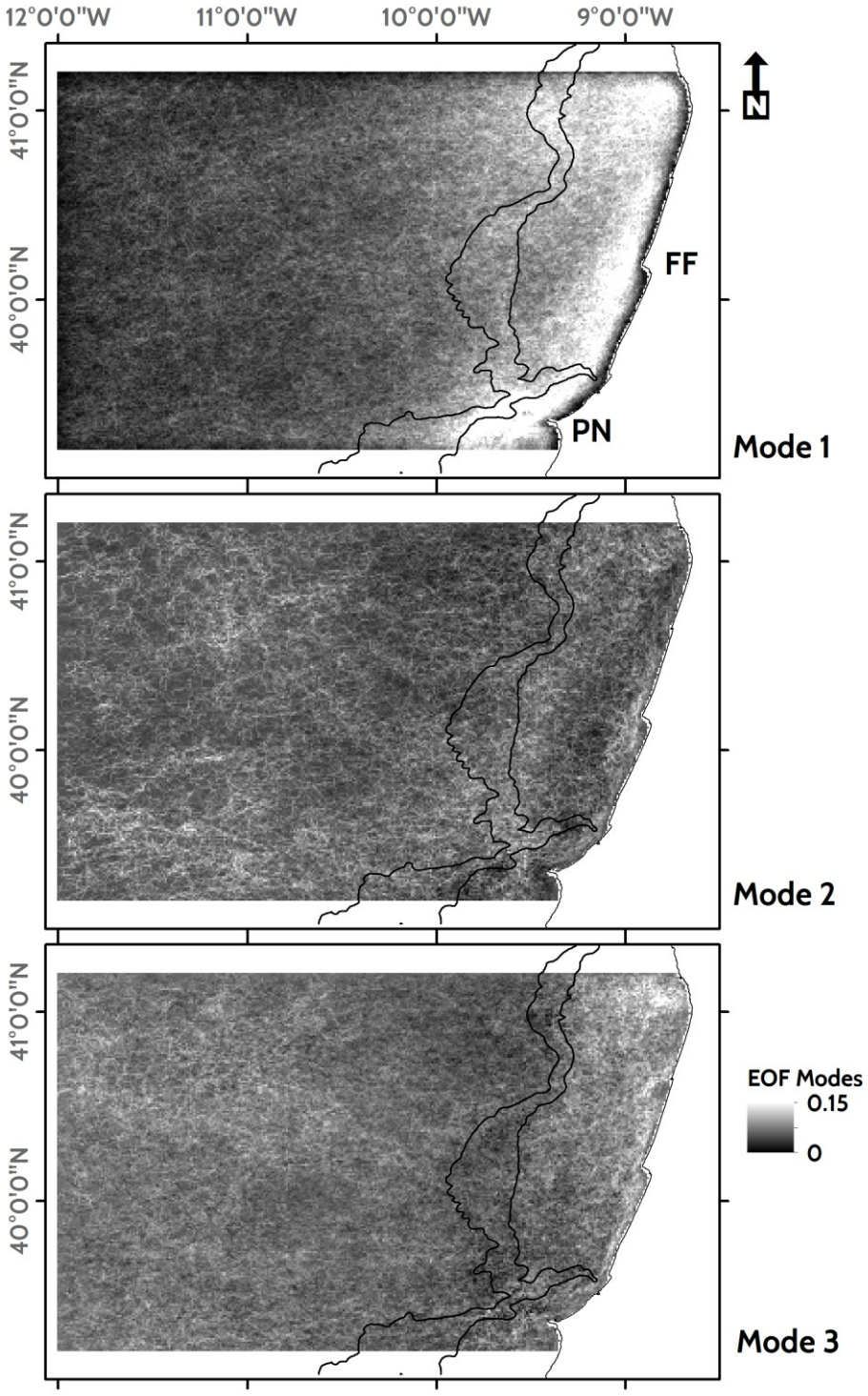


Figure 4.6.4. EOF Modes for the POD of thermal fronts, calculated from the mean POD for each calendar month (12 files). The 200 and 1600 m isobaths is displayed as a dark solid line. PN: penich; FF. Figueira da Foz.

The EOF modes represented in figure 4.6.4 were considered significant, since they describe, respectively, 42.1, 8.7, and 8.4% of total variance.

The different modes are responding to different aspects of frontal dynamics. Mode 1, for instance, is characterized by generally high values, with a spatial pattern close to that of the overall POD (figure 4.6.1). The spatial patterns of the amplitude values for this mode seem to represent the seasonal variability associated to upwelling months and the offshore movement of fronts, which we will describe in greater detail in the seasonal analysis of POD.

EOF 2, on the other hand, seems more closely correlated with the development of the poleward current, which is often characterized by edge-trapped tongues with significant mesoscale activity. In the map, in a way coherent with this interpretation, EOF mode 2 values increase nearshore and close to the shelf edge. Between both regions an area with lower values is visible. A transect at the latitude of Figueira da Foz (TFF) is depicted in figure 4.6.5, clearly showing the EOF amplitude drop.

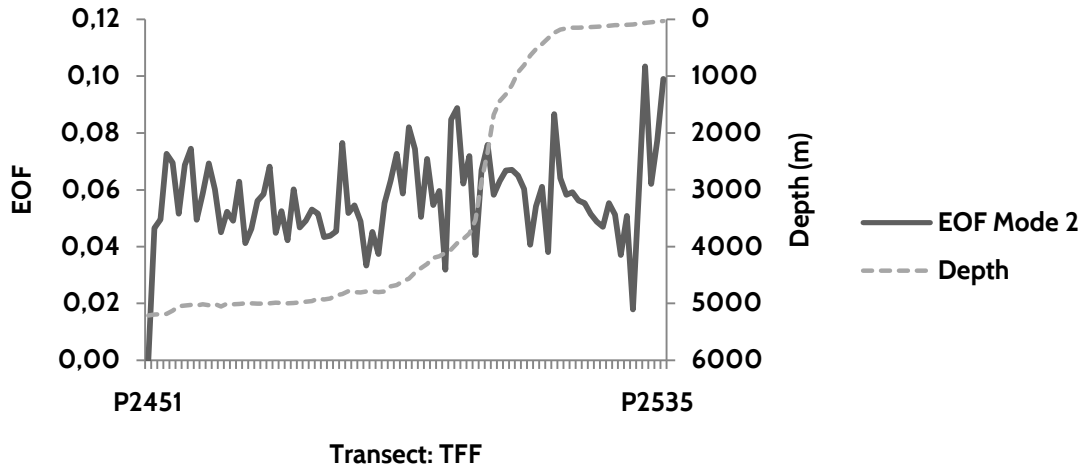


Figure 4.6.5. Plot depicting EOF Mode 2 and bathymetry in the TFF transect. Notice the lower EOF values in the shelf.

The southwestern regions of the study area are characterized by higher EOF 2 amplitude. This fact, together with the along-shelf values, may also be a signature of the action of the poleward current, and its interaction with other flows. Such interaction is responsible, as abundantly described thus far, for developed eddy fields and associated frontal boundaries.

Finally, EOF 3 presents interesting results, with higher values in the vicinity of estuaries, where river plumes are often active. In contrast, lower values were found for small regions southward of the Douro and Vouga river mouths. The nature of these ‘shadow’ zones is still unknown. Several hypotheses may be put forth. The formation of well developed and recurrent river plumes is amongst the possibilities. However, this would not explain why such a shadow zone is absent (or at least less conspicuous) at the Mondego river mouth.

The zonal bands of frontal activity described by Peliz et al. (2005) also seem to be present offshore in the third EOF mode, north of 40.5° N.

The spatial analysis can be supplemented by the depiction of the seasonal variability as suggested by the amplitude of EOF Modes' eigenvectors (figure 4.6.6).

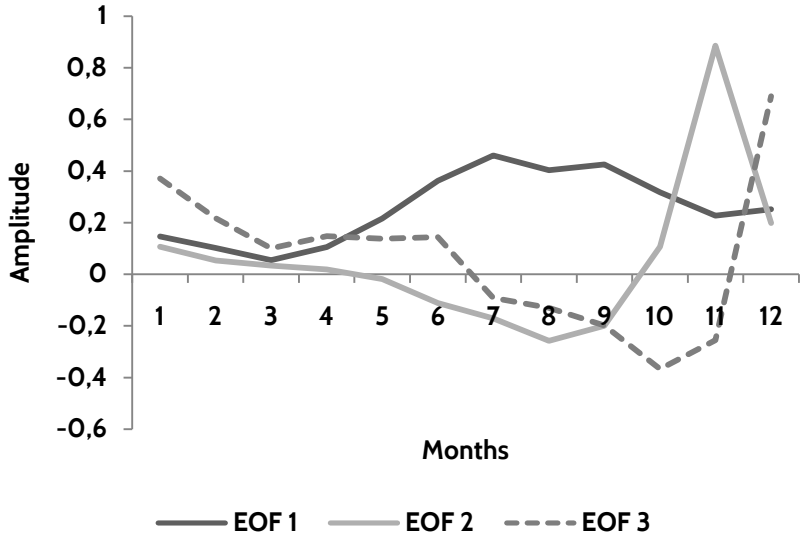


Figure 4.6.6. Amplitude of the eigenvectors of the first three EOF Modes, calculated from the mean monthly POD of thermal fronts.

The temporal variability of the amplitude of EOF 1 is consistent with the seasonal variability of frontal activity resultant of upwelling. The higher values are found in warm months (July-September), which coincide with the upwelling maxima recorded in our and previous studies. Simultaneously, the minimum amplitude recorded in March, occurs nearly simultaneously with the upwelling minima.

On the other hand, EOF 2's maximum amplitude occurred in November. The hypothesis of a combination of offshore front migration and the influence of the poleward current may not be discarded. The latter is supported by compelling albeit circumstantial evidence. The magnitude of EOF 2 in the southern sector of the study area and along the shelf edge, combined with the low magnitude calculated for the inner shelf, seem to suggest an important role of the IPC in this mode.

The amplitude of EOF 3 displays some similarities with that of EOF 2. However the minimum occurs when EOF 2 is peaking. The distribution of the higher values, north of the Mondego river, and the sharp increase of the amplitude in the winter months, suggests some correlation with the WIBP. As river discharge increases, so does EOF 3, which would also explain the 'hotspots' in the vicinity of the river mouths.

The analysis of the complete POD time series (n=144) yields a similar image, although the data gaps in the frontal maps (monthly) difficult the processing of the datasets.

The first three EOF modes were also significant (37.9, 15.7, and 14.9%). The time series of eigenvector amplitudes of EOF 1 is displayed in figure 4.6.7.

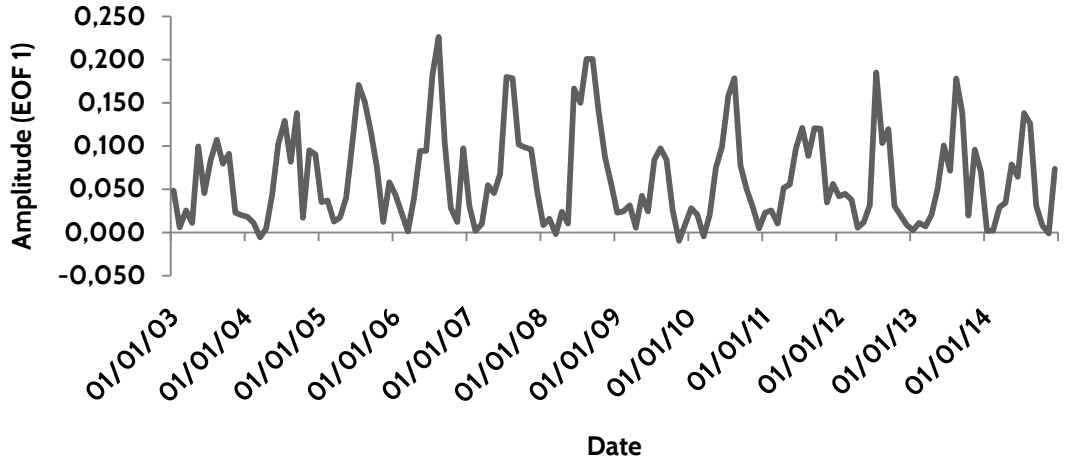


Figure 4.6.6. Amplitude of the eigenvectors of the first EOF Mode, calculated from the complete series of monthly POD of thermal fronts.

In the plot, the upwelling-related cycle is clearly visible. Although some differences are expected in the EOF modes of the mean monthly POD and full set, the pattern depicted in figure 4.6.6 seems to be represented as well in the new iteration.

The lower amplitudes are recorded more frequently in March (58.3%), followed by February (25%). Oppositely, the higher values are found in August (50%), followed by July (41.7%).

The time series resembles that of the UI_{SST} and in fact they are correlated (albeit weekly), with an $r = 0.69$ (significant for $p < 0.01$).

The use of EOF-based indices is widely used in oceanography as in the Multivariate El Niño/Southern Oscillation Index (ENSO MEI) (Wolter et al., 1993). In this case, the MEI is calculated as the first principal component derived from the bimonthly normalized fields of six variables. Previous studies, have dedicated a growing attention to the development of indices capable of describing, synoptically, the conditions of a given system. Such indices provide important insights into the highly dynamic coastal habitats, at adequate temporal and spatial scales (Tapia et al., 2009). The importance of the indices is directly correlated with its ability to describe in a simplified manner meaningful processes that impact geological or atmospheric phenomena and biological communities.

PCA-based indices are used extensively to this end, varying the datasets employed in the analysis and the objectives they serve (Black et al., 2014, Galindo et al., 2009).

Under this principle, and given the correlation of EOF modes (amplitudes and spatial patterns) with known physical processes in the study area, it is worth considering the development of a similar index to monitor relevant oceanographic processes off Portugal. Such index could replace or supplement existing (complex) models or, on the other hand, simple and region-specific metrics (e.g. UI_{SST}), by leveraging the wealth of data provided by a growing constellation of satellites and *in situ* assets.

The development of the index we now propose is outside the scope and objectives of our current work, requiring extensive analysis of multiple datasets in order to ascertain the meaningfulness and applicability. However, the datasets generated create a favorable setting to further develop the concept. The time series of thermal fronts (which can be updated seamlessly to continue the coverage into the future) offers an additional layer of information, which enables a spatialization of relevant processes and is itself one of the few oceanic habitats that can be efficiently monitored from space, even in deeper waters.

The analysis of the time series of monthly mean POD of thermal fronts and the analysis of individual months is likely to provide further insights into the dynamics of the these features.

In the following paragraphs we shall address the seasonal variability of POD of thermal fronts, through the analysis of the mean monthly POD maps generated from the dataset. The twelve files constitute the basis of the climatology, which presents an unparalleled vision of frontal dynamics off central Portugal.

Firstly, we shall analyze the quarterly Probability of Detection maps, which provide a synoptic view of frontal activity over the 11 years of study. The seasonal maps were created as the sum of probability of detection for three calendar months (January-March; April-June; July-September; October-December). The groups broadly match oceanographic and atmospheric processes that govern flows and features off central Portugal, including upwelling, the surface expression of the poleward flow or the WIBP.

Figure 4.6.7 displays the probability of detection of SST fronts in the period between January and March (2003-2014).

Frontal activity in the period is generally weak, with similar POD values within the shelf and in the open ocean. Fronts are largely influenced by the presence of the freshwater lens know as the Western Iberia Buoyant Plume, with its distribution mimicking that of the SST gradient that characterizes the feature. As it will be explored in greater detail in the next section, the high POD values are also found within the DTW-based R1 region (SST, k=15).

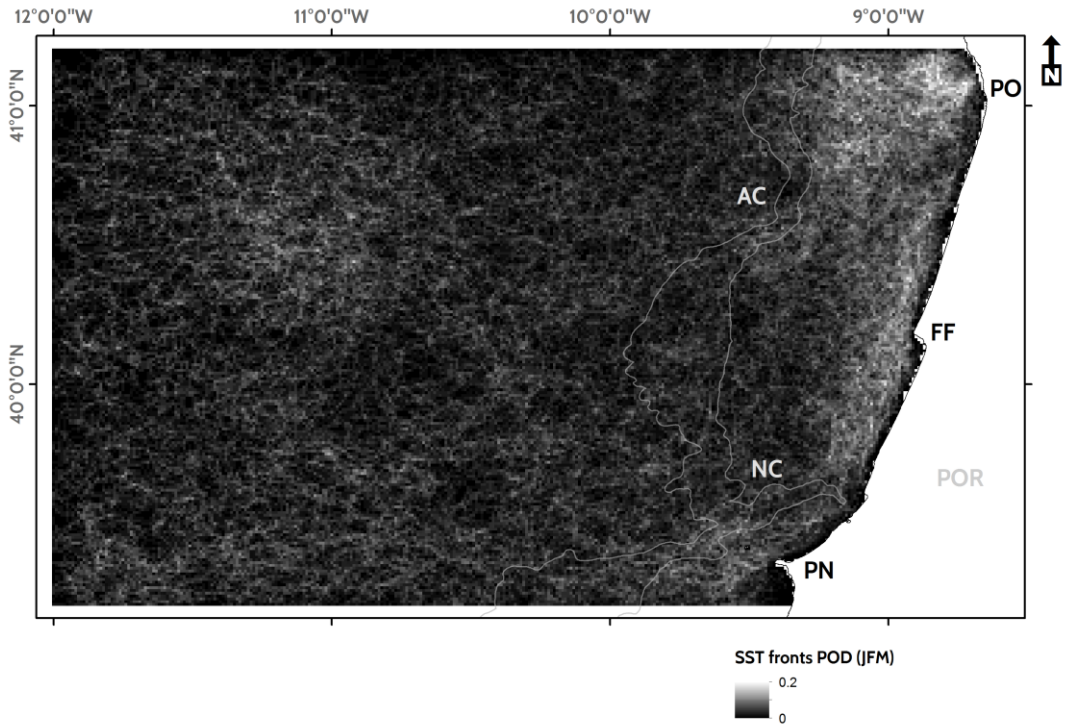


Figure 4.6.7. Seasonal POD of SST fronts (January-March) for the period of study (2003-2014). Please see text for more details.

Regions with higher POD are located North of the Nazaré Canyon, developing offshore in the proximity of the Mondego, Vouga, and particularly the Douro rivers. 'Shadow' zones are visible south of the Vouga and Douro rivers, which are reminiscent of the same areas signaled in the EOF Mode 3. In it, the same regions were characterized by lower amplitude values and it was hypothesized these matched areas influenced by river plumes (the halo within the tidal front). The seasonal map seems to support the hypothesis, as it replicates the same patterns. The presence of the plumes leads to the development of persistent frontal fields at the boundaries of the plume, and in the vicinity of the recirculation areas and far-fields.

The relevance of rivers in the POD during the period was further analyzed. Figure 4.6.8 depicts the same POD distribution mapped in figure 4.6.7, but a set of sampling points were added. The duplication of the map is intended to improve readability of the map, namely in the coastal area, where the points are now represented. These sampling points were used to extract POD values and generate an along-coast transect.

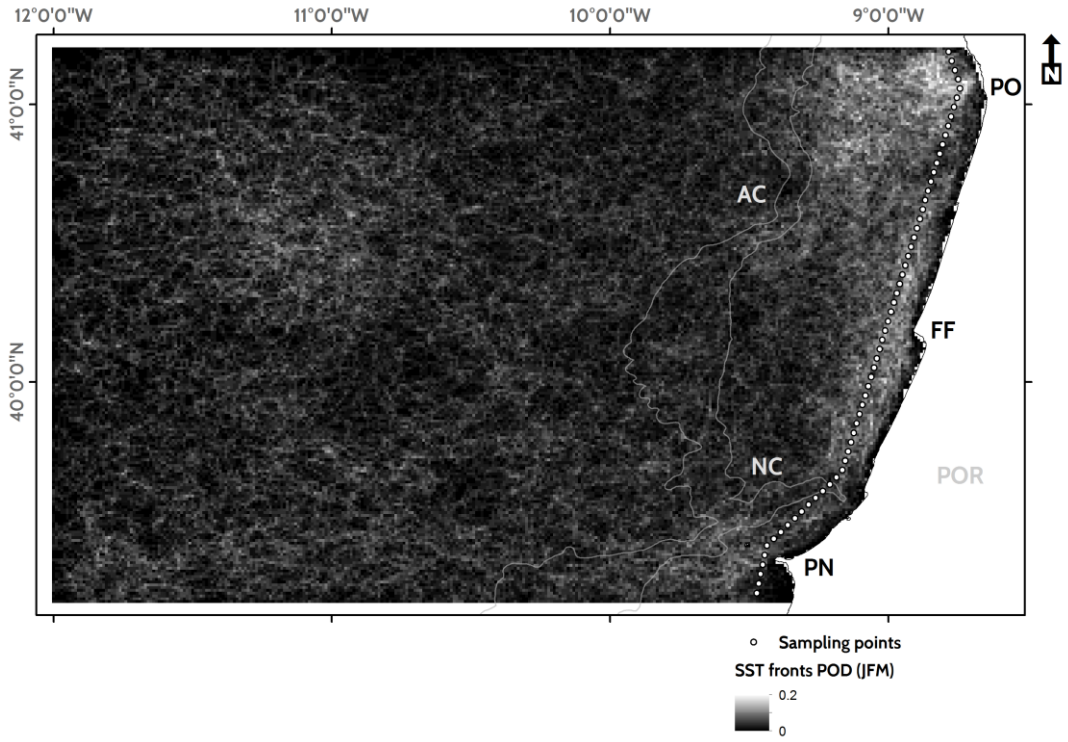


Figure 4.6.8. Sampling points used to retrieve seasonal POD of SST fronts (January-March) for the period of study (2003-2014). Please see text for more details.

The transect depicted in figure 4.6.9 highlights the role of river discharge in the establishment of frontal regions. In the regions between the area of direct influence of river plumes and in the neighborhood of Nazaré Canyon and Berlengas Islands, the values decreased. The minimum POD was in fact recorded in the Eastern terminal area of the axis of Nazaré Canyon.
The development of fronts associated to river plumes has important consequences to physical and biological processes with significant accumulations of material along the frontal axis (Otero et al., 2009). Sediment transport, for instance, is known to be affected by the development of these structures, which control the velocity and energy of the transport flow (Wang et al., 2007). This is thus another application, to which the study of frontal probability and distribution can provide important insights.

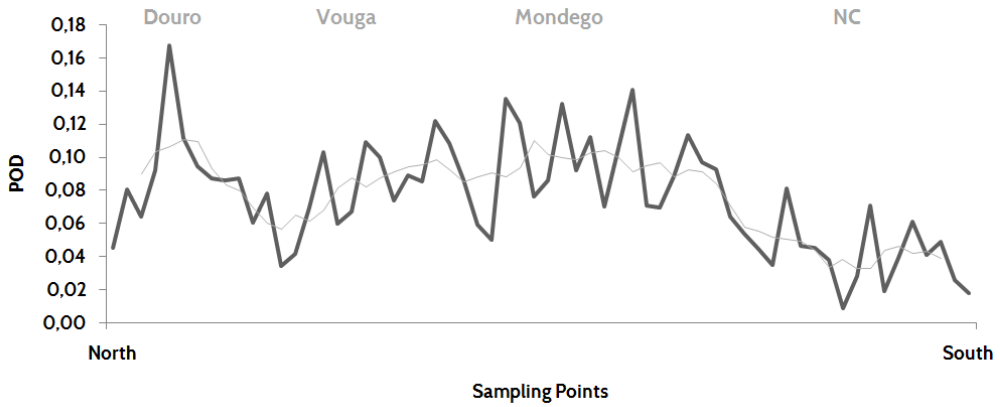


Figure 4.6.9. POD of SST fronts in the January-March interval (2003-2014) along the transect depicted in figure 4.6.8. Notice the increased POD values in the vicinity of major rivers (Douro, Vouga, and Mondego) and lower values in the southern end. NC: Nazaré Canyon. Grey line: moving average (period = 5).

POD values to the North of Nazaré reflect river discharge, as aforementioned, but it is important to emphasize that the presence of the WIBP is seemingly continuous along the coast. The width and thickness of the lens varies along and cross-shelf (Peliz et al., 2002), but the concentration of SST fronts in the region depends first and foremost on a combination of causes, including the occurrence of upwelling or downwelling events, and high-energy local processes (Otero et al., 2009). Therefore, although the distribution of fronts and the WIBP are correlated, localized maxima occur in response to specific interactions, with river plumes playing a fundamental role.

We shall now look into the separate SST frontal maps of January (figure 4.6.10), February (figure 4.6.11), and March (figure 4.6.12), in order to explore the monthly contributions to the overall POD.

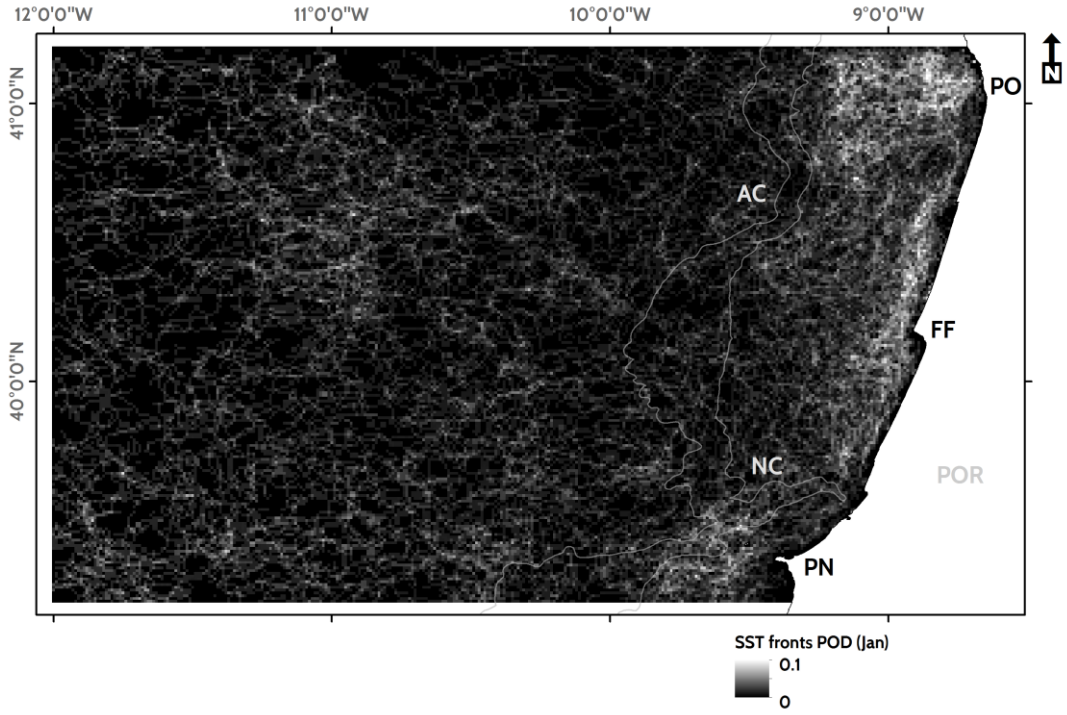


Figure 4.6.10. Monthly POD of SST fronts (January) for the period of study (2003-2014). Please see text for more details.

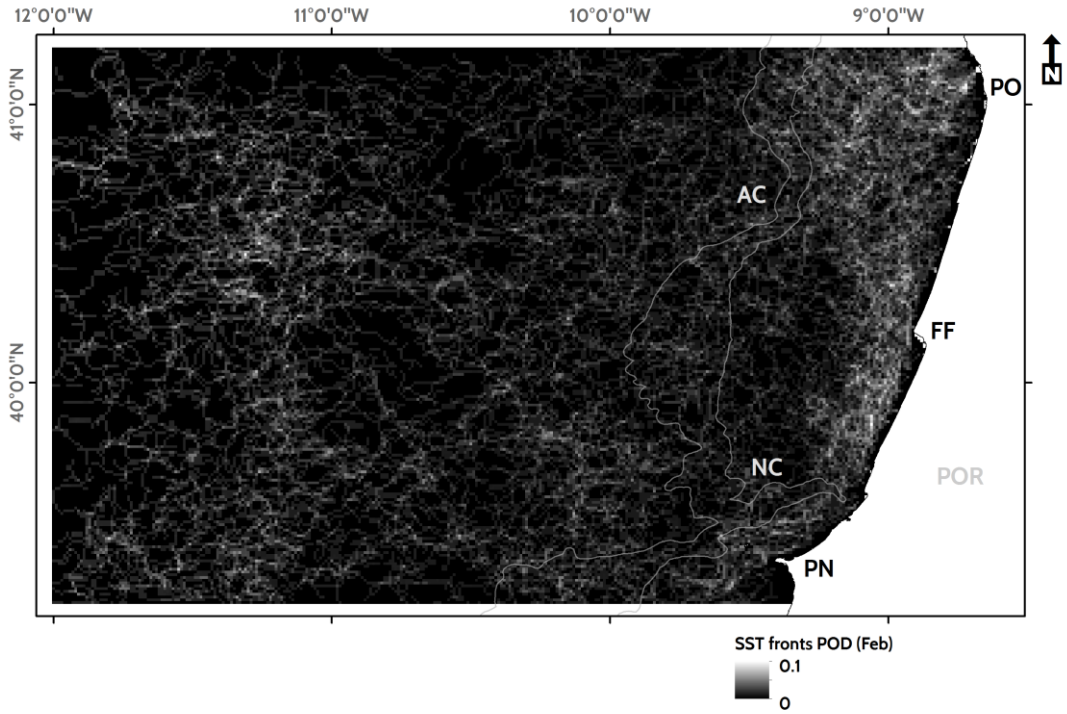


Figure 4.6.11. Monthly POD of SST fronts (February) for the period of study (2003-2014). Please see text for more details.

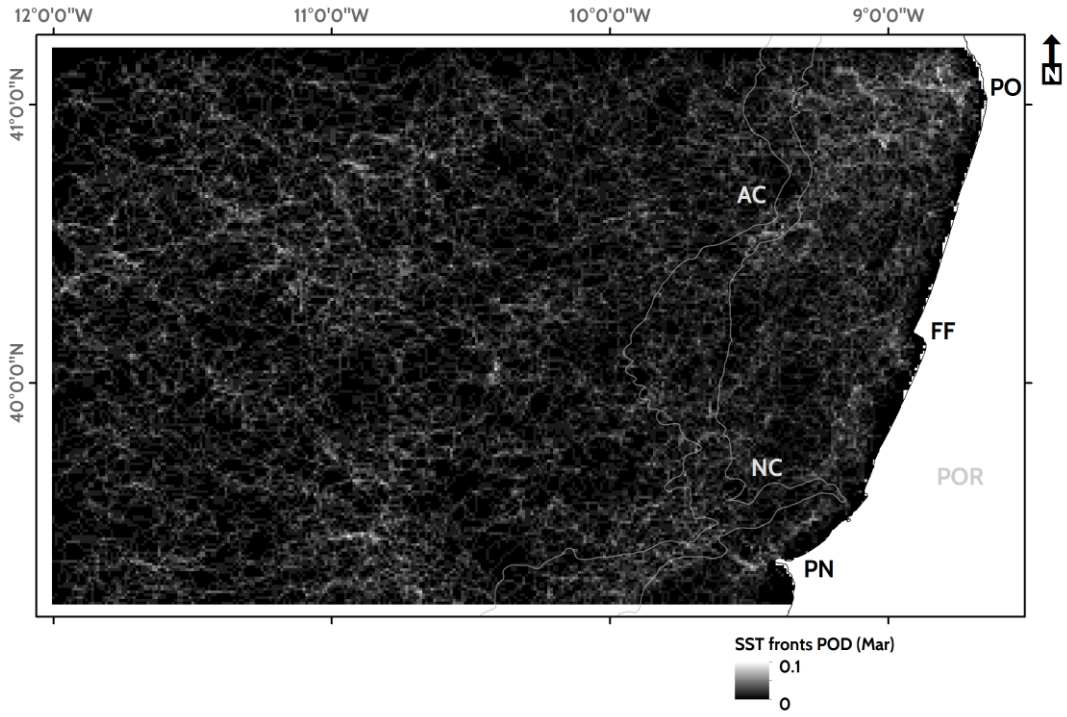


Figure 4.6.12. Monthly POD of SST fronts (March) for the period of study (2003-2014). Please see text for more details.

A comparative analysis of the three monthly POD maps, highlights the progressive decrease in frontal activity nearshore. In January, fronts are found in three main clusters. The first, to the North, in the vicinity of the Douro river, the second encompasses the Vouga and Mondego river regions, and the third envelops the Berlengas Islands. The first two clusters are the result of river discharge, while the third is likely the product of the interaction of coastal waters with the IPC, along the 200 m isobath. The January map also highlights the presence of the Western Iberia Winter Fronts (WIWiFs) suggested by Peliz et al. (2005).

The meandering winter front is present in the southern third of the study area, where it is zonally aligned until reaching the 200 m isobaths. At that point it turns northward along the shelf edge (IPC). According to the same author, the front establishes a boundary that divides the Iberian Basin into northern and southern sectors, which are influenced differently by the Azores Current. Because the position (and meandering) of the WIWiFs, the front is not depicted in the POD map as a solid, narrow region but as a broad area where front detection is likelier. This may also help explaining the high POD values recorded in the vicinity of the Berlengas, a possible point of contact and inflection of the WIWiFs near the shelf edge.

The February map shows a decrease in local POD values, due to the distribution of the WIBP and less pronounced effect or river fronts over the overall probability map. In March, the probability of frontal activity nearshore decreases even further and it is difficult to identify

regions in the shelf with POD values clearly above those recorded in the open ocean. Frontal activity is thus likely reduced in March, when the transition between seasons begins to take place and leave an imprint in the POD. Simultaneously, SST gradients, including of the WIBP (and river discharge) become less conspicuous, as described in earlier sections of the study. The conjunction of these facts creates the low POD and mostly uniform distribution of fronts across the study area, represented in figure 4.6.12.

The following period (April-June) presents a different image with the hallmark signs of the early stages of upwelling. Figure 4.6.13 represents the POD in this period, which brings together traits characteristic of both winter and summer.

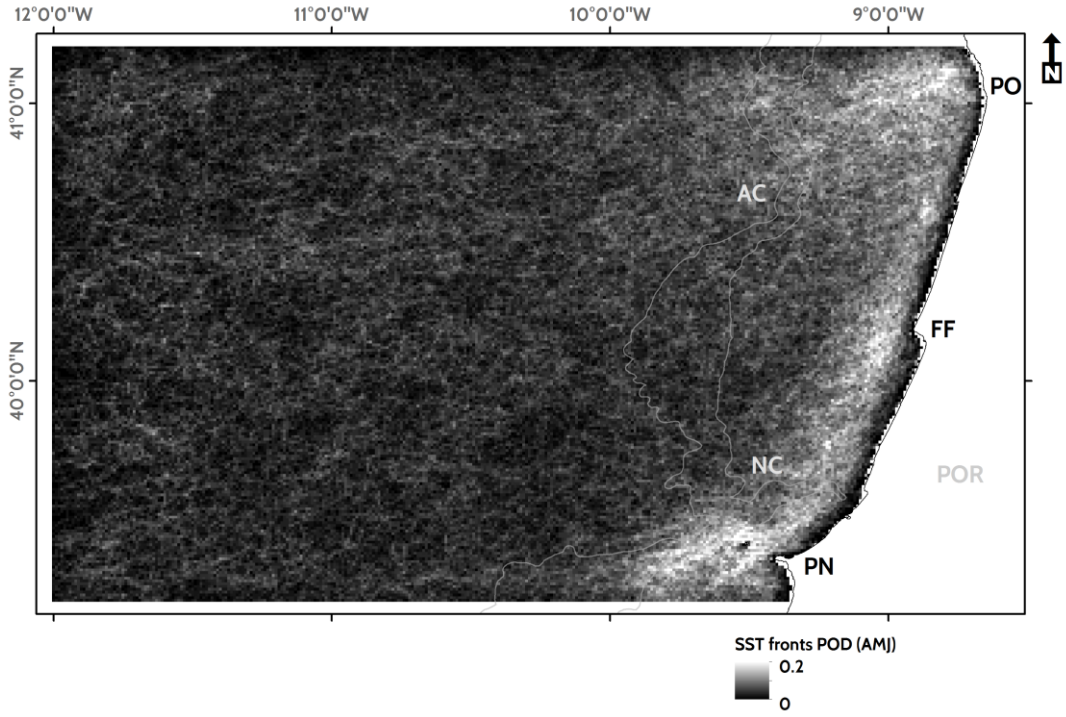


Figure 4.6.13. Seasonal POD of SST fronts (April-June) for the period of study (2003-2014). Please see text for more details.

The image highlights the higher POD near the Mondego, Vouga and Douro rivers. These are responsible for less pronounced frontal regions and create clusters similar to the F2, F3, and F4 of figure 4.6.2, which are embedded in a diffuse matrix of heightened POD created by the WIBP and upwelling. In the vicinity of the Berlengas Islands, the typical signature of the upwelling season (and tidal forcing) is already detectable, which translates into an increase of the POD (Castelao et al., 2008, Relvas et al., 2007).

The Carvoeiro-Berlengas area is a known hotspot of frontal activity generated by strong gradients and flows throughout the year (Peliz et al., 2002, 2003, 2005, Relvas et al., 2007).

The islands also seemingly create a small wake. Island wakes have been described before, namely in the work of Xie et al. (2001). In this case, the wake is suggested by the frontal decrease southward of the islands. The wake may be formed by coupled ocean-atmospheric action. Upwelling-favorable northerlies create an equatorward flow along the coast, which envelops the islands and generates a turbulent wake as they meet and contour the obstacle. This flow is also deflected by the nearby Cape Carvoeiro, possibly increasing the overall signature of the feature due to topographic deflection and forcing of the flow westward. A small patch of waters leeward of the islands displays a much lower POD value, although no discernable difference in SST values could be found.

Figure 4.6.14 displays the location of a zonal transect located southward of the Berlengas Islands. The POD of SST fronts was extracted for each point and plotted as depicted in figure 4.6.15.

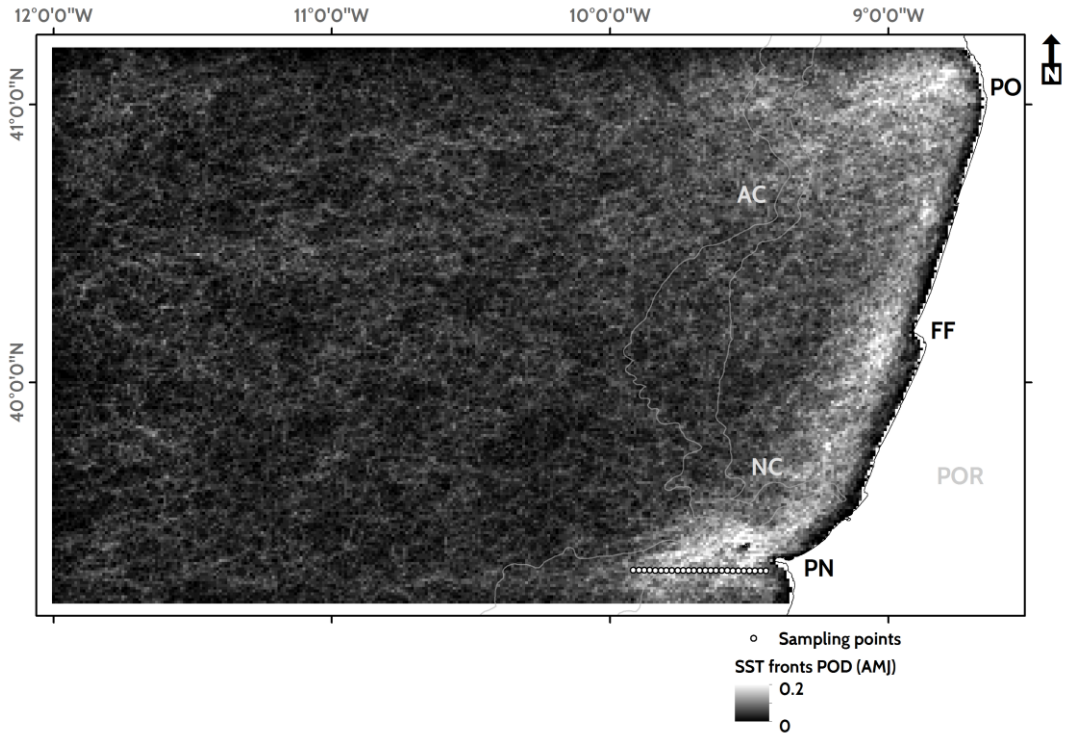


Figure 4.6.14. Sampling points used to retrieve seasonal POD of SST fronts (April-June) for the period of study (2003-2014). Please see text for more details.

Data from the transect simplifies the visualization of the gap between two high POD regions at either side of the small coastal island chain. In subsequent analysis to the following period (July-September) (figure 4.6.19) the wake-related gap in the POD field is less conspicuous, perhaps due to an increase in the turbulence as upwelling strengthens. More data is required

to fully characterize the potential effect of an island wake in the case of Berlengas, over front generation and persistence.

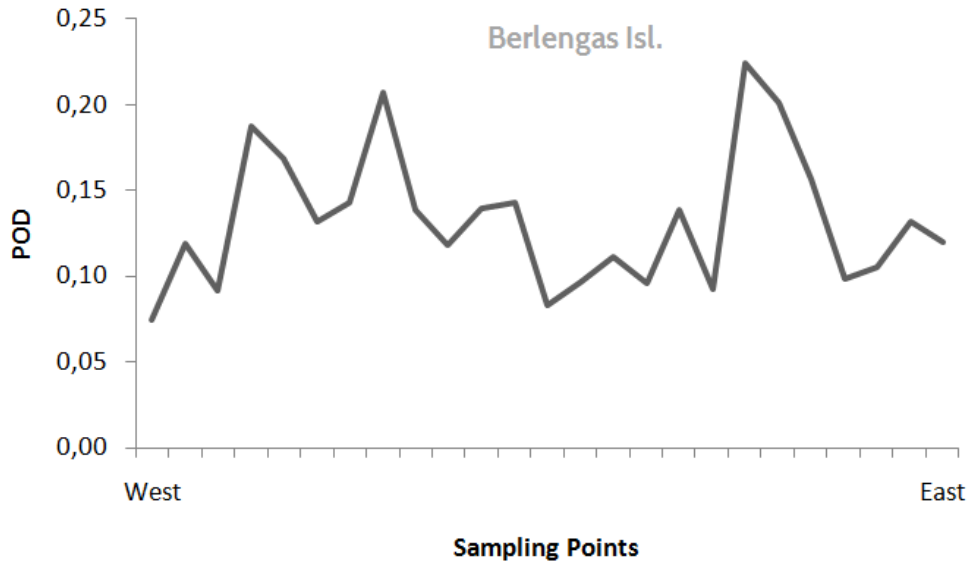


Figure 4.6.15. POD of SST fronts in the April-June interval (2003-2014) along the transect depicted in figure 4.6.14. Notice the decrease in POD values leeward of the Berlengas.

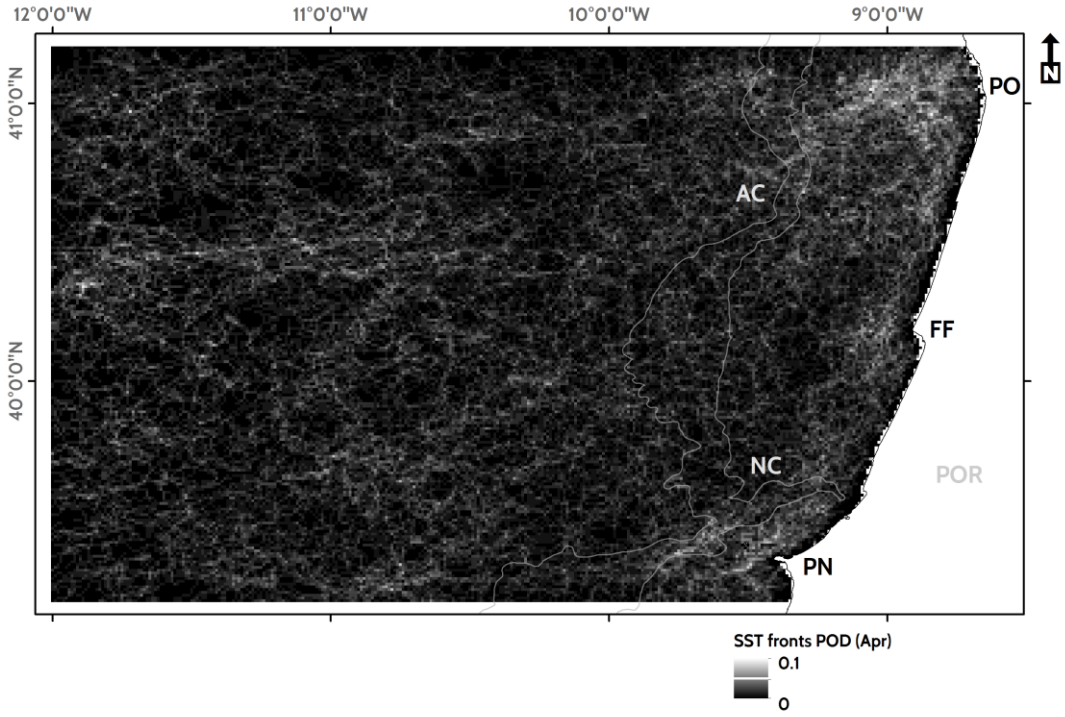


Figure 4.6.16. Monthly POD of SST fronts (April) for the period of study (2003-2014). Notice the developing high POD areas around Berlengas. Please see text for more details.

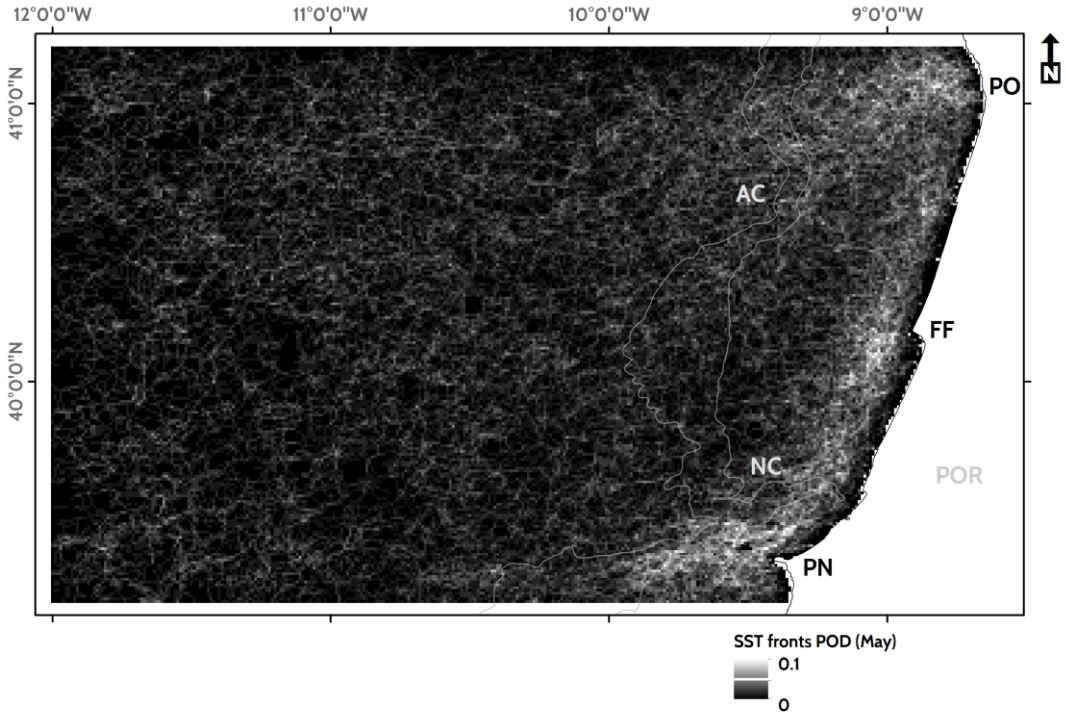


Figure 4.6.17. Monthly POD of SST fronts (May) for the period of study (2003-2014). Please see text for more details.

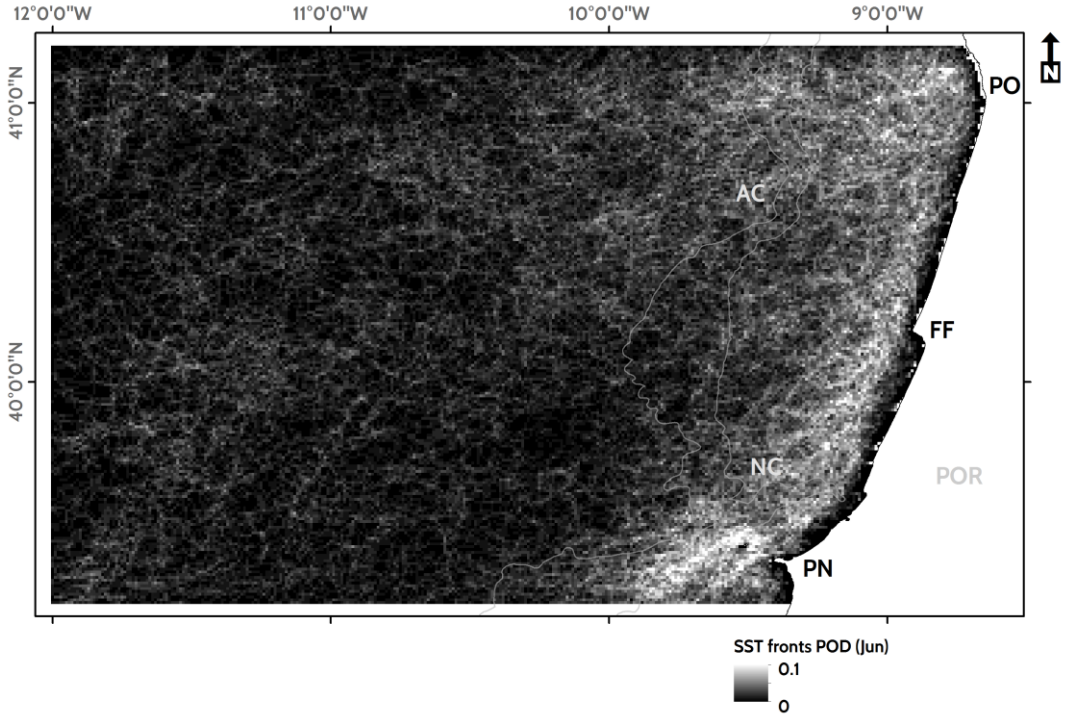


Figure 4.6.18. Monthly POD of SST fronts (June) for the period of study (2003-2014). Notice the wake formed by the Berlengas islands. Please see text for more details.

The sequence of monthly images shows the progressive expansion of the upwelling-dominated area from an initially narrow band near the coast (< 30km) visible in May. By June the offshore

migration of the fronts is noticeable, reaching a maximum distance of approximately 70 km offshore. The migration is connected to the evolution of the upwelling season (Breaker et al., 1986) which leads to the westward displacement of the prograde front created as a consequence of increased Ekman transport and the development of the coastal jet (Longhurst, 2007, Arístegui et al., 2006, Relvas et al., 2007).

However, in April, the image still showed the WIBP signature, particularly north of the Mondego and in the vicinity of the larger rivers. Furthermore, the zonal bands of frontal activity offshore continued to be present in the southern waters and at approximately latitude 40.5° N. However, POD in the shelf is already significantly higher than what was observed for the open ocean, throughout this period.

From July to September, the zonal difference increases even further (figure 4.6.19) as upwelling takes center stage as the dominant front generation process. The primacy is unequivocal and supported by all measures, from POD magnitude to the spatial distribution of the boundaries.

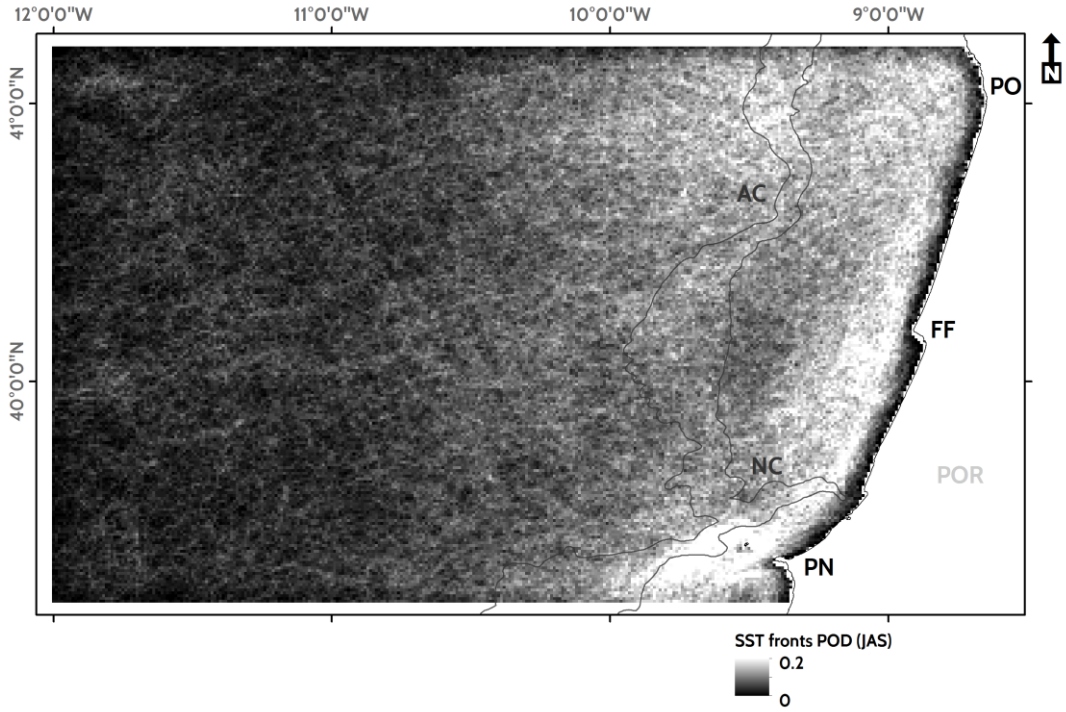


Figure 4.6.19. Seasonal POD of SST fronts (July-September) for the period of study (2003-2014). Please see text for more details.

SST front POD reaches a maximum during the summer period, stretching along the coast across the entire study area. Westward front migration continues and high POD areas were

now located over 150 km offshore. Coastline design and bottom topography appear to control – at least partially – the regional extent and distribution of the frontal regions.

Offshore (west of 10.5° W) POD values are lower and mostly homogeneous. The apparently high values of POD may be correlated with higher observation opportunities guaranteed by clearer skies.

Another relevant feature is the apparent formation of an important frontal region in the shelf edge, at the latitude of Figueira da Foz. In this region, it is perceptible a double front, the first near the coast, at a distance of 30 km and the second, located between the 200 and 1600 m isobaths. Such double systems have been identified in the past in this (Relvas et al., 2007) and other regions, and are held responsible for the development of coincident Chlorophyll a fronts (Longhurst, 2007). Relvas et al. (2007), despite the clear representation of a similar low POD area in the same region, does not offer any explanation to the process. We hypothesize the outer front is generated by the joint action of upwelling filaments and the mesoscale fields created in the interaction of the upwelling jets and offshore currents (cf. figure 4.4.1.12). The rich eddy fields created off shelf as highlighted in the analysis of the filaments in earlier sections provide a reliable and recurrent process that may help explain the POD values recorded in this period.

Inshore countercurrents may also contribute, albeit marginally in the study area to the development of a double frontal system, which creates a boundary between the coast and the upwelled waters that are limited to the west by another frontal system. Such solution has been suggested for regions to the North of the study area off Minho and Galicia (Spain) (Sordo et al., 2001, Relvas et al., 2007).

The sequence of monthly images (figures 4.6.20-4.6.22) provides further details on the sequence of events leading up to a fully developed upwelling system.

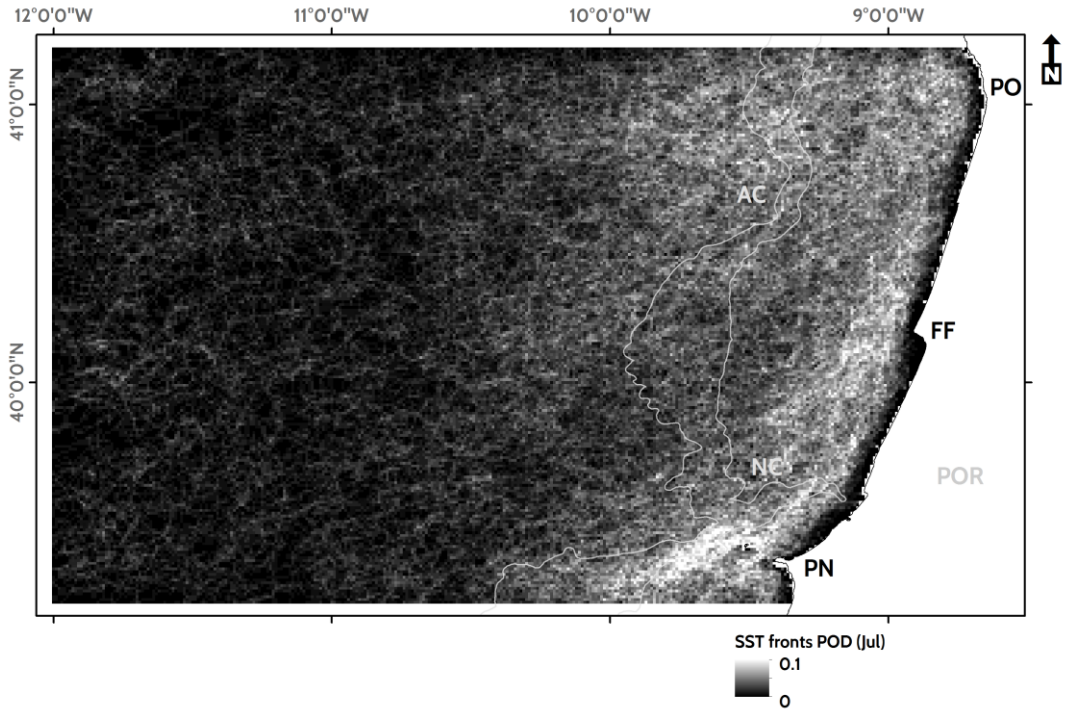


Figure 4.6.20. Monthly POD of SST fronts (July) for the period of study (2003-2014). Please see text for more details.

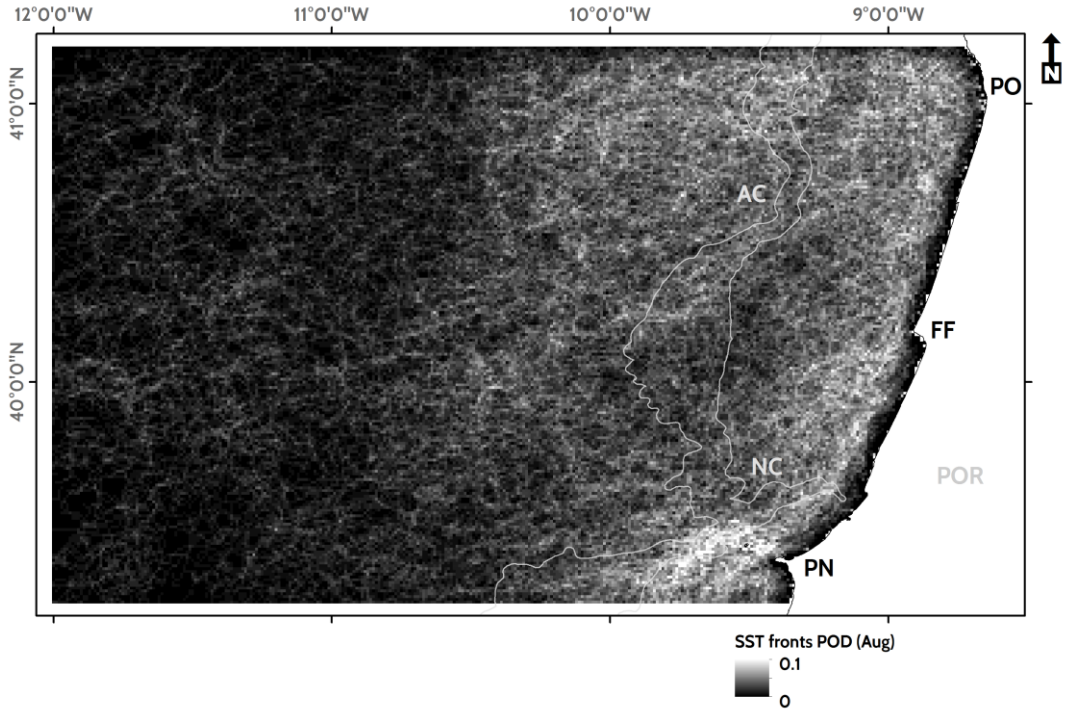


Figure 4.6.21. Monthly POD of SST fronts (August) for the period of study (2003-2014). Please see text for more details.

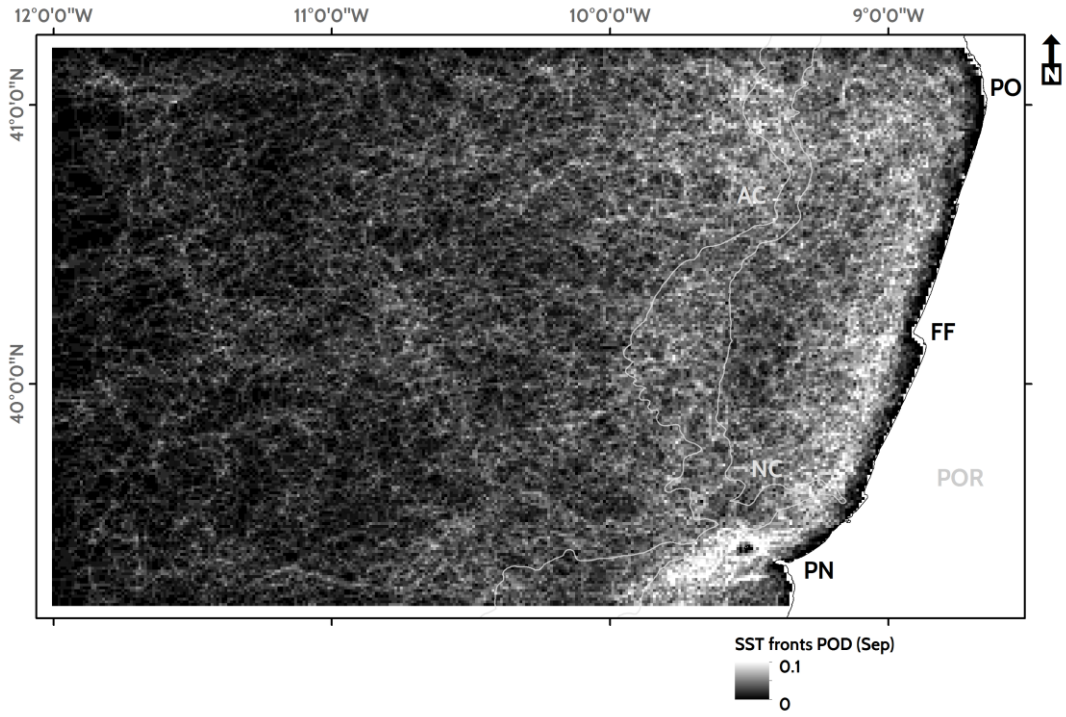


Figure 4.6.22. Monthly POD of SST fronts (September) for the period of study (2003-2014). Please see text for more details.

Frontal probability reaches its peak in July and August. It is difficult to determine when the peak occurs considering the inter-annual variability (Alvarez et al., 2011). Furthermore, different methods offer different estimates of the UI maximum, namely if the studies leverage satellite data or Ekman transport models (e.g. Picado et al., 2013).

In the study area, mean POD in July is of 2.4 (2.1) %, while reaching a similar 2.5 (2.0) % in August. The differences are not significant for the whole area. However, differences exist in the spatial distribution of the fronts. In July at the latitude of Oporto, the upwelling front is located *ca.* 100 km off the coast, and at a distance in excess of 150 km in August. The presence of filaments greatly contributes to these distances and the inter-annual variability, already described and illustrated (figure 4.4.1.12), calls for some caution in the estimation of maximum extents of high POD regions connected to upwelling.

The double-front system was discernable in the mid-section of the study area in all dates. Figure 4.6.23 depicts the location of a transect, at the latitude of Figueira da Foz, used to extract and compare POD values across the shelf.

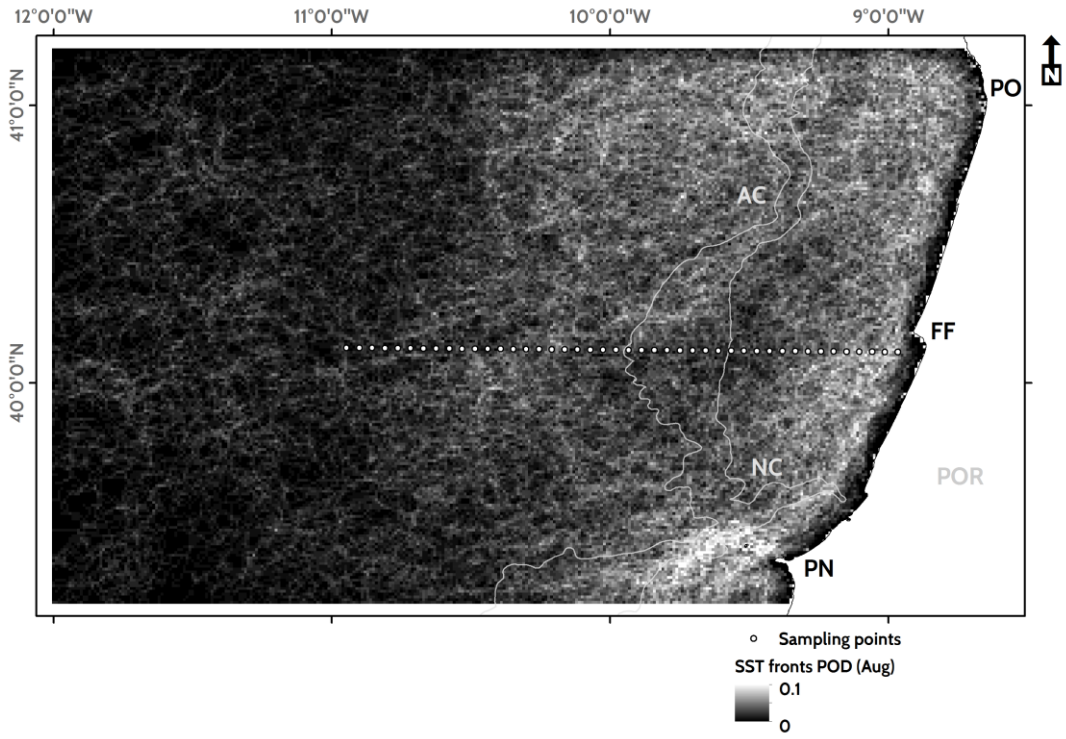


Figure 4.6.23. Sampling points used to retrieve seasonal POD of SST fronts (August) for the period of study (2003-2014). Please see text for more details.

The importance of upwelling over primary productivity and the role of fronts in localized increases of marine biodiversity are well established, including in the study area, where research on the subject is taking place for several decades (Otero et al., 2009, Santos et al., 2004, Peliz et al., 2004, Fiúza et al., 1983). However, it is relevant to understand whether SST and Chlorophyll a fronts occur at coinciding locations (at the work scale). The complexity of front generation processes and the existence of Deep Chlorophyll Maxima for instance, may create a mismatch between both frontal signatures. The relevance of visual cues in predatorial behavior (e.g. Nevitt, 1999) requires variables other than SST (possibly) to be used to identify favorable feeding areas along the frontal boundaries.

The data extracted from the transect represented in figure 4.6.23 is plotted in figure 4.6.24.

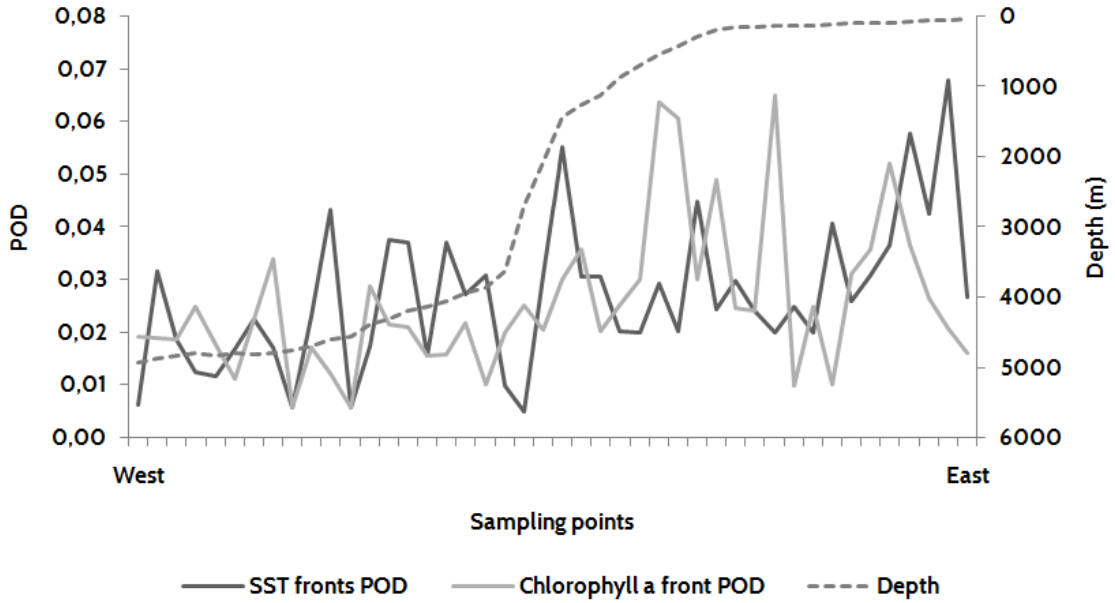


Figure 4.6.24. SST and Chlorophyll a front POD across a transect at the latitude of Figueira da Foz. Frontal POD is higher within the shelf with peaks nearshore and at the edge in the case of SST.

The noisy plot showed an overall agreement between the broad regions with high POD of both variables. General POD increases are identified nearshore (particularly for SST), mid-shelf, and shelf-edge. Chlorophyll a fronts reach maximum POD at the onset (coastal side) of the shelf slope, when depth begins to increase westward.

However the use of long term time series precludes the analysis of shorter-time correlations. The abundance of detections and the variability of features, even those displaying a certain recurrence (e.g. Aveiro Canyon filament), call for a more detailed analysis at the monthly scale or shorter. Figure 4.4.1.12 depicts a series of filaments anchored to bathymetric features. In August 2007, a large filament was detected that extended for more than 200 km westward, approximately at the latitude of Figueira da Foz.

Figure 4.6.25 depicts the SST fronts, Chlorophyll a fronts, and simultaneous detections of both, mapped from imagery of August 2007, using the methodology already described.

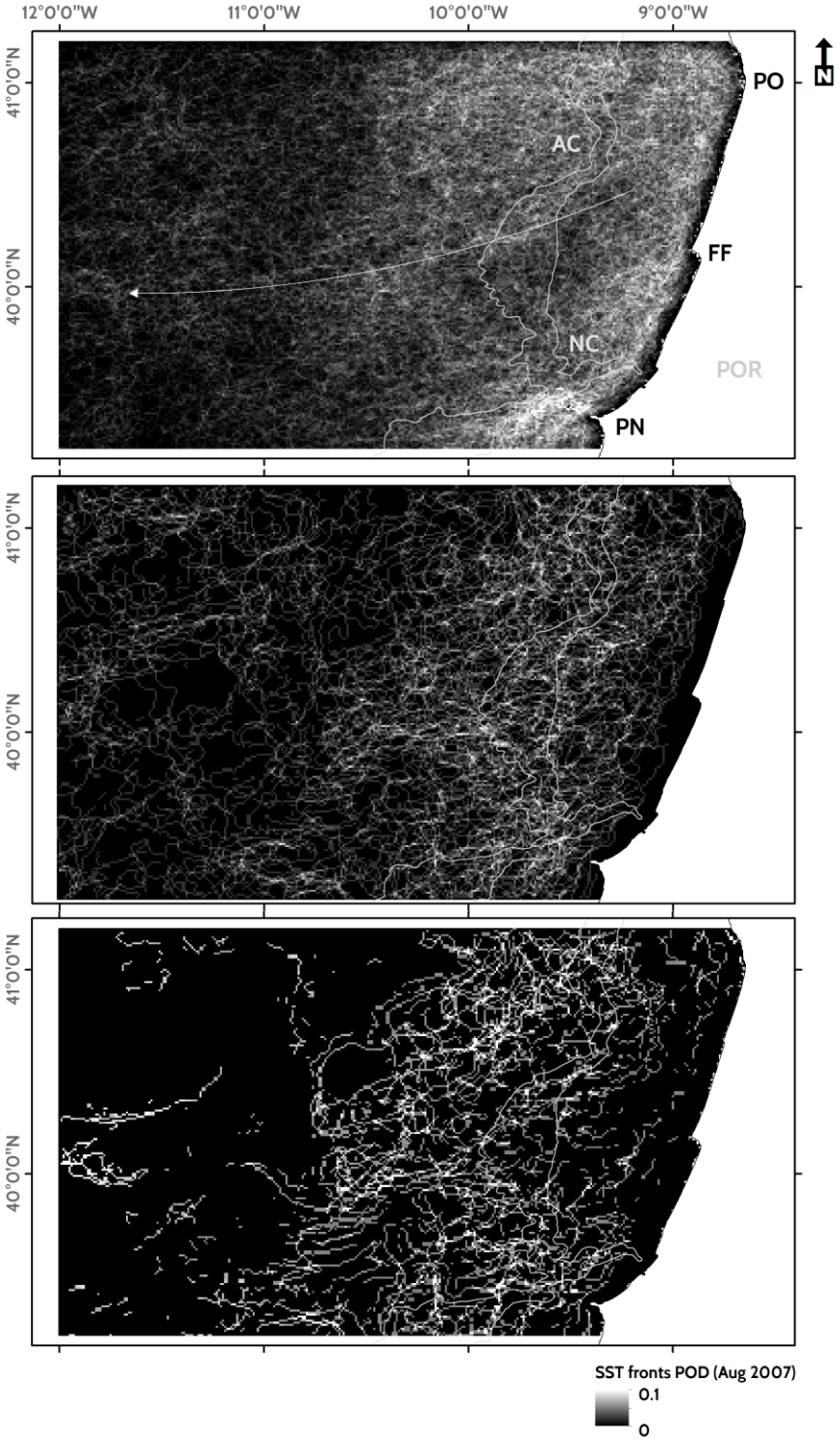


Figure 4.6.25. SST (Top), Chlorophyll a (Middle), and simultaneous (SST and Chlorophyll a) fronts POD in August 2007. The white arrow illustrates the approximate direction of propagation of the filament, determined from the analysis of MODIS AQUA SST imagery.

In this image it became clear that during the upwelling maximum (August), most Chlorophyll a fronts occur from mid to shelf-edge. SST fronts, in comparison, are more widely distributed

although tending to follow the regions influenced by major features, in particular the filament already mentioned.

Simultaneous fronts occur especially in the shelf edge region beyond the 200 m isobaths. The match-ups are also abundant within the area of influence of the filament and in longitudes as far as 11° W. The distribution of low SST values is coherent with the spatial pattern of Chlorophyll a and simultaneous fronts. This fact strongly supports the hypothesis suggesting a strong control by SST (and UI_{SST}) of Chlorophyll a fields and the role of thermal fronts as enhancers or promoters of local blooms and heightened productivity gradients (Kahru et al., 2012, Longhurst, 2007, Saraceno et al., 2005).

Another example of the strong correlation between upwelling features (filaments in particular) and frontal regions is offered by the already mentioned filament detected in July 2012 (Figure 4.4.1.12). Throughout the month, a developing filament progressed westward along the southern wall of the Aveiro Canyon. Simultaneous measurements by an ARGO platform enabled the analysis of the flow. The integration of a frontal map with the SST fields created a powerful image that highlights the importance of specific features to a monthly frontal field (figure 4.6.26).

Still, monthly time scales may prove excessive when the purpose of the application is to monitor frontal fields and correlate them with rapidly evolving phenomena, such as upwelling (Castelao et al., 2008). A difficult compromise is required between product frequency requirements and the inherent limitations of optical remote sensing over the oceans. The development of alternative front detection methodologies, resorting to SAR data for instance, could contribute to solve the conundrum, but the current state-of-the-art still precludes such a solution.

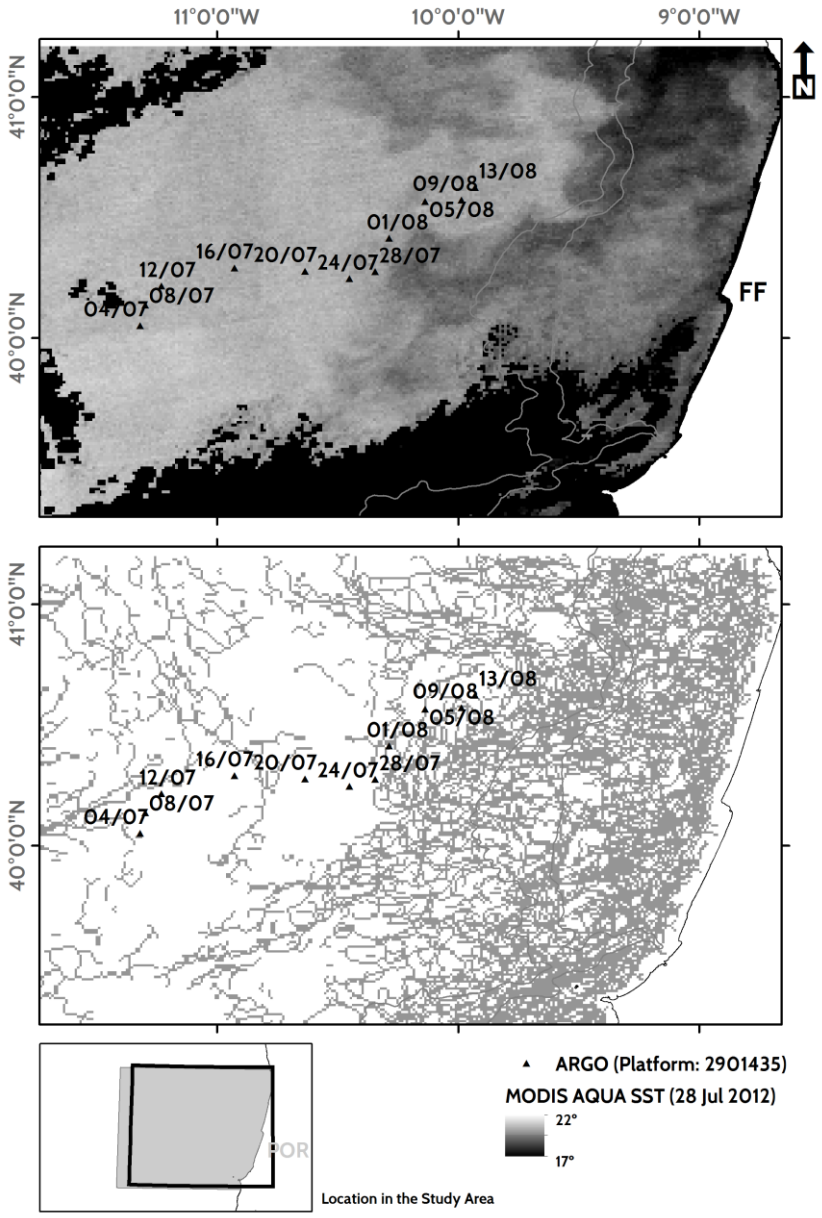


Figure 4.6.26. Upwelling filaments and frontal fields. Top: SST image acquired on July 28, 2012. The position of an ARGO float is depicted. Areas with no data are represented as black pixels. Lower image: Combined detections of SST fronts in July 2012 (monthly). Notice the correlation between the coastal upwelling (lower temperatures) and the distribution of fronts. FF: Figueira da Foz.

Finally, in the year's last quarter, the frontal fields rapidly decrease in magnitude and spatial distribution (figure 4.6.27). SST fronts tended occur in the vicinity of the 200 m isobaths and around the Berlengas Islands. The signature of river discharge is also visible, albeit discreet, in the vicinity of the Douro, Vouga, and Mondego rivers.

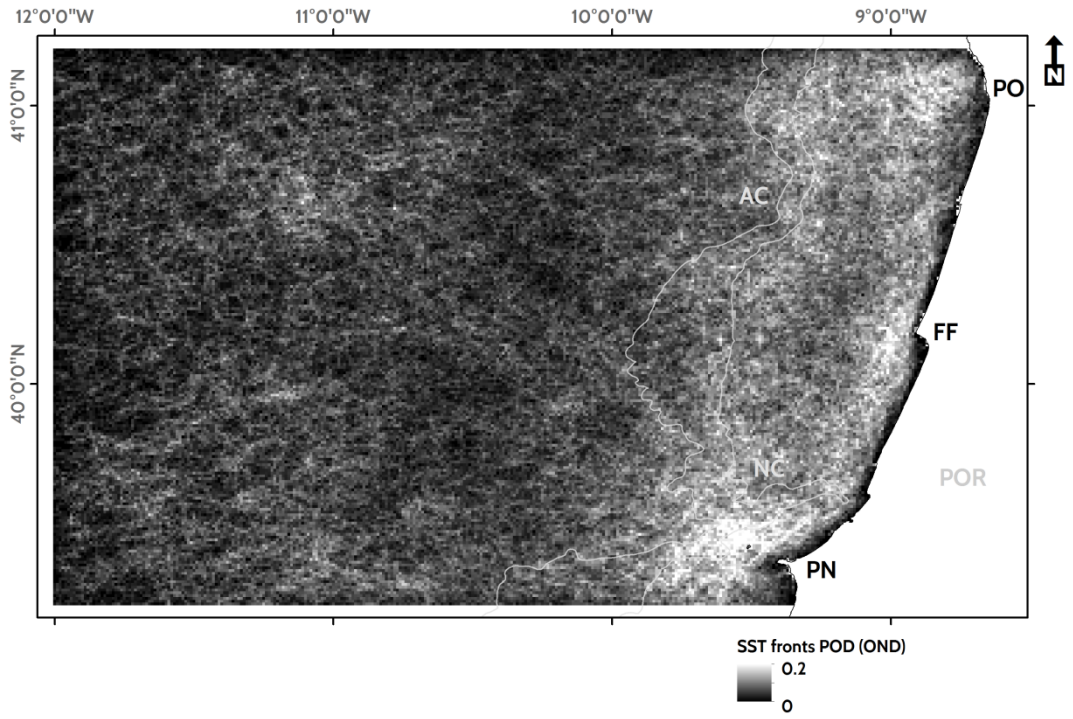


Figure 4.6.27. Seasonal POD of SST fronts (October-December) for the period of study (2003-2014). Please see text for more details.

The October frontal POD map (figure 4.6.28) shows a northward progression along-slope of these features, suggesting the IPC is resuming its surface expression upon the relaxation of upwelling-favorable winds.

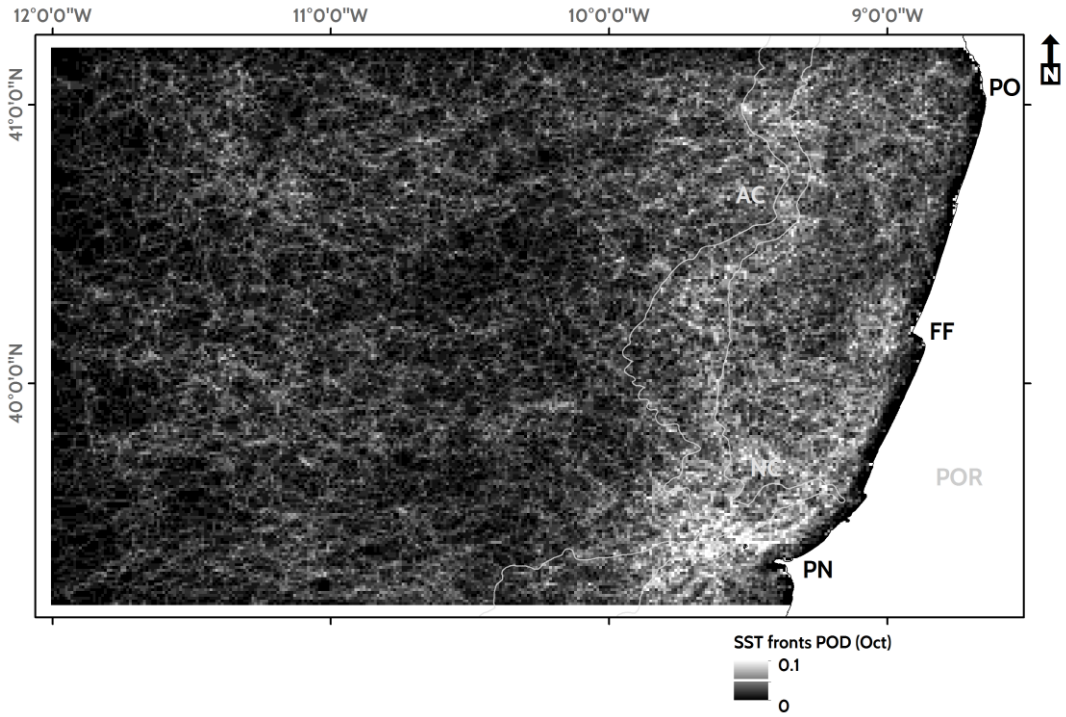


Figure 4.6.28. Monthly POD of SST fronts (October) for the period of study (2003-2014). Please see text for more details.

By November (figure 4.6.29), the fronts in the vicinity of the Mondego and Douro rivers become even more conspicuous, as the plumes generated by higher discharge values become an important driver of frontal activity. In the southern third of the study area, a region of higher POD values seems to suggest, once again, the importance of the WISiFs. These regions connects to the East with the shelf slope, creating a continuous (although inflected northward at Berlengas) of high POD values. We hypothesize this large feature is correlated with the development of the poleward current. Once again, as in the end-winter/early-spring POD maps, a region centered in 41° N; 11° W displays some persistent frontal activity. The bathymetry in this area does not support the development of fronts and therefore it is difficult to explain the nature of this anomaly, which could be connected to eddy activity in the interface between the Portugal Current and the IPC.

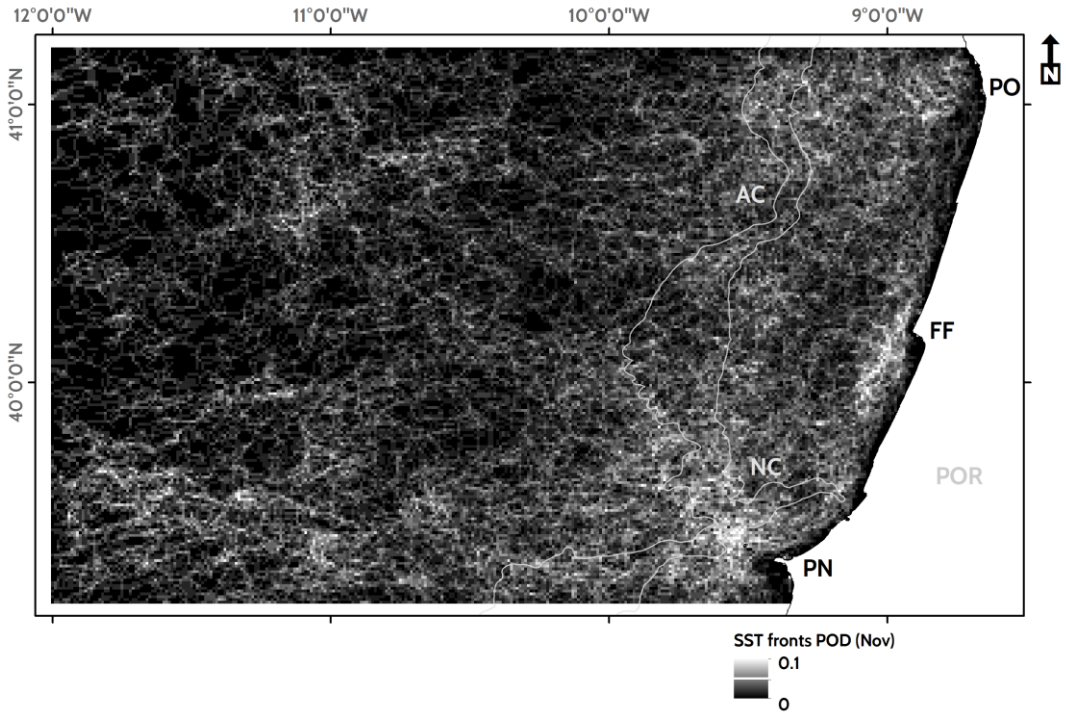


Figure 4.6.29. Monthly POD of SST fronts (November) for the period of study (2003-2014). Please see text for more details.

Finally, the December POD map shows a developed but not yet mature winter frontal distribution pattern. The WIBP becomes an important driver of fronts nearshore, with the IPC further contributing to the formation of frontal boundaries in the vicinity of the shelf edge. A sub-zonal band of higher POD extends westward from the 10° W emulating the division between the R3A and R3B regions detected by the DTW-based method (SST k=15).

The band appears to mark the northern limit of a region characterized by warmer water. Again, this may represent the WIWiFs as it meanders with the progression of cold months.

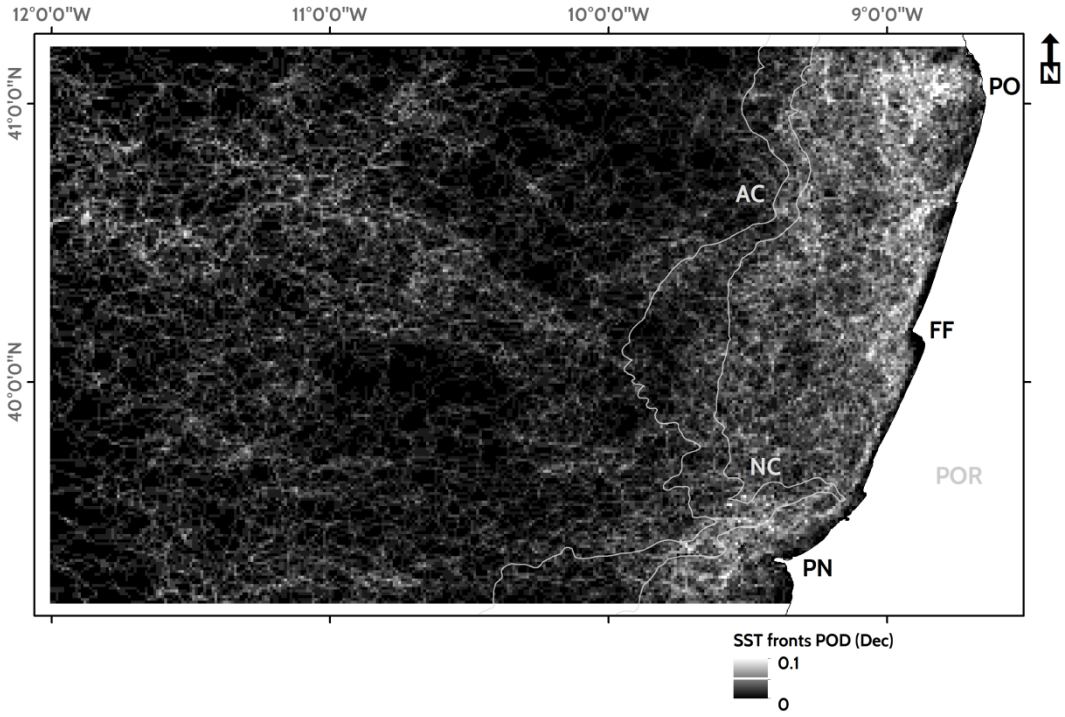


Figure 4.6.30. Monthly POD of SST fronts (December) for the period of study (2003-2014). Please see text for more details.

A final analysis was conducted to determine whether the fronts occurred at preferred directions. To test this hypothesis, one year of front detection maps was analyzed (2010). We limited the initial analysis to one year, in order to determine whether the complex and computationally-intensive process should be applied to the large number of segments created for the 11 years of study.

A processing chain was setup in order to recursively analyze the daily frontal maps and split front vectors into segments with uniform direction (41726 segments for 2010). The vectors were split at the vertices, and both length and direction were calculated. The resulting data were exported into tables, later aggregated per calendar month. The entire workflow was coded as a python script (arcpy) developed by the author.

Because the sheer number of fronts in a given direction would not reflect the reality, due to varying lengths of the segments, a corrective method was introduced. The total length of the segments was summed for angle intervals of 5°, from -90° to 90°. The summed length was then divided by the total frontal length of the month, thus returning a ratio, which characterized that particular bearing. The values were converted to radians and later to a range between $-\pi/2$ to $\pi/2$ for standardization purposes.

Consolidated monthly tables were created and the values plotted in rose diagrams, as depicted in figure 4.6.31.

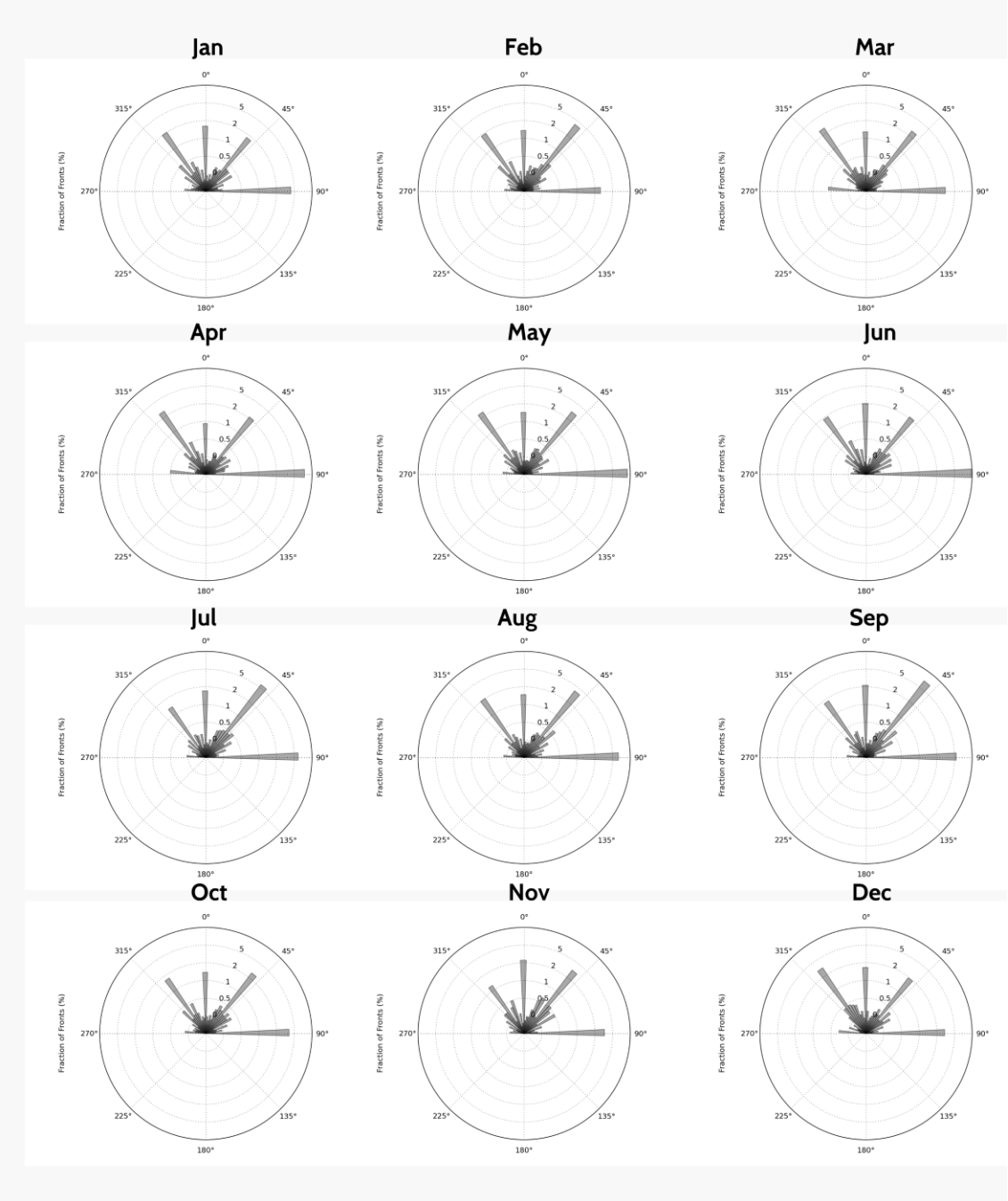


Figure 4.6.31. Rose diagrams created for the 12 months of 2010, depicting frontal bearings.

The tables were also tested for differences using ANOVA, but no significant differences were found between the months (for a $p < 0.01$). The values between 0 and $\pi/2$ were more frequent in all months but December. Still the differences were very small and probably irrelevant. The most frequent directions were found in the vicinity of -45°, 0°, 45° and 90°, highlighting the role of coastline geography in the definition of the broad frontal (and current) angles. The

recurrence of fronts in the 90° angle family suggests that the zonal bands described earlier may indeed play an important role, although not exceeding other sub-meridional bearings.
Considering the similarities between months, no further tests were conducted in the datasets describing the frontal fields of other years.
The complexity of frontal patterns was thus described using a wealth of data and cartographic representations, to attempt a deconstruction of the patterns describing the overall distribution of these features. The recurrent distribution of thermal fronts in certain regions, even if masked by the inherent inter-annual variability, is obvious. The influence of specific flows and features cannot be ignored and the climatology now available creates a powerful tool to support the study of the Portuguese sea.
A complex task in itself, the construction of the climatology spurred the development of various tools which introduced new approaches to the complex challenges of big data mining.
In the next chapter we shall attempt to reconcile frontal distribution data with the DTW-based regionalization effort presented in the previous section.

4.7. Integrating frontal boundaries and DTW

The casual observer may find it less than obvious the need to create divisions in a seemingly continuous domain, as the (fluid) ocean appears to be (IOCCG, 2009). But the analysis of time series of SST, Chlorophyll a, and other variables measured from space suggest otherwise.
The continuous nature of time series and the relation of each value with both its predecessor and subsequent field (Łuczak, 2016) creates a unique opportunity to analyze ecosystems not from static pictures but from dynamic datasets.
Such was the approach followed in this study, which compiled and analyzed a substantial dataset of SST and Chlorophyll a datasets for a region off central Portugal. The variables (alone and combined) were processed using a chain that included DTW to calculate a distance matrix and UPGMA as the clustering method.
But given the novelty of the methodology it is important to compare it to an established method employed to determine the boundaries of water masses and regions: thermal fronts. The relevance of these features in mechanisms of isolation or connectivity (and intermediate situations) is pivotal to create barriers or bridges affecting marine populations and recruitment (Woodson et al., 2012, Reese et al., 2011) furthering the case for a connection between the physical and biological boundaries of the sea.
Oceanic fronts separate water masses with different characteristics and depending on the gradient of the variable defining the front (temperature, salinity or other) have a different

magnitude. This means that fronts are not all alike and some may occur at minor gradients of local importance, while others establish barriers between major domains (e.g. upwelling front separating cold, nutrient-rich waters from warmer coastal ones).

Considering that thermal fronts, as addressed in this study, separate water masses that may display different origins (e.g. upwelling, poleward flows), these will have dissimilar nutrient contents, salinity, suspended matter concentration, flow direction and velocity, to name a few variables. These different traits are inherently what characterize ecoregions (Spalding et al., 2007).

Following the construction of a regionalization scheme, we must now describe how the regional map relates to oceanic fronts. To do so, both sets (frontal POD and regions) were compared. In particular we were interested in determining whether 1) frontal POD (overall and seasonally) differs across regions, and 2) whether different regions are physically isolated by thermal fronts (overall and seasonally).

Considering the degree of regionalization represented in the SST-based clusters at k=15, henceforth we shall adopt this system as the standard against which the distribution of frontal boundaries will be compared. The rationale behind this decision is set upon the level of segmentation of the study area and the correlation with oceanographic features and cycles of the WIP as described.

The first step in the analysis consisted in evaluating whether there was a difference in the mean number of fronts detected (with correction for number of observations) for pixels within the regions. To that end, the sampling grid depicted in figure 4.4.1.3 was used to extract the mean POD values (overall and quarterly mean over the study period). The data collected were then submitted to ANOVA to test for differences in the mean values.

For the overall mean POD, the regions did yield significant differences (for a $p < 0.01$) suggesting that the occurrence of fronts is not equal across the study area. This outcome was expected considering the spatial variability of frontal boundaries depicted in the POD maps.

The quarterly frontal POD analysis yielded similar results with significant differences obtained for all periods (for a $p > 0.01$).

Table 4.7.1 summarizes the mean POD values and standard deviation for the overall period, calculated for each region.

Table 4.7.1. Frontal POD within the regions defined by the analysis of SST time series (k = 15).

Region	Mean Frontal POD (fraction)	Standard Deviation
R1	0.03	0.01
R2A	0.04	0.01
R2B	0.03	0.01
R3A	0.02	0.01
R3B	0.02	0.01

The amount of detections across the study period limits the interest of the overall analysis. Furthermore, fronts are relevant as the boundaries of water masses inherent to specific seasons as processes. Therefore, although the recurrence of front detections in certain areas provides important insights into local oceanography and simultaneously on the ecological conditions of a region, the seasonal variability sheds further light on these processes.

Still, the overall analysis highlights the higher probability of frontal detections near the coast in the regions R1 to R2B and a decrease offshore (R3A and R3B).

Table 4.7.2 provides the mean and standard deviation of front POD for each quarter.

Table 4.7.2. Quarterly Frontal POD (mean and standard deviation) within the regions defined by the analysis of SST time series (k = 15).

Mean POD	R3B	R3A	R2B	R2A	R1
JFM	0,03	0,03	0,03	0,06	0,06
AMJ	0,04	0,04	0,07	0,10	0,08
JAS	0,05	0,06	0,13	0,15	0,11
OND	0,05	0,05	0,09	0,11	0,09
SD	**R3B**	**R3A**	**R2B**	**R2A**	**R1**
JFM	0,02	0,02	0,02	0,03	0,04
AMJ	0,02	0,02	0,03	0,04	0,05
JAS	0,03	0,03	0,04	0,05	0,06
OND	0,03	0,03	0,04	0,04	0,05

The quarterly analysis returned interesting results that are clearly in line with the body of information reported so far and the general knowledge on the study area. The first quarter presents the lowest mean POD values. However, the highest probability is found close to the shore in the R1 region. This is a direct consequence of the influence of river outflow and the WIBP. In the second quarter, with the onset of the upwelling season, the POD values increase across the coastal regions (R1 to R2B) but particularly in both R2 segments. This is in

agreement with the previously described analysis of transects and maps (including the EOF modes), which describe the westward development of the upwelling cells (and front).

At the peak of the upwelling season (July, August, and September) POD values continue to increase and the migration of fronts offshore also peaked. At R2B, which includes areas beyond the slope, the mean POD values are higher (for the first time) than those recorded for the coastal R1 region.

Finally, upon the relaxation of upwelling-favorable northerlies, POD values begin to decrease, approaching those obtained for April-June. Interestingly the mean POD at the R3B region increases, which is also agreement with the presence of the WIWiFs-related fronts. These were already discussed earlier but once again, the results seem to support our interpretation.

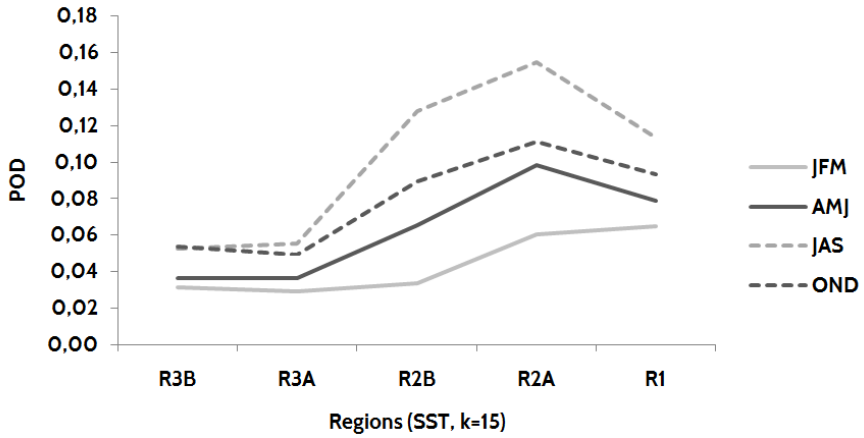


Figure 4.7.1. Quarterly Mean SST Front POD in each region. Notice front migration offshore and increase in activity in the summer months. JFM: January, February, March; AMJ: April, May, June; JAS: July, August, September; OND: October, November, December.

In light of the above, the answer to the first question placed at the beginning of this section is that the regions do display different frontal activity patterns, which are confirmed by different statistical measures. Furthermore, the mean POD values are in agreement with the hydrodynamic patterns of the region. The methodologies thus provide an adequate framework for zonal studies, creating regions that do have different oceanographic properties.

The correlation between regional boundaries and frontal activity was also studied. To that end, several transects (figure 4.7.2) were created in such a way that allowed the extraction of mean POD of thermal fronts across regional boundaries. If frontal activity increases in the vicinity of the boundaries, it means the regionalization scheme is indeed correlated with different water masses (Nieblas et al., 2014). In turn, the acknowledgement of this fact has important

implications for the design of management policies, as it opens new perspectives on bioregionalization and monitoring (Gregr et al., 2012, Rice et al., 2011).

The comparison of long term time series renders the analysis somewhat insensitive to inter-annual and seasonal variability. In this sense, it is not expected a perfect alignment of oceanic fronts and the boundaries of regions. In its place, a broad correlation of certain characteristics of the frontal field and the regions are expected. From the presence within the regions of a given family of thermal fronts, or the seasonal increase of thermal gradients, a spectrum of possible correlations is possible.

To assess potential correlations, several transects were designed to cross relevant areas within the study area (figure 4.7.2). The mean POD of thermal fronts were extracted and plotted against the location of the boundaries of each region.

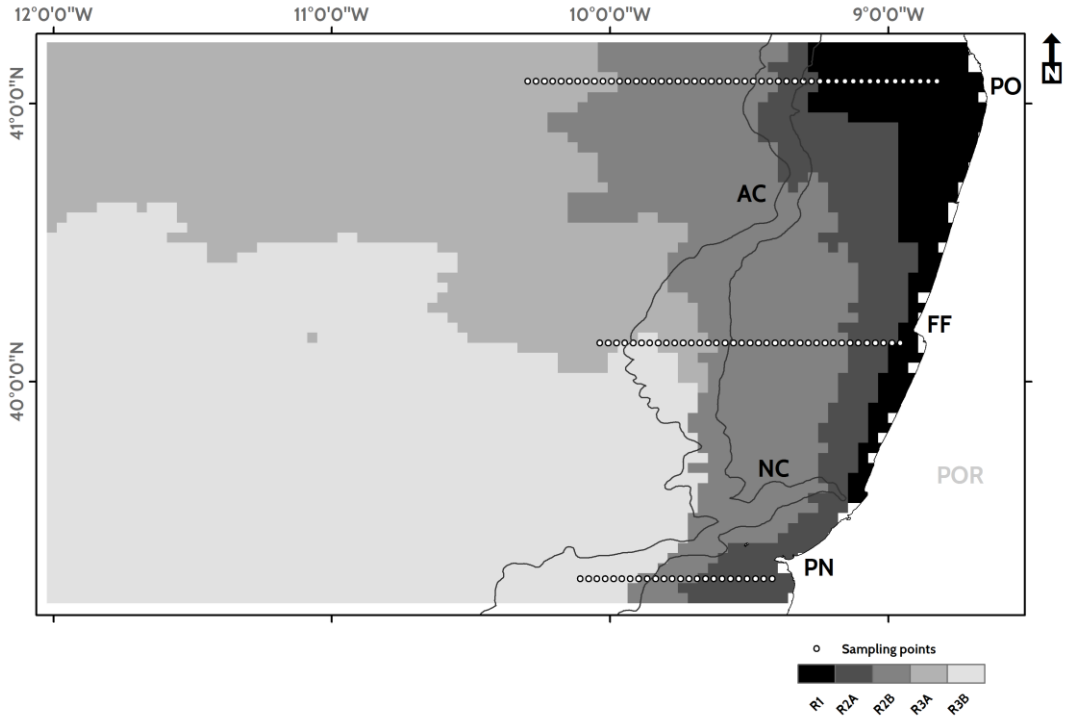


Figure 4.7.2. Location of the sampling points designed to compare the POD of fronts and regional boundaries. The sampling points are overlaying the SST k=15 clusters. For a color image of the regions, please see figure 4.4.2.4A. PN: Peniche; FF: Figueira da Foz; PO: Oporto; AC: Aveiro Canyon; NC: Nazaré Canyon; POR: Portugal.

Overall, it was found a good agreement between regional boundaries and the occurrence of heightened frontal activity. Although such agreement is not clearly visible throughout the entire border of the multiple regions it was identified with some clarity in the analyzed transects.

Figure 4.7.3 displays the POD values for the complete time series at the transect located at the latitude of Figueira da Foz.

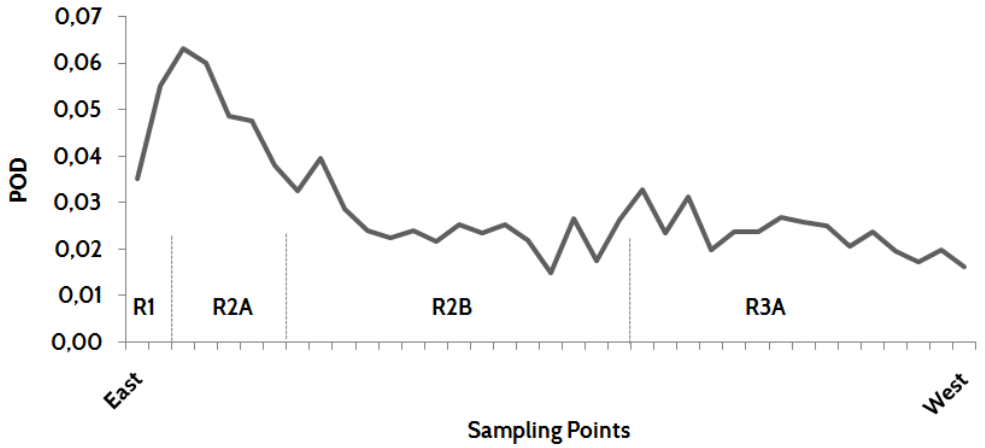


Figure 4.7.3. POD values at the Figueira da Foz transect, for the entire study period (2003-2014, all seasons). The SST k=15 regions (R1, R2A, R2B, and R3A) are annotated into the plot. Notice the coastline is represented to the left (East).

The regional boundaries in the depicted transect tend to occur in areas in or close to high POD areas. Specifically, the boundaries of R1 and R2A seem to be located in the vicinity of high POD areas, while the boundary R2B and R3A is collocated with such a region. The boundary between R2B and R3A marks the outer limit of the coastal ocean directly influenced by upwelling events. The upwelling fronts seen in the plot between R2B and R3A are thus amongst the most important and defining elements of the area. The stark contrast between coastal and open ocean waters is such that the increase in frontal POD is unmistakable. On the other hand, the inter-annual variability of frontal activity nearshore is likely to be significant. Variability in the WIBP and other features generates a complex frontal field, which is less likely to align perfectly with the regional boundaries.

Biogeographic units such as those created using our approach, may also depict the response to specific events or seasonal patterns. It is possible that certain events of particular importance in the yearly cycle may play an important role in the definition of regional boundaries.

In figure 4.7.4, the POD of fronts in the months of the last quarter (October-December, 2003-2014) seems to be well aligned with the boundaries of the regions (SST k=15).

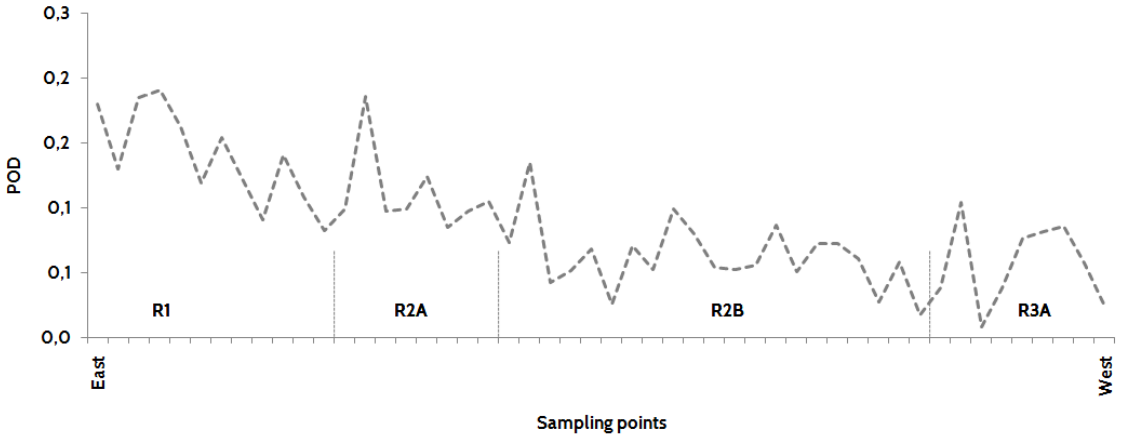


Figure 4.7.4. POD values at the Oporto transect, for the fourth quarter (October-December) of the study period (2003-2014). The SST k=15 regions (R1, R2A, R2B, and R3A) are annotated into the plot. Notice the coastline is represented to the left (East).

POD peaks can be traced to the boundaries of all regions, namely the R1-R2A and R2A-R2B frontiers. These are areas of intense mesoscale activity, influence of coastal processes, and where the WIBP has an important expression as a consequence of the Douro river discharge values. These are important structural elements of the coastal optical and thermal environment, which seem to be imprinted into the DTW classification. The R2B-R3A boundary also seems to be nearly coincident with a POD peak, which could be the consequence of the interaction of coastal waters with the IPC.

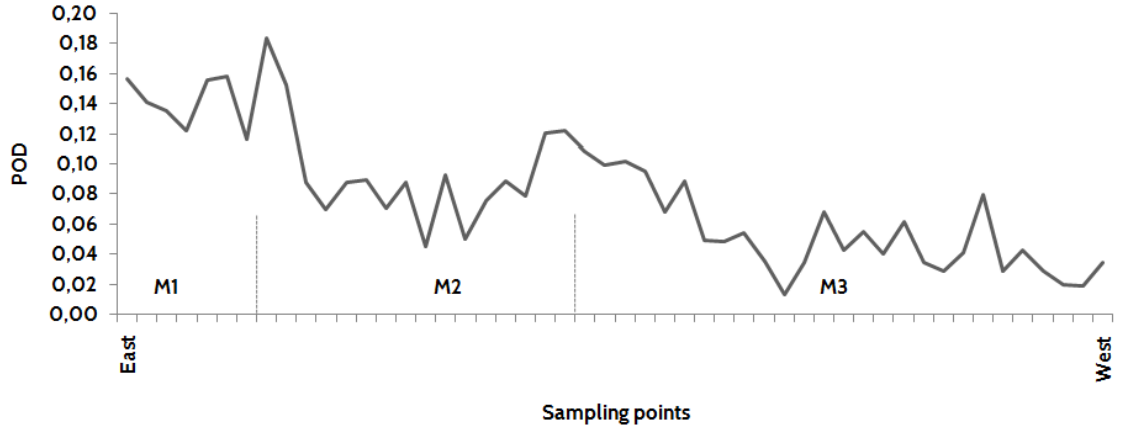


Figure 4.7.5. POD values at the Oporto transect, for the second quarter (April-June) of the study period (2003-2014). The Multivariate k=3 regions (M1, M2, and M3) are annotated into the plot. Notice the coastline is represented to the left (East).

The Oporto transect also provides an interesting correlation between the Multivariate regions (SST + Chlorophyll a, k = 3) and POD of thermal fronts for the second quarter of the year. The

three regions are clearly separated by areas of enhanced frontal activity, which are in contrast with the lower neighboring values.

The multivariate regions seemed to be strongly influenced by the Chlorophyll a fields and the results depicted in figure 4.7.5 may provide strong evidence of that relation. During the second quarter the northerlies that support upwelling take center stage (Fiúza, 1982) and the UI_{SST} becomes negative (Alvarez et al., 2011) after the April upwelling minimum already discussed.

The data from the transect thus seem to be suggesting the importance of the transition into the upwelling season and the rapid change in Chlorophyll a and SST fields within the shelf.

If this is indeed the case, the clustering is offering important insights into the enhancement of primary productivity along fronts and the importance of these events in the overall multivariate time series.

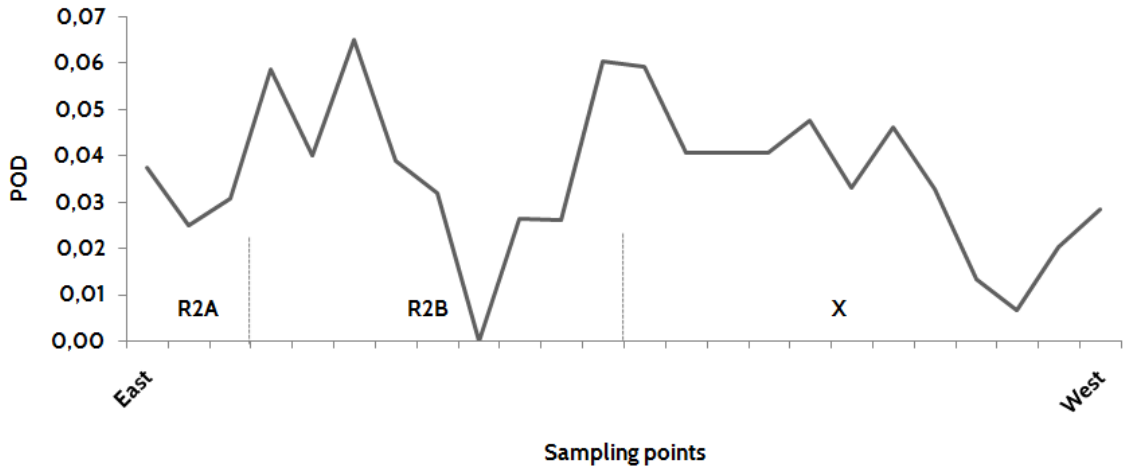


Figure 4.7.6. POD values at the Peniche transect, for the yearly frontal map of 2007. The 2007 SST k=15 regions (R2A, R2B, and X) are annotated into the plot. For a map of the regions, see figure 4.4.2.4D. Notice the coastline is represented to the left (East).

We also hypothesized that the yearly clusters could be more closely related to the frontal maps of the same year. Such correlation could, in theory, reduce the differences created by inter-annual variability.

Figure 4.7.6 depicts the relation between SST-based clusters calculated for 2007 and the POD map for that year. The regions seem to be separated by areas of intense frontal activity (notice the possible 'Berlengas Islands wake' in R2B) and thus it is indeed possible that yearly maps are more easily correlated with frontal maps. In fact, several transects yielded similar results when data from 2007 and other years were extracted.

Still, the difficulties inherent to yearly clusters render any conclusion difficult, considering the limitations that still persist in assigning standard region-codes to the clusters thus created.

In summary, the answer to the second question placed at the beginning of the section is also positive. Frontal regions are related to the regional boundaries of the SST and Multivariate clusters. The correlation is not simple as several different variables seem to play a role and even seasonal events can leave an important imprint in the regionalization effort.

The biogeographic Units generated by our approach are thus characterized by including different populations of fronts (with different mean POD fields) while being separated by regions of enhanced frontal activity, which can be either persistent (year-round) or seasonal.

Two new contributes to the study of regional oceanography were presented in this chapter. Firstly, the development of an objective, scalable, and reproducible bioregionalization scheme, based on uni- and multivariate time series created from the analysis of satellite imagery. The clusters are characterized by specific oceanographic conditions, which are highlighted by the differences in the patterns of SST variability across the year and inter-annually.

These abiotic forces have the ability to drive productivity and the overall structure of the ecosystem, including in the study area (Mason et al., 2006). The framework thus created is paramount to support the development of management policies that take into account the effective spatial boundaries of the systems (EU, 2007, Rice et al., 2012).

The approach suggested in this study is also relevant from a global perspective. Using DTW to generate distance matrices, which enable the development of a truly objective measure to compare time series at the pixel level, opens new possibilities in the field of biogeography. Successfully tested in the study area with uni- and multivariate datasets, the methodology can be scaled and applied globally. This is an important contribution to the body of methodologies available to meet the important goal of implementing classification systems in multiple domains in a coherent manner (Spalding et al., 2007).

The development of a long time series of frontal activity is not second to the development of the BC. In fact, the frontal maps provide yet another relevant tool that can help oceanographers, geologists, ecologists, and operational users to study the distribution of natural resources and plan their sustainable exploitation in new ways.

The frontal maps were compared against existing studies (e.g. Otero et al., 2009, Relvas et al., 2007, Peliz et al., 2005, Miller, 2009), and a wealth of new information was extracted from the complex and detailed maps. In fact the maps presented in this study outpace earlier versions available in the region by spatial and temporal resolution. This advancement was made possible by the use of a complex processing chain applied to MODIS AQUA data, which generated a substantial archive of frontal POD maps at various resolutions.

The California Current system, which shares considerable traits with the Portugal Current, both Eastern Boundary Currents, was up until now significantly more described in terms of frontal climatology. The present study helped close the gap and offer important details into the frontal climatology of the system. Throughout the text, we had the chance of comparing both systems and highlight the overall similarities, which are nonetheless, punctuated by important differences.

However some problems persist, which were also highlighted by Castelao et al. (2006) in a study of fronts off California. The challenges can be divided into two broad categories: operational and instrumental.

Operational challenges consist of the limitations to the monitoring efforts introduced by operational needs and cost considerations. The complexity and rapid evolution of several features call for the development of higher frequency maps. The time series of monthly data precludes (occasionally) the appropriate tracking of certain phenomena. The migration of upwelling fronts offshore, the development of filaments or the expansion of river plumes are some of the phenomena that should perhaps be monitored at a higher temporal resolution. This would also help prevent the introduction of significant noise into the POD maps that result of the short-term spatial variability of frontal boundaries.

However, if the objective is clear, implementing it is less straightforward. Frequent cloud cover precludes daily observations in many months (and regions of the world). Several studies rely on geostationary satellites to circumvent this problem (e.g. Hu et al., 2016, Legeckis et al., 2002, Peliz et al., 2002, Castelao et al., 2008). However geostationary sensors are often characterized by lower spatial resolution. This fact renders impossible the development of coastal front maps, with significant gaps in economically and ecologically relevant areas. The work by Castelao et al. (2006), despite its relevance to the field, did not cover the first 30 km off the coast due to the limitations imposed by the system.

These are thus a combination of operational and instrumental challenges, which are urgent to address. In the future, we aim at integrating multi-mission data from MODIS, VIIRS, Landsat, Sentinel-2/3, and other polar-orbiting and geostationary systems to create a truly synoptic and nearly continuous monitoring system. Although such effort is beyond the scope and timeline of the current project it is planned to be implemented shortly after the conclusion of this stage.

Finally, the dissemination of the datasets is often limited, restricting its use to a small community of scientific researchers. We also addressed this problem by introducing the development of the requirements and technologies necessary for the deployment of a comprehensive data portal. However, such portal must include not just the frontal maps, or

ecoregions but also information on the adjacent watersheds, which are responsible for so many of the features observed in the coastal system (Sherman, 2005).

The development of a stepwise strategy to develop such data portal included the development of adequate processing chains for ocean and land data, the automation of some of the processes, the lengthy interpretation process, and the final cartographic production. Thus far we have described the oceanic component of this workflow. In the following chapters we address the land cover products, the links rainfall and river outflow promote between terrestrial and oceanic systems, and the data distribution solutions devised. These solutions were developed with usability, cost effectiveness, and scalability in mind. These requirements are fundamental considering that all datasets produced, from frontal POD maps, to near-instantaneous frontal detection, or even the biogeographic classification maps are only as relevant as the use it is given to them by the end-user community.

PART III. The watershed-coastal continuum

Chapter 5. Watershed drivers of coastal ocean color

Combining coastal and watershed data into truly integrated management units is pivotal to ensure the continued success of policies aiming at the reduction of stressors and an overall improvement of the ecological status of ecosystems. The pressure exerted over coastal systems results of the sum of direct and indirect impacts, the latter often effected over watersheds that drain into the oceanic areas being analyzed (Lima et al., 2012).

The Large Marine Ecosystems approach offers an interesting example of a classification system, designed to organize coastal waters into coherent units, but connecting them to the adjacent watersheds (Sherman, 2005). This system was described in greater detail in an earlier section, but we mention it again because we consider it as an important template for future research.

In our work, we aim at developing a similar approach at the regional level. The aforementioned regionalization efforts, associated to the characterization of critical mesoscale phenomena can only be complete with the description of the adjacent watershed systems. Understanding the spatial distribution and intensity of human activities across the study region and how the rivers connect these sources of disturbance to the coastal waters is paramount to the future deployment of operational and integrated monitoring systems.

In the next chapters we shall present a set of case-studies of varying depth and scale that describe several metrics, methodologies, and visions for an integrated observation system monitoring the watershed-coastal continuum. Some of the results ought to be understood as exploratory seeking to open new perspectives and approaches to old, yet unresolved problems.

5.1. Impervious Surface Area (ISA)

5.1.1. Mapping land cover of watersheds: Impervious Surface Area

The present section consists of a paper originally published in Ecological Indicators (Mantas, V.M., Marques, J.C.M., Pereira, A.J.S.C., 2016. A geospatial approach to monitoring impervious surfaces in watersheds using Landsat data (the Mondego Basin, Portugal as a case study). Ecological Indicators, 71, 449-466. Permission for the unformatted reproduction was obtained from the editorial office of the journal.

The paper describes the development of an Impervious Surface Area (ISA) product for the Mondego river watershed, which serves as the precursor to a comprehensive Land Cover

dynamics product suite. The paper describes not only a new ISA product and the methodology towards its development but strategies to update the baseline map using medium resolution (Landsat) data. Sections and figure numbers were preserved to match those in the published paper.

Abstract

The urbanization of watersheds is a highly dynamic global phenomenon that must be monitored. With consequences for the environment, the population, and the economy, accurate products at adequate spatial and temporal resolutions are required and demanded by the science community and stakeholders alike. To address these needs, a new Impervious Surface Area (ISA) product was created for a Portuguese Watershed (Mondego river) from Landsat data (a combination of leaf-on multispectral bands, derived products, and NDVI time series), using Regression Tree Models (RTM). The product provides 30-meter spatial resolution ISA estimates (0-100%) with a Mean Average Error (MAE) of 1.6% and Root Mean Square Error (RMSE) of 5.5%.

A strategy to update the baseline product was tested in earlier imagery (2001 and 2007) for a subset of the watershed. Instead of updating the baseline product, the strategy seeks to identify stable training samples and remove those where change was detected in a time series of Change Vector Analysis (CVA). The stable samples were then used to create new ISA models using RTM. The updated maps were similar to the original product in terms of accuracy metrics (MAE: 2001: 2.6%; 2007:3.6%).

The products and methodology offer a new perspective on the urban development of the watershed, at a scale previously unavailable. It can also be replicated elsewhere at a low cost, leveraging the growing Landsat data archive, and provide timely information on relevant land cover metrics to the scientific community and stakeholders.

1. Introduction

More than half of the world's population inhabits urban areas (UN, 2014). In Europe, this figure is even higher, with 75% of the population (80% by 2020) living in cities (EEA, 2006). A similar urbanization trend was observed in Portugal in recent decades, where the urban population now reaches 42% of the total (INE, 2014).

In fact, urbanization is a global phenomenon that affects ecosystems, climate and the livelihood of human communities. In the United States alone, from 1982 to 2010, 17 million hectares of non-federal land were developed (USDA, 2010) and in Germany, 100 ha of land are

transformed into built-up/transportation classes every day (DESTATIS, 2007). The overwhelming pace of urbanization in the second half of the 20th century changed the face of European cities, with the emergence of 'sprawling' phenomena and profound transformations in the overall urban growth patterns (EEA, 2006).

This seemingly irreversible trend is not always accompanied by the development of efficient urban planning policies that could emerge from modern science and technology (Marinoni et al., 2013). Therefore, it is urgent to develop environmental indicators that can be employed routinely in the monitoring of urbanized ecosystems on a regional and global scale and at a rate consistent with that of development (Tiner, 2004). These rapid mutations of the landscape led to a profound change in the science of urban ecology (Pickett et al., 2008), which became increasingly integrated, with contributions from ecology, physical, social, and economic sciences (Grimm et al., 2000). This "human-environment system" calls for new perspectives (Turner et al., 2007), to foster the development of original management policies focused on sustainability, while accommodating growth and the population's legitimate aspirations (Marques et al., 2009).

Land Use and Land Cover (LULC) dynamics translate an otherwise abstract concept of change into measurable metrics, which can be objectively monitored to assess the impact over the health of land and aquatic ecosystems, the climate, and the territory (Loveland et al., 2002; Grove et al., 1997). However such monitoring is not without challenges, as the urban and peri-urban space is inherently complex and highly heterogeneous, truly a patchwork of multiple LULC classes (Xian, 2008; Bauer et al., 2008; Andrieu and Chocat, 2004). This complexity is the source of a multidimensional reality, with direct and indirect impacts over multiple biotic and abiotic variables, which may then influence social dynamics and interactions.

Stream hydrology and function are especially vulnerable to LULC (Morisawa and LaFlure, 1979) and the concept of 'urban stream syndrome' was devised to describe the set of changes associated to streams draining developed lands (Walsh et al., 2005). Streams in urbanized catchments are characterized by channel instability (Bledsoe and Watson, 2001), changes in the volume and duration of surface runoff (Weng, 2001), channel incision, loss of large wood and sediment load (Vietz et al., 2014). Furthermore, the increased input of pharmaceuticals (Veach et al., 2011), change in the nutrient load (Carey et al., 2013; Marinoni et al., 2013) and nonpoint-source pollution (Hurd and Civco, 2004; Brabec et al., 2002), or the change in sediment respiration rate (Feio et al., 2010) and in the benthic macroinvertebrate communities (Arco et al., 2012) can also be attributed to urbanization and imperviousness. In fact, urban riparian areas can lose the ability to act as nitrate sinks (Groffman et al., 2003) or even foster nitrification (Pickett et al., 2008).

Land development is thus tied to an intricate, complex web of stressors that are hard to replicate in a controlled setting and are specific to each water body (Feio et al., 2010).

Adding to the aforementioned effects of imperviousness, the urbanization of watersheds has been suggested to influence litter breakdown rates and species richness. This relation can be explained by the increased flow of watersheds with higher levels of imperviousness, which may promote the physical degradation of litter (Chadwick et al., 2006, Paul et al., 2006). In fact several papers address the impact of imperviousness over invertebrate and fish populations (e.g. Chadwick et al., 2006, Roy et al., 2003, Miltner et al., 2004, Klein, 1979), vegetation structure and health (White et al., 2006, Song et al., 2015), and bird populations, which are negatively associated to the presence of paved surfaces (McClure et al., 2015).

However, the impact of urbanization goes beyond the deterioration of water quality, and it is directly connected to a decrease in the distribution and fragmentation of natural vegetation. In turn, the new urban environment influences the thermal conditions, originating the Surface Urban Heat Island (SUHI) (Yuan and Bauer, 2007; Xian, 2008b) even in shrinking cities (Emmanuel et al., 2012). The SUHI consists in the increase of the surface temperature, observed in urban areas and that is strongly connected to the type of development and materials used (e.g. asphalt, rooftops) (Stathopoulou et al., 2007). The SUHI is the result of an accumulation of a significant amount of thermal energy, which may, in turn, generate a "thermal runoff response" (Kim et al., 2008) adding thermal pollution to the plethora of aforementioned stressors.

Interestingly, and despite the increasing global urban population, wide swaths of Europe and North America are subject to widespread shrinkage phenomena (Panagopoulos et al., 2013). In the case of Portugal, by 2060, the population is projected to fall to 8.6 million, from the current 10 million (INE, 2014). In parallel, and even prior to the country's population decline, changes at the city level led to the relocation of the inhabitants to peripheral communities. Portuguese municipalities tend to react to the decline through investment, especially when the decline is associated to the young population (Panagoupolos et al., 2013), which means that urbanization may not decline on par with population, but instead accelerate. This means that over the long term, in Portugal and elsewhere, the existing (or growing) infrastructure will have to be supported by a smaller population. This is thus a moment of particular importance in Europe, where the development of integrated monitoring strategies dedicated to developed lands takes center stage.

This reality is in contrast with that presented by many previous studies, which often focused on megacities (Duh et al., 2008, Grimm et al., 2008, Kraas, 2007) or regions of accelerated

growth (Davis and Schaub, 2005) that may not be representative of the regional or even national realities within parts of continental Europe.

In face of the increasingly complex and challenging scenario, cities and communities are struggling to improve environmental standards through the establishment of urban sustainability goals (Duh et al., 2008). Information, namely regarding the sealed or impervious surfaces that characterize urban environments, becomes central to the development of such planning and mitigation strategies (Bauer et al., 2008; Young et al., 2013). This context of extremes is motivating an increased awareness to the fact that data users need timely land cover products capable of offering an accurate depiction of the urban environment (Xian et al., 2010). The widespread use of remote sensing and Geographic Information Technologies (GIT) opened a new chapter in the characterization of land cover trends, watershed planning, and modeling (Tiner, 2004; Brabec et al., 2002).

Ridd (1995) first described the Vegetation-Impervious-Soil (V-I-S) model as a possible answer to quantitatively describe the urban morphology and therefore fulfill the emerging promise of accurate land cover mapping brought by multispectral satellite imagery. In recent years, and especially after 2000, a growing number of studies addressing the use of remote sensing of developed lands have been published although significant challenges persist (Weng, 2012).

Yang et al. (2003) describe the “anthropogenic impervious surface” as an indicator of the spatial extent and intensity of urban development. Rooftops, roads, sidewalks, and compacted soil constitute examples of impervious surfaces (Arnold and Gibbons, 1996), which are ecologically important because of their nature and arrangement (Schwarz, 2010). The Impervious Surface Area (ISA) – the impervious fraction within a pixel - can thus become the “self-consistent metric” enabling accurate comparative analysis of urban extents in transboundary regions (Xian, 2008).

Categorical classification systems (e.g. Anderson, 1976) can at times create artificial dichotomies between natural and developed lands, ignoring the continuum best described by pragmatic metrics that can be used in interdisciplinary studies (Cadenasso et al., 2007; Arnold and Gibbons, 1996). ISA does not describe function but only the amount of artificialization, which can be the driving force behind stream degradation and nonpoint-source pollution (Slonecker et al., 2001) at thresholds as low as 10% (Arnold and Gibbons, 1996). Although several authors suggest this threshold as key for stream degradation, permanent changes can be detected earlier, depending on the geographic context and parameter in analysis (Vietz et al., 2014; Chin, 2006; Bledsoe and Watson 2001).

Previous studies addressed the problem of ISA mapping using different or adapted strategies, sometimes relying in significant generalizations (Esch et al., 2009). Data sources range from

optical sensors, RADAR, and LiDAR and techniques include Regression Analysis (Sexton et al., 2013, Bauer et al., 2008), Classification and Regression Trees (CART) (Jiang et al., 2009), Spectral Mixture Analysis (Li et al., 2015, Zhang et al., 2014, Wu, 2004, Yuan and Bauer, 2007), Hyperspectral data analysis (van der Linden et al., 2009), Maximum Likelihood Classification (Jia et al., 2014), Multivariate Texture (Zhang et al., 2014b), and onscreen interpretation (Loveland et al., 2002). We suggest Weng (2012) for a comprehensive review of recent and historical methods used for ISA mapping.

Model performance is a function not only of the methodology but also of internal heterogeneity of the target area (Parece et al., 2013). Thus, the comparison of methodologies relying on multispectral data, including regression modeling, regression trees, and SMA is often site-specific, but performance differences are frequently negligible (Yuan et al., 2008). Hyperspectral data, on the other hand, promises to meet the spectral requirements of urban mapping (Herold et al., 2003), but the lack of systematic and affordable data limited the employment of these datasets in systematic mapping programs. Combining Synthetic Aperture Radar (SAR) and optical data may improve imperviousness estimates in all-weather conditions, through the reduction of the uncertainty associated with bare soil, shaded areas, and water surfaces (Zhang et al., 2014). Sentinel-1, a SAR-equipped satellite of the European Space Agency (ESA) is likely to give an important contribution to accelerating research on this topic, considering the high spatial and temporal resolution as well as the open access policy. This is particularly important considering that the use of a single SAR image yields poorer results when compared with the equivalent use of a unique Landsat scene (Zhang et al., 2012).

High-resolution imagery can broaden the perspectives on urban mapping and interpretation, namely through the use of image segmentation techniques (Thomas et al., 2003). However, data costs are still prohibitive for regional or national mapping initiatives, especially when unsupported by government agencies or sizeable private companies. Previous studies often stress the relevance of data accessibility and spatial resolution in the data selection process. The necessary compromise between technical requirements and cost fostered the adoption of multispectral optical data as the preferred source of information (Deguchi and Sugio, 1994). Landsat, in particular, offers several advantages to these studies, including the long-term digital archive, adequate spatial and spectral/radiometric resolution, and open data policy (Goward et al., 2006; Yang et al., 2003).

Because urban landscapes are highly heterogeneous, and most pixels composed of a mixture of different surfaces with different spectral properties (Wu, 2004), subpixel analysis is needed to quantify the contribution of each component to the overall structure (Yang et al., 2003; Xian, 2008). Therefore, ISA, as a continuous variable, avoids the problems inherent to nominal

classes, which is particularly relevant within the urban and peri-urban matrix (Bauer et al., 2008). The spatial resolution of current Landsat sensors is particularly suited to pursue this goal. A comparison with coarser products (e.g. CORINE 2012, Figure 1) emphasizes the importance of accurately depicting the connectivity of artificial surfaces and measure relevant metrics at the sub-pixel level. The spectral mixture of complex, large pixels in urban and peri-urban landscapes hampers the accurate retrieval of land cover metrics and impedes the depiction of the continuous urbanization gradient.

As aforementioned, the importance of monitoring urbanization derives in part of the rapid evolution of urbanized regions worldwide. However most land cover products are published several years after the original image acquisition, and are therefore outdated when published (Xian et al., 2010). It is therefore necessary to provide stakeholders and managers with high temporal resolution products focused on urban development, which requires a significant effort into the development of updating strategies (Jin et al., 2013; Weng, 2012; Sexton et al., 2013; Jensen, 1999).

Existing strategies used to update ISA maps differ, but often require some level of prior knowledge on land cover, the use of dense time series or of anniversary imagery, requirements that are not always easy – or possible – to fulfill (Jin et al., 2013; Jia et al., 2014; Xian; Huang et al., 2010).

A common trait to all visions is the requirement of high quality ground truth samples, which is one of the critical and most time consuming components of the process.

In Europe, a fast and cost-effective approach to ISA mapping is still missing (Esch et al., 2009). The CORINE Land Cover, a project of the European Union, partially fulfills this objective by providing cartography of European countries (39) using 44 thematic classes at spatial resolution of 100 and 250 meters (CORINE, 2012). These products are often used in numerous studies and applications where land cover is relevant (Martínez-Fernández et al., 2015; Suau-Sanchez et al., 2014; Pilli, 2012; Teixeira et al., 2014) , with urban classes being merged to represent the "sealed surfaces" (Schwarz, 2010). However, CORINE land cover database does not have the necessary thematic and spatial resolution for several critical hydrologic and watershed management applications (Macary et al., 2014) or biomass and carbon assessment studies in the urban context (Raciti et al., 2014).

The European Copernicus program also publishes five high resolution layers (HRLs) on land cover characteristics that can provide additional information on land cover, including imperviousness (Langanke et al., 2013). Validation of the Copernicus HRL is still ongoing and therefore it is difficult to ascertain the quality of the product in its current version. Data

distribution of the Copernicus layers is subject to restrictions, limiting access to the information by the potential user community.

1.1. Research Objectives

This study describes the development of an Impervious Surface Area (ISA) product for a baseline reference date of 2013 (ISA2013) covering the entire Mondego river Watershed, in Portugal (Figure 1). Focusing the study in a watershed is a strategic decision that aims to avoid the dissection of an otherwise geographically coherent unit into multiple ecologically-irrelevant administrative regions (Grimm et al., 2000).

We also present a processing scheme for semi-automated updates and test the approach for two periods (2001 and 2007) in two representative and contiguous counties (Coimbra and Condeixa-a-Nova), characterized by heterogeneous LULC. The update strategy differs from other mapping initiatives because instead of identifying the changed pixels in the baseline ISA map, it searches for stable training samples. These samples are then used as independent variables in a regression tree model, together with new samples that replace the changed ones if necessary. We hypothesize that this strategy, 1) originates a more homogeneous pool of training samples that can be used across different dates and even as mapping methodologies evolve, 2) results in more homogeneous and comparable products, 3) updated ISA maps do not necessarily carry the errors of previous model runs. Finally, the advantages of a regional approach to imperviousness mapping are discussed along with plans for the operational distribution of the generated datasets in a platform designated as “Smart Basins”.

2. Materials and methods

The methodology can be summarized as the following sequence (Figure 2): a) acquire and pre-process remote sensing and ancillary data (section 2.2); b) create detailed samples of pervious/impervious surfaces through the visual interpretation of high resolution imagery (section 2.4); c) create the baseline Impervious Surface Area model for 2013 using Landsat and ancillary data (section 2.5); d) characterize change in a subset of the study area (section 2.6); e) update high resolution samples (created in b)), through the removal of the changed areas from the pool using a method based on annual time series of Change Vector Analysis (section 2.6); f) update the ISA models (2001 and 2007) using the updated training sample database (section 2.6).

2.1. Study area

The Mondego river is the largest under exclusive Portuguese administration (Figure 1) with a basin covering 6670 km^2. The watershed is often divided into three different regions including the estuary, the "Lower Valley" (to the West) and the "Upper Valley" (to the East) (e.g. Teixeira et al., 2014).

The "Upper Valley" is characterized by topographically complex terrain where the headwater is located ("Serra da Estrela", with a maximum elevation of 1993 m). This area is geologically dominated by igneous (granites) and metamorphic rocks (mainly schist and graywackes). The western sector of the basin ("Lower Valley") is characterized by gentle slopes and mild elevations with sedimentary rocks (mostly sandstones and limestones) of Mesozoic and later ages, often supporting valuable farmland along the flat floodplains. The river meets the Atlantic Ocean in a small mesotidal estuary, subject to strong anthropogenic pressure (Marques et al., 2002; Cunha and Dinis, 2002). Agricultural, industrial, and other human activities upstream are responsible for frequent eutrophication episodes (Marques et al., 2003).

The basin is home to an estimated permanent population of 712969 (ARH Centro, 2012) in 46 counties. The most significant urban centers (counties) are Coimbra (138 058 inhabitants in 2013), Viseu (98 601), Figueira da Foz (61 291), and Pombal (54 413) (INE, 2014). Central Portugal, a statistical unit that comprises the Mondego river Watershed is subject to a significant population decline, having lost 58 111 inhabitants between 2001 and 2013 (INE, 2015)

According to the CORINE 2012 land cover product, only 3.3% of the watershed is occupied by artificial areas (Figure 1). Forest and semi-natural areas (64.3%) and agricultural uses (31.8%) occupy the bulk of the surface cover across the basin. Geology and topography control the distribution of these classes, which are, nonetheless, found evenly interspersed in the watershed.

The socio-economic relevance of the territory supports the integration of the imperviousness products with the human dimensions of the basin, in support of the development of a future, integrated, data analysis platform.

2.2. Remote Sensing datasets

Remote sensing data were used to generate the training samples (used as 'ground truth') and as the main input to the ISA model and change detection protocol. The delineation of the training samples followed a protocol based on the visual interpretation of high resolution satellite and aerial imagery, while the ISA models were built using Landsat data. Both protocols are described in the following sections.

Three different high resolution data sources were used as 'ground truth' across the different steps of the workflow. Notice that the high resolution imagery was not employed directly in the implementation of the ISA model, but only as the source of information for the delineation of training samples and validation of the methodologies.

For the *circa* 2001 coverage, we obtained aerial photographs in natural color mode acquired by the *Instituto Nacional de Intervenção e Garantia Agrícola* (an Institute of the Portuguese Ministry of Agriculture) in the year 2000. The images are provided as orthophotomaps with a spatial resolution of 0.5 or 1m, depending on the geographic area. In this study samples from both collections were used.

For the 2007 regional ISA model, we relied on aerial photographs acquired by the Portuguese Geographic Institute (IGEO) in the same year. The mosaics are distributed as 'true color' orthophotomaps with a spatial resolution of 0.5 meters. Finally, the calibration and validation of 'ground truth' datasets for the 2013 baseline ISA product were created using high resolution satellite imagery acquired in 2012 and 2013 from multiple sources including QuickBird 2 and WorldView 1/2. The data were accessed as 'true color' pan-sharpened, orthorectified images with a spatial resolution between 0.5 and 0.65 m depending on the sources.

All high resolution images were resampled to a common spatial resolution of 1 meter to standardize mapping procedures across regions and dates, and allow a common Minimum Mapping Unit, as described in the following section. True color composites were used in the interpretation process, since it was the only band combination common to the datasets from all periods. The use of additional bands would have contributed to improve the delineation of the training samples, by reducing the uncertainty in the classification as pervious/impervious. However, in the absence of data with common spectral attributes, it was considered preferable to homogenize the data sources used for classification.

The development of the ISA model relied on the use of Landsat Surface Reflection (SR) products for both the baseline 2013 research-grade product and the 2001 and 2007 updating tests. For the 2013 product, all imagery was acquired by the Operational Land Imager (OLI) equipping Landsat-8. For both the 2001 and 2007 tests, a combination of TM and ETM+ data

from Landsat-5 and 7 were used. Two scenes are required to map the entire watershed (Path 204 Row 032 and Path 203 Row 032).

Previous studies differ on the atmospheric correction methods (if any) employed to normalize the Landsat imagery. The Multi-Resolution Land Characteristics Consortium 2001 protocol (Chander et al., 2009) limits processing to at-sensor reflectance. This protocol is then used in by other authors (e.g. Xian *et al*., 2009; Homer *et al*., 2004; Jin et al., 2013; Yang et al., 2003; Wu, 2004). Other studies use more sophisticated atmospheric correction methods including the ATCOR2 algorithm (Esch et al., 2009) or 6S (Sexton et al., 2013). Huang et al. (2010) reports there are little differences in the detection of disturbance and classification using Surface Reflectance (SR) and At-Sensor imagery.

Nonetheless, in this paper, because we used a long time series acquired by different sensors (Pacifici, 2014), we adopted a conservative approach, relying on SR data, made available by the United States Geological Survey for Landsat 4-5 TM, -7 ETM+ and -8 OLI (Masek et al., 2013, USGS, 2015).

The imagery selection process included a visual inspection and the removal of scenes with cloud cover exceeding a pre-defined threshold of 8% for the relevant land area. The Quality Assessment band (QA) distributed with the Landsat SR product was relevant in this context by providing a simple and objective way to determine the usable area within the scene. The presence of dust and smoke, wildfires or haze was also considered in the quality assessment. For selected scenes, areas affected by clouds were filled by temporally close acquisitions (within 1 month of the target image).

Although a single image was used as reference for each product date (2013 and the 2001 and 2007 updates), a larger number of images were acquired for the extraction of ancillary data, including phenology information. We hypothesized that the fluctuations in the Normalized Difference Vegetation Index (NDVI) across a phenological cycle could improve the performance of the ISA model. Urban areas have generally stable and low NDVI values, whereas forests and semi-natural areas have higher NDVI amplitudes across the year and generally higher values. Cropland displays significant NDVI fluctuations across the growing season, which can be captured in the time series. As such, the highest NDVI value was extracted for each reference year (per pixel) as well as the amplitude value, using the highest and lowest values within the $\mu \pm 2\sigma$ range, in order to minimize the impact of image artifacts. This expedite phenological characterization was based on 8 images for 2013, 15 images for 2007, and 6 images for 2001. The number of images used in 2007 reflects the need to compensate the localized data loss caused by Landsat-7's SLC-OFF malfunction. Missing values were not filled in this instance.

On the other hand, for change detection, and considering that we were interested in inter-annual change, one image was selected for each year between 2000 and 2013. The anniversary images were selected using as reference the baseline 2013 image used to build the ISA model (acquired in 2013/251). The SLC-OFF malfunction was addressed in this case, given that for the years 2004, 2005, 2008, 2009, and 2012 no alternate Landsat dataset was available. The missing values of the anniversary image were filled using same-season (summer) imagery acquired in the closest viable acquisition(s) of two preceding months. The replacement data were first histogram-stretched to reduce the phenological differences between both acquisitions, using the intact pixels as reference. It is also important to notice that the test areas, where the Change-Vector Analysis was applied are only marginally affected by the SLC-OFF problem, due to their location within the scene. With the increasing availability of Landsat-8 and Sentinel-2 data, the application of the method in the future should be streamlined.

The multispectral bands were used to compute several products including Principal Component Analysis (PCA), Normalized Difference Vegetation Index (NDVI), Normalized Burn Ratio (NBR), Normalized Difference Moisture Index (NDMI), and the ratios $B6.B7^{-1}$, $B5.B6^{-1}$ and $B4.B2^{-1}$ (band numbers relative to the Operational Land Imager).

2.3. Ancillary Data

To support the development of the ISA models, ancillary data were used to provide context information, which may be relevant in the construction of the models and interpretation of results.

Topographic contour lines were extracted from the military maps of the watershed at a scale of 1 : 25 000 and used to build a 30m Digital Elevation Model (DEM) collocated to the Landsat imagery. Slope maps (percentage), slope aspect (degree), Valley Depth and Dominant Landform rasters were generated from the original DEM using SAGA GIS. The watershed was also segmented into sub-basins, using automated stream delineation and catchment area calculation tools, available with the ArcHydro package for ArcGIS.

A simplified Geological Map at a scale of 1 : 100 000 was also used in the construction of the ISA models. The vector map was converted to cells, collocated to the Landsat imagery. Each cell was assigned to the dominant geological formation within the pixel area.

A land cover product distributed by the Portuguese government (COS2007) and produced from the interpretation of aerial photography, with a Minimum Mapping Unit of 1 ha (IGP, 2010) was used in support of change-vector analysis.

Census records were downloaded from the Portuguese Statistics Institute (INE) for the counties and parishes ("Freguesias") of the Mondego river Watershed.

Road and street data were downloaded from the Open Street Map website (OSM, 2015). To each road segment, attributes on speed limit and direction were added to the geodatabase. The street data were used to calculate average transit times (service areas) from central Coimbra, using the Network Analyst tool available in ArcGIS.

2.4. Training Samples

The development of the ISA model (2013) requires the delineation of high spatial resolution training samples, which are used as ground truth in the ISA regression tree model. The samples were manually mapped by the authors upon interpretation of the high resolution satellite and aerial imagery. A common Minimum Mapping Unit of 4 m^2 was used in the delineation of pervious and impervious surfaces, which were stored as vectors in a geodatabase. Each contiguous mapping block covered a minimum area of 8100 m^2 (equivalent to 3x3 Landsat pixels), i.e. no smaller and disconnected samples were mapped.

The resulting binary (pervious/impervious) vector file was then converted to a regular grid with a spatial resolution of 30 m and collocated to the Landsat pixels. Each cell thus generated contains the percentage of pervious and impervious fraction within the pixel area. A total of 148055 cells (covering 133.25 km^2) were created using this method and used in the calibration of the ISA regression tree model. An additional set covering 14 000 pixels (*ca*. 10% of the calibration set) was used to validate the model, using the same mapping methodology.

The samples are representative for different land cover classes and intra-class variability. Particular attention was given to different types of urban infrastructures and housing types to account for the wide range of materials and development intensities. Given the area of the watershed, and despite its internal diversity, no stratification of the AOI was made based on ecoregions or topography. Nonetheless, sample data were stratified by vegetation cover type to account for the seasonal variability in greenness between vegetative cover types (Bauer et al., 2008). The stratification resorted to the classes included in the CORINE 2012 land cover map. An equivalent number of sampling points were delineated for each class.

The method described in section 2.6, addresses the selection of the training samples thus created, from stable areas, for ISA product updates (in the study for the reference years of 2001 and 2007).

2.5. ISA Regression Tree Model

Regression Tree Models (RTM) are machine-learning algorithms described in great detail by Breiman et al. (1984). RTM algorithms partition the original sample and create a tree that includes a combination of a descriptor and a linear regression (Yang et al., 2003; Xian, 2008). The combination of all the rules and equations describe the sample used to recursively build the model, simplifying nonparametric and nonlinear relationships between the different variables. In the case of the ISA model, each linear regression enables the estimation of the percentage of impervious surface, per pixel, for a subset of the complete sample. The model relies on high quality training samples which are used as a continuous dependent (target) variable. The satellite imagery and other ancillary data (topography and geology) become the independent or explanatory variables during model development. The imagery used includes anniversary, leaf-on Landsat-8 OLI multispectral bands and derived products as well as a set of two rasters containing the Maximum and Minimum NDVI value for the hydrological year. The ISA2013 model was built on Weka using the M5P algorithm (Quinlan, 1992) and translated into GeoTIFF rasters (one per rule) using a custom python script developed by the authors for ArcGIS. Each raster was then combined into a single image file containing the ISA estimates for the entire watershed. The product describes the imperviousness level of the Watershed, at a spatial resolution of 30 meters and for the reference year of 2013.

The validation of the model outputs relied on traditional accuracy metrics (Janssen et al., 1994) applied to independent reference data (equivalent to 10% of the calibration set). Amongst the metrics used is the Mean Average Error (MAE), which is calculated as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |ISA_m - ISA_g| \quad (1)$$

Where ISA_m is the estimated impervious surface area and ISA_g is the ground truth for the same pixel.

The Root Mean Square Error (RMSE) was also calculated for all outputs. The RMSE is calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ISA_m - ISA_g)^2} \quad (2)$$

The Pearson correlation coefficient was used to assess the linearity of the correlations between ISA estimates and ground truth data. The Pearson correlation coefficient (r) is calculated as follows:

$$r_{xy} = \frac{Cov(ISA_m Y)}{\sqrt{Var\,(ISA_m)}\sqrt{Var\,(Y)}} \quad (3)$$

where Y is the variable under study.

Univariate least squares regression analysis was used to determine the correlation between ISA2013 and census data. Analyses were conducted in several variables including population at county and parish ("freguesia") levels as well as county purchasing power (PP). For each variable, the hypothesis $H_0 : \beta_1 = 0$, where β_1 is the model's slope parameter, was tested for a significance level of 1% at Microsoft Excel. Considering the mismatch between watershed limits and the administrative boundaries of counties and parishes, it was not possible to extend the analysis to the entire watershed. Instead, only counties in which 70% of the territory was located within the Mondego watershed and contained the most significant urban center were considered, totaling 24 municipalities. Of these, 10 counties were entirely located within the watershed and 10 had over 90% of the territory inside the study area. In the case of parishes, the smaller area simplified the process, and all jurisdictions entirely within the watershed were considered (n=241). Taking into account the large sample it was deemed unnecessary to include parishes that spread across multiple watersheds, thus limiting the sources of error.

Using the ISA2013 dataset, the imperviousness of sub-basins found within the watershed was determined, in order to compare the estimated values with the 10% and 30% ISA thresholds suggested by Arnold and Gibbons (1996). The sub-basins were delineated from the regional Digital Elevation Model using ArcHydro (at ArcGIS). The minimum drainage area (flow accumulation) threshold required for stream and sub-watershed delineation was set to 10 000 pixels for a raster with a spatial resolution of 25 m. The ISA2013 was then calculated for each sub-basin using zonal statistics and compared with the predefined stream status thresholds.

2.6. Change detection and updates to the baseline product (ISA2013)

The development of high quality training data is one of the barriers preventing frequent updates to land cover maps (Yang et al., 2003). Therefore, previous studies discussed the advantages - and disadvantages - of manual photo interpretation against automated classification (Xian, 2008) and strategies for sample selection (Jia et al., 2014).

One of the objectives of the study is to assess the feasibility of future near real-time updates to the baseline 2013 product (forward and backward analysis). Instead of updating the changed pixels in the ISA map, our methodology identifies the stable training samples and runs the model using this set. New samples can be created to replace the flagged ones as needed, in order to maintain the variability of ISA within the pool of training areas. For this purpose, a simple strategy was devised that relies on the construction of a time series of change, based

on the analysis of anniversary multispectral Landsat Surface Reflectance data. To reduce the amount of change due to phenological differences, a maximum tolerance of 1 month from the reference date was allowed (Xian et al. 2010). The reference anniversary data were acquired over the summer to reduce the effect of topographical shadows. The test was conducted in the contiguous Coimbra and Condeixa-a-Nova Counties, which are characterized by a diverse land cover, terrain, and geology. Furthermore, the two counties display different population growth trajectories, with significant population movements from Coimbra to Condeixa. This fact poses interesting opportunities to study the influence of population dynamics over ISA.

Image pairs (yearly) were analyzed using the Change Vector Analysis (CVA) methodology (Malila, 1980) from the year 2000 up to 2013.

Change Vector Analysis magnitude is calculated as:

$$\rho = \sqrt{\sum_{n=1}^{6}(X_{n,r} - X_{n,s})^2} \qquad (4)$$

where, the number of reflectance bands of Landsat is represented as n, X_r is the early date image and X_s the late date image. The CVA yields a change magnitude and it is through the comparison against a pre-defined or empirically derived threshold that change is detected and flagged. Average CVA magnitudes differ for different land cover classes and can be adjusted for automated change detection (Xian et al., 2009, 2010). However, in this study, no prior land cover information was available at the same scale and therefore a different approach was required.

After this initial step, the highest value in the time series is assigned to the pixel (CVA_{max}). The CVA_{max} can then be compared against a threshold to determine whether change occurred over the period. In summary, the CVA_{max} is calculated from all anniversary image pairs available from the date of the initial baseline product (not the last update) to the date of the new ISA map. The process can be applied both to earlier or later dates. The identification of an optimum threshold for change detection was made empirically as follows.

The Portuguese COS2007 land cover map was used to randomly create 400 points for each of the consolidated land cover classes (artificial, forest and semi-natural, agricultural areas, wetlands). For each point, the CVA_{max} was retrieved along with a change/no-change flag based on the interpretation of the 3 sets of high resolution imagery already described in section 2.2. Using the information thus acquired, a conservative (common) value was determined that enables the selection of unchanged areas across the different land cover types. The maximum CVA for the entire study period (CVA_{max}) admitted for a pixel to be considered as no-change and maintained in the pool of training samples matches the 2.5% lower values found within the watershed (or a CVA_{max} of 0.1 in the study area).

A training sample located in a pixel marked as changed is flagged in the database and is not used in subsequent product updates. It can, nonetheless, continue to be used in the reprocessing of earlier-date products, before change was signaled by the method.

This approach is only possible because we are not updating the cartographic products but removing the changed training samples used in the classification process. This may lead to an over extraction, which in this case is preferable over the use of erroneous inputs to the model. By updating the training samples and not the earlier-date product (i.e. as in the National Land Cover Database approach), it becomes possible to update the mapping methodologies as new sensors and methods become available, without creating hybrid cartographic products.

A confusion matrix was built to calculate the accuracy of the concept, using an additional independent set of 316 random points, which were visually analyzed for change detection using the three sets of high resolution imagery (2001, 2007 and 2013). The interpretation was then compared with the automated change detection protocol for validation purposes.

Upon validation, the methodology was applied to the 2013 baseline training sample described in section 2.4, to create products for 2001 and 2007. A subset of the complete pool of training samples (watershed-wide) was generated from the data located in the two test counties for the purpose of this test. The ISA2013 product was then reprocessed (herein named $ISA_{sub}2013$) using the training sample subset to guarantee comparability between the baseline subset and the updated products. The sample subset was then screened for change using the aforementioned CVA-based methodology and ISA products generated for 2001 and 2007, using the Regression Tree Model method described in section 2.5.

The three ISA datasets were analyzed in detail in the two test counties and changed pixels were extracted after identification using the aforementioned methods. The changed pixels were quantified and then characterized according to a three-class system, upon visual interpretation of the high resolution imagery. The categories considered included residential areas (e.g. apartment buildings and single-family housing), services (e.g. commercial areas, schools), and infra-structure (e.g. roads, parking lots). These metrics were calculated for the subsetted test area, individual counties (n=2), parishes (n=41), and service area rings (i.e. driving times from Coimbra's city center, as described earlier).

Two hypotheses were tested using the ISA time series, including whether 1) the expansion of ISA was correlated with population dynamics and 2) whether the type and intensity of new developments were correlated with initial imperviousness levels.

To evaluate the impact of population change, the parishes were segmented into negative population growth (n=22) and neutral or positive growth (n=19) groups. For each parish, the

total area of newly developed pixels was calculated for each development class. Finally, single-factor ANOVA was applied to the set, for a significance level of 5%, using Microsoft Excel.

The influence of initial conditions (pre-development, ISA2001) over subsequent ISA (ISA2013) values and development typology (residential, services, and infra-structures) were tested using ANOVA, for a significance level of 5%. This test was followed by Tukey-Kramer's test, both of which were applied at Microsoft Excel. ANOVA was applied to test the equality of means between all groups, after which the Tukey-Kramer test was used to evaluate the pairwise differences between all three development classes. This is adequate as a post-hoc multiple comparison method, which can be applied when the sample sizes are unequal, which was the case in the current application. To this end, the pre-development (ISA2001) and post-development (ISA2013) imperviousness values were extracted for all the changed pixels, which had been manually segmented into the three different development classes.

3. Results

3.1. Impervious Surface Area 2013 (ISA2013)

The ISA2013 provides basin-wide imperviousness surface fraction estimates at a spatial resolution of 30 meters for circa 2013. This model was successfully built for the entire Mondego river Watershed (Figure 3) and illustrates the different urbanization patterns across the watershed. The ISA2013 map highlights how the densely populated centers of Coimbra and Viseu contrast with the smaller county seats and the forested regions of the Southeast.

The model, based on 148055 training samples (or pixels), included 302 different rules and associated linear regressions, for a predefined minimum of 100 training samples per rule. Several error metrics were calculated to assess the performance of the model. The Correlation Coefficient (*r*) was of 94.5%. The Mean Absolute Error (MAE) was of 1.6% and the Root-Mean-Square-Error (RMSE) of 5.5%.

The correlation between the ground truth samples (validation subset) and the ISA2013 output is depicted in figure 4, which further attests to the good performance of the model.

A derived, categorical map was created from the original ISA2013 with 5 classes: Undeveloped (ISA < 10%), Developed Open Space (DOS) (10% ≤ ISA < 20%), Developed Low Intensity (DLI) (20% ≤ ISA < 50%), Developed Medium Intensity (DMI) (50 ≤ ISA < 80%) and Developed High Intensity (DHI) (ISA > 80%).

The ISA2013 was also visually inspected against the only existing comparable product, the Copernicus Imperviousness High Resolution Layer 2012 (CIHRL). The CIHRL is available as a web service only and no official validation has been published yet.

Figure 5 compares the imperviousness as estimated by both products for the urban area of Viseu. The CIHRL and the ISA2013 share the common outline of the urban area. Despite the innovative nature of the Copernicus layer, it depicts impervious pixels in a much more discontinuous way, even when ground truth suggests otherwise. In the map, we highlight such situations in A) and B), where the roads are only partially represented or not at all in the CIHRL. The urban center is also depicted in a less realistic manner in the CIHRL, with numerous isolated patches of imperviousness present within the urban matrix. The cumulative impact of these errors is difficult to estimate from a visual interpretation, but may lead to an important underestimation of the imperviousness at the local or regional level.

3.2. Change Detection and product updates

Determining a threshold for CVA_{max} was an important step of the product update to earlier dates (2001 and 2007). Two contiguous counties (Coimbra and Condeixa-a-Nova) were selected to test the method. Coimbra is the most populous county in the watershed with 138058 inhabitants in 2013, INE, 2015) and is connected to Condeixa-a-Nova (17346 inhabitants in 2013) by numerous routes. In fact, the city of Condeixa-a-Nova is the only county seat located within 20 minutes of Coimbra's city center, under optimum driving conditions, as demonstrated by network analysis at ArcGIS (Figure 6). Furthermore, the population of Condeixa-a-Nova is increasing, unlike that of Coimbra. It thus offers and interesting comparison of two population centers with a different structure and population dynamics, which might be reflected in the ISA patterns and evolution.

The validation of the CVA-based method suggests that the use of the 2.5% lowest values found within the area (CVA_{max}) is appropriate for the selection of a pool of stable training samples. The overall accuracy of the method is of 94.6% (91.9% Producer's Accuracy and 95.5% User's Accuracy), with a Cohen's Kappa of 0.89. The methodology was thus used to select the suitable pixels for model calibration using 6-year steps (2001 and 2007).

The ISA models were produced after the training samples were screened for change in the subsetted test area using a total of 24158 points for 2013 (extracted from the original pool created for the watershed-wide Research Grade ISA2013), 20997 for 2007, and 18532 for 2001. This means that over the course of 12 years, 76% of the original training samples were still stable, despite the accelerated changes in some areas within the study region. A

percentage split of 10% of the sample was chosen to validate the model, using the same error metrics as in the watershed-wide ISA2013.

The ISA2001 had a Pearson Correlation Coefficient (*r*) of 96.6%, with a MAE of 2.6% and RMSE of 7%. In comparison, the ISA2007 had an *r* of 95.7%, a MAE of 3.6%, and RMSE of 8%.

The ISA2013 was reprocessed for the purpose of validating the product update strategy using the pool of points located within the subsetted area of interest ($ISA2013_{sub}$) (see Figure 2 for a description of the workflow). In this instance, the Pearson correlation coefficient was of 96.4%, with a MAE of 2.8% and RMSE of 7.4%. All error metrics show, to some extent, an inferior performance when compared against the 2013 watershed-wide model, possibly as a consequence of the smaller number of ground truth samples used in the calibration.

The production of the 3 models shows the rapid increase of impervious area within the counties (Table 1, Figure 6). In both cases the change is more abrupt between 2007 and 2013. A simultaneous increase in the diversity of imperviousness fraction values occurred in both counties.

In Coimbra, the average county ISA increased from 8% in 2001 to 9% in 2007, reaching 12% in 2013. From 2001 to 2007, the developed open space and medium intensity ISA classes increased 21.5% and 22.4% respectively. In the period between 2007 and 2013, the fastest growing class is the High Intensity ISA, which increases 51% in comparison with the area estimated for the previous period.

On the other hand, the development in Condeixa-a-Nova is taking place at a faster rate. From 2001 to 2007, developed high intensity pixels increase by 102% and again 101% from 2007 to 2013. The second highest growth rate takes place in Developed Open Space pixels, with an increase of 59% in 2001-2007 and 48% in the later period.

A total of 8944 pixels (or 1.75% of the total area of both counties) were flagged for change and visually inspected. The pixels were characterized in terms of the prevailing function of the newly developed areas, from interpretation of the high resolution imagery and field surveys. Three categories were considered including residential areas (apartment buildings and single-family housing), services (e.g. commercial areas, schools), and infra-structure (e.g. roads, parking lots). Residential areas account for 32.9% of the flagged pixels, Services cover 30.6%, and infrastructure the remainder 36.4% of the total. To determine whether the expansion of impervious areas was driven by population dynamics, the three categories were analyzed for each of the 41 parishes ("Freguesia", the smallest administrative unit of the country) that form the two counties (Figure 6). The data from the parishes were segmented into positive(and neutral) and negative population change groups as aforementioned. No significant differences

were found in all three cases, including for residential areas, after analysis using ANOVA ($p<0.05$) was applied.

Nonetheless, the starting Imperviousness (ISA2001) in the new developments was found to be different between all categories (ANOVA with a $p<0.05$, followed by the Tukey-Kramer Test). The lowest initial ISA (before change) was identified in pixels where change was dominated by the construction of new infra-structures (mean ISA2001 of 5.4 (standard deviation: 10.3)%. Development for services took place in pixels with an initial ISA (2001) of 9.1 (12.9)% and residential areas in cells with a mean ISA2001 of 11.5 (12.6)%. The final ISA (2013) in flagged pixels was, unlike the starting ISA, different for Residential areas (65.5 (15.9)%) when compared with the other categories, but not between Infra-structure (69.2 (16.6)%) and Services (74.4 (17.6)%) (ANOVA, $p< 0.05$, followed by Tukey-Kramer).

The distance from the center of Coimbra, measured as the optimum driving time (service rings, Figure 6), offers additional insights regarding the dynamics of urban development. 72.8% of all pixels flagged as changed are found within the two inner rings (0-15 minutes), even though the ring area accounts for just 45% of the total area of the combined counties. On the other hand, the 20-25 minute travel-time ring, which covers 29.6% of the two counties, includes only 4.1% of the pixels flagged for change.

The analysis of ISA2013 patterns within the service area rings, highlights the importance of urban centers in attracting development (Figure 6). Development in the inner service ring (0-10 minutes from the city center) is dominated by ISA levels over 20% (DLI: 12.9%, DMI: 12.9%, and DHI: 11.5%), when undeveloped areas (ISA < 10%) are discarded. However, from the second ring onwards, developed pixels are predominantly of the low intensity class (39% of developed pixels), with a decreasing proportion of medium or high intensity ISA2013 (DLI: 6.4%, DMI: 3.4%, and DHI: 1.8%).

3.3. Imperviousness in the Watershed

Impervious Surfaces in the Mondego river watershed are distributed in a heterogeneous manner, with local high density areas and vast, mostly unpopulated areas.

With an average ISA2013 of 3.8 (12.3)%, only 8.9% of the pixels in the watershed had an imperviousness fraction greater than 10% and are hence considered as developed. From these, 50.1% were located within 2 km of all motorways and trunks (figure 7), even though this buffer covers only 28.6% of the watershed. The coarser CORINE 2012 land cover product, with a spatial resolution of 100 m, points to a much lower estimate of artificial areas (3.3% of the basin, figure 1).

The aforementioned development pattern was also perceivable at the sub-basin level. Figure 7 represents the sub-basins within the Mondego river watershed and highlights the 10% and 30% critical imperviousness levels. 95.4% of the watershed is composed of sub-basins with an average ISA2013 under 10%. However, 4.1% exceeded that threshold and 0.5% were in average more than 30% impervious. The critical areas were mostly coincident with the buffer applied to the main road network. 82.9% of all sub-basins (in area) with an average ISA2013 greater than 30% were located inside the aforementioned buffer. Similarly, 71.6% of the pixels located in sub-basins where 10 % ≤ ISA2013 ≤ 30%, were found within the buffer.

Several Protected Areas, including natural parks, RAMSAR and Natura 2000 sites are found within the watershed, totaling 893 $sq.km^{-1}$ of conservation zones. ISA2013 within these areas was lower than average at 1.2 (5.7) %, reflecting legal and topographic limitations to construction in these regions. Nonetheless, in the vicinity of the parks, within a 5 km buffer, average ISA2013 reached 3.9 (12.8)%. The high imperviousness value (higher than the watershed average) underlines the pressure exerted over some of these important conservation areas.

The census data offers a different perspective on the relation between urbanization and population in the Mondego watershed. The mismatch between the natural and administrative or census boundaries hinders the integration of both components of an otherwise interconnected system.

County population (2013) was compared against ISA2013, when at least 70% of the county was located within the Mondego watershed (and includes the most significant urban center of the county), totaling 24 municipalities. The data suggested that county population and imperviousness were correlated significantly (r = 0.9, significant for a $p<0.01$). Nonetheless, it seems that the correlation is more expressive for counties with a population exceeding 20 000, but the limited number of counties with a population beyond this threshold in the study area (8) precludes further testing of this hypothesis at present.

At the parish level, population density was moderately correlated with ISA2013 (r = 0.69, significant for a $p<0.01$), suggesting other variables influence the spatial distribution of the population at the local level.

ISA2013 was also found to be correlated with the purchasing power (PP) (2011) at the county level. However, and even more interestingly, the Developed High Intensity Fraction of ISA (ISA > 80%) was strongly correlated with PP with an r of 0.96 ($p<0.01$) (Figure 8).

4. Discussion

4.1. Performance of the methodology

The strategies adopted in this study provide a straightforward and integrated workflow for the production of baseline impervious maps and later updates in a cost-effective way.

Differences in the size, topography, and land cover of the study areas limit the usefulness and reliability of accuracy comparisons with previous studies. Nonetheless, the very low MAE (ISA2013=1.6%), RMSE of 5.5%, and high Correlation Coefficient (r = 94.5%) suggest that the ISA2013 product can be considered research-grade. The overall RMSE of the ISA2013 product was lower than those reported in several works of reference, including Wu (2004) (10.1%) or Xian et al. (2010), which ranged from 8.5 to 11.15%. The MAE reported by Li et al. (2015) (5.44%) and Esch et al. (2009) (15.0% for the best case) were also higher than the values calculated for the ISA2013. The same applies to the updated prototype maps created for a subset of the watershed.

The comparison against the Copernicus Imperviousness HRL is problematic, since the validation of the product is not yet available. The visual comparison of ISA2013 and CIHRL suggests our product provides a superior representation of the typical connectivity of developed pixels. This advantage of the ISA2013 over the CIHRL is particularly important because it is increasingly relevant to assess not only the prevalence of impervious surfaces (Total impervious Area, TIA) but also the connectivity of these surfaces with the drainage system (Effective Impervious Area, EIA) (Brabec et al., 2002).

Still, the implementation of the mapping workflow was not without challenges. As in previous studies (Bauer et al., 2008; Wu, 2004), tree cover and urban heterogeneity complicated the retrieval of high quality samples and spectral signatures. The use of NDVI time series, with the extraction of the maximum and minimum value of the phenological cycle within a 2σ range, played an essential role in the reduction of uncertainty, namely in cropland and forested areas. Still, densely forested areas (namely in evergreen forests) can obscure impervious surfaces found under the tree canopy and lead to a discontinuity of developed areas that should otherwise be connected in the maps.

The large number of training and validation samples required a long process of delineating polygons and classify them as pervious/impervious. This step, essential for the production of the baseline product, is the most time-consuming step of the processing chain, requiring a trained expert capable of classifying the surfaces from high resolution imagery according to permeability (Yang et al., 2003; Xian, 2008; Jia et al., 2014). The high resolution imagery required to create the training samples that feed the regression tree model may also become a

barrier to the adoption of the methodology elsewhere. However, the increasing availability of high resolution imagery may limit this problem, at least for recent reference dates and into the future.

It is important to emphasize that an important aspect of urban imperviousness studies is the ability to compare the estimates for different dates. The methodology presented in this paper, a prototype at this stage, offers a simple, cost-effective way to estimate imperviousness in time series of Landsat data and equivalent sensors. The methodology is based on the selection of stable training samples from the pool of data used in the production of an earlier version. By updating the pool of training samples, and not segments of the baseline products as in previous studies, it is also possible to change the mapping methods, to reflect advances in science and technology. In this study, the proposed method to identify stable samples enabled the production of two updated maps (2001 and 2007) without the production of new samples.

The overall accuracy of the method and the stability of the training samples in the subset of the watershed (76% of the samples were stable during a 12-year period) support its use in this region, as well as in others with similar characteristics. It is also important to emphasize that, despite the stability of the samples, the Coimbra and Condeixa-a-Nova counties were characterized by accelerated LULC changes driven by processes of urban expansion and renewal.

Regardless of the smaller number of training samples used in each update, these products, upon validation, yielded similar accuracies to the original research-grade ISA2013, with a marginal degradation of MAE and RMSE values. The method is thus sufficiently robust for annual updates of ISA maps (at anniversary dates), which seems an adequate periodicity for both scientific studies and management applications. Furthermore, the training sample database can be updated and the model generated shortly after data acquisition, reducing the lag between the reference date and product release. Still, it is important to emphasize that the workflow depends on the availability of compatible data across the entire study period (e.g. Landsat time series).

Ultimately, the seamless updates are only possible because the method is extremely conservative as it excludes all pixels that exceed a very low change magnitude threshold. This is a possible shortcoming, in the sense that it may eliminate a significant number of training samples from land cover classes with naturally high intra and inter-annual variability (e.g. croplands) or disturbed areas with no changes to the anthropogenic imperviousness (e.g. burn scars). On the other hand, it is capable of detecting even small changes to the LULC, as highlighted by the performance metrics. The use of multiple change thresholds (as in Xian et al., 2009) could offer a viable alternative to the presented method, but only when LULC data

are available at a common resolution, which is not the case for the study area. The development of additional products, using similar methodologies and spatial resolution, but focusing on other land cover metrics is currently underway and will enable the development of new and improved change detection techniques.

4.2. ISA in the watershed and future work

The Mondego river watershed is characterized by low-to-moderate anthropogenic imperviousness (average ISA of 3.8%), which is distributed asymmetrically across the study area.

Such heterogeneity ensures the existence of vast and mostly undeveloped areas, some of which legally protected but measurably encroached by urbanized regions. ISA values above the watershed average were found in the fringes of the parks and are a cause for concern, as some of these areas are effectively surrounded by nearly-continuous impervious barriers. The ISA product, together with other similar data layers and tools could support the development and enforcement of low-imperviousness corridors in selected, ecologically-significant streams and limit the number of sub-basins exceeding the 10% critical ISA threshold. This is particularly important in a watershed where a significant amount of sub-basins had poor biological or chemical quality (Ferreira et al., 2004) and important areas and water reservoirs are affected by recurrent eutrophication events (Marques et al., 2003, Oliveira and Monteiro, 1992).

Considering the anthropogenic nature of ISA, it is important to combine environmental metrics with the knowledge provided by social, behavioral, and economic sciences (Grimm et al., 2000). This connection is reflected in the increasing demand, by government and managers, of tools capable of quantifying the impact of urban change (Carlson, 2003; Hebble et al., 2001; Young et al., 2013). The resulting information is often on the basis of legislation capable of mitigating the impact of artificialization and internalize the externalities through the taxation of impervious surfaces to promote the protection and restoration of basins (Feitelson et al., 2004, Maryland House Bill 987). Such schemes may leverage an ISA layer such as that suggested in this paper, to implement cost-effective zonal taxation schemes at the sub-basin level, incorporating both objective environmental metrics as well as economic variables at the community level.

In Portugal, several legally binding plans at national, regional and local level exist. These generally set admissible land uses and maximum imperviousness levels and can play a relevant role in the management of urban expansion, as reported for the study area (Monteiro et al., 2015, Tavares et al., 2012). Nonetheless, accurate data on impervious surfaces were not

publicly available at 30 m (or better) resolution and thus a careful and critical scrutiny of urbanization trends was difficult.

Because ISA is the enduring imprint of human activities in a territory, understanding the population dynamics of the watershed is relevant to comprehend the spatial patterns of that variable. Both variables in the watershed mirror the national reality, in which 50% of the population is concentrated in 11% of the counties, predominantly located in coastal regions (INE, 2012). Despite the segmentation of the watershed into 47 counties, approximately 33.2% of the watershed's population resided in Coimbra and Viseu, the largest urban centers of the region, which also account for 20.9% of the total developed pixels in the ISA2013 product.

This dichotomy between developed coastal areas and the depressed hinterland is meaningful from both an ecological and social standpoint. The ISA profiles of the most and least populous counties (Figure 9) translate the differences in the urban structure, access to services, and infrastructure, which, in the end, influence the livelihood of the communities.

The connection between income structure and land cover processes has been reported to influence – and be influenced by - urban development and imperviousness levels in other regions in similar ways (Hietel et al., 2007, Lu et al., 2006). Gross Domestic Product per capita for instance, was demonstrated to be correlated with imperviousness in European cities (Schwarz, 2010; Kraas, 2007). A similar trend seems to be at play in the Mondego Watershed as suggested by the correlation between High Intensity Development (ISA > 80%) and county Purchasing Power (Figure 8). This was an expected correlation considering the nature of constructions associated to this development class (e.g. highways, dense urban fabric, retail areas, and public venues). The correlation between purchasing power and urban development is in line with findings by Glaeser et al. (2001), which point to quality of life (including access to cultural and commercial spaces) as well as wages as determinant factors capable of ensuring the success of cities.

Nonetheless, shrinkage phenomena affecting cities across Europe (EEA, 2006) may lead to gradual changes in these relations. In fact, the simultaneous acceleration of rural flight phenomena and abandonment of traditional centers, foretell an "everwidening disparity" between different urbanized areas (Glaeser et al., 2001). These changes may also erode the observed correlation between county population and ISA (r = 0.90). At the parish level, the population density is already only moderately correlated with ISA (r = 0.69) in the Mondego watershed.

The widespread regional population decline is thus only locally reversed as a result of internal rearrangements of the population within the sphere of influence of the main urban centers in a process that was still active after 2001. Relocation to the fringes of the city is often driven by

economic and social reasons and perceptions on issues such as the cost of housing or general standards of living (EEA, 2006, McGranahan et al., 2010). However, it is also important to emphasize that these figures already reflect some of the economic, social and population changes caused by the recent sovereign debt crisis, clearly outside the scope of this paper.

A consequence of the complex set of aforementioned mutations, which were recognized in the imperviousness maps produced, is that the new development *loci* become scattered along transportation axes, enabling the daily commute of the new residents towards the traditional economic centers (INE, 2015). It is worth noticing that nearly 73% of pixels flagged for change between 2001 and 2013 in the subsetted study area were located within 15 minutes of Coimbra's city center. Furthermore, the overwhelming amount of developed pixels in the entire watershed (50.1%) located at a short distance (up to 2 km) of the (asymmetrical) major road network, exemplifies the importance of transportation routes in the construction of ISA spatial patterns in the study area.

Nonetheless, population decline is not always necessarily correlated with ISA dynamics at the temporal scale of the study. Coimbra, despite a significant and persistent population shrinkage between 2001 and 2013 (-10202 inhabitants), witnessed an increase in ISA in the same period. In fact, at the parish level, urban development in Coimbra and Condeixa-a-Nova does not seem to be driven by population changes. Parishes with declining population appear to be as likely to sustain urban growth (including the construction of residential areas) as those with an increasing population. The apparent contradiction can be explained by two different drivers. On the one hand ISA increases are in line with the findings of Panagopoulos et al. (2013), which suggest that counties with persistent population decline tend to increase the public expenditure per inhabitant, including in infra-structure. In effect, two thirds of the pixels flagged as changed in the studied counties, comprise either infrastructure or service developments. On the other hand, new residential areas may continue to be constructed as a result of an increase in the per capita use of land as purchasing power improves (Haase et al., 2013), as reported for several emerging counties of the area of study (INE, 2015). Simultaneously, vacant residential buildings are rarely demolished, especially when located in the city center, preventing a reduction of ISA values at the county level.

Interestingly, the starting imperviousness fraction seems to be an important factor driving the type of subsequent development, which could be motivated by land availability and socio-economic factors (Hietel et al., 2007). ISA2001 was lower (5.4%) in pixels where infrastructure was subsequently developed, compared to others where service (9.1%) or residential areas (11.5%) were constructed. The final ISA (2013) of residential areas was, on the other hand, lower than that of Service or Infrastructure pixels. The natural, cultural, and economic drivers

of urban development are clearly at play (Antrop, 2005) and could be monitored and characterized with the ISA product. Determining the location of the development is based on the evaluation of a series of requirements, which differ according to the type of construction. The proximity to existing roads, other dwellings or conversely, the availability of large swaths of undeveloped land influence the initial ISA value per category, which is also characterized by typical final imperviousness levels.

This combination of requirements and characteristics seem to support future attempts to model the sprawl of urban areas in the study area, in response to different scenarios.

The data accuracy and scale as well as the high frequency nature of the methodology thus suggest that specific events, redevelopment projects, and changes to the economic profile of a region can be traced to ISA values while, or shortly after, they take place (e.g. Euro 2004 Soccer Championship or the 'Polis' redevelopment program in Coimbra). Estimating and mapping ISA is thus significant to a broad range of issues in environmental science as well as to characterize the human dimension of this territory. The rapidly evolving landscapes and the spatial scale at which these processes take place call for high temporal resolution of imperviousness products (Weng, 2012) as those enabled by the methodology presented in this paper.

The mapping of the Mondego river Watershed ISA builds on these trends, offering a new, reproducible, and affordable tool to scientists, managers, and decision makers, interested in the local and regional impact of urban development over environmental and socio-economic variables alike.

The development of mapping initiatives focused on watersheds yields information on critical variables for meaningful territories and not just within the often arbitrary and mutable administrative divisions. The geographic consistency of the cartographic products is thus a fundamental aspect of this work, as proposed by previous studies (e.g. Arnold and Gibbons, 1996).

To make the data available to the scientific community and stakeholders, an online platform is currently in development. The project, named 'SmartBasins', is focused on the development of an online cross-platform service where the ISA layers, as well as other geospatial information will be made available in their native and derived formats, with no costs to the users. ISA2013 is the precursor layer to this platform, which aims to address the data needs of a wide range of scientific and operational users. The data thus provided (30-m spatial resolution, with updates every 2 years along with similar retrospective analysis) is in line with identified needs and now a validated protocol enables its operational production (Carlson, 2003; EEA, 2006; Teixeira et al., 2014). A later paper will describe the platform and the different data layers it will provide

at the time of launch, supplementing ISA with other critical remotely sensed and modeled metrics.

Additional research will focus on the development of other LULC layers, at a similar or improved spatial resolution and with analogous potential for regular updates. Introducing details into structure and function, analyze connectivity and expand the study of water quality indicators is planned, along with a reinforcement of the classification of agricultural and forested land.

5. Conclusion

At the dawn of the 'age of cities', the momentous challenges ahead call for innovative solutions that combine science, technology and an understating of the social dynamics.

The optimization of urban development policies must be based on sound metrics that spur social and economical development while addressing the needs and hopes of the populations (Marinoni et al., 2013; Duque et al., 2015; Krellenberg et al., 2014).

The ISA2013 was successfully developed and validated to address identified needs not only at a regional level, but Europe-wide. The high accuracy levels of the baseline 2013 product combined with the seamless updating strategy opens new possibilities for the continuous monitoring of an important environmental indicator.

Although the focus of this paper is on the development of ISA mapping methodologies it became clear that the impact of this variable over environmental and social variables deserves further research.

The societal challenges created by a rapidly evolving population structure and changing cultural values are permeating into imperviousness dynamics.

The population is expanding into counties where, traditionally, both population and ISA levels were lower, increasing the pressure over the infrastructure and forcing adjustments, often *ad hoc*, to the management of the territory. The decentralization of cities will also lead to a multiplication of stressors and the expansion of nonpoint-source pollution as the natural space becomes increasingly fragmented and the soils sealed. The ISA data suggests counties in the study area have thus far failed to adapt to the new reality of declining population, insisting in the expansion of the urbanized surface in an attempt to reverse the shrinkage phenomenon (as also reported by Panagopoulos et al., 2013). The increasingly heavy infrastructure will necessarily have to be maintained by a smaller, older population under a more restrictive budgetary environment.

The development of land use and land cover products that provide reliable and objective metrics was never as relevant as now. Map producers must also consider the appropriate methods and formats for data distribution to guarantee that the data is used by the stakeholders and environmental scientists with no impediments. This has been a central concern of the authors, whom are now developing the SmartBasins platform for ISA (and other) data distribution, integration and modeling. The new opportunities created by the launch of new satellite missions, including ESA's Sentinel constellation, creates interesting perspectives for the future of LULC and the development of sub-pixel, continuous metrics for a better understanding and management of the territory.

FIGURES

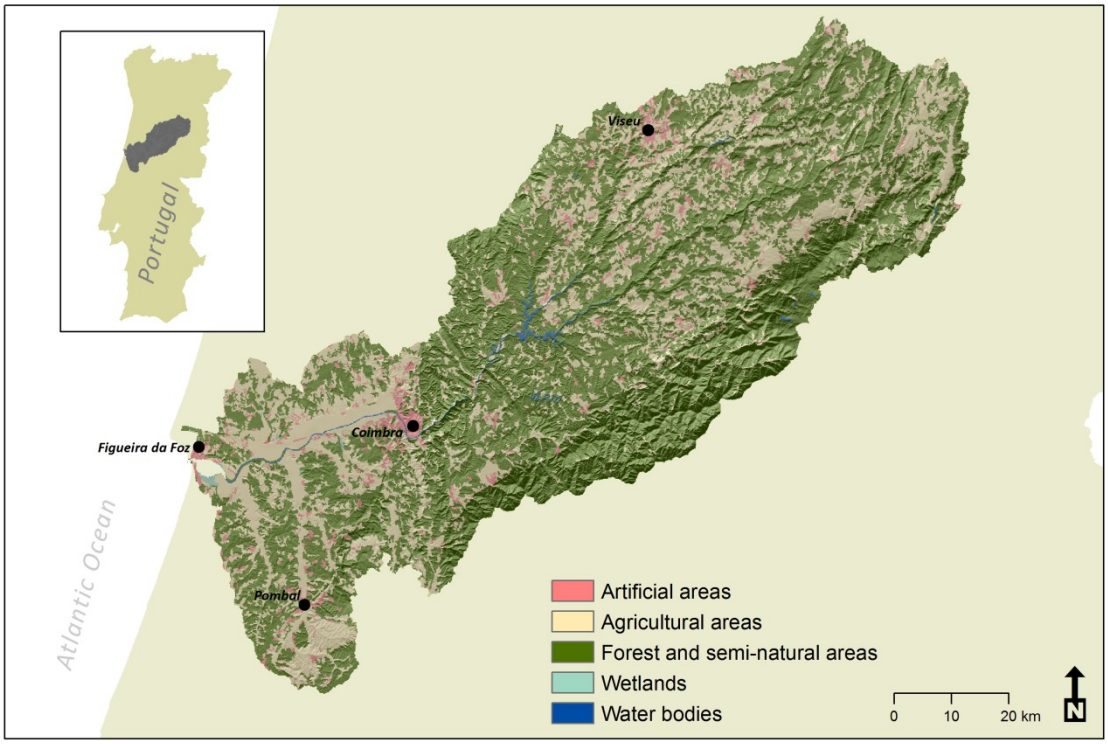


Figure 1. Map of the Mondego river Watershed, depicting a simplified land cover map (CORINE 2012), location of major urban centers, and terrain.

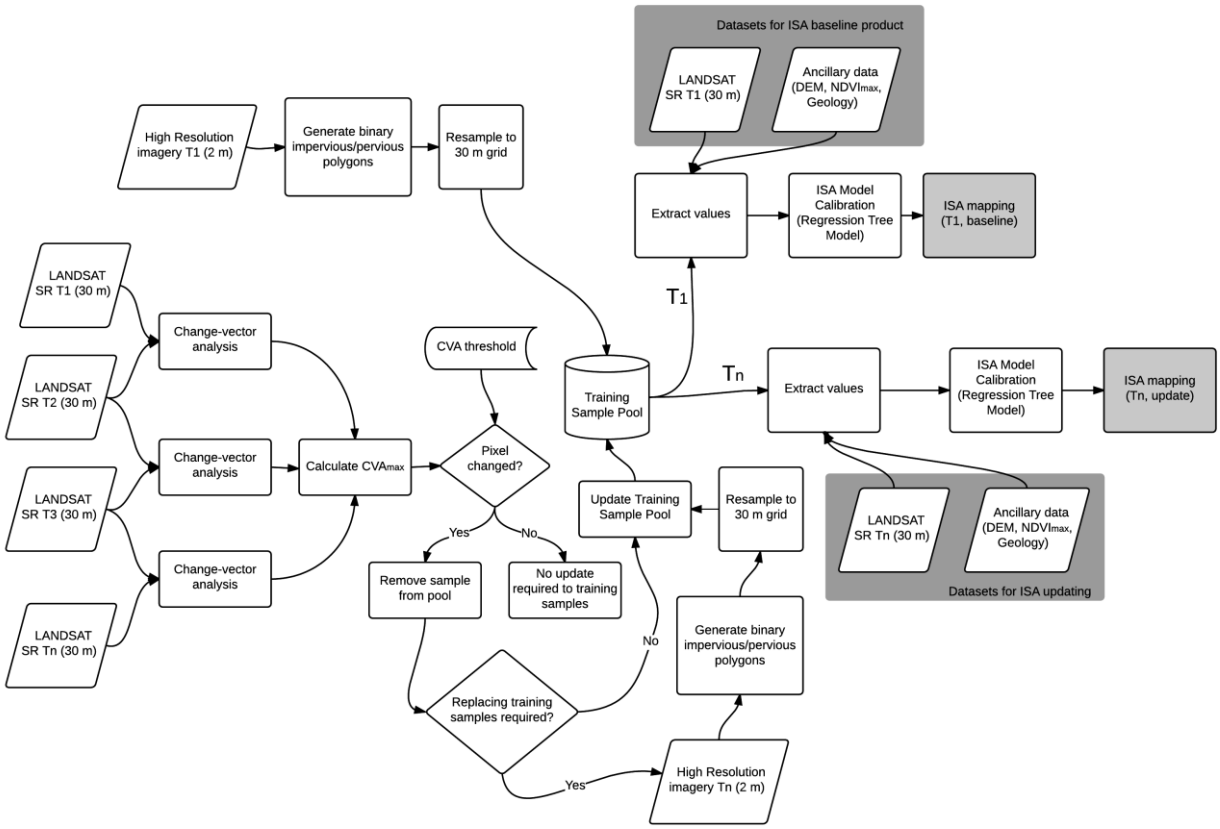


Figure 2. Generalized workflow of the methodology employed to map Impervious Surface Area (ISA) using Landsat data, and subsequent updates through Change-Vector Analysis (CVA). Baseline product (T1) matches ISA2013 in this paper, while the update (at an earlier or later date) (Tn), corresponds to the ISA2001 and ISA2007 tests.

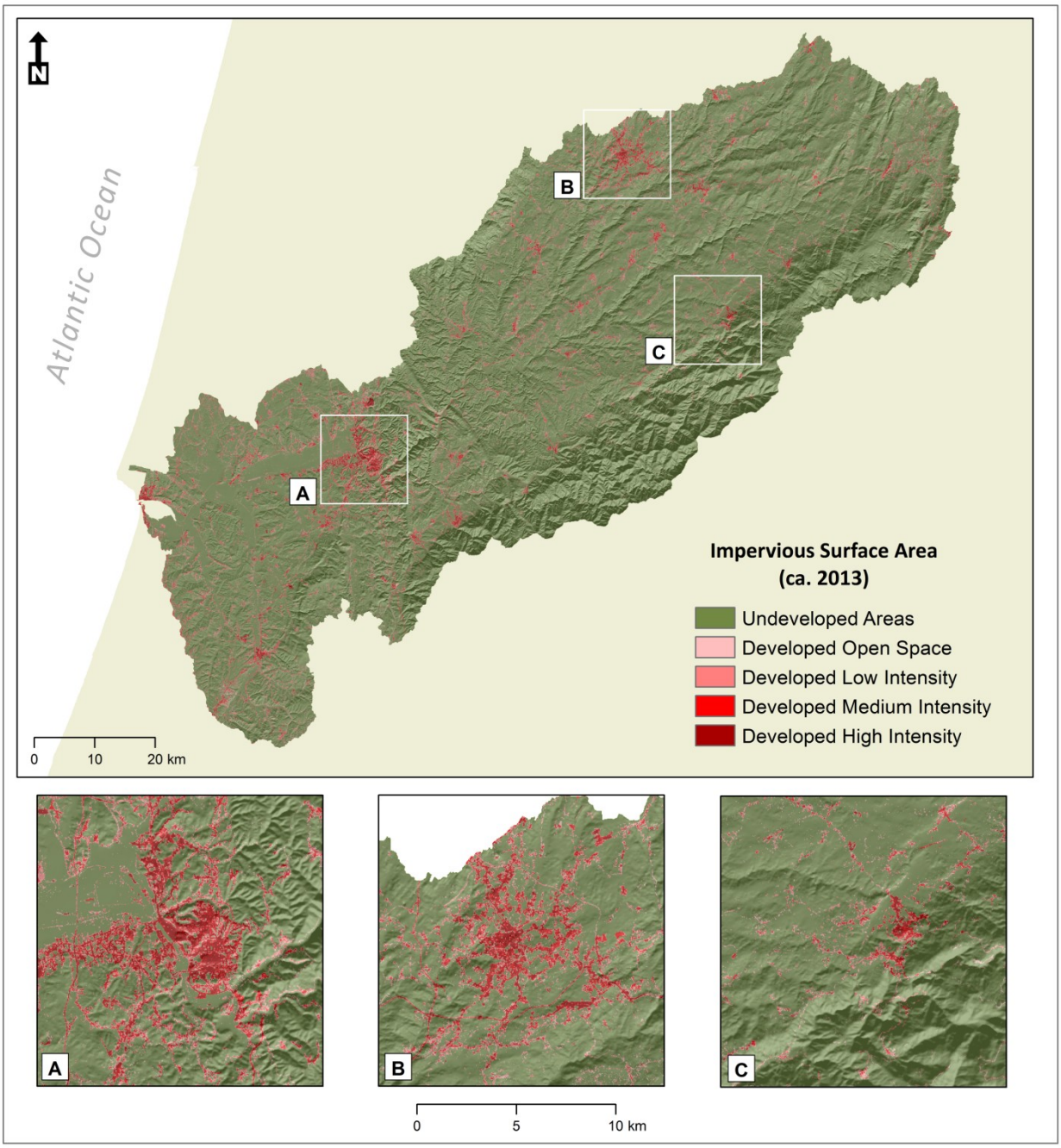


Figure 3. The Mondego river Watershed Impervious Surface Area for *circa* 2013 (ISA2013). Three representative areas are displayed in greater detail including Coimbra (A), Viseu (B), and Seia (C). The ISA classes are described in the text.

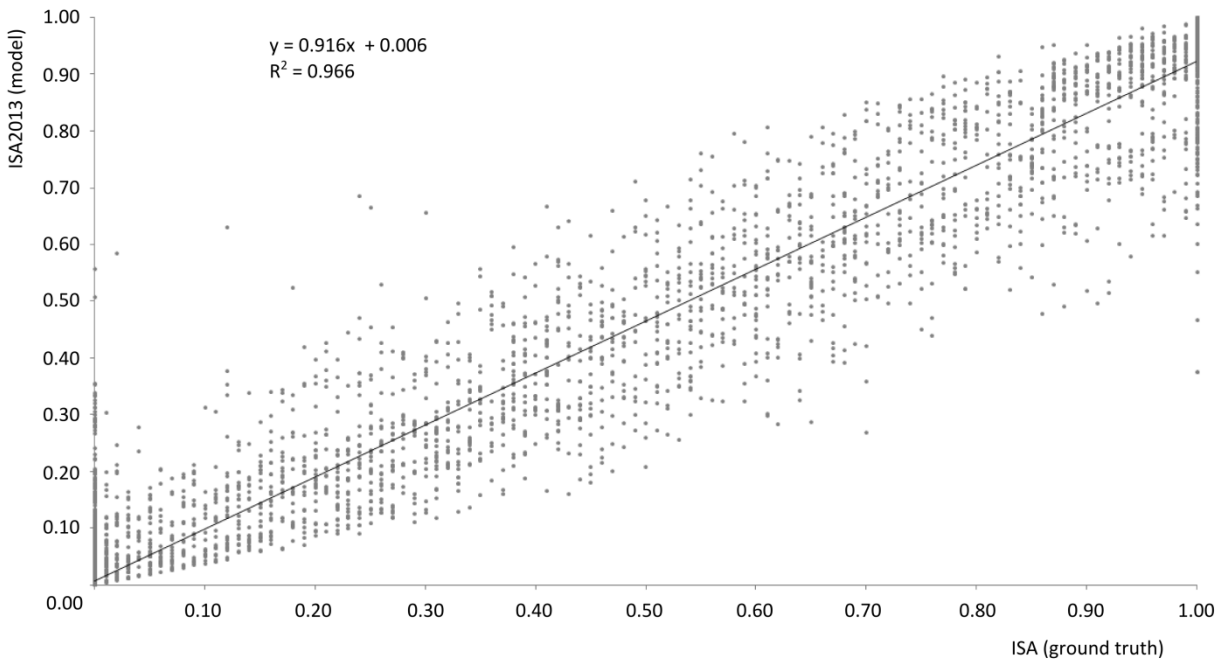


Figure 4. Correlation of independent ground truth Impervious Surface Area samples and the ISA2013 model.

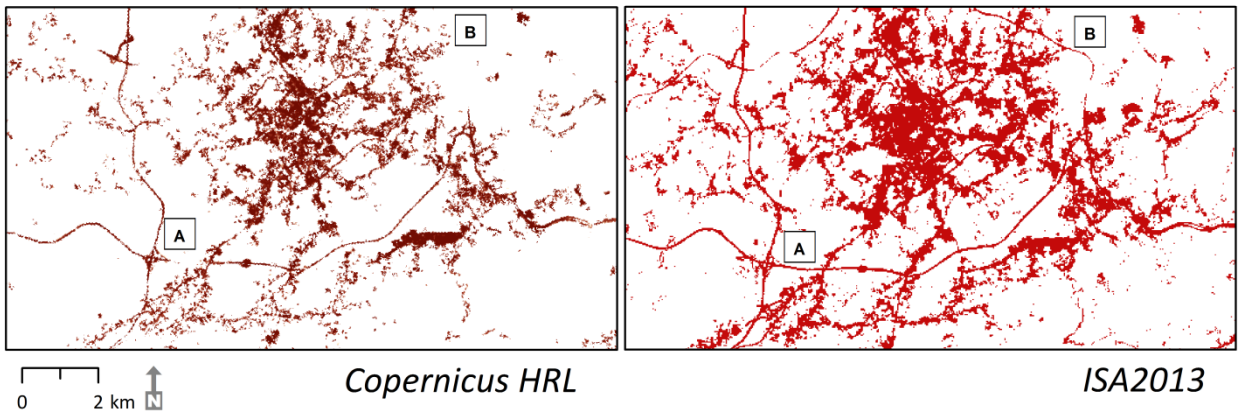


Figure 5. Comparison of the Copernicus Imperviousness High Resolution Layer and the ISA2013 (depicting pixels with ISA>10%). A) and B) highlight regions where the ISA is represented erroneously (patchiness) in the Copernicus product in contrast with the more accurate ISA2013 equivalent.

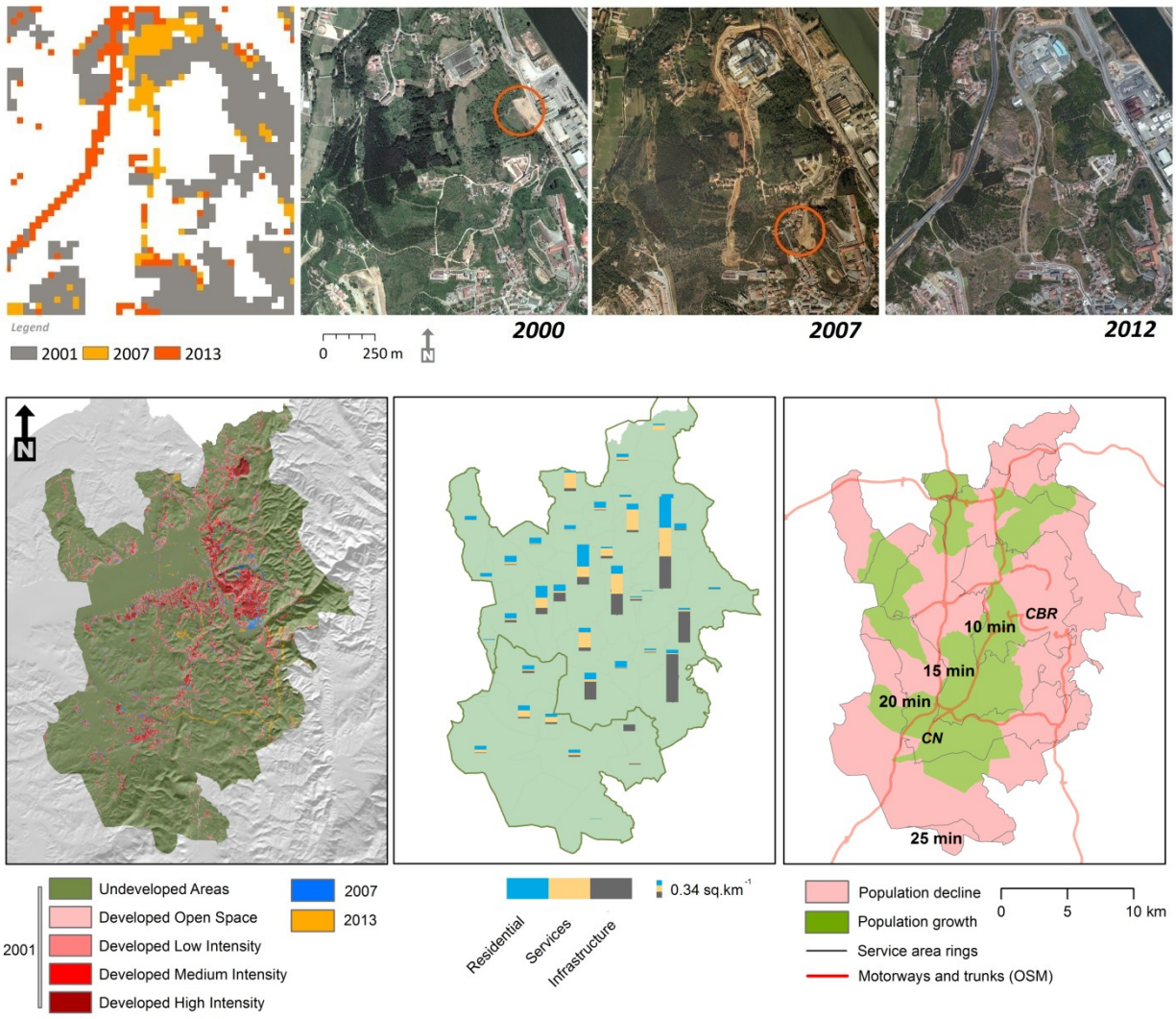


Figure 6. Top: The models can accurately capture urban change as depicted in the subset. The map shows the evolution of developed land (ISA > 10%) from 2001 to 2013. Transient change (highlighted by the red circles in the aerial images) was accurately recognized as such by the methodology and not classified as impervious surfaces. Bottom: (Left) Imperviousness map showing the development classes (2001) and subsequent developments (2007 and 2013) in the Coimbra and Condeixa-a-Nova counties. (Center) Imperviousness change (2001-2013) characterization into Residential, Services, and Infrastructure categories, per parish. (Right) Population change per parish and service area rings (with traveling times). Motorways and trunks are displayed as well. CN: Condeixa-a-Nova, CBR: Coimbra.

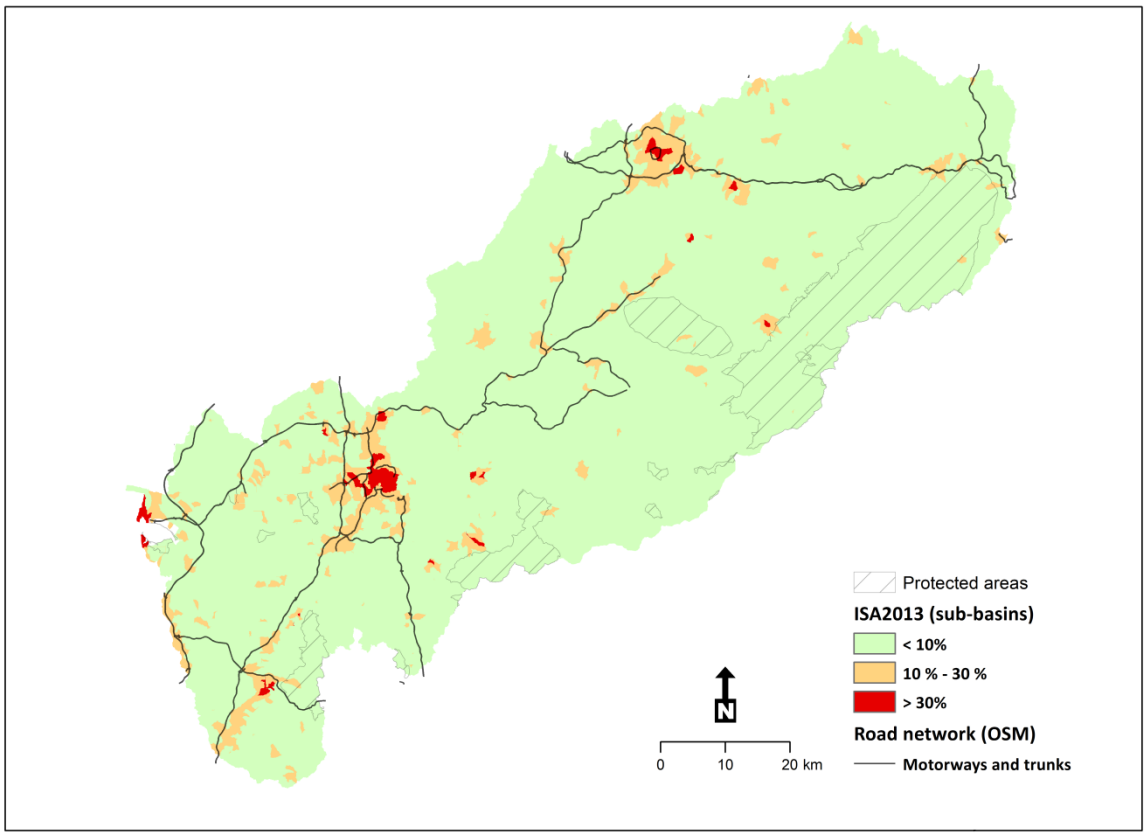


Figure 7. Catchments within the Mondego river Watershed depicted in function of ISA thresholds (see text for explanation). Protected areas and main road network displayed for reference purposes.

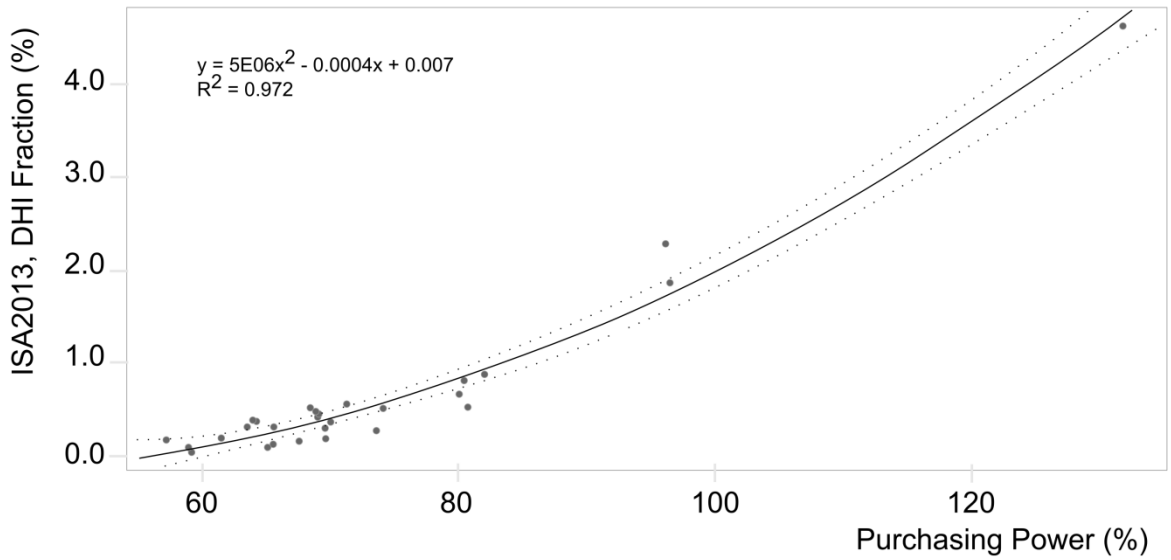


Figure 8. Purchasing Power of counties as a function of the average Developed High Intensity (DHI) fraction (ISA2013 > 80%) and regression line (solid). Dotted line: 95% Confidence interval.

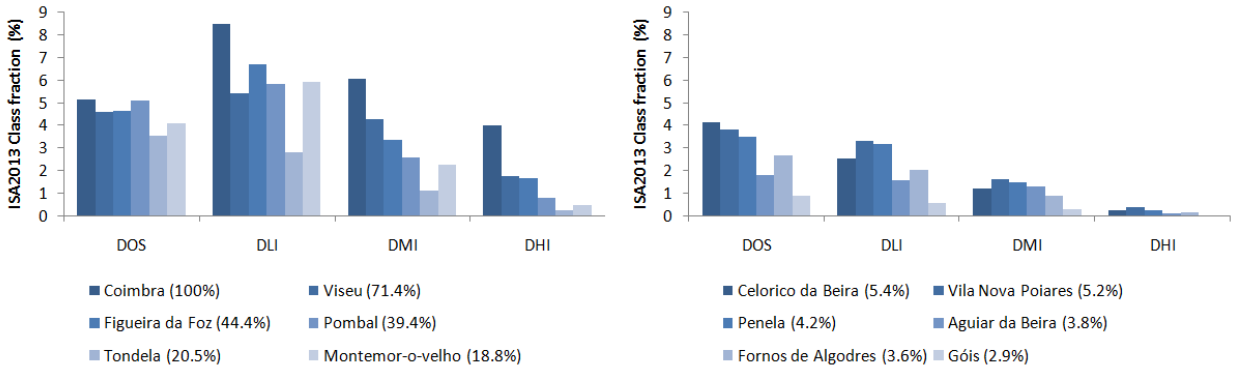


Figure 9. ISA2013 fraction per development class for the most (left) and less (right) populous counties of the watershed. Population as a percentage of the number of inhabitants of the most populous county (Coimbra) is included in the legend. DOS: Developed Open Space, DLI: Developed Low Intensity, DMI: Developed Medium Intensity, DHI: Developed High Intensity (see text for details).

Tables

Table 1. Impervious Surface Area fraction (%) statistics (mean and standard deviation) for the two test counties (Coimbra and Condeixa-a-Nova) where the ISA updating strategy was evaluated.

	2001		2007		2013	
County	Mean	SD	Mean	SD	Mean	SD
Coimbra	8.4	19.4	9.1	20.7	11.8	23.5
Condeixa-a-Nova	2.3	9.6	3.1	11.0	4.0	13.1

5.1.2. Introducing SAR data to map ISA – a preliminary assessment

In this section, we briefly describe the development of a new ISA layer for a Portuguese watershed (Mondego river) based on the fusion of Landsat-8 OLI data with Sentinel-1A multitemporal information. The approach supplements the efforts presented in the previous section, and was first presented at the European Space Agency's Living Planet Symposium 2016, in Prague, Czech Republic.

In this approach, a Regression Tree Model was built using a combination of Optical (Landsat-8 OLI) and Synthetic Aperture RADAR (SAR) (Sentinel-1A) data. The ground truth database used for the development of the layer was the same used for the optical-only approach.

The need to develop an ISA layer based in a combination of spectral and textural information drove the development of the prototype layer. It also served as a first step towards the development of SAR-only products, enabling all-weather mapping capabilities, essential for multiple critical applications, which often require near real time data on the status of infrastructures.

In the preliminary study, SAR and optical data were fused, in order to evaluate improvements in the output ISA layer. Such improvements, if verified, could be leveraged in the development of subsequent land cover layers (e.g. forest cover, agricultural land, etc…)

One Landsat-8 OLI image was fused with four Sentinel-1 (VV/VH polarization) images, as described in table 6.1.2.1.

Table 5.1.2.1. Earth Observation datasets used in the study for the preliminary assessment of a data fusion approach for the development of a new ISA layer.

Satellite	Sensor/Polarization	Acquisition Date
Landsat-8	OLI	2015/06/26
Sentinel-1A	VV/VH	2015/05/19
Sentinel-1A	VV/VH	2015/05/31
Sentinel-1A	VV/VH	2015/06/12
Sentinel-1A	VV/VH	2015/06/24

Landsat data were acquired as a Surface Reflectance product (Masek et al., 2013) from the Earth Resources Observation and Science (EROS) Center of the United States Geological Survey (USGS). A number of derivative products were generated from the dataset, including Normalized Difference Vegetation Index (NDVI), Normalized Burn Ratio (NBR), Normalized Difference Moisture Index (NDMI), and the ratios $B6.B7^{-1}$, $B5.B6^{-1}$ and $B4.B2^{-1}$.

The S1A data were acquired from the European Space Agency's Scientific Data Hub and processed using the open software SNAP (Sentinel-1 Toolbox).

A processing chain (figure 5.1.2.1) was created for the SAR data, which included the sequential processing of data including standard steps. Amongst these were calibration, terrain correction, and coregisteration. This step was fundamental to guarantee the reduction of speckle noise through multitemporal filtering (figure 5.1.2.2).

Landsat data were used as the spatial reference for the development of the ISA layer, which would thus have a 30-m spatial resolution, with all layers collocated to the original pixels of the optical imagery. As such, the Sentinel-1 calibrated products were rescaled and collocated to the Landsat grid, generating a product with a common spatial resolution and pixel distribution.

All products were exported into ArcGIS 10.2 as GeoTIFF files, and data fusion steps initiated, with the application of Principal Component Analysis (PCA) to the Landsat multispectral and Sentinel-1A bands.

Upon the application of PCA, a lower number of input variables can be used as inputs to feed the Regression Tree Model, with significant advantages from a computational and operational standpoint (Ban et al., 2014).

As in the original layer, which was previously described, non-EO datasets were included in the process, namely slope (as percentage), slope aspect (degrees), Valley Depth, and Dominant Landforms. Furthermore a simplified geological map of the watershed was included as well.

The *in situ* dataset is described in detail in the previous section, including the methodology to update the training samples for recent years.

Once again, the M5P algorithm was used and applied over Weka, with the results being exported and mapped into GeoTIFF at ArcGIS 10.2, using a custom-built python program (arcpy), developed by the author.

The Mean Absolute Error (MAE) and Pearson correlation coefficient were also calculated in order to compare the output with that of the optical layer.

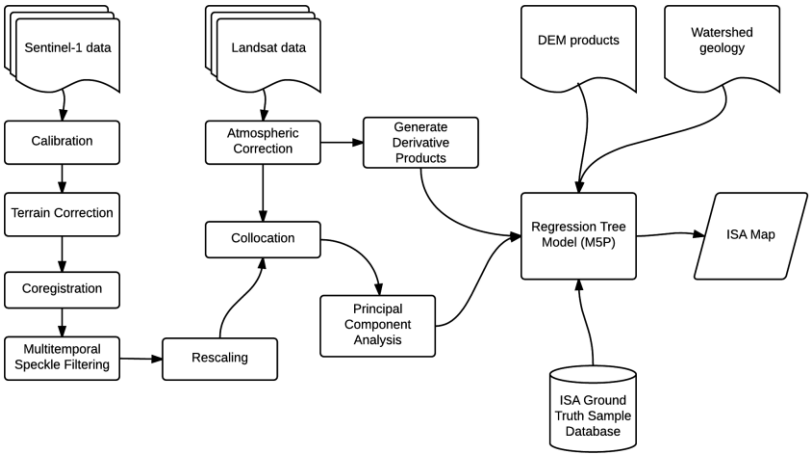


Figure 5.1.2.1. The workflow adopted to generate the data fused ISA product.

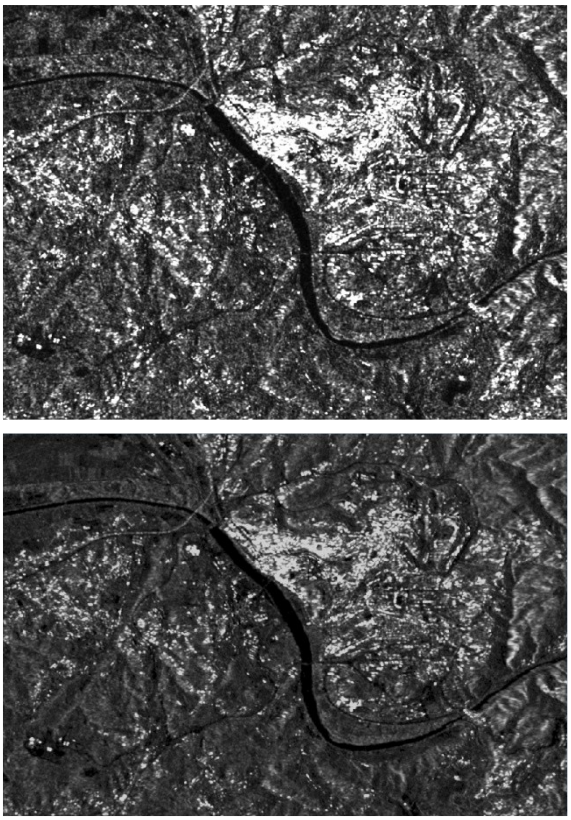


Figure 5.1.2.2. Speckle noise reduction of the SAR imagery through multitemporal filtering. Top: original unfiltered image of Coimbra, Portugal. Bottom: Spekcle fitered image. The reduction of speckle noise is evident.

The PCA produced a simplified set of raster files, from which the data were extracted at the calibration sites, to generate the RTM.

The first principal component (PC) explained 76.26% of the variance in the dataset (99.37% in the first three PC), which was driven essentially by the optical data, as concluded upon visual inspection of the output.

14 layers were thus used to generate the RTM, including the first three PCA outputs and the ancillary layers created from Landsat, topographical, and geological data.

These layers enabled the production of an ISA layer for the entire Mondego river watershed (figure 5.1.2.3). The model was described by 256 rules (and linear regressions), and characterized by a Pearson Correlation Coefficient of 96.2% and a Mean Absolute Error of 2.9% (figure 5.1.2.4).

For the reference date (2015), 9.0 % of pixels in the Mondego river watershed were developed, with the target area having an average ISA of 3.9 (11.6) %.

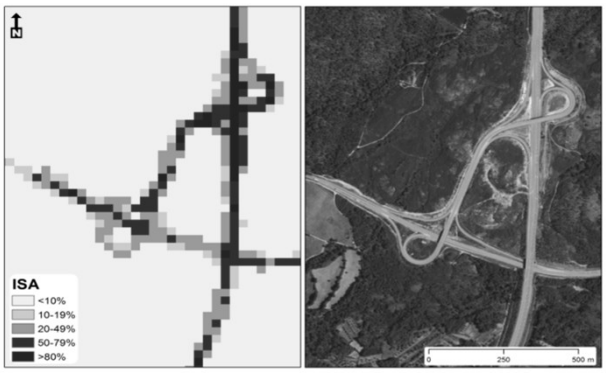


Figure 5.1.2.3. Subset of the data fusion ISA layer for the Mondego watershed, showing a comparison of the new layer and a true color high resolution satellite image. Notice the accurate representation of the developed land.

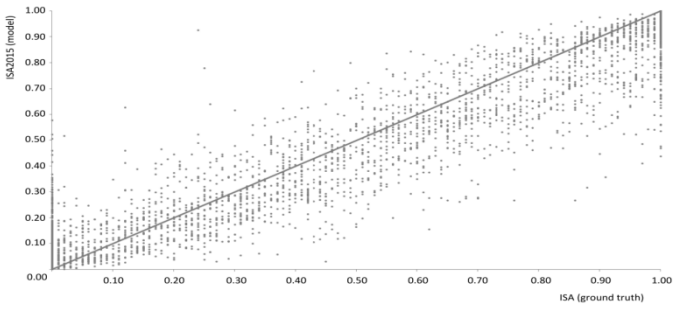


Figure 5.1.2.4. Plot of the correlation between validation (*in situ*) data and the data fusion ISA layer. (Grey line: x=y).

The SAR/Optical preliminary data fusion approach thus enabled the development of a more accurate ISA layer, when compared with the optical-only product previously described. We

must emphasize that this analysis was focused on a preliminary product, developed as a supplementary project spurring from the main ISA layer development effort and dedicated to the evaluation of the new Sentinel-1 datasets made available by ESA.

The output generated by the new model was within the set of requirements set for the original optical layer. In fact few differences were identified in both sets, which may result of the dominance of optical data in the PCA step. The value of *r* was higher in the new version (96.2%), when compared with the original product (94.5%). However the MAE was 1.3 points higher, suggesting that the data fusion approach introduces challenges to the correct estimation of ISA.

It was determined, for instance, that in agricultural fields, the original product outperforms the data fusion test layer. Still, SAR data provided important textural information (Zhu et al., 2012), which was preserved in the PCA and used by the RTM, improving model performance locally.

Several advantages of the use of the data fusion approach were identified, including the lower computational cost of development at the RTM stage (less rules and regressions), and the simplicity of selection of a time series, considering the fact SAR data is mostly unaffected by cloud cover. On the other hand, The need to apply multitemporal speckle filtering calls for the acquisition of several SAR acquisitions from a period of time consistent with the objective of the application. In regions developing rapidly, the acquisitions cannot be acquired over a long period, or they will represent a different reality. In this test, the acquisition of imagery over a 2-month period was deemed acceptable, especially considering the large area of study.

Both datasets (Landsat and Sentinel) are available for free under efficient open data policies, which are paramount to guarantee that applications such as those we described become operational products in the future.

In fact, the replacement of Landsat-8 with Sentinel-2, or the development of SAR-only datasets shall provide users with new options for data analysis. High frequency updates of land cover maps of the Mondego river watershed, and other similar systems, are no longer unattainable considering the steps taken in this study to develop a set of comprehensive and complementary solutions. The development of an efficient *in situ* training sample database, the ability to update said set, and the ability to leverage the information to create accurate land cover layers were successfully completed.

Next, we shall explore succinctly how the ISA layer can be compared with water quality variables. In the future, additional layers, already in development by the author, will provide a holistic, high resolution view of the watershed, unparalleled to what is currently available. Yet, and considering the development life cycle, analysis is still restricted to imperviousness, which is, nonetheless, an important driver of water quality in urbanized watershed worldwide.

5.1.3. Water quality in the watershed-coastal continuum

The ecological status of streams in watersheds is a relevant aspect of the daily lives of communities worldwide, which often have to struggle to maintain an adequate supply of potable water (UN, 2012). In the context of our study, we were also interested in evaluating the potential impact of land cover dynamics and manmade structures over the water quality and color in a watershed-coastal system (Lima et al., 2012, Sherman et al., 1990). Considering the central location within the study area, the moderate size of the basin, and the wealth of information available, the Mondego river watershed was selected as the main target area.

As addressed in a previous section, the Mondego river generates important and persistent buoyant river plumes, which can be detected and monitored using Earth Observation Data. These plumes can be detected because the water color and thermal properties are different from the surrounding medium. These differential characteristics also impede – to a certain extent – the mixture of waters, creating a sub-system that is fed predominantly by the river input generated by the Mondego.

A pressing question is whether water feeding these features can potentially affect the optical environment, biological communities, and economic activities in the region?

In recent decades, the increased nitrogen load (and that of other pollutants) has increased steadily in coastal waters worldwide (Howarth, 2008, Rabalais et al., 2009). This fact has lead to the increase in the occurrence and duration of Harmful Algal Blooms (HAB) and a general disruption of the ecosystems due to a combination of factors including anoxia, changes to habitats and the food-web (Howarth, 2008).

The expansion of eutrophicated coastal waters is a serious concern, with communities worldwide increasingly aware of the required corrective measures that must be rapidly implemented to contain the problem (Anderson et al., 2002).

Despite the urgency of comprehensive global studies, regional research and models are required to adequately respond to local needs while safeguarding the social and economic stability of the territories.

Land cover trends (and regulations) are thus intimately connected to the protection of watersheds and of the coastal systems to where they drain.

Considering the time and budget constraints inherent to a PhD project, our activities were limited to the development of an Impervious Surface Area product. The development of this complex product not only provided sound basis for the study of urbanization in the watershed, but allowed the development of the required methodologies and technologies to further advance land cover mapping in central Portugal and elsewhere.

As such, and considering that at present, ISA is the primary variable available at an adequate spatial scale and with a robust validation for the region of interest, we shall attempt to understand the correlation of the variable with water quality. Albeit exploratory in its nature, it shall provide the basis for subsequent analysis already employing the full suite of land cover products to be made available at SmartBasins, described latter in this work.

The previous section already highlighted several important effects of imperviousness to water quality in watersheds. In this section we shall summarize the current state of the water quality across the Mondego river watershed and highlight some of the connections between the measured variables and estimated imperviousness levels at the sub-basin scale. Hopefully, this section will offer novel information from which further studies may be conducted, stimulating a debate on the potential of integrated watershed monitoring systems, interlinked to coastal observation systems.

To this end, water quality data were acquired from the Portuguese Hydrologic Information System (SNIRH) (Table 5.1.3.1) for the period between January 1, 2011 and December 31, 2013. The location if the stations are depicted in figure 5.1.3.1. For those measurements below the limit of detection (LOD), the value was replaced by half the LOD whenever the total number of replacements does not exceed 15% (EPA, 2000). For each site (n=19), the associated sub-basin was delineated for the calculation of the average ISA.

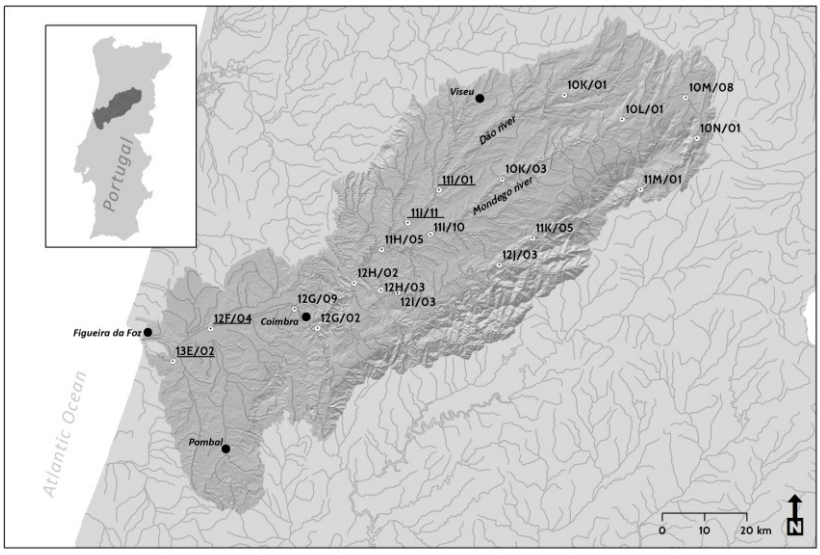


Figure 5.1.3.1. Map of the Mondego river Watershed, depicting the topography, major streams, and the location of major urban centers. The water quality sampling sites (SNIRH) are depicted and labeled (see Table 5.1.3.1). Stations belonging to cluster A are underlined.

Table 5.1.3.1. Water Quality parameters and results for the period between January 1, 2001 and December 31, 2013. The data were downloaded from SNIRH and processed as described in the text. *: Sites in the Mondego sub-basin, excluding the Dão river watershed and sites downstream of Coimbra.†: Sites in Cluster A (X-means clustering).

Gauge Name	Site ID	Mean ISA2013	Total Suspended Solids	pH	Total Nitrate	Conductivity 20ºC	Fecal Coliforms	Chloride	Total Nitrogen
Pai Diz*	11M/01	0.3	1.7(0.6)	6.5(0.2)	1.0(0.0)	14.4(5.1)	70.9(72.8)		0.5(0.3)
Aldeia Viçosa*	10N/01	1.1	1.5(0.0)	6.8(0.1)	2.2(0.8)	39.1(8.9)	107.5(108.1)		0.8(0.3)
Ratoeira*	10M/08	1.7	6.3(7.3)	6.8(0.2)	1.5(0.7)	56.8(11.2)	104.7(82.9)	5.9(1.9)	0.9(0.1)
Ponte Juncais*	10L/01	2.3	3.6(2.8)	7.2(0.5)	1.2(0.5)	55.8(11.8)	96.3(151.1)	6.3(2.3)	0.8(0.1)
Nelas*	10K/03	2.4	4.9(3.7)	7.1(0.4)	1.6(1.0)	73.3(15.5)	160.4(185.3)		1.0(0.4)
Aguieira Dam Pinheiro Ázere*	11I/10	2.7	3.8(2.6)	7.9(1.0)	1.7(1.0)	80.6(6.0)	10.1(22.5)	10.5(1.2)	0.8(0.2)
Agueira Dam*	11H/05	3.3	4.6(2.8)	7.9(1.1)	2.1(1.3)	82.9(7.7)	21.0(33.8)	10.8(1.7)	1.1(0.3)
Penacova*	12H/02	2.9	4.0(5.1)	7.0(0.3)	2.8(0.9)	78.4(9.8)	36.9(45.3)	10.1(1.5)	1.0(0.3)
Açude Coimbra*	12G/09	3.0	2.5(2.1)	7.2(0.2)	2.5(0.9)	83.4(10.3)	123.4(2710.6)	10.3(1.8)	1.1(0.5)
Aguieira Dam Santa Comba†	11I/11	4.3		8.1(1.2)	4.4(3.2)	115.5(19.9)	183.0(138.1)	16.0(3.4)	1.9(0.7)
Ferreirós†	11I/01	3.1	4.3(3.4)	7.0(0.2)	4.5(3.1)	111.8(24.7)	143.6(307.3)		1.6(0.5)
Ponte Santa Clara	10K/01	1.4	4.8(6.0)	6.8(0.1)	1.4(0.9)	71.6(9.8)	36.3(49.4)		0.6(0.2)
Ponte Formoselha†	12F/04	3.5	5.7(3.2)	7.5(0.2)	5.3(1.7)		148.3(145.3)		2.5(0.9)
Ponte Cabouco	12G/02	1.7	5.1(6.2)	7.3(0.2)	2.4(1.3)	92.9(17.2)	237.1(131.3)		1.0(0.1)
Sandomil	11K/05	1.2	2.3(2.9)	6.6(0.2)	2.2(1.1)	34.6(14.9)	96.6(89.4)		0.9(0.2)
Aldeia Dez	12J/03	0.7	2.3(3.2)	6.9(0.2)	1.1(0.4)	43.3(9.7)	255.6(587.7)		0.7(0.1)
Alb. Fronhas	12I/03	1.4	2.0(1.1)	7.7(0.9)	1.0(0.0)	52.7(3.6)	10.0(12.8)	6.7(1.4)	0.5(0.2)
Ponte Mucela	12H/03	1.4	2.0(1.8)	6.9(0.1)	1.1(0.5)	47.6(5.2)	31.8(54.8)	5.8(0.4)	0.6(0.1)
Lares†	13E/02	3.7	15.6(10.5)	7.7(0.1)	4.2(0.6)	[4556(3230)]	43.3(35.3)		1.7(0.4)
Full Basin		3.8	4.3(3.2)	7.2(0.5)	2.3(1.3)	66.7(39.2)	100.9(73.6)	9.2(3.4)	1.0(0.5)
Units		%	$mg.l^{-1}$		$mg.l^{-1}$	$uS.cm^{-1}$	MPN.100 ml^{-1}	$mg.l^{-1}$	$mg.l^{-1}$

Regression analysis was performed on the water quality data, which were compared against ISA values for each sub-basin. To this end, the medians of each station (2011-2013) were calculated and log-transformed to mitigate seasonality (Mehaffey et al.,2005). Since no other product provides other land cover variables at a comparable spatial resolution, the regression analysis was limited to impervious surface area, as aforementioned. In the future, we expect to extend this analysis to other variables, using a set of products developed using a common set of requirements. This effort is described in a latter section of the present work.

Amongst the objectives was to understand the underlying affinities between the different sites. This is an important step towards the establishment of an objective segmentation methodology to organize the watershed into coherent units. Up until now, the Mondego river

watershed is divided into three regions, the 'upper valley', the 'lower valley', and the estuary. But these regions are still diverse and are not based on objective metrics. The somewhat subjective 3-region system fails, for instance, to differentiate the Mondego and Dão sub-basins, characterized by different geology, topography, and land cover patterns.

To accomplish a segmentation of the sites, the water quality stations were clustered using the X-Means methodology (Pelleg et al., 2000) implemented in the software Weka. X-means, an improvement over the popular k-means clustering method, offers several advantages including the automated definition of the number of clusters and optimization the measurements of the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). The X-means is also more efficient computationally, when compared with the k-means methodology.

It must be emphasized that the X-means methodology reflects more the state of the stream segments than actual natural divisions found within the study area, but these are still valuable data that can be used in support of management activities.

The clustering algorithm resorted to a set of regular measurements of several variables including Total Suspended Solids (TSS), pH (Laboratory), Total Nitrate, Conductivity (20º C), Fecal Coliforms, and Total Nitrogen. The algorithm determined the existence of 2 cluster centers, with a Bayesian Information Criteria (BIC) score of -24.21. The smallest cluster (Cluster A) groups 3 stations, including 'Ferreirós' (11I/01), 'Agueira Santa Comba Dão' (11I/11), and 'Ponte Formoselha' (12F/04). All other stations were grouped in the second cluster (B). The 'Lares' (13E/02) was not included in the analysis initially because it is strongly influenced by the estuarine environment and no other comparable stations are available for analysis. Nonetheless, when included, this site is a member of Cluster A.

The stations of Cluster A are located in different sectors of the watershed (Figure 5.1.3.1) but share the proximity (upstream) to large urban centers and high ISA average values. Stations 11I/01 and 11I/11 are located immediately downstream of Viseu (the second largest city in the watershed) and Mangualde (an important industrial center), while station 12F/04 is located downstream of Coimbra (the largest city in the watershed). Other possible explanations of the results were explored, including the contribution of farming activities to water quality. According to CORINE2012, the fraction of farmland in the sub-basins is significantly higher in the regions drained by the sites of Cluster A (ANOVA, $p < 0.01$).

Nonetheless, it is impossible to assess the impact of other land cover classes in this study given the absence of products with a comparable spatial resolution. Furthermore, a coarse comparison of ISA and % Agricultural ($r = 0.83$, $p < 0.005$ land in the sub-basins yields a strong correlation as well, suggesting a co-variance and further complicating the analysis. Considering the different spatial resolution and methodologies used to create both datasets, the analysis

is, in itself, statistically meaningless. However it suggests the hypothesis of a correlation between both variables, which may hinder the successful separation of both contributes when determining the source of water pollution, unless higher resolution studies are conducted.

Considering these limitations and bearing in mind that none of these results are final, the potential impact of ISA over water quality variables at sub-basin level was further explored. Several variables show a strong correlation with the average Impervious Surface Area of the sub-basins (Table 5.1.3.2). Of these, conductivity at 20°C is the variable with the strongest linear correlation (positive) with ISA2013 described in the previous section. The Pearson correlation coefficient for the entire set of sites is of 0.90 (significant at a $p < 0.01$), but excluding the Dão river sub-basin and sites downstream of Coimbra, the value of *r* increases to 0.97.

Table 5.1.3.2. Results of a regression analysis of the water quality variables (2011-2013) and Impervious Surface Area (2013).

	Regression Analysis (Log transformed)		
Variable	β	R	Significant for p<0.01
Total Suspended Solids	0.64	0.62	Yes
pH (laboratory)	8.48	0.69	Yes
Total Nitrate	0.51	0.50	No
Conductivity (20° C)	1.01	0.90	Yes
Fecal Coliforms	-0.02	0.03	No
Total Nitrogen	1.10	0.74	Yes
Chloride	1.06	0.91	Yes

Total Suspended Solids, Total Nitrogen, and pH are all correlated with ISA2013 in a significant manner ($p < 0.01$). The *r* values calculated in the regression analysis are similar for the three variables, and not particularly high. On the other hand, Total Nitrate and Fecal Coliforms are not significantly correlated with ISA2013 in the Mondego Watershed, at the work scale.

In the case of Total Nitrates ($mg.l^{-1}$) and Total Nitrogen ($mg.l^{-1}$), two sites are worth analyzing in greater detail. The first is located in the central section of the 'lower valley' (station 12F/04, "Ponte da Formoselha") and has the highest average Total Nitrate concentration and Total Nitrogen of the entire set. This is in line with previous research for the area, which is known to play an important role in the eutrophication of the estuary due to agricultural runoff (Marques et al., 2003).

The second interesting location is the Aguieira Dam, a major structure (water storage capacity of 423000×10^3 m^3 (IA, 2005)) located in the 'Upper Valley' where the Dão and Mondego rivers

meet. The sampling station located in the 'Dão river' sub-basin (station 11I/11, "Agueira Santa Comba") has an average Total Nitrate value of 4.43 (± 3.2) $mg.l^{-1}$ and Total Nitrogen of 1.9 (± 0.69) $mg.l^{-1}$. On the other hand, the station located in the Mondego sub-basin section of the dam (station 11I/10, "Pinheiro Ázere"), has an average Total Nitrogen value of 0.78 (± 0.24) $mg.l^{-1}$ and 1.694 (± 1.03) $mg.l^{-1}$ for Total Nitrate.

Previous studies suggest the occurrence of eutrophication and algal blooms at Agueira Dam (IA, 2005; Oliveira and Monteiro, 1992), which prompted us to analyze the Landsat time series used in this study to search for signs of eutrophication through the analysis of 'water color'. The imagery confirmed a recurrent change in the reflectance values, which are consistent with an increase of the algal biomass in the surface waters of the Dão section of the dam (Figure 5.1.3.2).

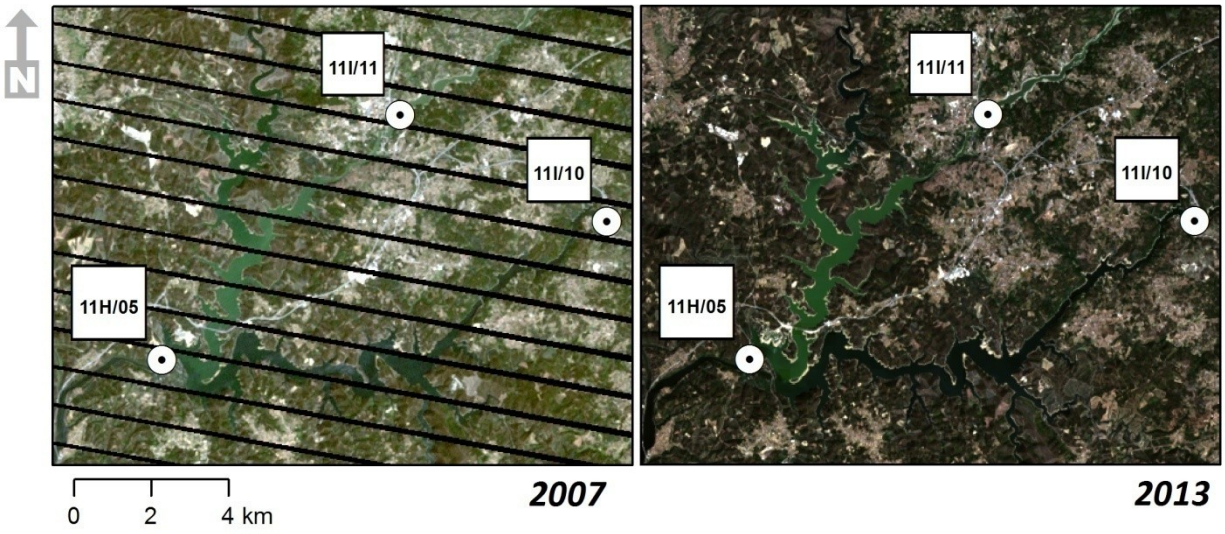


Figure 5.1.3.2. True color Landsat imagery of the Aguieira Dam (31 August, 2007 and 8 September, 2013), showing algal blooms in the Dão section of the waterbody. The water sampling sites are represented and labeled. The site at the Dão river is identified as 11I/11 and the site at the Mondego river is identified has 11I/10. The dam wall is located close to site 11H/05. The black stripes are caused by a sensor malfunction (SLC-Off).

It was not possible to correlate Landsat reflectance values with *in situ* algal biomass or attempt an estimation of surface concentration.However the author was able to verify on the field the occurrence of a surface accumulation of algae in the vicinity of Santa Comba Dão matching the 2013 anomaly imaged by Landsat. Future research could be directed towards the development of an semi automated monitoring tool relying on multi-mission data to deliver early warnings

on the occurrence of blooms in relevant reservoirs to stakeholders and the scientific community.

Considering the small number of stations with Chloride concentration measurements, and the lack of data for low ISA2013 sub-basins, a regression analysis was performed but it is hardly representative of the watershed.

This path finding approach was made possible by the application of novel mapping strategies, which enabled the production of the highest resolution imperviousness map of the Mondego river watershed to date. The open archive of SNIRH was also essential, attesting to the importance of the implementation of open data policies by government agencies managing data with potential scientific applications.

The first step of the analysis included the clustering of water quality stations, using an array of variables, in order to spatially organize the different sites. The data from clustered elements can be compared or further explored in order to support the identification of the drivers motivating the shared traits. This work follows the approaches adopted in previous studies, which employed clustering techniques (e.g. k-means) to identify water quality zones and sources of pollution (Mandel et al., 2015, Sharif et al., 2015).

The use of the X-means algorithm enabled the automated identification of 2 clusters. The automation of the method, which enables the selection of an optimum number of clusters is a significant advantage when compared with the k-means alternative (Pelleg et al., 2000).

The first cluster (Cluster A) grouped a set of stations located downstream of two urban centers (Coimbra and Viseu). The stations are characterized by high values of Total Nitrogen and Total Nitrates, as well as high values of conductivity. ISA2013 in these sites is also amongst the highest average values as highlighted by Table 5.1.3.1. According to Corine 2012, 5 stations not included in this cluster, drain areas with equivalent agricultural fractions. Thus, agriculture alone seems an unlikely cause driving the clustering results. Water quality is determined by the interplay of different drivers, which can be captured by clusterization in different ways (Akbar et al., 2013).

In this case, urban runoff from Coimbra, Viseu, and Mangualde are likely to originate a higher pollutant load, to which industrial effluents and agricultural activities further contribute. Nevertheless, additional research is required to confirm this hypothesis, reinforcing the need for additional layers to be included in a suite of products describing continuous land cover metrics at a spatial resolution of 30 m (or better). Pollution sources from main urban areas include sewage and nonpoint source pollution from surface runoff. Sewage contamination is likely to be mitigated by water treatment plants, under the heavy investments executed in the sector in recent decades across the country.

Regression analysis was explored as an additional information source regarding the impact of ISA over water quality. The observed correlation between conductivity and ISA, may result of secondary salinization sources, such as sewage and industrial activities (Cañedo-Arguelles et al., 2013), since natural sources are unlikely given the geologic setting of the 'upper valley'. Conductivity can be influenced by ions such as chloride and nitrate (Rose et al., 2001, Hatt et al., 2004), both of which were succinctly analyzed in this study. Chloride concentration was linearly correlated with conductivity ($R=0.98$, $p<0.01$), suggesting this ion is responsible for the measured values. Previous studies highlight the interplay of landscape and anthropogenic variables in the definition of chloride concentration in water samples, with urban runoff playing an important role (Morgan et al., 2012, Corsi et al., 2015).

Road deicing is another common factor explaining this correlation, but not in the Mondego watershed. The use of road salt is somewhat limited in Portugal, due to the infrequent and typically localized snow cover. As such, other secondary sources including wastewater effluents, the treatment of drinking water, and fertilizers are more likely to account for the measured values (Steele et al., 2011; Kelly et al., 2012; Mullaney et al., 2009). In fact, wastewater effluent inputs contribute to about half the total chloride load in watersheds with significant levels of urbanization (Steele et al., 2011). Nonetheless, the chloride sample is too small (9 stations) to allow a careful and exhaustive analysis.

Total Nitrogen, pH, and Total Suspended Solids are correlated significantly with ISA2013. As in the previous cases, it is difficult to determine a causal relation between imperviousness and these variables, especially when other land cover metrics are not currently available. Nonetheless the increasing pH levels with ISA2013 are in line with previous studies (Coulter et al., 2004). Similarly, the role of urban land cover was recognized as a significant contributor to increased Total Nitrogen concentration in streams both recently (Ding et al., 2015, Sun et al., 2013) and historically (Cohen et al., 1976). Untreated sewage, urban runoff, and discharges from industrial and agricultural facilities can drive the relation between ISA and these variables.

Previous studies provide evidence that imperviousness and loss of naturalness impacts water quality variables in the region (Ferreira et al., 2004; Feio et al., 2010; Arco et al., 2012). The application of the widely recognized 10% and 30% imperviousness thresholds to the sub-basins found within the Mondego river watershed, highlighted the extent of the most vulnerable areas. It also showed how these areas are located predominantly in the areas of influence of major motorways and trunks, with the few exceptions found in the vicinity of county seats in the hinterland. This information can be leveraged to prioritize conservation and restoration

efforts in the watershed, using accurate disturbance metrics, previously unavailable at this scale.

In summary, and although the Mondego river Watershed is traditionally divided into 3 regions (i.e. 'Upper Valley', 'Lower Valley' and Estuary, e.g. Teixeira et al., 2014), land use and land cover (including ISA) and water quality dynamics such as the persistent eutrophication of the Dão-segment of the Aguieira dam suggests the 'Upper' watershed should be further divided into the Dão and Mondego sub-basins for environmental research purposes. Such a division has the advantage of covering an area that partially overlaps with the NUTS III statistical polygons.

The environmental relevance of the ISA product was succinctly tested through the comparison against several water quality variables across the watershed. Although the impact of ISA over water quality was not in question (Schueler et al., 2009), we were interested in the specific signatures of disturbance for the region and at the scale of the study.

This is particularly important to understand the ecological (and economic) relevance of the water inputs from the watershed (*lato sensu*) into the coastal system, particularly in the estuarine and nearshore environment off Figueira da Foz. Eutrophication events reported at the estuary (Marques et al., 2003, Lillebø et al., 1999, Pardal et al., 2000), the fact that the sites in the terminal course of the Mondego are clustered together with the station at Aguieira (often eutrophicated), and the special conditions in the coastal system near Figueira da Foz (e.g. persistent buoyant plume) jointly contribute to the importance of this topic. As such, it must be considered whether, under appropriate conditions, riverine water quality conditions may influence key ocean color variables in the vicinity of the river mouth, namely within the area of influence of the riverine plume generated by the Mondego river.

Although it is unlikely that the dramatic blooms seen in the imagery covering Aguieira are seen in the coastal waters off Figueira da Foz, it is possible that localized effects (in time and space) are detectable using Earth Observation data. We most strongly encourage the development of future research projects to address this issue.

Other disturbances, including in oceanic turbidity fields can be introduced not only by river discharge but also by manmade structures, including those coming from domestic and industrial marine outfalls.

In the study area, several of these marine outfalls were in operation, some producing clearly visible signatures in the spectral data collected by MODIS and Landsat sensors. Amongst these outfalls we can emphasize the one at São Jacinto (Aveiro) and the Leirosa Industrial Outfall, south of the Mondego river estuary. The output of both can be detected in EO data, without

the employment of complex techniques, hinting at the impact over the spectal signature of waters in the vicinity of the outfall.

However, we attempted to quantify the impact of the outfall over turbidity, relying on a proxy variable – K_d490, the Diffuse attenuation coefficient (K_d) for downwelling irradiance at 490-nm. This is a product systematically generated for MODIS data, which can be used as a proxy for water turbidity and visibility because it measures light attenuation by absorption and scattering effects created by particles in the water (organic and inorganic).

The algorithm in use to generate K_d490 (m^{-1}) products relies on the empirical model suggested by Mueller (2000). The model uses a band ratio of the normalized water-leaving radiances (nLw) at 490 and 555 nm. Kd490 is then calculated as:

$$K_d 490 = K_w 490 + A(\frac{nLw(490)}{nLw(555)})^B$$

Where, A, B, and K_w490 are model coefficients.

Because the spatial resolution of MODIS might preclude the production of a detailed view of the outfall's signature, we applied an adaptation of the General Empirical Relation Model (GERM) (Gao et al., 2008) to create a high resolution K_d490 image from Landsat-8 OLI data.

The Leirosa industrial outfall was selected for the test, considering its size and characteristics. The structure was constructed by ETERMAR for CELBI, and serves the industrial facility (paper pulp production) located onshore. The pipe has a width of 1200 mm, a length of 1546 m, and is equipped with a diffuser, with the outflow entering the marine environment at a depth of 14.2 m (CELBI, 2013).

A pair of same-day (23 August, 2013) match-ups of MODIS AQUA level-2 and Landsat-8 OLI were acquired. The Landsat image was processed to Surface Reflectance using FLAASH, using the methodology already described in an earlier section.

Afterwards, the Landsat image was downscalled to the same resolution (~1 km) of the Level-2 K_d490 MODIS AQUA image. The pixels of both datasets were collocated at SeaDAS and a low pass filter (3x3 kernel) was applied to the Landsat data.

After completion of this step, the values from each filtered Landsat band and the MODIS K_d490 product were extracted, whenever the quality flags of the latter suggested the fields were reliable. A total of 3326 sample pixels covering the region between Figueira da Foz and Oporto were selected and extracted in that manner.

The extracted data were exported to Weka, where a Regression Tree Model using the M5P algorithm (Hall et al., 2009) was used to create an empirical Landsat K_d490 product. The RTM

was created using a 66% partition of the data, producing 8 different rules, each describing a regression model. This step departs from the original GERM methodology, but returned a model with a Pearson correlation coefficient of 0.95 and a mean absolute error of 0.0189 m^{-1}.
The different rules and equations were mapped as GeoTIFF files using the same python script designed to create the Imperviousness Surface Area map, described in the previous section.
The results are depicted in figure 5.1.3.3, which highlight the size of the turbidity plume generated by the outfall.

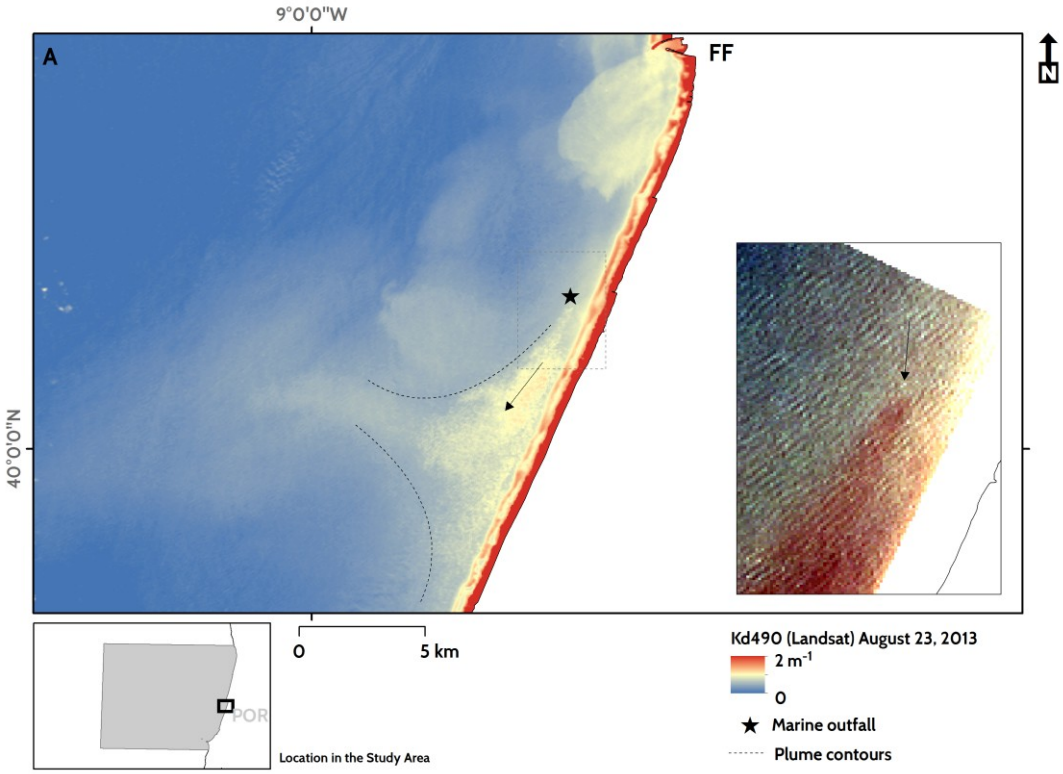


Figure 5.1.3.3. Landsat-8 OLI K_d490 image of a subset of the study area, South of Figueira da Foz, acquired on August 23, 2013. The image depicts the turbidity plume created by an industrial marine outfall in the vicinity of Leirosa. The inset (B) depicts the 'true-color' Landsat-8 OLI image of the area near the outfall, showing the point-source of the plume (black arrow).

The plume can be detected as far as 10 km southward of the source, while dispersing offshore, aided by a filament protruding westward and extending 5 km from the coastline. This filament creates favorable dispersion conditions that enhanced and spread the impact of the plume to a wider area than what would be expected under different conditions.
The K_d490 reaches values as high as 1.3 m^{-1}, while the Mondego river plume doesn't exceed 0.9 m^{-1}. The higher the value of K_d490, the lower is the visibility in the ocean's surface layers, and shallower the photic depth. Additionally, the turbidity plume may also interfere with the

detection of frontal activity considering the sharp gradient in ocean color (and possibly SST as well) between plume waters and the surrounding environment. This effect is likely to be more prominent in months when the overall turbidity is less significant and detection is thus simplified.

This simple example shows adequately how even seemingly minor infrastructures, such as marine outfalls can have a sizeable impact over the optical environment of broad regions. It would be interesting to evaluate the impact of the altered optical environment over the biological communities of the area, especially those who are known to have specific requirements of underwater visibility or photic depth.

Monitoring watersheds, including both natural and manmade discharges is paramount to guarantee that products heavily reliant on ocean optics are produced and interpreted adequately, in light of varying conditions and characteristics of the flows.

To interpret the signatures of these features it is required some level of knowledge regarding the nature of the inputs. In the case of watersheds, land cover data is required to understand the contributions of different variables into the overall water quality. For manmade structures, knowledge of the type of effluent, flow, and source is relevant, but often difficult to obtain.

A significant limitation to the mapping and monitoring attempts described in the present section arise from the absence of comparable products describing multiple land cover metrics (e.g. forest area, farmland) as well as of maps identifying and characterizing relevant infra-structures (e.g. outfalls).

In summary, water quality is capable of influencing critical variables that may be retrieved from EO data including Chlorophyll a concentration, SST, and turbidity. Examples of these effects were exemplified throughout the text, highlighting the need to carefully evaluate the impact of natural and manmade water quality drivers upon inspection of satellite imagery of large riverine basins and coastal systems.

The dramatic water color changes due to algal blooms in the Aguieira Dam, the turbidity plume in the vicinity of the industrial marine outfall, and the traits of the Mondego river plume close to the estuary are but a few examples of the influence of water quality over the optical environment of the watershed-coastal continuum.

Connecting watersheds to the coastal environment requires the action of rivers or artificial systems, the former governed by precipitation levels. If water color may be influenced by watershed-based activities, the dimension (and again water color) of river plumes is essentially determined by rainfall values across the basin. As such, combining information sources on land metrics and rainfall is essential to develop an integrated image of the system. That is the subject of the following sections.

5.2. Rainfall and runoff contributions to coastal ocean color

A proper management of the coastal system requires an adequate knowledge of inland dynamics and a clear understanding of the connectors linking both domains. Water is the perhaps the most prominent binding element in the movement of sediments, nutrients, life forms, and pollutants from land to ocean.

In our study we emphasized, persistently, the connection between watersheds and the coastal system and attempted to offer compelling evidence to the impact of riverine systems (or land-based activities) over coastal water masses and ocean color. The river plumes generated by rivers are perhaps the most emblematic representation of the coupled land-ocean dynamics. As described earlier, the Mondego river plume size, spatial patterns, and detectability using EO data are all modulated by the interaction of several variables. The complex patterns generated by the plume are similar to those found elsewhere (Horner-Devine et al., 2009) and illustrate the complex processes at play (Cunha et al., 2002).

The ecological relevance of these features (Zamon et al., 2014), and of the broader freshwater lenses created by the cumulative impact of western Iberian rivers (Peliz et al., 2002) suffices to justify the integrated study of both systems.

Yet, previous studies also emphasized the relevance of watersheds as important contributors to the overall stressors affecting shelf ecosystems (Lima et al., 2012). In fact, it has been suggested, and rightfully so, that marine biogeography should incorporate watersheds in oceanic partitioning efforts (Callejas-Jimenez et al., 2012).

Biogeographic classification systems like the Large Marine Ecosystems, to which we alluded to earlier (Sherman, 1990), offered the conceptual basis for the approach put forth in this study, which aims at integrating data from different sources to characterize the riverine-marine continuum, with an emphasis on spatial discontinuities. Rivers are, in this context, ideal candidates to play the role of significant sources of boundaries in the coastal system. Still, challenges persist to the establishment of widespread, reliable, and cost-efficient monitoring networks.

In particular, rainfall gauge networks are often limited, or the data is not easily available at reasonable costs, thus preventing the widespread development of user-driven solutions either for research or operational applications.

Satellite Rainfall Estimates (SRE) can mitigate these limitations and provide a solution (or soure of redundancy) for the problems including the spatial distribution of gauges or its availability and survivability (Tapiador et al., 2012, Chen et al., 2013).

The Portuguese portal SNIRH (SNIRH, 2016) offers a good example of how SRE can indeed replace or supplement existing networks, even in developed countries. When users search for weather data on the portal, 680 stations are available and provide daily rainfall measurements if no other constraints are added. Yet, only 399 have recorded data in the last 365 days (as of October 2016) and 76 in the last 5 (4 in the Mondego river watershed). Furthermore, the spatial distribution of the gauges is irregular, with wide swaths of the country lacking data. Between 2008 and 2009 the maintenance of the stations was also suspended, compromising data quality throughout the period.

Other government agencies, like the Instituto Português do Mar e da Atmosfera (IPMA) have similar networks but the data is not provided to users at a low cost. As such, data sources such as that provided by SNIRH are pivotal to a wide range of applications but often fail to meet the requirements of users including timeliness, quality control, and coverage.

In this study, we shall evaluate whether SREs can be employed in the Mondego river watershed, given its central location within the overall study area and relevance to the WIBP (marking its southern border). To this end, several performance metrics shall be used to compare satellite and *in situ* measurements. The outcome of the validation study, shall drive the development of the rainfall segment of the multivariate data portal described in a latter section – the SmartBasins.

The SmartBasins initiative aims at developing a holistic watershed-coastal monitoring system, in a stepwise progression, to integrate the information into synoptic reports, meaningful to a broad end-user community.

Weather data, and rainfall information in particular is commonly available in European countries. However, as mentioned earlier, much of this information is often provided by private and public operators at a fee, a fact that often constrains the regular and cost-effective monitoring of these variables outside official channels. As such, it is pressing to devise and deploy efficient rainfall data distribution systems capable of being introduced into scientific and operational settings (Meng et al., 2014; Huffman et al., 2010).

Reducing the costs associated to these data would favor the development of new applications capable of addressing pressing societal challenges by a more transversal set of end-users. Such applications may include a broad range of applications from vegetation health monitoring (Du et al., 2012), hydrological modelling (Tapiador et al.,2012, Scofield et al., 2003), and land cover studies (Mantas et al., 2015) to state only a few. We expect that, providing validation of alternative rainfall estimates is completed and awareness is raised with the potential end-user community, innovative solutions leveraging these data sources may be rapidly identified.

The most efficient alternative to the use of rainfall gauges or land-based systems (e.g. NEXRAD), is in the adoption of Satellite Rainfall Estimates (SRE) (Huffman et al., 2007).

The same principle applies to our study area with a caveat. The Mondego watershed is covered by a set rainfall gauges maintained by the Portuguese government (SNIRH, 2016). Data from these gauges is freely available and accessible in open, online archives. Still, data coverage from these stations has been occasionally disrupted due to maintenance issues. Although other systems do exist, users of these data would be left with few (if any) available data sources offered for free with comparable quality frequency.

The SmartBasins portal, which includes in its set of requirements the exclusive use of proprietary in-house datasets or third-party information made available in free and open archives, would certainly be amongst the systems affected by such data shortage.

For most of the study period (2003-2014), data from the Tropical Rainfall Measuring Mission (TRMM) were the main source of freely available SREs with a near global coverage and distributed from an easy (and free) access point. The mission was a joint venture of NASA and the Japanese Space Agency (JAXA) and was responsible for a significant leap in rainfall studies across the globe. Completed in 2015, after 17 years in space, TRMM was followed by alternative systems launched before its fiery plunge into Earth. The most prominent of these new generation missions is the Global Precipitation Measurement (GPM) Core Observatory (Liu et al., 2016b), again a partnership of NASA and the Japanese counterpart.

The core observatory's data are supplemented by information acquired by a constellation of other satellites from multiple agencies and organizations worldwide. Although it introduces important improvements to the field, data from GPM were not used extensively in this study considering that only a few months of post-orbit calibration data were available. Furthermore, significant data gaps hindered the use of SNIRH's rainfall gauge data. Although future plans include extensive validation efforts of the newer GPM mission products, in the context of this study it was considered to be out of context. Still, a preliminary and highly exploratory comparison of results from TRMM and GPM for selected gauges is also presented.

The Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) provided the core products made available to the public using the satellite's data (Huffman et al., 2007, 2010). These products were available in two versions (research grade and real-time) and released every three hours at a spatial resolution of 0.25°. The research grade product was the result of a complex processing chain that included a step in which the SREs were merged with data from *in situ* gauges. The workflow introduced a temporal lag between data acquisition by the satellite and final release to the public of approximately 1 month.

On the other hand the real-time product was available only 9 hours after data acquisition, making it a suitable option for ongoing monitoring and operational applications.

Data from GPM, particularly the Integrated Multi-satellitE Retrievals for GPM (IMERG) product, follows a similar approach, despite the release of three different versions ("early", "late", and "final") (Huffman et al., 2015).

The IMERG products are divided into three categories, with different latencies and characteristics. The 'Early Run' (6 hours), 'Late Run' (18 hours), and 'Final Run' (4 months) all serve different purposes and can be employed by users with different objectives and requirements.

These data products are made available at GIOVANNI at a spatial resolution of 0.1° and half-hourly (in the best cases) temporal resolution (Liu et al., 2016b).

The study area is characterized by variable patterns, which translate the topographic variability, distance to the coast, and the boundary between Köppen-Geiger's Csa and Csb regions. To the East, where annual precipitation values are higher (> 1500 $mm.year^{-1}$), it is also where the terrain is more complex, offering additional challenges to the retrieval of SRE (Mantas et al., 2015).

Previous studies employed the (now legacy) TRMM datasets to describe rainfall patterns in the region of interest and correlate it with other important metrics (e.g. Immerzeel et al., 2009).

Topography is known to play an important role in the performance of SREs worldwide (Mantas et al., 2015, Ward et al., 2011; Dinku et al., 2008). Complex terrain and rainfall characteristics (Conti et al., 2014, Kidd et al., 2012) create challenging conditions for the accurate retrieval of rainfall values.

Dissimilarities in the regional conditions translate into differing performance results, as reported in a variety of studies (e.g. Li et al., 2014, ICIMOD, 2013, Almazroui, 2011, Chen et al., 2013, Sapiano et al., 2009). The disparities in the literature strengthen the case in favor of regional validation studies, as that we are presenting succinctly in this work.

Still, most studies agree in several aspects, including an improved performance of the TMPA research grade algorithm (against the real-time counterpart), and with temporal aggregation (Dinku et al., 2010b, Almazroui, 2011; Scheel et al., 2011).

As such, and considering the relative density of rainfall gauges in the study area, a brief regional performance analysis was conducted to determine (preliminarily) the feasibility of employing TMPA or IMERG datasets for rainfall studies and applications in the study area. Both datasets tested are higher level products, which include multi-sensor data, namely from a network of *in situ* gauges. These are therefore amongst the best options available for rainfall studies, although limited in terms of real time applicability due to large latencies.

5.2.1. Data sources and validation metrics

Daily rainfall data from a set of 52 rainfall gauges maintained by a Portuguese agency (Autoridade Nacional da Água) and made available at the website of the *Sistema Nacional de Recursos Hídricos* (SNIRH, 2016) were downloaded and stored. The data were submitted to a simple quality control, which was aided by the descriptive tables characterizing the stations (including a quality index) made available at the website. Only gauges that collected valid measurements for at least 80% of the time series were selected to ensure quality standards were met. This requirement limited the analysis to a total of 15 gauges spread across the watershed.

The data were then stored as daily and monthly fields, which would be used as the ground truth information for TMPA validation. Because TMPA is generated at a spatial resolution of 0.25°, several SNIRH gauges were located within a single SRE grid cell, but after the removal of lower quality sites, such overlapping did not occur. In such cases, the different gauges would have been evaluated separately but were also averaged, to test the impact of a variable number of ground truth datasets over the final validation outcome at daily and monthly time scales (Mantas et al., 2015, Porcù et al., 2014).

Information on the elevation of the gauging sites was available in the metadata. The gauges were located at sites in a broad range of elevation values (minimum=15 m, maximum=1560 m (776 m after the removal of low quality gauges), mean= 305.1m), thus supporting a comprehensive validation and intercomparison of TMPA values in different orographic settings.

TMPA data were downloaded from NASA's Giovanni (Berrick et al., 2009) as daily 3B42 Version 7 Research grade products (Huffman et al., 2007, 2010). The data is provided at the spatial resolution of 0.25°, as aforementioned, and a match-up table was generated to correlate the ground truth gauges with the TMPA fields. The SRE were then summed to generate products with the same temporal resolution for comparison purposes (daily, 8-days, monthly).

The SNIRH archive was further searched to identify rainfall gauges containing data coincident with GPM operations. A total of 11 stations were selected (at least 90% of the data from January to December 2015) and the matching IMERG Final Run fields downloaded from GIOVANNI. The same quality control was applied to these data, and none of the SNIRH gauges met the 80% threshold and were thus eliminated. Because no other sources of data were available, it was not possible to validate the IMERG product.

Several metrics were used to compare the SREs and *in situ* measurements. No interpolation techniques were used in the study, and instead the data comparisons were conducted using data from the SRE grid cell matching the location of the gauge.

Amongst the statistical measures calculated were the Unconditional Bias, Pearson correlation coefficient, and Root Mean Square Error (RMSE).

Unconditional Bias is calculated as:

$$Bias = \sum SRE \,/\, \sum gauge$$

RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(R^{SRE}(xi) - R^{gauge}(xi))^2}$$

Where, R^{SRE} is the satellite precipitation estimate and R^{gauge} is the rainfall value recorded in the SNIRH gauge.

Pearson's Correlation Coefficient, despite several conflicting arguments regarding the normality of rainfall distributions, was employed, considering the long-tradition of including this metric in performance assessments (Wang et al., 2009).

Finally, several categorical performance metrics were applied to daily match-ups, using contingency tables (table 5.2.1) to determine whether rainfall detections were successful. Rainfall was considered to be detected whenever a value over 0.1 mm was detected by other satellite or *in situ* sensors.

Table 5.2.1. Contingency table used to evaluate the performance of the rainfall estimates.

	In situ rain	*In situ no rain*
SRE rain	a	b
SRE no rain	c	d

In particular, three different metrics shall be calculated. These include the Frequency Bias Index (FBI), which is the ratio of the number of events in the satellite product against the number of field-validated targets. FBI is calculated as:

FBI = (a+b)/(a+c)

The False Alarm Ratio (FAR) shall also be calculated and consists in the number of false alarms for the total number of satellite-based detections. FAR is calculated as:

FAR = b/(a+b)

Finally, the Probability of Detection (POD) is to be calculated as well. PAR is the number of true detections against all *in situ* detections. POD is calculated as:

POD = a/(a+c)

5.2.2. SRE validation results and discussion

The TMPA product performs poorly at daily scales (mean R = 0.35), especially when compared with the monthly aggregated product (mean R = 0.81). This is in line with previous findings as aforementioned and offers important challenges for high frequency monitoring of watersheds.

The limited size of the sample limits the options available for interpreting results. In fact, we encourage the development of future research projects to expand the validation to the entire Iberian Peninsula and include information from newer products, including IMERG, as attempted in this study. Table 6.2.2.1 summarizes the performance results obtained for the selected gauges.

Table 5.2.2.1. Summary of the validation of TMPA data using *in situ* measurements acquired from SNIRH. At monthly scales, the TMPA product performs at adequate levels for monitoring purposes.

Gauges	R (Monthly)	RMSE (Monthly)	FBI	PAR	POD	Bias (Monthly)	Elevation
Aguiar da Beira	0.9	41.2	0.4	0.4	0.3	1.1	776.0
Alagoa	0.8	48.9	0.4	0.3	0.3	1.2	182.0
Almaça	0.8	51.2	0.4	0.3	0.3	0.9	116.0
Carregal do Sal	0.9	45.7	0.7	0.5	0.4	0.8	275.0
Casal do Rato	0.8	40.7	0.3	0.4	0.2	1.3	22.0
Condeixa	0.9	41.4	0.4	0.3	0.3	0.8	90.0
Degracias	0.7	60.3	0.3	0.3	0.2	1.4	326.0
Fragosela de Baixo	0.7	61.9	0.4	0.4	0.3	0.8	376.0
Mourilhe	0.8	58.7	0.8	0.6	0.4	0.7	548.0
Oliveira do Hospital	0.9	43.7	0.7	0.5	0.4	0.8	468.0
Penela	0.8	39.7	0.4	0.3	0.3	0.9	253.0
Pombal	0.8	37.6	0.3	0.3	0.2	1.2	95.0
Santa Comba Dão	0.9	39.5	0.5	0.4	0.3	1.0	289.0
Soure	0.8	40.1	0.3	0.3	0.2	0.9	18.0
Sátão	0.8	47.8	0.5	0.4	0.3	1.0	570.0
Mean	0.8	46.6	0.5	0.4	0.3	1.0	293.6
Minimum	0.7	37.6	0.3	0.3	0.2	0.7	18.0
Maximum	0.9	61.9	0.8	0.6	0.4	1.4	776.0

There is no clear trend regarding the relation between TMPA estimates and SNIRH measurements. If at some gauges TMPA tends to overestimate rainfall values, in others the opposite was recorded.

Because elevation is often a key explanatory variable to SRE performance, the effect of this variable was assessed through linear regressions. However no significant correlation was found for any of the performance metrics. Similarly, distance to the coast was assessed and non-significant results were returned as well.

To further assess these relations the gauges were organized according to elevation, with the creation of 2 groups (threshold at 250 m) with a roughly equal number of elements. ANOVA was then applied to search for differences in the performance metrics calculated for both groups with the results mimicking those aforementioned. No significant differences were found between groups. The samples were also separated in two groups coastal (including all sites from the shore to the longitude of Penela) and inland. ANOVA was applied to the validation metrics but, and statistically signicant differences were found for the FBI ($p < 0.01$). The FBI provides information on the ability of the SRE to detect rainfall events in comparison with the total number of actual events. In this case it seems that at the coastal sites, FBI was significantly lower than in the inland (mountainous) regions. This translates into an improved ability of the TMPA product under evaluation to detect rainfall events inland, which is in line with previous findings elsewhere (e.g. Mantas et al., 2015), although due precautions must apply when comparing different regions.

Despite the findings concerning differences in FBI, we hypothesize that the size of the watershed is perhaps too small for complex regional patterns to be detected in the limited time series. Therefore, additional research is once again recommended, preferably with a larger dataset (nationwide or even Iberian).

The spatial distribution of key metrics (Pearson correlation coefficient and POD) is depicted in figures 5.2.2.1 and 5.2.2.2.

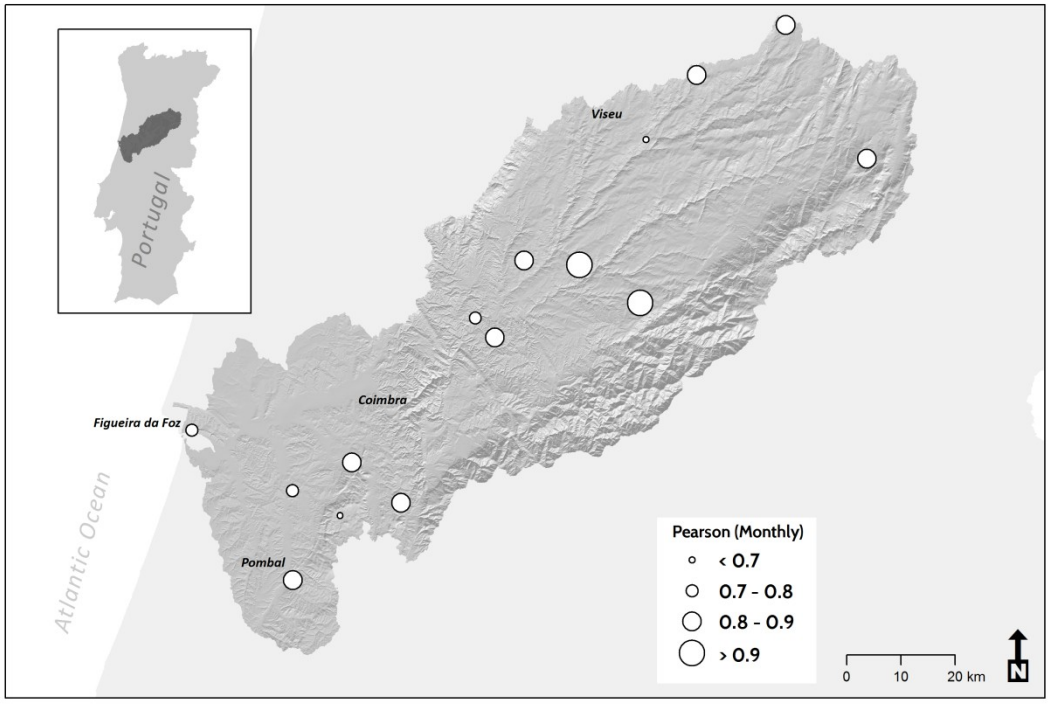


Figure 5.2.1.1. Pearson Correlation Coefficient (R) values for the monthly aggregated TMPA product calculated for the selected SNIRH rainfall gauges in the Mondego river watershed.

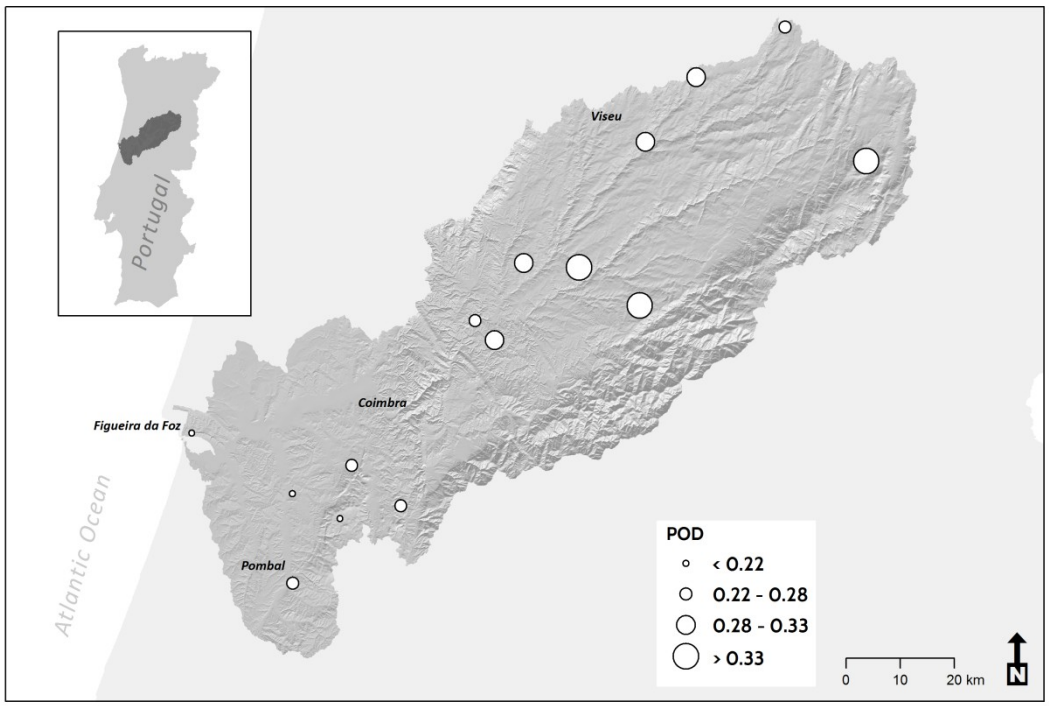


Figure 5.2.1.1. Probability of Detection (POD) values calculated for the selected SNIRH rainfall gauges in the Mondego river watershed.

The maps further add to the hypothesis suggesting that the spatial distribution of the validation results does not obey to a clear logic as elsewhere. The apparent lack of spatialization, which may be only apparent or the result of undersampling, makes it difficult for an in-depth analysis of the dataset and implications for regional watershed monitoring.

Differences in precipitation levels across gauges may also be insufficient to generate significant patterns across regions. This is an important aspect governing the performance of SREs and therefore of the validation metrics, unlikely to play a central role in the Mondego Watershed.

Because the culling of gauges for the sake of quality limited the amount of datasets that could be used, it was also not possible to validate the effect of multiple gauges in a single TMPA cell. This was expected to contribute to an improvement of the metrics, as the heterogeneity of rainfall is reduced by averaging the measured fields over multiple sites.

However, and for the purpose at hand, it is safe to consider the TMPA product adequate for monthly time scales but not daily. As performance improves with aggregation, it is indeed possible that for time scales from 15-days, the product is adequate for monitoring purposes. However, it must also be emphasized that the product that was evaluated is a science-grade version, not the near real-time counterpart. The latter would have performed more poorly, highlighting the persisting challenges in rainfall monitoring, especially at the local scale.

However, for coastal applications, in which the total rainfall value over a period of over 15-days is usually adequate for the evaluation of the development of features such as the WIBP, TMPA (or equivalent) is likely to be adequate and provide estimates at the watershed level with sufficient accuracy. Other applications including phenological studies or agricultural monitoring may also benefit exceedingly of these products.

However, several limitations still apply, namely the seemingly unspatialized distribution of the performance metrics. Other effects, namely of the locally complex terrain (Ward et al., 2011, Matzler et al., 2000) and seasonal snowfall (Huffman et al., 2008) should be examined in greater detail. At the present scale, however, it appears as if the Mondego watershed is a suitable region for the employment of SREs, with no statistically significant differences across most performance metrics. This is an important aspect to consider if the operational deployment of these products is to be operationalized.

5.2.3. SREs as a critical data source

Satellite Rainfall Estimates in the Mondego watershed can be considered as a reliable source of precipitation data at variable temporal resolutions, especially upon aggregation into longer periods. The lack of continuous data sources from *in situ* sources, which hampered the

validation attempts highlight the need to develop well calibrated rainfall products to supplement traditional sources.

The science-grade products evaluated emphasize the importance of incorporating data from *in situ* assets into the rainfall products. At the daily scale, which may be relevant for critical civil defense applications, real-time (Mantas et al., 2015) and even science-grade products are less reliable, as once again demonstrated in this study. Merging different sensors and decentralizing data sources through the incorporation of data from multiple providers is likely to be the way forward. This objective has been set as a priority for future studies, namely in the context of the SmartBasins initiative. Additionally, the development of comprehensive SRE validation software packages would probably constitute an important contribution to the field, streamlining performance studies and standardizing approaches internationally.

The emerging open access trend promoted by governments worldwide is expected to cover rainfall datasets in countries – such as Portugal - where access to such information is still restricted or expensive. Private weather stations, including by hobbyists, are also important sources of information, certainly capable of enhancing the available network significantly. Provided quality control methodologies are implemented, such decentralized sources may play an important role in revolutionizing rainfall studies and applications worldwide.

The variable performance of SRE algorithms with varying rainfall intensities in daily scales was once again observed in this analysis. These can introduce significant challenges to the development of early warning systems, which often rely on multivariate and multi-temporal (including hourly or daily scales) pre-defined thresholds to determine the risk to population and infra-structures.

Still, for the applications described in this study, which include rainfall-runoff modeling for WIBP research, water color (and quality) monitoring in watersheds and coastal waters, and phenological studies, SREs are certainly a useful and reliable tool. These applications will often rely on 8-, 15-, or 30-day intervals (or even seasonal), with the correlation of several variables occurring at variable spatial and temporal resolutions well beyond the daily or local scale. As such, incorporating SREs, as currently available and in hybrid satellite/*in situ* models is not only possible, but recommendable.

TMPA data were only succinctly analyzed in this study, which is not focused on SRE research. However, the analysis offers important elements, in line with those found in previous studies that support the use of the dataset for operational and scientific applications. The IMERG product was not evaluated because *in situ* datasets were not available in sufficient quantity and quality. The performance of the newer products cannot be described and must therefore be used with caution within the study area.

In the following chapters we describe the development of the data integration and distribution systems, which were developed throughout this project in a stepwise manner. Rainfall estimates took center stage along with oceanographic and land cover variables in the integrated approach to describe the watershed/coastal continuum.

Chapter 6. Integrated Watershed-Coastal management

Watersheds are the structural units of a territory. Delimited by objective geological boundaries, few other natural systems can be so adequately (and simply) defined and mapped. Nonetheless, political jurisdictions worldwide are still inadequately designed, failing to mimic the natural boundaries that should be acknowledged as first step towards a sustainable use of resources.

In this chapter, we attempt to offer contributions towards the dissemination of actionable information at the watershed level, promoting the reinforcement of cultural processes leading to the development of strong 'watershed identities'.

As aforementioned, watershed characteristics and stressors propagate downstream and can potentially affect water quality (and water color) in coastal areas. As such, we devised a setp-wise approach to establish a self-sustainable, integrated data distribution and science literacy program focused on watersheds and coastal systems. The Smart Basins initiative (www.smartbasins.com) aims at being a focal point for stakeholders in watersheds worldwide. The Mondego river watershed, central to the study area addressed by our study, serves as one of several implementation test-beds, where the solutions can be prototyped prior to global launch.

Rainfall data, compilation of information available from multiple sources, but also the dissemination of novel products including the ISA or frontal density cartography are some of the objectives to be accomplished in this first stage of development. A seamless integration of all products, to create a true information and decision support platform is expected to take place in the future, as outlines in subsequent sections.

6.1. Data distribution platforms, a watershed-based approach

The present section consists of a paper originally published in Computers & Geosciences (Mantas, V.M., Liu, Z., Pereira, A.J.S.C., 2015.A web service and android application for the distribution of rainfall estimates and Earth observation data. Computers & Geosciences, 77, 66-76.). Permission for the unformatted reproduction was obtained from the editorial office of the journal.

The paper describes the first of a series of steps towards the development of an integrated watershed/coastal information system. At the time, the subset of the platform was named "TRMM Explorer", after the core observation satellite from which it derived the rainfall

information. Sections and figure numbers were preserved to match those in the published paper.

Abstract

The full potential of Satellite Rainfall Estimates (SRE) can only be realized if timely access to the datasets is possible. Existing data distribution web portals are often focused on global products and offer limited customization options, especially for the purpose of routine regional monitoring. Furthermore, most online systems are designed to meet the needs of desktop users, limiting the compatibility with mobile devices.

In response to the growing demand for SRE and to address the current limitations of available web portals a project was devised to create a set of freely available applications and services, available at a common portal that can: 1) simplify cross-platform access to Tropical Rainfall Measuring Mission Online Visualization and Analysis System (TOVAS) data (including from Android mobile devices), 2) provide customized and continuous monitoring of SRE in response to user demands and 3) combine data from different online data distribution services, including rainfall estimates, river gauge measurements or imagery from Earth Observation missions at a single portal, known as the Tropical Rainfall Measuring Mission (TRMM) Explorer.

The TRMM Explorer project suite includes a Python-based web service and Android applications capable of providing SRE and ancillary data in different intuitive formats with the focus on regional and continuous analysis. The outputs include dynamic plots, tables and data files that can also be used to feed downstream applications and services.

A case study in Southern Angola is used to describe the potential of the TRMM Explorer for SRE distribution and analysis in the context of ungauged watersheds. The development of a collection of data distribution instances helped to validate the concept and identify the limitations of the program, in a real context and based on user feedback.

The TRMM Explorer can successfully supplement existing web portals distributing SRE and provide a cost-efficient resource to small and medium-sized organizations with specific SRE monitoring needs, namely in developing and transition countries.

1. Introduction

Precipitation has a far-reaching influence over ecosystems and most fields of human activity from food production to engineering and safety (IPCC, 2013, Tapiador *et al*., 2012).

However, measuring and monitoring rainfall events remains challenging and costly. The global rain gauge network is still unevenly distributed, resulting in important data gaps or

representativeness problems. The gauge coverage is particularly poor in developing countries and remote areas and the insufficiencies are accentuated by extreme events (Liu *et al*., 2014).

To address such shortcomings, several agencies designed and launched Earth Observation (EO) missions to collect near-global Satellite Rainfall Estimates (SRE) using passive microwave and optical (infrared) sensors. The launch of the Tropical Rainfall Measuring Mission (TRMM) in 1997 influenced SRE research in subsequent decades and opened a new age for global precipitation studies. Since then, the growing constellation of EO missions contributing to the production of SRE generated important inputs to numerous critical applications with undeniable downstream societal benefits (Meng *et* al., 2014, Hou *et al*., 2013, Kidd and Levizzani, 2011). The newly launched Global Precipitation Measurement Core Observatory (GPM) builds on this legacy and promises important advances in the unification of the different precipitation datasets currently available (Hou *et al*., 2013).

SRE can mitigate the uneven distribution of rainfall gauges across the globe (Tapiador *et al*., 2012, Collier, 2007), become a viable alternative in gauge-denial scenarios (Chen *et al*., 2013) and offer new options for near real time (NRT), continuous monitoring at moderate costs.

Some of these developments were made possible after the introduction of algorithms that aggregate the data collected by different sensors, using a variety of technologies, each with inherent advantages and limitations. The TRMM Multisatellite Precipitation Analysis (TMPA) is one such algorithm (Huffman *et al*., 2007, 2010). It aggregates data from different passive microwave and infrared sources for most of the planet (50ºN-S) and generates both a near real time and a research grade (RG) product.

TMPA data have been subject to extensive independent validation efforts that attest to the good performance of both the NRT and RG versions (Almazroui, 2011, Scheel *et al*., 2011, Dinku *et al*., 2010, Su *et al*., 2008). The quality of TMPA products spurred the use of these data in a broad range of applications from hydrological modeling and ecosystem research to disaster mitigation and response. Because of this diverse range of applications, SRE are often used together with other datasets such as vegetation indices (e.g. Du *et al*., 2013). For this reason, several online services provide SRE and other EO products simultaneously, often relying on map-based user interfaces. Examples of such systems include CropScape (Han *et al*., 2012), the “Heavy Rain, Flood and Landslide Estimates” portal (Hong *et al*., 2007), the River Watch webpage (Brakenbridge *et al*., 2012) or the GSMaP (Ushio *et al*., 2009), each addressing data distribution and analysis needs from a different perspective.

However, the growing demand for an easy point of access to SRE and, in particular, to TMPA data led to the development of NASA’s TRMM Online Visualization and Analysis System

(TOVAS) (Liu *et al*., 2007, 2012), which is one of the instances of the online system GIOVANNI (Berrick *et al*., 2009).

TOVAS is an online interface to access rainfall estimates that supports the production of high-quality visualizations, which can be exported to different customary outputs (e.g. plots, ASCII, maps). Other related prototype services have since been developed using the technology powering GIOVANNI and the data it generates (Liu *et al*., 2014, 2009). These spin-off projects respond to specific science problems and user demands while demonstrating the viability of using GIOVANNI to create new services downstream from primary TMPA data sources.

Nevertheless, while previous studies suggest that TMPA data may efficiently replace traditional data sources used in Early Warning Systems (EWS) - most notably in developing nations - (Wardah *et al*., 2008, Collier, 2007), TOVAS was not designed to serve such a purpose. The advantages of EWS and routine monitoring tools can only be fully realized if efficient, automated systems are in place, continuously analyzing data and making them available to end-users in convenient formats (UNEP, 2012). The absence of options for off-the-shelf continuous data analysis and the impossibility of retrieving values from an irregular area or moving target are some of the current limitations of TOVAS.

The emergence of multiple cloud-based solutions suitable for scientific computing (Prodran *et al*., 2013) and the introduction of smartphones as tools for science (Weng *et al*., 2012) herald new opportunities (and challenges) to the distribution of EO data to the end-user. The combination of existing data distribution and visualization systems, such as TOVAS, with new cloud-computing services and smartphone 'apps', creates new, interconnected and reusable data processing and mining options (Evangilidis *et al*., 2014, Díaz *et al.,* 2012).

This paper describes the effort to create a set of freely available, scalable applications and services that can 1) simplify cross-platform access to TOVAS data (i.e. run in a similar way across platforms with different architectures, Operating Systems and software, including desktop computers or Android devices), 2) provide customized and continuous monitoring tools for SRE and 3) combine different services at a single portal. Collectively, the applications are developed and distributed through the TRMM Expedite Exploration of Rainfall Estimates or TRMM Explorer Project.

The solution combines a Python Web Service (TRMM Explorer Web Service) for data processing tasks with Android smartphone apps (TRMM.Mobile and the "User Development Kit") for rapid data access.

The TRMM Explorer (TE) solutions address existing and potential user needs through the redistribution of TOVAS data in new ways.

This paper first describes the TMPA data, distributed by TOVAS and TE, followed by a description of the Web Service and Android App, their development and respective features. We then present a case study where the different solutions were deployed to monitor watersheds in southern Angola. We conclude with remarks on persisting challenges and opportunities, including in the context of the recent launch of the GPM (Huffman *et al*., 2013, Hou *et al*., 2008, 2013).

1.1. TMPA data

TMPA products are the main data source of the TRMM Explorer project. The TMPA algorithm generates near-global (50ºN-50ºS), 3-hourly rainfall estimates at a spatial resolution of 0.25x0.25º by combining the data from different sensors. Sensors employed include the TRMM Microwave Imager (TMI), Special Sensor Microwave/Imager (SSM/I), Advanced Microwave Scanning Radiometer-EOS (AMSR-E), Advanced Microwave Sounding Unit (AMSU-B) as well as sensors from the National Oceanic and Atmospheric Administration (NOAA) and the Defense Meteorological Satellite Program (DMSP).

Two versions are made available to users, a research grade (3B42V7) and a Near Real Time alternative (3B42V7_RT). The 3B42V7 is generated after additional steps are taken to calibrate the final output including the use of data from the TRMM Combined Instrument (TCI) and from a global rainfall gauge network (Global Precipitation Climatology Project and Climate Assessment and Monitoring Systems). Unlike the research grade product, which is distributed with a delay (*ca*. 2 months), the Real Time version is disseminated with a latency of 9 hours. The 3B42V7_RT is thus an important input to operational activities but because of the processing scheme used, it is outperformed in common validation metrics by the Research Grade equivalent (e.g. Dinku *et al*., 2010). Both versions are available at TOVAS, currently in version 7. We refer the reader to Huffman *et al*. (2007) for an in-depth description of the algorithm.

2. Web Service development

The TRMM Explorer Web Service (herein called the 'Web Service') was designed to provide SRE and supplementary data for specific geographic regions including watersheds or ecoregions, as requested by the users.

The Service was designed to include an online portal developed in Hypertext Markup Language (HTML), a cloud-based Python (Rossum *et al*., 2001, www.python.org) program that performs

the data-related tasks and generates the outputs and, finally, an online data storage tier (Figure 1).

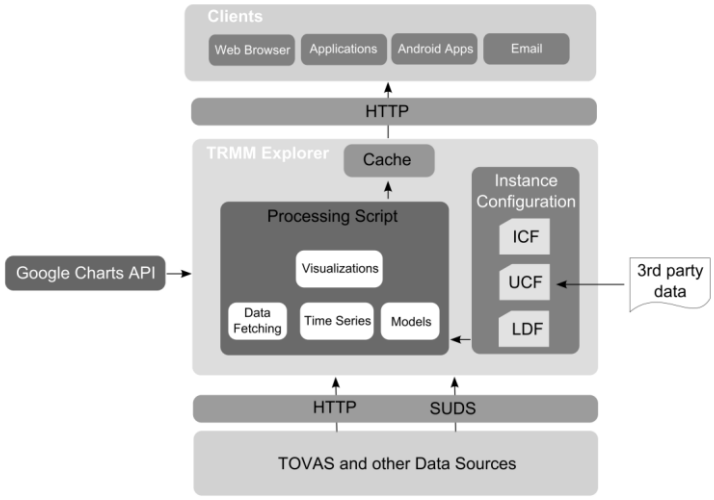


Figure 1. Simplified system architecture of the TRMM Explorer Web Service. At the bottom, TOVAS and other data sources, which are fetched by the processing script using different methods. The processing script calls the data according to instructions stored in the Instance Configuration File (ICF). The visualizations are created and stored in the cloud, or sent to the users via email. Users can access data using web browsers or applications, as displayed at the top.

Python as an object-oriented, extensible and high-level programming language is especially suitable for projects such as the TRMM Explorer. Numerous modules are available that simplify the overall development effort, making it a popular choice in scientific projects (Oliphant, 2007).

Development of the Python program was conducted using the Integrated DeveLopment Environment (IDLE).

The service uses Google Maps Application Programming Interface (API) and the Google Charts API to enhance the data visualizations and queries. Both APIs support a richer user experience while reducing development costs and hurdles.

Current and anticipated users include researchers from the broad geoscience community, educators and students from all levels interested in studying or monitoring SRE for a given region and where *in situ* data are unavailable, limited or unreliable. Technical staff from regional stakeholder organizations may also benefit from the capabilities offered by the TRMM Explorer. The project presents new options that relieve the strain put on the resources of small to medium organizations with specific SRE monitoring needs, namely in developing and

transition countries. The anticipated diverse user community will be able to rely on the flexible nature of the service, which supports different data interaction options and access levels. In particular, the program was designed to streamline the integration of new areas of interest (AOI) or different observation lengths without a substantial increase in the amount of resources consumed by the service or complex programming tasks.

In summary, the TRMM Explorer Web Service acts as data mediator, powering different visualizations and outputs capable of feeding downstream applications. The ultimate goal is to empower users, especially those with limited access to robust data processing facilities, to realize the full potential of SRE and other, ancillary products.

The service can a) download TOVAS data and reprocess them, b) generate dynamic plots, c) generate tabular and file outputs, support d) alert configuration and e) Customization and regionalization options (Table 1).

Table 1. Data outputs generated by the TRMM Explorer Web Service.

Type of Output	**Links to imagery**	**User contributions (UCF)**	**Cross-platform**
Table	No	Yes	Yes
CSV file	No	Yes	Yes
Plots	Yes	Yes	Yes
KML	Yes	Yes	Yes
Email	No	Yes	Yes

a. TOVAS data download and Instance configuration

Satellite Rainfall Estimates are a form of geospatial data dependent on location, time and observation length. These three variables are used at TOVAS for the retrieval and display of TMPA data. The location is given by a rectangular grid covering a number of 0.25x0.25º pixels that the user sets interactively. The time is given by a start date (no earlier than 1997/12/31) and an end date. The near real time product is distributed with a latency of 9 hours, in contrast with the approximately 2 months of the research grade equivalent. Finally, the data can be distributed at different observation lengths including 3-hourly, daily or monthly.

The TRMM Explorer uses this combination of location, time interval and observation length to construct an Instance. But unlike TOVAS, in which the user defines the key parameters manually in every session, the settings of a TRMM Explorer Instance are stored and retrieved

automatically. Once created, the Instances download and generate visualizations at specified hours without human intervention using job schedulers.

A single Python program (herein called 'processing script') downloads and processes the data but the variables (area of interest, observation length amongst others) are fetched from multiple Instance Configuration Files (ICF) (Table 2 and Figure 2). Therefore, an upgrade of the processing script becomes immediately available to all instances.

ICF	IName = "RCBS1" #Sets the Instance Name
	LDF = "PRT" #Sets the Instance Language using ISO 3166-1
	defOutput = "Plot" #Sets the default output
	Prod = "3B42RT"
	TGC = 4
	NGA = {1:-13.2,2:-13.2,3:-13.1,4:-13.1} #Grid Cells Coordinates, North
	SGA = {1:-13.4,2:-13.4,3:-13.2,4:-13.2} #Grid Cells Coordinates, South
	WGA = {1:13.8,2:14.1,3:13.3,4:13.6} #Grid Cells Coordinates, West
	EGA = {1:13.9,2:14.2,3:13.4,4:13.7} #Grid Cells Coordinates, East
	Gweight = "Empty"
	NumberDays = 15
	Email = "True" #Sets whether the Instance can send emails to users
	UDF = "URL" #Sets the approved UDF file path
	Thresholds = "Empty"
	Models = "Empty" #Sets whether the Instance uses Models (Decision Tree)
	ModelRules = "Empty" #Enumerates the Rules being used

UCF	A;2013;9;13;Landsat-7 Image;Click here B;2013;10;10;Damage Report;For a report click here

LDF	varA1 = "Precipitação" # Precipitation
	varA2 = "mm" # units, mm
	varA3 = "Grelha" # Grid

Figure 2. Three files are important to configure an instance: the Instance Configuration File (ICF) in Python, the User Contributions File (UCF), which is a text file containing HTML tags stored at third-party servers and the Language Definitions File (LDF) a Python file that stores the variable strings for different languages. The processing script imports the proper LDF following the instructions set in the ICF (variable 'LDF').

The ICF is itself a Python file, containing values for several different variables including the coordinates of the area of interest (AOI), list of products, start and end dates, observation length and links to external, ancillary data sources (User Contribution Files, UCF). Each line of the ICF sets the values for a specific variable. The first line of the ICF, for instance, defines the name of the instance (variable IName), while the second line sets the Language Definition File (LDF). Also important is the 'Prod' variable, which establishes the main product to be retrieved (e.g. for daily rainfall estimates in near real time, the variable is set to '3B42RT'). If the instance requires other products to be downloaded, additional 'Prod' variables are added to the ICF (i.e. 'Prod2', 'Prod3', etc…). These variables are used by the processing script to create the URL and download the SRE from TOVAS, process the values and generate the outputs. Because the data are aggregated after download of the individual cells, unlike TOVAS, the TE can generate

visualizations for an irregular area or moving target (Figure 3). It fetches the data from individual TOVAS cells and then combines them, generating an area-averaged output.

Table 2. Customization options available at the TRMM Explorer.

Configuration File	Description
Instance Configuration File (ICF)	(Python file) Sets all required parameters for instance deployment including area of interest, observation length, products, preferred visualization mode and path to User Contributions File.
User Contributions File (UCF)	(Text file) Stored at third-party servers from 'trusted users' enables the annotation of plots, tables and maps. Supports HTML tags and links to a variety of media formats.
Language Definitions File (LDF)	(Python file) Includes the translation of instance nomenclature, alarms, data quality flags and other relevant elements.
String Queries	Enable the customization of the data visualization through the use of simple strings in the URL.

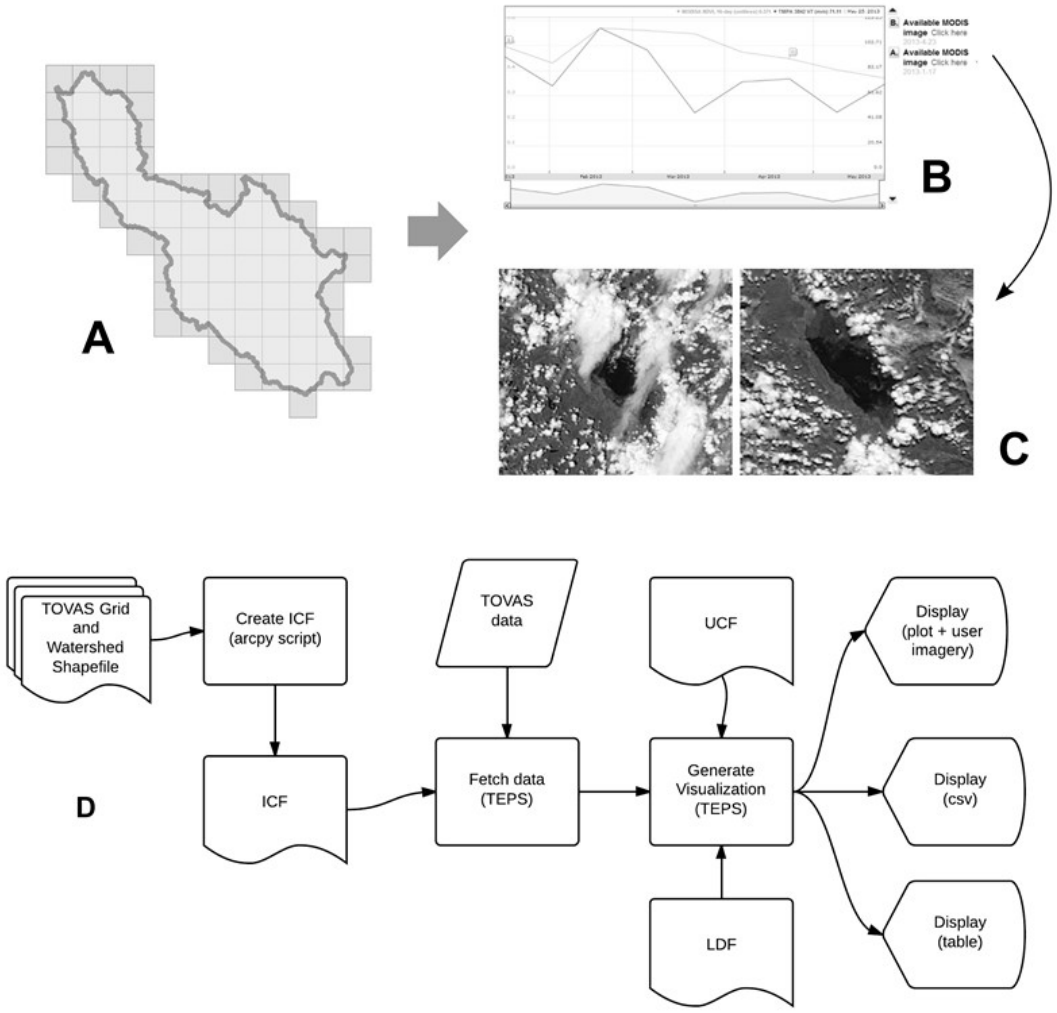


Figure 3. The development of new instances of the TRMM Explorer Web Service is simple. It starts with the identification of relevant TOVAS grid cells (0.25ºx0.25º) using an ArcGIS arcpy custom script which generates an ICF (A). The processing script (TEPS) then fetches the definitions from the ICF and constructs URLs to get the data from TOVAS and other sources. The visualization is built using further definitions (optional) from the UCF and LDF (B). Ancillary data such as imagery, reports, and photographs can be displayed if the user clicks on the annotations (C). The simplified workflow is presented in D.

The script can also aggregate the data to different observation lengths. Daily rainfall estimates can, for instance, be combined into 16-day averages. This aggregation can be useful, for instance, to users interested in comparing rainfall estimates with MODIS vegetation indices (e.g. MOD13Q1). In fact, the TRMM Explorer can download and display additional data from other sources including the USGS river gauge network and MODIS Global Subsets from the Oak Ridge National Laboratory (ORNL, 2011).

Another important feature of the service is the ability to implement simple models. The downloaded SRE from TOVAS can be used to feed simple linear or regression tree models and generate new products based on prior research by the users. The implementation of the models was validated successfully in a test designed to implement automated adjustment of the rainfall estimates using regional calibration equations derived from the comparison of the SRE with data from rain gauges. The models are described in greater detail in later sections, namely as a possible method to configure alerts.

The Instances are created upon submission and approval of a simple, 2-page Instance Proposal (IP) which must describe the objectives and requirements sought by the users. The user must indicate the location of the area of interest (or submit with the IP an ESRI Shapefile (.shp) or raster file defining the area), list of products, observation length and latency. The ICF can be built manually or by a custom, semi-automated arcpy script (preferred method) which uses the data submitted by the user to identify the relevant 0.25ºx0.25º cells. This process is performed by the technical team of the TRMM Explorer project. After the ICF is configured, no other intervention is required to set the area of study.

Upon approval of the IP, an ICF is created by the TRMM Explorer team and the data begins to be regularly processed by the system. This data policy favors a rational use of resources and opens new venues for tailored user-support, feedback and the development of customized and rich experiences.

b. Dynamic Plots

The visualization of long time series namely of datasets with marked seasonality can be difficult when displayed in static plots. It is also impossible to query the data from static plots in conventional image file formats such as JPEG or PNG.

The TRMM Explorer uses the Google Charts API to deliver SRE in a dynamic format that can be queried while allowing up to 2 different variables to be displayed simultaneously. Users can zoom in and out of the time series, get the values for each data point and navigate to external data sources by clicking on annotations (Figure 3 and 4). The chart is created by the processing script, using the definitions of the ICF and data from TOVAS (and eventually ancillary data sources). The program generates an HTML file (containing the plot) that is stored in the server. The HTML file is embedded in the instance page at the TRMM Explorer portal, which is displayed when a user requests it (Figure 5).

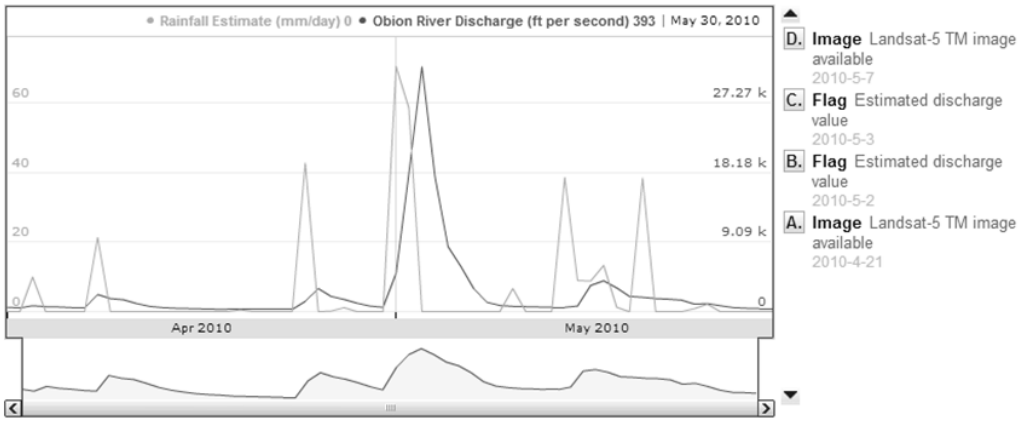


Figure 4. In this example, 3B42V7 data from TOVAS are displayed together with USGS river discharge gauge data (gauge ID 07024500, Obion River, near Greenfield, TN). The archived instance provides historical data on a record rainfall event in Tennessee, USA from May 2010, for reference and education purposes. Quality flags and Landsat imagery are also accessible. The data are also provided in csv and table formats at the portal.

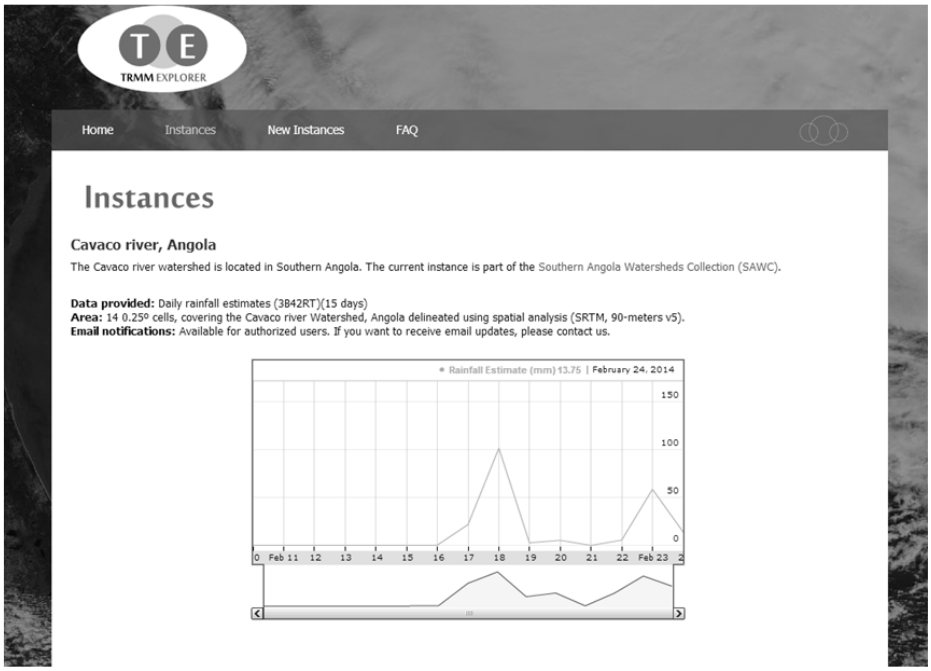


Figure 5. An instance displaying 15 days of rainfall estimates for an Angolan watershed (SAWC collection) displayed at the TRMM Explorer project portal. The dynamic plot generated by the processing script is updated on a regular basis and stored in the cloud.

c. Tabular and file outputs

The inter-connectivity of Web Services is a cornerstone of the modern internet (Evangilidis *et al*., 2014, Díaz *et al*., 2012, O'Reilly, 2007). To support the use of TRMM Explorer data by applications downstream, the processing script can create tabular outputs or generate files with the same data used in the dynamic plots.

The Service can create tables in HTML, Extensible Markup Language (XML), comma separated values (CSV) file or Keyhole Markup Language (KML) file without additional changes to the Python processing program or ICF (Table 1). The default output is the Dynamic Plot but the user can call a Uniform Resource Locator (URL) containing a query string that sets the desired output. The URL follows a generic format of the type: 'http://www.trmmexplorer.org/?instancename&display=CSV'. The query string portion of the URL ("display=CSV") passes the data to the processing script that interprets it and generates the selected visualization. The URL (containing the query string) can be manually inserted in the address bar of the browser or called from an application automatically. When the user interacts with the instance using the browser, the "display=CSV" query string triggers the automatic download of the csv file. Applications can also retrieve the file using the same URL, but the method to do so depends on the language in which the software is programmed. The csv file is, in the context of the TRMM Explorer, the preferred method for data sharing and intake by downstream applications.

At the TRMM Explorer website, a simple example of a web application developed in python for Google App Engine (https://cloud.google.com/appengine/docs), demonstrates the concept ("TE_HelloWorld"). The application calls an instance using the urllib2/urllib modules and downloads the csv file before parsing it. These data can then be analyzed by the third party, downstream software.

On the other hand, the 'display=table' option calls an HTML file containing the data arranged in a table. The data can be visualized in a browser or downloaded from an application for other uses, including display, parsing and analysis. Other query strings include the plot visualization (display=gv) and xml/kml file (display=kml).

Each instance is documented at a specific webpage, which is available at the website. At the page, users can find which query strings are enabled for the dataset along with instructions regarding their use.

The multiple options maximize the potential of the TRMM Explorer as a data mediator.

d. Alert configuration

The continuous monitoring of near real time SRE is only possible if the data are downloaded, analyzed and the results distributed automatically to the end users.

The TRMM Explorer instances can be configured to compare the SRE to predefined rainfall thresholds. If the conditions set in a pre-defined decision tree are met, then the system adapts the visualizations using specific color codes and annotations or sends emails to a list of pre-approved users. The decision tree must be submitted by the user upon request of a new instance. The trees consist in a set of mutually exclusive conditions that are tested on one or multiple variables, all of which must be available within the instance. The outcome of each test determines whether the program tests another condition or initiates an activity.

The tree may have a single level, and evaluate whether the rainfall estimates fall below a certain threshold or evaluate complex conditions including, for instance, different products (multi-level decision tree). The 'ModelRules' field in the ICF (Figure 2) stores the information that must be evaluated. The "ModelRules" follow a simple structure ("if…then" statement) and could be written as: "ModelRules = if (estimate<=5): / flagModel = "L" /elif(estimate > 5) & (estimate <=10): / flagModel = "M" / else: / flagModel = "H"". In this case, the instance compares the rainfall estimate against three intervals (<=5; >5 and <=10, > 10) and assigns a value to the string variable named "flagModel", which is used by the processing script to store special warnings.

The 'Thresholds' option in the ICF simplifies this process by setting a single value against which the rainfall estimates are compared. Future work may lead to the development of seasonal thresholds, configurable directly in the ICF in a similar way.

e. Customization and Regionalization

The TRMM Explorer Web Service is focused on the regional analysis of SRE. As such, the users may be interested in adapting the interfaces to local languages or share additional information with the end-users.

The User Contributions File (UCF) is a step towards that objective. The UCF, a text file containing HTML tags, can be hosted at third-party servers maintained by 'trusted users'. The UCF supports 6 comma-separated fields including an identifier, the date (year;month;day), a text string and an hyperlink if available (Figure 2b). In the UCF, users can store ancillary data (one per line of the UCF) that can be displayed together with the core instance data at the specified dates. The ancillary data (videos, links to satellite imagery and reports, photographic

records) can be displayed as annotations or in separate tables attached to the visualization. An experimental variant of the UCF enables users to set the coordinates of the grid cells from which the data should be retrieved. This can be particularly useful to track moving targets. The user can store the coordinates of the target feature into the UCF, which is read by the processing script.

Finally, the entire experience can be adapted to local languages. The instances can be translated/localized using an optional Language Definitions File (LDF) (Figure 2c). Not only can the GUI elements be translated but so can the data quality flags and alerts. The LDF sets the values for a number of string variables that are used by the processing script to create the visualizations (excluding variables). If no LDF is set (LDF="empty"), or if the LDF does not contain the translation of all available variables, the processing script uses the default values. Figure 2C shows an example of a subset of an LDF file, containing three variables (e.g. varA3 is set to 'Grelha', the Portuguese translation of the word 'Grid').

3. Android App development

3.1. TRMM.Mobile development

Android is an operating system based on a Linux kernel (Android developers, http://developer.android.com) that is available on smartphones and tablets of most major manufacturers. Android, a Google product, currently holds a dominant market position and is expected to consolidate that advantage in coming years (IDC, 2013).

The mobile TRMM Explorer application, named TRMM.Mobile, was developed in Java (Java, http://www.java.com) using Eclipse (www.eclipse.org), an IDE that can be used together with the Android Development Tools and Android Software Development Kit (SDK) distributed by Google to simplify the development of new Android applications.

The app was tested in emulators and real smartphones from different brands, to improve the user experience and minimize compatibility issues across devices.

The TRMM.Mobile takes advantage of built-in capabilities available in most devices, to expedite the access to TOVAS from a simplified Graphical User Interface (GUI). The GUI simplification allows circumventing the constraints inherent to mobile devices including the smaller screen size, limited processor power, memory and battery life (Weng *et al*., 2012).

Amongst the most immediate advantages of SRE is the ability to obtain timely data that can be used for multiple purposes, such as flood prediction. Timely access to data requires adequate data distribution solutions that can take advantage of today's ubiquitous mobile platforms.

The application is able to a) Set the variables for TOVAS data download and b) store key variables and data for later use.

a) SRE data download and display

To access TOVAS data it is necessary to provide the coordinates of the AOI, the start and end date and the product to be downloaded as aforementioned. Unlike the TRMM Explorer Web Service, the Android App is still unable to download data from irregular areas directly from TOVAS.

The user can take advantage of the integrated GPS often found in smartphones and set the coordinates for the AOI automatically reducing the need to type long strings (Figure 6 and 7). If the user is interested in a different region, it is also possible to manually insert the bounding coordinates of the AOI box (North, South, East, West). Future versions will support the selection of pre-defined areas (e.g. watersheds or ecoregions) and geocoding to simplify site selection and enable the communication with existing TRMM Explorer Instances.

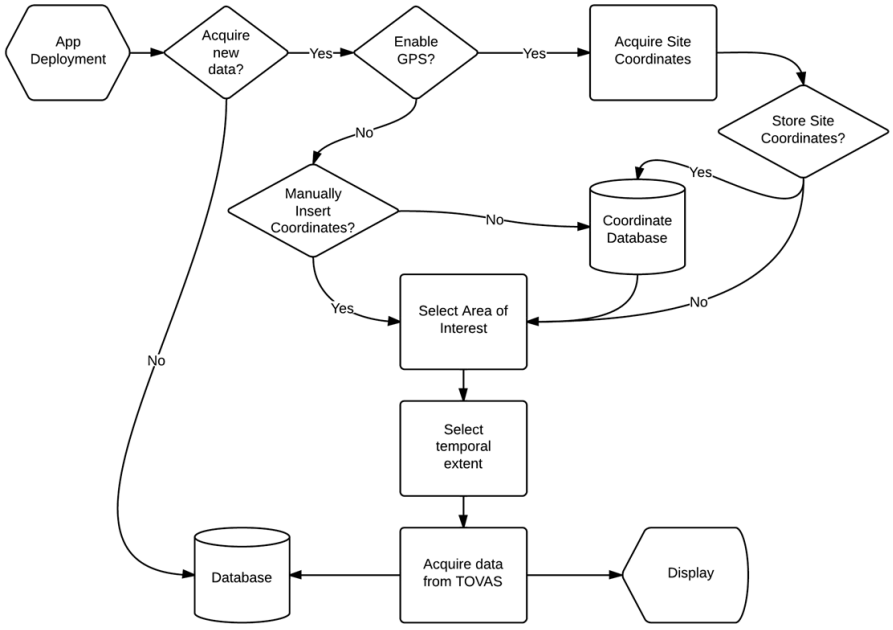


Figure 6. Simplified flowchart depicting the operation of the TRMM.Mobile smartphone application.

The user can then proceed to select the start and end dates from a 'Time Picker' dialog using an Android 'Fragment' (Figure 7).

Finally, the user can select the product (Research Grade or Near Real Time) and the application will fetch the data by constructing a URL and calling it. The app then downloads an ASCII file

from TOVAS that contains the values for the selected combination of variables. The data thus downloaded is displayed in a tabular format to simplify the analysis of the values.

Figure 7. The initial data selection screen of the TRMM.Mobile smartphone application. The coordinates may be inserted manually, if the user deactivates the GPS.

b) Data and variable storage

The simplification of smartphone app interfaces is a recurrent topic in papers, textbooks and Android documentation. To minimize the need for manual data inputs; i.e., coordinates, site name or dates, the key variables can be stored in an internal storage file for later use. The data output from TOVAS can also be stored but using the External Storage. In this case the user can export the data into a text document, which can be saved to a removable storage media (e.g. SD card).

3.2. User Development Kit (UDK)

The abundance of simple mobile app programming solutions can transform almost anyone into a developer (e.g. Turbak, 2012). There are also undeniable advantages in the use of TOVAS-like data in educational contexts, especially in secondary and undergrad levels (Passow, 2010). At the same time, in the classroom environment, it is often adequate to use simpler programming languages or even a visual approach to the development of applications. As such, the use of

MIT's App Inventor (http://appinventor.mit.edu) to create a flexible, scalable template for the distribution of TRMM Explorer data were also studied and a project template created, named the "TRMM Explorer User Development Kit" (UDK),.

The MIT's App Inventor (AI) is an online development environment that uses building blocks that mimic traditional code snippets. The UDK is not a functioning Android App but a downloadable project file (.aia) containing a scalable project. The project includes the necessary code blocks to establish communication with the TRMM Explorer Instances, unlike the TRMM.Mobile solution, which fetches data from NASA's TOVAS website. The user can then change the blocks as needed and build the Android Application Package (.apk file) using MIT's App Inventor and run it on any supported device.

The User Development Kit empowers users, even those who are not programmers, to develop highly customized applications using data processed by the TRMM Explorer Instances.

Unlike the Android app described in the previous section (TRMM.Mobile), the UDK is a template for MIT's App Inventor that was designed to access the Tabular and CSV outputs of the TRMM Explorer Web Service and not TOVAS.

The template includes files that once uploaded to the MIT's App Inventor Development Environment, can be modified by the user. Together with the project files, the users can access, at the TE portal, tutorials on how to adapt the template to specific objectives. Although the UDK is especially suitable for classroom projects, it is an innovative way to engage the user community. Still, this approach does not invalidate the traditional approach to app development in a conventional IDE such as Eclipse.

Abundant information in the form of official MIT's App Inventor documentation and user-made tutorials are also available.

The UDK can a) access data from TE Instances and b) parse and analyze the data.

a) Data access

The UDK uses a "Web.Get" block to call the URL of a specific instance. The URL must contain the name of the instance and query string that specifies the output format (CSV). The query string follows a common structure for all instances as aforementioned and defines a set of variables that are passed to the server that, in turn, returns the appropriate response. In this case, the UDK requires the data to be served in a CSV file. The structure of the string query is thus in the form: ?instancename&display=CSV, which is added to the web address of the website (i.e. www.trmmexplorer.org/?instancename&display=CSV).

The data can then be stored as a string or as a list and analyzed or displayed using text labels. The UDK is set to call a pre-defined URL (which can be changed) associated with a static instance. This instance (Obion river, USA) includes data from an extreme precipitation event that took place in Tennessee, United States in 2010. Since this instance is not updated, users can test their application with a stable dataset, allowing for the reproduction of results, even in different dates.

A traditional Java Android app created with an IDE (e.g. Eclipse) can access the TE Instance data in the same way. UDK users can, nonetheless, bypass the steep learning curve associated with programming Android apps.

b) Data Analysis

The optimization of the GUI is an important aspect of scientific computing. The screen resolution and size of typical smartphones further adds to the need of designing high-quality interfaces.

The parsed CSV file downloaded from the TRMM Explorer and stored as a list can be split and the values compared against pre-defined thresholds, set by the user. If certain conditions are met, the application can show an alert icon or change the font color and size thus alerting the user to the relevance of the values.

Using other code blocks, the user can create new, innovative ways of interacting with TRMM Explorer rainfall data. Users may thus create highly customized applications independently, without straining the resources of the TRMM Explorer Project.

4. Case Study - Southern Angola Watersheds Collection (SAWC)

It is difficult to capture the inherent variability of precipitation in both space and time without a capable monitoring network in place. Numerous watersheds worldwide where available *in situ* data are limited or even inexistent have, nonetheless, a significant human population (Alcántara-Ayala, 2002).

A considerable number of these ungauged watersheds are located in developing or transition countries, where the lack of data is the result of an endemic shortage of scientific resources, both human and material. During extreme weather events, existing monitoring networks may also fail, leaving managers and responders with insufficient data on which to act (Chen *et al*., 2013).

Building on work by the authors and respective organizations, a collection of TRMM Explorer Instances was created for several neighboring river basins in Southern Angola (river Cavaco, Bentiaba and Curoca) (Figure 8). The Southern Angola Watersheds Collection (SAWC) was designed to validate the concept of an instance-based software, driven by user needs and capable of cross-platform data delivery. These watersheds were selected because despite recent studies (e.g. Dinis *et al*., 2013) the body of scientific information available is still limited and the collection of *in situ* data next to inexistent.

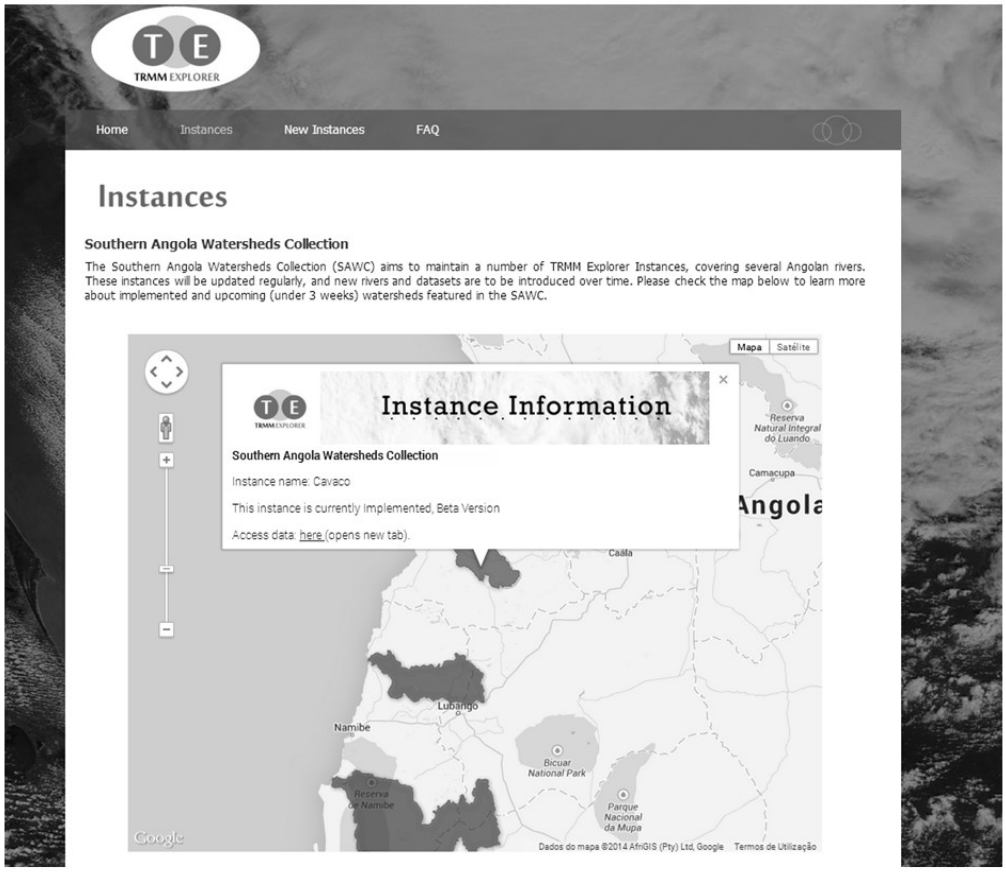


Figure 8. The Southern Angola Watersheds Collection data access map at the TRMM Explorer. The map takes advantage of the Google Maps API to display information on the location of the selected river basins and enable immediate access to the instances and ancillary information.

The rivers are located in a region dominated by a sharp contrast between the dry and wet seasons and marked by frequent episodes of destructive flooding. According to Koppen's classification, the region is characterized by a climate of the types: BSk, BSh, BWh, BWk and Cwb (Kottek *et al*., 2006) with aridity increasing towards the coast and the south. A significant amount of the total rainfall occurs at the head of the basins, but the most significant population centers are found on the coast (e.g. Benguela and Namibe). This fact contributes to

an uneven distribution of rainfall and hydrologic gauges across the basins, often biased towards the urban centers, and thus far from the main water sources.

To address these limitations, the SAWC was created to collect and distribute satellite rainfall estimates in a systematic and continuous manner for these regions in near-real time.

The development of an instance starts with the definition of the Area of Interest (AOI) and the selection of the list of products. These variables are stored in the ICF for later use by the processing script.

The river basins were delineated using 90-meter Shuttle Radar Topography Mission (SRTM) elevation data acquired from the EROS center of the United States Geological Survey and manually corrected using ancillary data sources.

Using a customized script designed to generate the Instance Configuration Files (ICF) of the TRMM Explorer at ArcGIS (arcpy), the relevant TOVAS grid cells (irregular grid) were selected by intersecting the basin extents with the 0.25 x 0.25º TOVAS grid.

The 3B42V7_RT product was selected, to allow for the acquisition of near real time data. With a latency of 24 hours (set in the ICF as well), the data are automatically downloaded from TOVAS, using a job scheduler that calls the processing script at a predefined hour. The data thus generated are stored and displayed in different formats (dynamic plot, tabular, CSV), which can be accessed by calling the URL containing the proper query string. Email alerts and updates are also available for a list of selected users, which includes researchers working in the area.

After implementing the instance code, the TRMM Explorer portal was updated to display the new instances. A map depicting the basins, created using the Google Maps API, is the entry point for the users. A link in the watershed's information balloon redirects the user to a page containing a brief technical description of the instance and a dynamic plot with the rainfall estimates for the last 16 days (which can be extended if necessary through a specific query string).

New instances are typically 'private' for a variable period of time after implementation (inferior to 6 months) and accessible only to the users submitting the Instance Proposal (IP). After that initial Quality Control (QC) period, the information becomes publicly available to all users. When the instances are being used to test or implement unpublished models, the private period may be extended until the publication of results in order to safeguard authorship.

Because the SAWC collection of instances was developed in response to existing needs identified by research teams studying the area, the basin-wide SRE are contributing to accelerate research activities. The data now available can be used to create and test rainfall-

runoff models and existing models, especially if based in non-parametric regression tree models, such as the M5P (Quinlan, 1992). Recent research demonstrated the feasibility of modelling surface runoff and river flow using M5P models fed by rainfall values and assimilated gauge data (e.g. Stravs et al., 2008, Iorgulescu et al., 2004). Such models, because they are a simple collection of conditional statements, can be easily translated into a python-based processing program. This is an important research topic that future versions of the TRMM Explorer, or other similar programs, should consider, namely when selecting the data inputs and output formats.

The TRMM Explorer does not currently support models of this level of complexity. However the data generated by the program (rainfall estimates with different observation lengths) is made available in convenient formats (e.g. csv), which can be used to feed other applications.

Rainfall thresholds can also be set for routine flood surveillance in preparation for a future EWS.

Users can also take advantage of the User Development Kit in order to receive the data in an Android device in a fully customized GUI.

The SAWC helped validate the concept driving the development of the TRMM Explorer. The implementation of a new TRMM Explorer instance (or collection of instances) is simple because a single processing script and one configuration file per instance suffices to create new opportunities for data exploration using different platforms. Deployment of the SAWC and feedback from users and partners contributed to identify existing problems and limitations of the system. In particular, users asked for the introduction of forecasting tools and the integration of other products such as vegetation indices and optical imagery.

5. Conclusion

Free and easy access to data can contribute to the rapid advancement of science (Wulder *et al*., 2012). Still, the analysis of long time series can put strain on the resources of small to medium organizations (e.g. de Beurs and Henebry, 2004, Yue *et al*., 2013, Cook *et al*., 2006). Furthermore, data portals are often designed to meet the requirements of desktop users (e.g. Houska, 2012, Houska *et al*., 2012) and TOVAS is not an exception. Ergo, many users can still experience difficulties in translating the free data policy into operational advantages and convert SRE into actionable information in a timely manner. Off-the-shelf solutions are thus important to users with technical or budget constraints, which limit the resources available to develop applications from the ground up.

The TRMM Explorer addresses these problems successfully by providing a suite of cross-platform solutions for immediate data visualization or to feed other applications downstream as described.

Users from the geoscience community, in a broad sense, can take advantage of the Web Service and mobile application to rapidly access and explore data, monitor rainfall in preparation of dedicated EWS or compare SRE to other datasets for specific regions and observation lengths. Visualizations of relevant historical events can be created easily for educational purposes and models implemented through the configuration of the ICF.

The free, scalable and low-maintenance Web Service is an adequate solution for long-term monitoring efforts of wide areas by small organizations without the constraints inherent to other EO data portals. The Service can also serve as a test bed for users to test and prove concepts and design new applications. The TRMM Explorer offers new options to data analysis that were not available and because the maintenance and upgrade of the processing script is simple, the Web Service can be continuously improved, in response to user feedback and identified opportunities.

The service can be provided with no costs to the user, because of the simple methodology used to integrate new areas of study or generate different visualizations and the low-maintenance requirements. The adopted data policy, which requires the submission of a formal proposal and a subsequent regular evaluation of the active instances also contributes decisively to the sustainability of the project. The policy actively prevents AOI overlaps, optimizes resource usage and website traffic and contributes to identify the scientific and societal impact of the instance.

The integration of new products and *in situ* data sources as well as the introduction of new tools for time series analysis are planned for the future. The improvements will contribute to a sustainable growth of the user community in response to new science opportunities.

The ultimate objective of the TRMM Explorer is to give users simple access to SRE data in all steps of a research project from data collection and regional characterization, to time series construction and analysis, near real time monitoring, model implementation and outreach activities. The launch of the GPM Core Observatory will also contribute to a continued interest in SRE research and applications, which the TRMM Explorer Project cannot ignore and must integrate in future work.

6.2. The SmartBasins initiative

Data distribution is the logical and ultimate step of any research project aiming at developing new datasets that can be adopted by stakeholders. This principle applies to the SmartBasins initiative, a portal designed to provide data (geospatial and other) on watersheds and coastal systems.

The platform adopts the design and data distribution principles and solutions developed for the pathfinder TRMM Explorer project but is driven by a more robust and versatile engine.

SmartBasins is not just a data access point but an encompassing, cross-platform information system capable of providing near real-time data on multiple variables derived from multi-mission data.

The development of the system is driven by a series of requirements including the ability to run automated analysis on data (UNEP, 2012) using web-based solutions for data analysis (Prodran et al., 2013). The outputs of the service must be flexible and enable cross-platform visualization and analysis, from desktop and mobile systems, as increasingly demanded by users (Weng et al., 2012).

Sherman et al. (2009) highlighted the importance of analyzing coastal waters together with adjacent watersheds. This principle was adopted in the establishment of the Large Marine Ecosystems with practical implications, namely in the assessment of nitrogen fluxes into the coastal environment (Sherman et al., 2004).

However the 64 LMEs are naturally a global dataset in nature and function, lacking the level of detail required for the development or enforcement of regional or local management policies and monitoring strategies. Portuguese (continental) waters, for instance, are assigned to a single LME (25: Iberian Coastal), but numerous (some relevant) watersheds drain to this area. These watersheds are administratively assigned to groups based on their geographic area and characteristics. Although it simplifies administrative procedures and the overall management, this system still presents several inconveniences, including the sheer size of the clusters, detachment of the population, and the lack of a unifying agency or council capable of developing and enforcing adequate watershed-focused management strategies.

As aforementioned, namely when describing the Imperviousness Surface Area products, we stand for an integrated administrative-ecological approach to the management of the territory. Watersheds are the structural unit of land masses but they are equally the drivers of several processes in the coastal sea (and beyond). The Western Iberia Buoyant Plume is an interesting example of such an influence. The freshwater lens alters the marine medium, with profound ecological impacts that were abundantly described earlier.

The plume's signature is also revealed by the regionalization scheme put forth in this study, clearly separating the regions where the rivers play an important, albeit seasonal role from others where other processes take center stage.

Although the WIBP becomes the combined effect of the action of multiple rivers, regional and local processes and features (e.g. thermohaline fronts) are clearly influenced by individual watersheds.

As this was one of the conclusions of our study of the distribution of thermal fronts in the coastal sea off Portugal, the development of a data distribution portal based on these concepts was deemed the most adequate solution.

Furthermore, several methodologies and data products generated in the course of this study are relevant to a wide range of scientific and operational applications. Releasing the data and exploring ways of promoting its integration in 'real-world' applications seemed the most reasonable approach. As mentioned by Breaker et al. (2005), the "lack of marketing" of frontal maps in California was a major obstacle to the adoption of this indicator across a range of fields. Since potential users were unaware of the existence of such datasets, these were not fully exploited compromising the sustainability of such studies in the future.

The SmartBasins initiative builds on the lessons learnt from the development of the earlier TRMM Explorer and applies the concepts explored throughout the project, especially that of the inherent connectivity of land and coastal systems.

Two main requirements were defined, including the need to establish a simple access point, driven by a robust processing chain; and the focus of the portal is information rather than raw data, while also offering the latter to users who wish it.

These guidelines were also adopted in other projects, most notably in response to growing demands for straightforward data access points (Liu et al., 2012).

Several examples exist of data distribution systems focused on the coastal (including land) and atmospheric variables. However a common aspect of most (if not all) platforms is the lack of interconnectivity between spatial domains and different variables. NASA's Giovanni (Berrick et al., 2009), especially after Version 4 was released, does offer simple (and limited) sample correlation analysis options. Still, it is presently still focused on data dissemination and quick visualization, often of downscaled data instead of the original higher resolution sets.

The Oregon Coastal Atlas (Oregon, 2016) and the Portuguese SNIRH (SNIRH, 2016), are interesting - although quite different - examples of online coastal information systems. Both provide information to both the public and specialized audiences in multiple formats. The SNIRH in particular is an important and rare source of open data, most commendable and invaluable for scientific research and environmental awareness alike. The system is,

nonetheless, quite focused on *in situ* assets, lacking much of the information that could be derived from alternate data sources such as satellites. In this sense, SmartBasins could provide an interesting additional data source from which users could access information unavailable elsewhere.

These websites are also focused on large areas (a country and a state), making it perhaps overwhelming for non-specialized publics interested in local data alone. This is perhaps even better exemplified by CORIOLIS (Coriolis, 2016) or Copernicus Marine environment Monitoring Services, platforms specifically designed for the distribution of complex oceanographic scientific data available in massive amounts.

The need for purely scientific data distribution systems will hardly subside. Yet, the underlying design concepts could alter, shifting to usability and multi-mission/instrument paradigms. It would also be useful for the portals to be able to dynamically adapt, or recognize, specific spatial domains, limiting the need for complex queries.

The Chesapeake Bay Program's website (CBP, 2016) is a good example of a data portal focused on a small, well defined region. This is a class of websites that is opposite to the previously described examples. The website provides information on a range of environmental issues and the data collected I support of environmental monitoring efforts of a single, well defined system.

Specialization can also occur in terms of topics addressed. That is an alternative approach that may be useful to specific user communities. The Indian Potential Fishing Zones (INCOIS, 2016) advisories are provided to fishermen across the Indian coastline upon the analysis of Earth Observation data.

The system is renowned and used as a template for similar system elsewhere, due to its originality and usefulness. However, the portal is hardly an example in terms of design or employment of state-of-the-art webmapping tools. Although this may not be a limitation *per se* it can, nonetheless limit the effectiveness of the portal and hamper the accomplishment of the objective for which it was designed. In a world increasingly connected, it is necessary to explore emergent technologies, not for the sake of fashion but of user friendliness.

The SmartBasins portal was developed in a two-tier approach, as it shall be described later. The processing chain fed a database, which was then queried to generate the visualizations, akin to those offered by TRMM Explorer.

The more advanced portal is characterized by significantly improved processing chains capable of acquiring, processing, and displaying a wider range of data. It also guarantees continuity, as the TRMM mission drew to an end, calling for alternate data sources to be identified and leveraged.

In fact, the new core rainfall product is provided by the IMERG suite. This product is complemented by MODIS (SST, Chlorophyll-a concentration, Thermal Fronts), Landsat (land cover layers), and modelling data (DTW).

Considering that the core functions and workflows are based on the programs created for TRMM explorer we shall not delve, once again, into the details of the data retrieval and analysis. Yet, in several aspects, both services differ in key aspects, which warrant further attention.

The SmartBasins service is constituted by two segments as aforementioned: data processing and data distribution. Both segments are developed to serve specific, and different, objectives, thus requiring equally dissimilar technologies to power each element.

Data from different sensors are made available by agencies in different processing levels and formats. Considering the multi-sensor nature of the solutions presented, it is necessary to design a processing chain capable of handling such variety of data sources. This is a key difference from TRMM Explorer, which relied on a much more reduced set of missions to retrieve and provide data to the public.

To address this challenge, a Python-based workflow (Rossum et al.,2001) was designed and programmed by the author, and several enhancements were introduced, when compared with the original TRMM Explorer Processing Script (TEPS). The processing chain is capable of search, download, and process the data over pre-defined regions resorting to the Geospatial Data Abstraction Library 2.1.0 (GDAL) (GDAL, 2016) as the core package powering the workflow. The previously used grid cells, matching TRMM pixels were discontinued in the current iteration. The Instance Configuration Files (ICF) were kept, but instead of containing the location of the cells, these fields were replaced with a path to a shapefile, which delineates the area of interest. Entire watersheds (polygon), buoy locations (point), or river segments (lines) for instance may be used as regions of interest from which data are to be extracted and analyzed. This represents an important evolution from the earlier TRMM Explorer made possible by profound changes in the code, which was made lighter and computationally more efficient. A unique ICF is capable of storing the settings of each target area independently of the data source from which information is to be retrieved.

Figure 6.2.1 describes the overall architecture of the system, which retains certain similarities with TRMM Explorer.

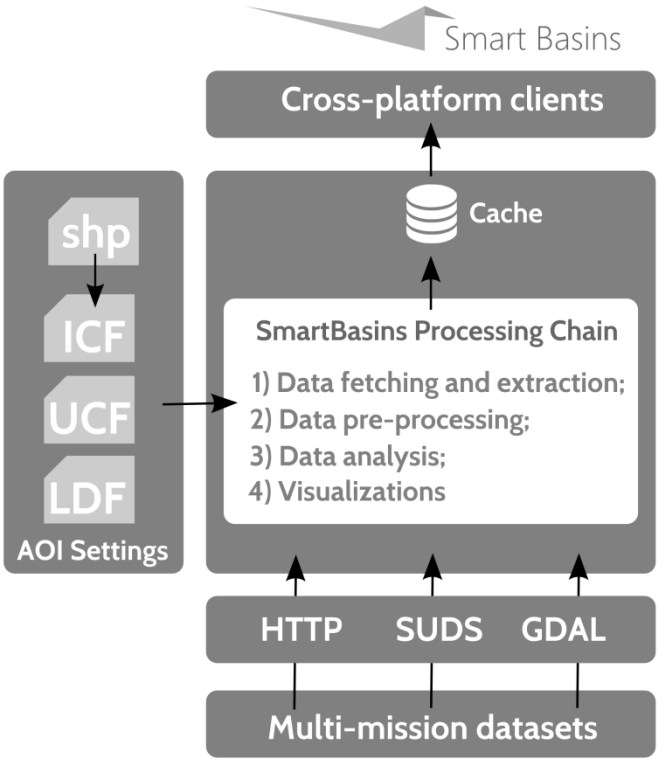


Figure 6.2.1. Simplified architecture of SmartBasins. The TRMM Explorer was used as testbed for the development of the platform, while most algorithms employed in this study were developed in Python to enable the subsequent integration in the platform as needed.

The development focus at this stage was centered on the development of the processing chains, with limited investment in the user interface or outputs (figure 7.2.1). The data generated by the system can be exported seamlessly into different formats, which can then be represented in cartographic, tabular or graphical formats, depending on the user needs it seeks to serve.

The experience acquired during the development of the TRMM Explorer was critical to guarantee a smoother development of the SmartBasins processing chain. A significant part of the code used in the development of the processing chains of SmartBasins leverages code developed for its pilot-predecessor. The stepwise approach favored an interative development effort, supported by ample feedback either during the submission of scientific papers, in scientific meetings (e.g. ESA Living Planet 2013 (Mantas et al., 2013b) and IV Meteorological Symposium in Lima, Peru), or directly from users.

To provide a seamless access to the data by users, a number of different Content Management Systems (CMS) were tested. Of these, Joomla was selected as the most efficient and adequate to the objectives of the initiative.

Data distribution is organized into watershed-level and coastal. In both cases, preference is given to placemark-driven data distribution, providing users without significant expertise with a simplified way to interact with geospatial data. To accomplish this task, simplified maps were made available using the FocalPoint extension to Joomla. Static pages containing predefined fields are generated for each area of interest. The processing chain is capable of digesting the data into meaningful values that can be published online through File Transfer Protocol (FTP). Data download is also possible for those users who wish to analyze the datasets offline, for instance within a Geographic Information System.

The User Configuration Files (UCF) were also retained in the new version. These files are stored in a third party server and called from the SmartBasins portal. Furthermore, page curators may also access and edit selected pages, fostering the development of a true community of watershed and coastal scientists and stakeholders.

The Language Definition Files, on the other hand, were only partially retained. The files are now used exclusively by the processing chain, to generate language-specific outputs including tables and reports. The portal is designed in a way that users can access contents in different languages, benefiting of embedded translations of the interfaces. Specific content, unique to SmartBasins can also be customized, and the same content displayed in different languages according to preferences set by the user. This offers greater flexibility than previous solutions, especially as the project is scaled to cover increasingly diverse regions.

Figure 6.2.2. The SmartBasins website contains both the numeric outputs already available at the TRMM Explorer and new spatial features, described in detail in specialized pages. In the example, the imperviousness changes (2001-2013) are made available to users, along with a technical description of the methodology.

We must emphasize the relevance of adding contents in a format and language familiar to the users, under the penalty of developing a data portal stakeholders cannot use. Figure 6.2.2

depicts such a case, where a specific land cover product is characterized in Portuguese, considering that it covers a region in central Portugal.

In the near future, the SmartBasins portal shall include WMS functionality to enable direct and seamless data distribution and integration with industry-standard geospatial software packages.

Finally, summarized reports and alerts can also be emailed directly to relevant user groups, providing near real-time actionable information on a variety of critical variables. These capabilities supplement the existing Android “App” and “User Development Kits” first developed in the context of the TRMM Explorer project.

As aforementioned, user interfaces were not at the heart of development efforts thus far and more research and development is required prior to a public release of the entire portal.

The TRMM Explorer website (www.trmmexplorer.org) is still online and serves as a repository of information and legacy datasets. With the introduction of IMERG datasets and other innovations in the processing chain and data distribution capabilities of the program behind the visualizations, the TRMM Explorer ceased to be updated or upgraded.

Development efforts were definitely moved to the SmartBasins platform as of late 2015, enabling a rapid advance and integration of the different scripts created and described in this study.

Currently, the processing chain is fully operational and the outputs validated, but the data visualization was not yet published. This strategy was adopted (as it had been earlier with TRMM Explorer) to prevent the full release of the portal prior to publication.

Currently, the portal release (with data functionality) is expected to take place before April 2017.

Still, the SmartBasins portal has also been used successfully as a focal point for environmental data distribution and visualization experiments, even prior to the full release of the visualization options. Amongst these we can highlight the Story Maps projects that provided information to users on international projects of recognized scientific and educational relevance.

The GLOBE ENSO Student Research Campaign was the first such initiative (figure 6.2.3), prompting students and educators worldwide to contribute with stories, photos, videos, and other media related to their view of the then ongoing El Niño event. The project, suggested to GLOBE officials by the author, has been an important experiment on the sharing of scientific data and personal experiences in near real-time.

The El Niño Story Maps (http://www.smartbasins.com/storymaps/) have a profound connection to the overall theme of this work. This was an attempt to reconcile different scales

and data sources (including people's experiences) into integrated visualization solutions that cover both land and sea as a natural continuum.

Figure 6.2.3. The GLOBE ENSO Student Research Campaign page describing the Story Maps project (available online at: https://www.globe.gov/web/el-nino/el-nino-campaign/el-nino-story-maps)

The success of the El Niño story maps led to the development of another similar project, this time connected to the Rising Voices program. The program (Collaborative Science with Indigenous Knowledge for Climate Solutions) aims at promoting cross-cultural dialogue and find solutions to emerging challenges created by climate change.

The Rising Voices 4 workshop took place in Hawai'i in July 2016 and addressed *storytelling for solutions*. Through a number of activities which included group discussions and field trips, participants were able to debate and exchange views on relevant topics. Field trips included a visit to Mauna Loa's observatory, an important scientific landmark of global relevance.

The challenge from a development standpoint consisted in providing information on the outcome and conclusions reached in the workshop while keeping a geospatial connection. Furthermore, the human side of the activity could not be neglected.

The portal built for the El Nino Story Maps was thus expanded, with new functionality added to it. The Odissey.js library was used to power the geographic content, while HTML5 was employed to build the website itself. Both the story contents and website can be linked

(although were not in this instance) to data processed server-side using the Python processing chains. Perfectly integrated within the SmartBasins framework, the Rising Voices 4 Story Maps (www.smartbasins.com/RV4) was thus an important milestone in the development of a fruitful relation between the potential user community and the portal.

In the future, we expect to expand the SmartBasins portal worldwide and provide near real-time data on multiple variables (land, ocean, and atmosphere) while pursuing the active participation of the scientific community and the public.

It is also desirable to expand the 'starter kits' initially developed for the TRMM Explorer website. These can be used in classroom and 'real-world' settings, presenting an opportunity for users to create their own visualization solutions, but based on a common framework.

Chapter 7. Concluding remarks

This thesis is focused on the distribution and dynamics of key oceanic variables and features, with the emphasis on mesoscale structures. However, the coastal environment is influenced by numerous forces, which drive the complex interactions and patterns we can observe so clearly in satellite imagery. To explore thermohaline fronts and biogeographic units off Portugal we also addressed the near and distant processes capable of influencing the imaged waters. Land cover dynamics, rainfall values and even the development of data distribution systems – critical to provide data to end-users – were all addressed in different parts of this work.

Nonetheless, and despite the ancillary information, at the core of the research project is the organization of the coastal sea off central Portugal. The study contributed decisively to answer important questions regarding spatial patterns of variables such as Sea Surface Temperature or Chlorophyll-a Concentration. The patterns were also correlated with SST fronts, which constitute natural barriers and ecotones separating water masses in an otherwise illusory continuous ocean.

A considerable amount of challenges were found throughout the project, partly due to the large amount of data employed from different sources and characterized by very different formats, spatial, temporal and spectral resolution, and, ultimately, objective.

Oceanographic data are still scattered across multiple websites from different organizations and despite recent efforts to reduce the patchiness of oceanic data sources, comprehensive data archives are still elusive. The difficulties are not limited to countless hours of browsing and querying data portals. Automated data transfers and even the development of processing chains are hampered by numerous imagery formats and access requirements. In the end, the development of complex processing chains successfully tackled the challenge. However, we must recommend researchers and decision-makers to avoid the multiplication of data sources and work towards the establishment of truly global, multi-sensor, multi-parameter information access portals.

Such simplification holds the promise of a true revolution in oceanography. Open archives translate into more users, enhanced dynamics, and therefore a more active community, transcending that of the traditional ocean data users. The difficulties encountered – in the age of big data - became one of the most revealing experiences recorded during this endeavor.

Scientifically, several conclusions are worth mentioning. Although we addressed these topics before, we must not conclude without a final reference to at least some of the answers found and questions raised in the past 4 years.

The use of Dynamic Time Warping as a method in biogeography has the potential to offer game-changing options for the segmentation of wide areas using relatively easy to acquire datasets. The use of this method was only made possible by recent breakthroughs in remote sensing, which opened a completely new window on the dynamic nature of the ocean.

Our analysis enabled us to conclude that critical data, such as that of SST and Chlorophyll-a Concentration were available where it was needed the most – the coastal environment. Although it is also true that current state-of-the-art algorithms are still affected significantly by the Case-1/Case-2 dichotomy, SST fields in particular offer an important and irreplaceable source of information on the spatial structure of ecosystems.

The sensors we employed (MODIS) are unable to provide users with much detail on the coastal environment given the spatial resolution (1 km) and adjacency effects created by the shoreline. As such, multi-scale methodologies would be helpful to accurately map specific, nested domains in the dynamic regions of the inner shelf.

In fact that is an important aspect to take into consideration when planning the development of a biogeographic classification. Can nested domains be classified using higher resolution data? And how can dissimilarities in the sensors (signal-to-noise ratios, specific algorithms) affect the coherence of the results in the different scales? This analysis was beyond the scope of the current study, which aimed at evaluating –and did so successfully – the usability of DTW with EO data to partition the ocean. Nonetheless, and considering the initiatives that emerged from this thesis (e.g. SmartBasins) these were but the first steps towards continuing the journey towards a global classification system.

The emergence of new machine-learning (and deep learning) tools, or their simple massification due to improving (and cost-efficient) cloud computing solutions is also likely to have a profound impact on biogeography in the near future. It will be up to the marine research community to identify the IT tools available to pursue these objectives and employ them in new contexts. As with all revolutions, we expect significant changes to disrupt how marine ecology is perceived, leading to a transition between traditional approaches and newer, eminently computational methods.

This was the principle guiding our work and the results seem to confirm our understanding of biogeography as a computational problem, for which we need adequate inputs.

The same applies to the frontal maps generated. The maps are unprecedented for the study area, offering a good balance between spatial (1 km) and temporal (daily) resolution for an important sector of the Portuguese coast and for a long period of time. Still, challenges persist. The spatial resolution still limits the delineation of near shore fronts and cloud cover hampers the continuous monitoring of fronts. Yet, the frontal probability maps not only confirmed the

important role of topography and coastline design in the development of spatial patterns, but highlighted new regions of strong activity that were previously uncharted.

The role of the upwelling, WIBP, of poleward currents, the interaction of different water masses and even river discharge produce – jointly – a complex setting where numerous fronts generated by different processes co-exist. Such maps are invaluable to researchers and operational users alike, considering the importance of these features to marine life, the propagation (or containment) of pollutants, and overall water quality.

Several questions persist, some perpetuated by the lack of data. ARGO-like systems are paramount to our understanding of the ocean. Yet, such systems are still limited to operate in the open ocean. Perhaps it is time to consider the development and deployment of an array of durable coastal drifters capable of profiling shelf waters reliably.

Several features warrant further analysis, including the Aveiro filament and the WIBP. In this study we were able to revise the boundaries of this feature, using different methods including frontal probability maps and DTW-based regions. These features also appear to influence the correlation between SST and Chlorophyll-a Concentration. Field surveys are necessary considering the difficulty inherent to the retrieval of accurate estimates in coastal (Case-2) waters. Yet, once again, the coherent spatial patterns that seem to organize the ocean off central Portugal reemerge. Future management policies and research activities should not ignore such organization but instead explore and exploit it as a useful tool for multiple purposes.

In the context of a changing world it is also relevant to determine whether the regions can be used for the purpose of tracking changes and depict their evolution from a spatial perspective. Furthermore, Biogeographic Classifications also provide useful and powerful visualizations capable of supporting decision-making and educational efforts considering that they represent reality in bounded and easy to follow units.

Fronts are yet another layer, albeit a relevant one, in the complex biogeographic map of the ocean. As natural borders and ecosystems in their own right, it is perhaps time to conceive the development of comprehensive maps of the ocean that, just like their land counterparts, provide information on a number of key variables in accurate and captivating manners.

As in the seminal work by Sherman et al. (1990), we agree with the integration of watersheds in biogeographic classification schemes. Coastal ecosystems are directly affected by adjoining watersheds and a true continuum is formed by both. As such, mapping watersheds in detail and understanding the processes at play that have the potential to disrupt marine life (and processes) is of the utmost importance. Water is the vector that carries this influence from land to the Ocean and therefore it must be tracked closely. In this study we aimed at

contributing as much as possible within the time and resource constraints inherent to this kind of work to the establishment of such a continuous perspective.

Mapping (in higher resolution and with greater accuracy than ever before) the imperviousness surfaces of the Mondego river watershed was the first of such steps. The ISA map offers new data that can be leveraged to understand how water quality (and color) is impacted by the anthropogenic modifications of the territory. The capabilities and know-how amassed during this period enabled the development of other experimental projects, including trials with SAR-based LULC classification.

The build-up of such capabilities seemed relevant to the stakeholder community and motivated the development of a stepwise sub-project dedicated to data acquisition, processing, and distribution.

The earlier TRMM Explorer addressed the development of dedicated rainfall and land cover products (which included vegetation phenology analysis) and enabled a first approach to the immense challenges influencing the development of cloud-based EO processing chains. Still, the overwhelming response by end users and the need to incorporate new products (including oceanographic data) led to the development of a more comprehensive and lasting concept – the SmartBasins Initiative. This project, which will not die with this thesis, is expected to become an important locus of development in the field of data integration and visualization for the watershed-coastal system for years to come. The involvement of stakeholders, researchers, and operational users is an objective that seems to be the logical conclusion of this study. An open conclusion, which is never really completed.

Many questions remain unanswered and many other were raised by this work. It is our hope, nonetheless, that it may offer a contribution to the analysis of important variables and processes, while introducing new methodologies and procedures to map our common resources threatened by a changing world.

We opened this thesis with the famed words written by Wallace more than a century ago – “Nothing like a perfect zoological division of the earth is possible”; and although this is true, advances in computation, the increasing amount of data sources available, and our improved understanding of living systems let us hope, with realistic albeit careful optimism that in the future we may come to the conclusion that the perception of such impossibility was but an oddity of past times.

References

Acha, E.M., Mianzan, H.W., Guerrero, R. A., Favero, M., Bava, J., 2004. Marine fronts at the continental shelves of austral South America: Physical and ecological processes. J. Mar. Syst. 44, 83–105.

Acker, James, 2015. Color of the Atmosphere With The Ocean Below: A History of NASA's Ocean Color Missions. CreateSpace Independent Publishing Platform. 372 p.

Agência Portuguesa do Ambiente / ARH do Centro IP, 2012. Plano de Gestão das Bacias Hidrográficas dos rios Vouga, Mondego e Lis integradas na Região Hidrográfica 4. Ministério daAgricultura, Mar, Ambiente e Ordenamento do Território. http://www.apambiente.pt/?ref=16&subref=7&sub2ref=9&sub3ref=834, accessed May 2015.

Akbar, T.A., Hassan,Q.K., Achari, G., 2013. Clusterization of Surface Water Quality and Its Relation to Climate and Land Use/Cover. Journal of Environmental Protection 4, 333-343.

Alcántara-Ayala, I., 2002. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries, Geomorphology 47, 107 – 124.

Almazroui, M., 2011. Calibration of TRMM rainfall climatology over Saudi Arabia during 1998–2009. Atmos. Res. 99, 400–414.

Almonacid-Caballer, J., Sánchez-García, E., Pardo-Pascual, J.E., Balaguer-Beser, A.A., Palomar-Vázquez, J., 2016. Evaluation of annual mean shoreline position deduced from Landsat imagery as a mid-term coastal evolution indicator. Mar. Geol. 372, 79–88.

Alpers, W., 1985. Theory of radar imaging of internal waves. Nature 314, 245–247.

Álvarez-Salgado, X. a., Figueiras, F.G., Pérez, F.F., Groom, S., Nogueira, E., Borges, a. V., Chou, L., Castro, C.G., Moncoiffé, G., Ríos, a. F., Miller, a. E.J., Frankignoulle, M., Savidge, G., Wollast, R., 2003. The Portugal coastal counter current off NW Spain: New insights on its biogeochemical variability. Prog. Oceanogr. 56, 281–321.

Álvarez-Salgado, X., Herrera, J.L., Gago, J., Otero, P., Soriano, J.A., Pola, C.G., Garcia-Soto, C., 2006. Influence of the oceanographic conditions during spring 2003 on the transport of the Prestige tanker fuel oil to the Galician coast. Marine Pollution Bulletin 53, 239–249.

Alvarez, I., Gomez-Gesteira, M., DeCastro, M., Lorenzo, M.N., Crespo, a. J.C., Dias, J.M., 2011. Comparative analysis of upwelling influence between the western and northern coast of the Iberian Peninsula. Cont. Shelf Res. 31, 388–399.

Alvarez, I., Lorenzo, M.N., Decastro, M., 2012. Analysis of chlorophyll a concentration along the Galician coast: Seasonal variability and trends. ICES J. Mar. Sci. 69, 728–738.

Ambar, I., Fiúza, A.F.G., 1994. Some features of the Portugal Current System: A poleward slope undercurrent, an upwelling related southward flow and an autumn winter poleward coastal surface current. *In*: 2nd International Conference on Air-Sea Interaction & Meteorology & Oceanography of the Coastal Zone, American Meteorological Society: 286-287.

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., Peck, D.L., 1976. A Land Use And Land Cover Classification System For Use With Remote Sensor Data, U.S. Geological Survey Circular 671. USGS, Washington, DC.

Andrieu, H., Chocat, B., 2004. Introduction to the special issue on urban hydrology. J. Hydrol. 299, 163–165.

Antrop, M., 2005.Why landscapes of the past are important for the future. Landscape and Urban Planning 70, 21-34.

Apel, J. R., 1987. Principles of Ocean Physics. Academic Press, Ltd., London, 634 pp.

Arco, A.I. del, Verónica, F., S., G.M.A., 2012. The performance of biological indicators in assessing the ecological state of streams with varying catchment urbanisation levels in Coimbra, Portugal. Limnetica 31, 141–154.

ARH do Centro IP, 2012. Plano de Gestão das Bacias Hidrográficas dos rios Vouga, Mondego e Lis integradas na Região Hidrográfica 4. Ministério daAgricultura, Mar, Ambiente e Ordenamento do Território. Retrieved May 2015, Available at: http://www.apambiente.pt/?ref=16&subref=7&sub2ref=9&sub3ref=834.

Arístegui, J., Álvarez-Salgado, X., Barton, E., Figueiras, F., Hernández-León, S., Roy, C., Santos, A., 2006. Oceanography and fisheries of the Canary Current/Iberian region of the eastern North Atlantic. In: Robinson, A., Brink, K. (Eds.), The Sea, . *In*: The Global Coastal Ocean: Interdisciplinary Regional Studies and Syntheses, vol. 14 (Part B). Harvard University Press, Cambridge, MS, pp. 877–931.

Armstrong, D.A., Mitchell-Innes, B.A., Verheye-Dua, F., Hutchings, L. ,1987. Physical and biological features across the upwelling front in the southern Benguela. *In*: The Benguela and Comparable Ecosystems. A. Payne, J. Gulland, K. Brink (Eds.), Sea Fisheries Research Institute, Cape Town, pp. 171–190.

Arnold, C.L., Gibbons, C.J., 1996. Impervious Surface Coverage: The Emergence of a Key Environmental Indicator. J. Am. Plan. Assoc. 62, 243–258.

Austin, J.A., Lentz, S.J., 2002. The inner-shelf response to wind-driven upwelling and downwelling. Journal of Physical Oceanography 32, 2171-2193.

Baith, K., R. Lindsay, G. Fu, McClain, C.R., 2001. Data analysis system developed for ocean color satellite sensors. EOS Transactions AGU, 82, 202-203.

Bakun, A., 1994. Climate change and marine populations: Interactions of physical and biological dynamics. Workshop on the Scope, Significance, and Policy. Implications of Global Change and the Marine Environment, Greenpace/University of Rhode Island, USA, 14– 17 May, pp. 1– 16.

Bakun, A., 2006. Fronts and eddies as key structures in the habitat of marine fish larvae: opportunity, adaptative response and competitive advantage. Scientia Marina 70 (Suppl. 2), 105-122.

Ban, Y., Yousif, O., Hu, H., 2014. Fusion of SAR and Optical Data for Urban Land Cover Mapping and Change Detection. Chapter 18 *In*: Global Urban Monitoring and Assessment through Earth Observation, Ed: Q. Weng, CRC Press, pp. 353-386.

Barber, R.T., Hilting, A.K., 2000. Achievements in biological oceanography. *In*: 50 Years of Ocean Discovery. Ocean Studies Board, National Academy Press, Washington DC, pp 11-21.

Batteen, M.L., Rutherford, M.J., Bayler, E.J., 1992. A numerical study of wind and thermal forcing effects on the ocean circulation off Western Australia. Journal of Physical Oceanography 22, 1406–1433.

Bauer, M.E., Loffelholz, B.C., Wilson, B., 2008. Estimating and Mapping Impervious Surface Area by Regression Analysis of Landsat Imagery, in: Weng, Q. (Ed.), Remote Sensing of Impervious Surfaces. CRC / Taylor & Francis, pp. 3 – 19.

Baumann, K-H., Andruleit, H., Bockel, B., Geisen, M., Kinkel, H., 2005. The significance of extant coccolithophores as indicators of ocean water masses, surface water temperature, and palaeoproductivity: A review. Palaontologische Zeitschrift 79,93–112.

Belanger, C.L., Jablonski, D., Roy, K., Berke, S.K., Krug, a. Z., Valentine, J.W., 2012. Global environmental predictors of benthic marine biogeographic structure. Proc. Natl. Acad. Sci. 109, 14046–14051.

Belkin, I.M., 2002. Front. *In*: Nybakken, J.W., Broenkow, W.W., Vallier, T.L. (Eds.), Interdisciplinary Encyclopedia of Marine Sciences. Grolier, Danbury, CT, pp. 433–436.

Belkin, I.M., Cornillon, P.C., Sherman, K., 2009. Fronts in Large Marine Ecosystems. Prog. Oceanogr. 81, 223–236.

Belkin, I.M., O'Reilly, J.E., 2009b. An algorithm for oceanic front detection in chlorophyll and SST satellite imagery. J. Mar. Syst. 78, 319–326.

Benazzouz, A., Mordane, S., Orbi, A., Chagdali, M., Hilmi, K., Atillah, A., Lluís Pelegrí, J., Hervé, D., 2014. An improved coastal upwelling index from sea surface temperature using satellite-based approach - The case of the Canary Current upwelling system. Cont. Shelf Res. 81, 38–54.

Berrick, S.W., Leptoukh, G., Farley, J.D., Rui, H., 2009. Giovanni: A Web service workflowbased data visualization and analysis system. IEEE Trans. Geosci. Remote Sens. 47 (1), 106–113.

Berk, A., Bernstein, L.S., Anderson, G.P., Acharya, P.K., Robertson, D.C., Chetwynd, J.H., Adler-Golden,S.M., 1998. MODTRAN Cloud and Multiple Scattering Upgrades with Application to AVIRIS, Remote Sensing of Environment 65, 367-375.

Bernstein, R.L., Breaker, L., Whritner, R., 1977. California Current Eddy Formation: Ship , Air , and Satellite Results. Science 195 (4276), 353–359.

Berrick, S.W., Leptoukh, G., Farley, J.D., and Rui, H., 2009. Giovanni: A Web service workflow-based data visualization and analysis system, IEEE Trans. Geosci. Remote Sens. 47(1), 106-113.

de Beurs, K. M., Henebry, G. M., 2004. Land surface phenology, climatic variation, andinstitutional change: analyzing agricultural land cover change in Kazakhstan. Remote Sensing of Environment 89, 497–509.

Black, B. a., Sydeman, W.J., Frank, D.C., Griffin, D., Stahle, D.W., García-Reyes, M., Rykaczewski, R.R., Bograd, S.J., Peterson, W.T., 2014. Six centuries of variability and extremes in a coupled marine-terrestrial ecosystem. Science 345, 1498–502.

Bledsoe, B.P., Watson, C.C., 2001. Effects of urbanization on channel instability. J. Am. Water Resour. Assoc. 37, 255–270.

Borges, M.F., Santos, A.M.P., Crato, N., Mendes, H., Mota, B., 2003. Sardine regime shifts off Portugal: a time series analysis of catches and wind conditions. Scientia Marina 67 (Suppl. 1), 235–244.

Bower, A., Armi, L., Ambar, I., 1997. Lagrangian observations of meddy formation during a Mediterranean undercurrent seeding experiment. Journal of Physical Oceanography 27, 2445–2575.

Bower, A., Serra, N., Ambar, I., 2002. Structure of the mediterranean undercurrent and mediterranean water spreading around the southwestern iberian peninsula. Journal of Geophysical Research 107 (C10), 3161.

Brabec, E., Schulte, S., Richards, P.L., 2002. Impervious Surfaces and Water Quality: A Review of Current Literature and Its Implications for Watershed Planning. J. Plan. Lit. 16, 499–514.

Brakenridge, G.R., Cohen, S., Kettner, A.J., De Groeve, T., Nghiem, S.V., Syvitski, J.P.M., Fekete, B.M., 2012. Calibration of orbital microwave measurements of river discharge using a global hydrology model. Journal of Hydrology 475, 123-136.

Brander, K., Blom, G., Borges, M., Erzini, K., Henderson, G., MacKenzie, B., Mendes, H., Santos, A., Toresen, R., 2003. Changes in fish distribution in the eastern North Atlantic: Are we seeing a coherent response to changing temperature?. ICES Mar. Sci. Symp. 219, 261–270.

Breaker, L.C., Mavor, T.P:, Broenkow, W.W., 2005. Mapping and Monitoring Large-Scale Ocean Fronts Off the California Coast Using Imagery from the GOES-10 Geostationary Satellite, Publ. T-056, California Sea Grant College Program, University of California, San Diego, La Jolla. (Available at http://repositories.cdlib.org/csgc/rcr/Coastal05_02).

Breitbarth, E., Oschlies, A., LaRoche, J., 2007. Physiological constraints on the global distribution of Trichodesmium—Effect of temperature on diazotrophy. Biogeosciences 4, 53–61.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees, Wadsworth International Group, Wadsworth Statistics/Probability Series. Wadsworth.

Brink, K.H. and T.J. Cowles, 1991. The coastal transition zone program. J. Geophys. Res. 96 (14),637-647.

Breaker, L.C., Mooers, C.N.K., (1986). Oceanic variability off the Central California coast, Prog. Oceanogr. 17, 61–135.

Brown, O.B., Minnett, P.J., 1999. MODIS Infrared Sea Surface Temperature Algorithm: Algorithm Theoretical Basis Document Version 2.0., NASA Contract No. NAS5-31361, University of Miami, 91 p.

Brown, S., Nicholls, R., Woodroffe, C., Hanson, S., Hinkel, J., Kebede, A.S., Neumann, B., Vafeidis, A.T., 2013. Sea-Level Rise Impacts and Responses: A Global Perspective. *In*: Finkl CW (Ed.). Coastal Hazards. Netherlands: Springer; 2013. pp. 117–149.

Cadenasso, M.L., Pickett, S.T. a, Schwarz, K., 2007. Spatial heterogeneity in urban ecosystems: Reconceptualizing land cover and a framework for classification. Front. Ecol. Environ. 5, 80–88.

Callejas-Jimenez, M., Santamaria-del-Angel, E., Gonzalez-Silvera, A., Millan-Nuñez, R., Cajal-Medrano, R., 2012. Dynamic Regionalization of the Gulf of Mexico based on normalized radiances (nLw) derived from MODIS-Aqua. Cont. Shelf Res. 37, 8–14.

Canny, J., 1986. A Computational Approach to Edge Detection. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8, 679–698.

Carder, K. L., Chen, F. R., Lee, Z., Hawes, S. K., Cannizzaro, J. P., 2003. MODIS ocean science team algorithm theoretical basis document MOD-ATBD-19, case 2 chlorophyll a. College of Marine Science: University of South Florida, Florida, USA.

Carey, R.O., Hochmuth, G.J., Martinez, C.J., Boyer, T.H., Dukes, M.D., Toor, G.S., Cisar, J.L., 2013. Evaluating nutrient impacts in urban watersheds: challenges and research opportunities. Environ. Pollut. 173, 138–49.

Carlson, T., 2003. Applications of remote sensing to urban problems. Remote Sens. Environ. 86, 273–274.

Carpenter, E. J., Subramaniam, A., Capone, D. G., 2004. Biomass and primary productivity of the cyanobacterium Trichodesmium spp. in the tropical N Atlantic ocean. Deep-sea research part I — Oceanographic research papers 51, 173−203.

Carr, M.E., Kearns, E., 2003. Production regimes in four Eastern Boundary Current systems. Deep-Sea Res. II 50 ,3199–3221.

Carton, X., Chérubin, L., Paillet, J., Morel, Y., Serpette, A., Le Cann, B., 2002. Meddy coupling with a deep cyclone in the Gulf of Cadiz. Journal of Marine Systems 32 (1-3), 13–42.

Casey, K., Cornillon, P., 2001. Global and regional sea surface temperature trends, J. Clim. 14, 3801– 3818.

Castelao, R.M., Mavor, T.P., Barth, J. a, Breaker, L.C., 2006. Sea surface temperature fronts in the California Current System from geostationary satellite observations. J. Geophys. Res. 111, C09026.

Cayula, J.-F., Cornillon, P., 1992. Edge detection algorithm for SST images. J. Atmos. Ocean. Technol. 9, 67-80.

Cayula, J.-F., Cornillon, P.C., 1995. Multi-Image Edge Detections for SST Images. J. Atmos. Ocean. Tech. 12, 821–829.

Chesapeake Bay Program (n.d.). Chesapeake Bay Program. Retrieved May, 2016, Available at: http://www.chesapeakebay.net/track/tools

CELBI, 2013. Declaração Ambiental. Celulose Beira Industrial (CELBI) Figueira da Foz, 36 p.

Chadwick, M., D. Dobberfuhl, A. Benke, A. Huryn, K. Suberkropp and J. Thiele (2006). Urbanization affects stream ecosystem function by altering hydrology, chemistry, and biotic richness. Ecological Applications 16(5), 1796-1807.

Chander, G., Huang, C., Yang, L., Homer, C., Larson, C., 2009. Developing Consistent Landsat Data Sets for Large Area Applications: The MRLC 2001 Protocol. Geosci. Remote Sens. Lett. IEEE 6, 777–781.

Chapman, C., 2014. Southern Ocean jets and how to find them: Improving and comparing common jet detection methods. J. Geophys. Res. Ocean. 119, 1-22.

Charlock, T.P., 1982. Mid-latitude model analysis of solar radiation, the upper layers of the sea, and climate. J. Geophys. Res. 87, 8923-8930.

Chen, S., Hong, Y., Cao, Q., Kirstetter, P., Gouorley, J.J., Qi, Y., Zhang, J., Howard, K., Hu, J., Wang, J., 2013. Performance evaluation of radar and satellite rainfalls for Typhoon Morakot over Taiwan: Are remote-sensing products ready for gauge denial scenario of extreme events? J. Hydrol. 506, 4–13.

Chin, A., 2006. Urban transformation of river landscapes in a global context. Geomorphology 79, 460–487.

Copernicus Marine Environment Monitoring Service (CMEMS). Retrieved January 01, 2016, Available at: http://marine.copernicus.eu/

Cohen, W.F., Sirisinha,K., Lee, G.F., 1976.Nitrogen availability in urban runoff. Journal of the Water Pollution Control Federation 48(2), 339-345.

Colas, F., 2003. Circulation et dispersion lagrangiennes en Atlantique Nord-Est. Ph.D. thesis, Université de Bretagne Occidentale, 943, 253 p..

Collier, C.G., 2007. Flash flood forecasting: what are the limits of predictability?, Quart. J. Roy. Meteorol. Soc. 133(622), 3–23.

Cook, R. B., Margle, S. M., SanthanaVannan, S. K., Holladay, S. K., Beaty, T. W., 2006. Subsetting Tools for MODIS Land Products: Time-series data for field sites, Global Vegetation Workshop, Missoula MT, August 8-10, 2006, Available at: http://daac.ornl.gov/MODIS/MODIS-menu/other/modis_200605.pdf, January 20, 2014.

Conti, F., Hsu, K., Noto, L., Sorooshian, S., 2014. Evaluation and comparison of satellite precipitation estimates with reference to a local area in the Mediterranean Sea. Atmos. Res. 138, 189–204.

Cordeiro, N.G.F., Nolasco, R., Cordeiro-Pires, A., Barton, E.D., Dubert, J., 2015. Journal of Geophysical Research : Oceans Filaments on the Western Iberian Margin : A modeling study. J. Geophys. Res. Ocean. 120, 5400–5416.

CORINE, CORINE Land Cover, Retrieved December 2015, Available at: http://www.eea.europa.eu/publications/COR0-landcover.

Coriolis: In situ data for operational oceanography. (n.d.). Retrieved January 1, 2015, from http:// www.coriolis.eu.org/

Corsi, S.R., De Cicco, L.A., Lutz, M.A., Hirsch, R.M., 2015. River chloride trends in snow-affected urban watersheds: increasing concentrations outpace urban growth rate and are common among all seasons. Sci. Total Environ. 508, 488–97.

Costanza, R., Arge, R., Groot, R. De, Farberk, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V, Paruelo, J., Raskin, R.G., Suttonkk, P., van den Belt, M., 1997. The value of the world ' s ecosystem services and natural capital. Nature 387, 253–260.

Costello, J.H., Sullivan, B.K., Gifford, D.J., Van Keuren, D., Sullivan, L.J., 2006. Seasonal refugia, shoreward thermal amplification, and metapopulation dynamics of the ctenophore *Mnemiopsis leidyi* in Narragansett Bay, Rhode Island. Limnol Oceanogr 51, 1819−1831.

Coulter, C. B., Kolka, R. K., Thomson, J. A., 2004. Water quality in agricultural, urban, and mixed land use watersheds. Journal of the American Water Resources Association 40 (6), 1593–1601.

CCS - The Center for Climate and Security, 2016. Briefing Book for a New Administration.The Climate and Security Advisory Group, Washington, DC, 27 p.

Cunha, P.P., Dinis, J., 2002. Sedimentary dynamics of the Mondego estuary. *In*: M.A. Pardal, J.C. Marques, M.A. Graça (Eds.), Aquatic Ecology of the Mondego River Basin: Global Importance of Local Experience. pp. 43-62, Imprensa da Universidade de Coimbra, Coimbra, Portugal.

Cushing, D.H., 1990. Plankton production and year-class strength in fish populations: an update of the match/mismatch hypothesis. Adv. Mar. Biol. 26, 249−293.

Daniault, N., Mazé , J.P., Arhan, M., 1994. Circulation and mixing of Mediterranean Water west of Iberian Peninsula. Deep Sea Research I 41, 1685–1714.

Davis, C., Schaub, T., 2005. A transboundary study of urban sprawl in the Pacific Coast region of North America: The benefits of multiple measurement methods. Int. J. Appl. Earth Obs. Geoinf. 7, 268–283.

deCastro, M., Gómez-Gesteira, M., Lorenzo, M.N., Alvarez, I., Crespo, A.J.C., 2008. Influence of atmospheric modes on coastal upwelling along the western coast of the Iberian Peninsula, 1985 to 2005. Climate Research 36, 169-179.

DESTATIS, 2007. Umweltnutzung und Wirtschaft. Bericht zu den Umweltökonomischen Gesamtrechnungen. Wiesbaden, Statistisches Bundesamt. 120 p.

Deguchi, C., Sugio, S., 1994. Estimations for Percentage of Impervious Area by the Use of Satellite Remote Sensing Imagery. Water Science and Technology. 29, 1-2, 135-144.

Devred, E., Sathyendranath, S., Platt, T., 2007. Delineation of ecological provinces using ocean color radiometry. Mar. Ecol. Prog. Ser. 346, 1–13.

Díaz, L., Granell, C., Huerta, J., Gould, M., 2012. Web 2.0 Broker: A standards-based service for spatio-temporal search of crowd-sourced information, Applied Geography 35, 448 - 459.

Ding, H., Trajcevski, G., Scheuermann, P.,Wang, X., Keogh, E., 2008. Querying and mining of time series data: experimental comparison of representations and distance measures. VLDB, 1542–1552.

Ding, J., Jiang, Y., Fu, L., Liu, Q., Peng, Q., Kang, M., 2015.Impacts of Land Use on Surface Water Quality in a Subtropical River Basin: A Case Study of the Dongjiang River Basin, Southeastern China. Water 7, 4427-4445.

Dinis, P.A., Mantas, V.M., Andrade, P.S., Tonecas, J., Kapula, E., Pereira, A.J.S.C., Carvalho, F.S., 2013. Contribution of TRMM rainfall data to the study of natural systems and risk assessment.Cases of Application in SW Angola.Estudos do Quaternário 9, 33 – 43.

Dinku, T., Ruiz, F., Connor, S.J., Ceccato, P., 2010. Validation and intercomparison of satellite rainfall estimates over Colombia. J. Appl. Meteorol. Climatol. 49, 1004–1014.

Dinku, T., Chidzambwa, S., Ceccato, p, Connor, S.J., Ropelewski, C.F., 2008. Validation of high-resolution satellite rainfall products over complex terrain. Int. J. Remote Sens. 29 (14), 4097–4110.

Donlon, C. J., et al. (2007), The Global Ocean Data Assimilation Experiment High-Resolution Sea Surface Temperature Pilot Project, Bull. Am. Meteorol. Soc. 88, 1197–1213.

Du, L., Tian, Q., Yu, T., Meng, Q., Jancso, T., Udvardy, P., Huang, Y., 2013. A comprehensive drought monitoring method integrating MODIS and TRMM data. Int. J. Appl. Earth Obs. Geoinf. 23, 245–253.

Duh, J.-D., Shandas, V., Chang, H., George, L.A., 2008. Rates of urbanisation and the resiliency of air and water quality. Sci. Total Environ. 400, 238–56.

Duque, J.C., Patino, J.E., Ruiz, L. a, Pardo-pascual, J.E., 2015. Landscape and Urban Planning Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. Landsc. Urban Plan. 135, 11–21.

Durant, J.M, Hjermann, D., Falkenhaug, T., Gifford, D.J., Naustvoll, L., Sulivan, B.K., Beaugrand, G., Stenseth, N.C., 2013. Extension of the match-mismatch hypothesis to predator-controlled systems. Marine Ecology Progress Series 474, 43-52.

Earth Explorer, United States Geological Survey. (n.d.). Retrieved January, 2014, from http:// www.earthexplorer.usgs.gov/

EC, 2007. European Parliament Legislative Resolution of 11 December 2007 on the Council Common Position for Adopting a Directive of the European Parliament and of the Council Establishing a Framework for Community Action in the Field of Marine Environmental Policy (Marine Strategy Framework Directive) (9388/ 2/2007 e C6-0261/2007 e 2005/0211(COD)).

EEA, 2006. Urban sprawl in Europe - The ignored challenge. Publications of the European Communities, No 10, Copenhagen, 60 p.

Emmanuel, R., Krüger, E., 2012. Urban heat island and its impact on climate change resilience in a shrinking city: The case of Glasgow, UK. Build. Environ. 53, 137–149.

Enfield, D.B., Mestas-Nuñes, A.M., Trimble, P.J., 2001. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. Geophysical research Letters 28(10), 2077-2080.

EPA, 2000. Guidance for Data Quality Assessment: Practical Methods for Data Analysis, United States Environmental Protection Agency, EPA QA/G-9, QA00 Update.

Esch, T., Himmler, V., Schorcht, G., Thiel, M., Wehrmann, T., Bachofer, F., Conrad, C., Schmidt, M., Dech, S., 2009. Large-area assessment of impervious surface based on integrated analysis of single-date Landsat-7 images and geospatial vector data. Remote Sens. Environ. 113, 1678–1690.

ESSO - Indian National Centre for Ocean Information Services. Retrieved January, 2016, Available at: http://www.incois.gov.in/MarineFisheries/PfzAdvisory .

Evangilidis, K., Ntouros, K., Makridis, S., Papatheodorou, C., 2014. Geospatial services in the Cloud, Computers & Geosciences 63, 116 - 122.

Fairbridge, R.W., 1980. The estuary: its definition and geodynamic cycles. In: Chemistry and Biogeochemistry of Estuaries. E. Olausson, I. Cato (Eds.), New York, pp. 1-35.

Feio, M.J., Alves, T., Boavida, M., Medeiros, A., Graça, M.A.S., 2010. Functional indicators of stream health: a river-basin approach. Freshw. Biol. 55, 1050–1065.

Ferreira, V., Graça, M. a S., Feio, M.J., Mieiro, C., 2004. Water quality in the Mondego river basin: Pollution and habitat heterogeneity. Limnetica 23, 295–306.

Le Fèvre, J., Viollier, M., Le Corre, P., Dupouy, C., Grall, J.R., 1983. Remote sensing observations of biological material by LANDSAT along a tidal thermal front and their relevancy to the available field data. Estuar. Coast. Shelf Sci. 16, 37–50.

Fiúza, A.F.G., Macedo, M., Guerreiro, M., 1982. Climatological space and time variation of the Portuguese coastal upwelling. Oceanologica Acta 5, 31–40.

Fiúza, A.F.G., 1983. Upwelling patterns off Portugal. In: Coastal upwelling: its sediment record, E. Suess, J. Thiede (Eds.), Plenum Publishing, New York, pp. 85- 98.

Flament, P., Armi, L., 2000. The Shear, Convergence, and Thermohaline Structure of a Front. J. Phys. Oceanogr. 30, 51–66.

Fofonoff, N. P., Millard, R.C., 1983. Algorithms for computations of fundamental properties of seawater. UNESCO Technical Papers in Marine Science No. 44, 53 pp.

Frank, K., Petrie, B., Fisher, J., Leggett, W., 2011. Transient dynamics of an altered large marine ecosystem. Nature 477, 86-89.

Franz, Bryan A., Kwiatkowska, Ewa J., Meister, Gerhard, McClain, Charles R., 2007. In James J. Butler, Jack Xiong (Eds.), Utility of MODIS–Terra for ocean color applications, Earth observing systems XII. Proceedings of the SPIE, 6677. (pp. 66770Q).

Fraser, L., 1985. High Altitude Research at the Applied Physics Laboratory in the 1940s. John Hopkins APL Technical Digest. 6 (1), 92–99.

Frouin, R., Fiúza, A.F., Ambar, I., Boyd, T.J., 1990. Observations of a Poleward Surface Current off the Coast of Portugal and Spain during winter. Journal of Geophysical Research 95, 679–691.

Galindo-Cortes, G., De Anda-Montañez, J., Arreguín-Sánchez, F., Salas, S., Balart, E., 2010. How do environmental factors affect the stock-recruitment relationship? The case of the Pacific sardine (*Sardinops sagax*) of the northeastern Pacific Ocean. Fish. Res.102, 173–183.

Gao, F., Masek, J.G., 2008. Normalizing ASTER data using MODIS products for land cover classification. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37, B7.

Garrido, S., Cunha, M.E., Oliveira, P.B., van der Lingen, C.D., 2006. Diet composition and feeding behaviour of Iberian sardine (*Sardina pilchardus*). ICES CM 2006/F17.

Garvine, R.W., 1982. A steady state model for buoyant surface plume hydrodynamics in coastal waters. Tellus 34, 293–306.

GDAL, Retrieved March 2014, Available at: http://www.gdal.org/

Giorgino, T., 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. Journal of Statistical Software 31(7), 1-24.

Glaeser, E. L., J. Kolko, Saiz, A., 2001. Consumer city. Journal of Economic Geography 1, 27–50.

Gordon, H. R., & Wang, M., 1994. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm. Applied Optics 33, 443−452.

Gordon, H. R. and Morel, A., 1983. Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery. A Review, *In*: Lecture Notes on Coastal and Estuarine Studies, R. T. Barber, N. K. Mooers, M. J. Bowman and B. Zeitzschel (Eds.), Springer-Verlag, New York, 114 p.

Gould, J., Roemmich, D., Wijffels, S., Freeland, H., Ignaszewsky, M., Jianping, X., Pouliquen, S., Desaubies, Y., Send, U., Radhakrishnan, K., Takeuchi, K., Kim, K., Danchenkov, M., Sutton, P., King, B., Owens, B., Riser, S., 2004. Argo profiling floats bring new era of in situ ocean observations. Eos, Trans. Am. Geophys. Union 85, 185.

Goward, S., Arvidson, T., Williams, D., Faundeen, J., Irons, J., Franks, S., 2006. Historical record of landsat global coverage : Mission operations, NSLRSDA, and international cooperator stations. Photogramm. Eng. Remote Sensing 72, 1155–1169.

Gregr, E.J., Ahrens, A.L., Ian Perry, R., 2012. Reconciling classifications of ecologically and biologically significant areas in the world's oceans. Mar. Policy 36, 716–726.

Grimm, N.B., Grove, J.M., Pickett, S.T. a., Redman, C.L., 2000. Integrated Approaches to Long-Term Studies of Urban Ecological Systems. Bioscience 50, 571–584.

Grove, J.M., Burch Jr., W.R., 1997. A social ecology approach and application of urban ecosystem and landscape analysis: a case study of Baltimore Maryland. Urban Ecosytems 1, 259–275.

Groffman, P.M., Bain, D.J., Band, L.E., Belt, K.T., Brush, G.S., Grove, J.M., Pouyat, R. V, Yesilonis, I.C., Zipperer, W.C., 2003. Down by the riverside: urban riparian ecology. Front. Ecol. Environ. 1, 315–321.

Haase, D., Kabisch, N., Haase, A. 2013.Endless Urban Growth? On the Mismatch of Population, Household and Urban Land Area Growth and Its Effects on the Urban Debate. PLoS ONE 8(6): e66531.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA Data Mining Software: An Update, SIGKDD Explorations 11(1), 10-18.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T., Selig, E.R., Spalding, M., Steneck,R., Watson, R., 2008. A global map of human impact on marine ecosystems. Science 319, 948–952.

Han, W., Z. Yang, L. Di, and R. Mueller. 2012. CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Computers and Electronics in Agriculture 84, 111–123.

Hatt, B. E., T. D. Fletcher, C. J. Walsh, Taylor, S.L., 2004. The influence of urban density and drainage infrastructure on the concentrations and loads of pollutants in small streams. Environmental Management 34,112– 124.

Haynes, R., Barton, E.D., 1990. A Poleward Flow along the Atlantic Coast of the Iberian Peninsula. Journal of Geophysical Research 95, 11425–11441.

Hebble, E.E., Carlson, T.N., Daniel, K., 2001. Impervious Surface Area and Residential Housing Density: A Satellite Perspective. Geocarto Int. 16, 15–20.

Hedley, J.D., Roelfsema, C.M., Chollett, I., Harborne, A.R., Heron, S.F., Weeks, S.J., Skirving, W.J., Strong, A.E., Mark Eakin, C., Christensen, T.R.L., Ticzon, V., Bejarano, S., Mumby, P.J., 2016. Remote sensing of coral reefs for monitoring and management: A review. Remote Sens. 8.

Herold, M.E. Gardner, D.A. Roberts, 2003. Spectral resolution requirements for mapping urban areas. IEEE Transactions on Geoscience and Remote Sensing 41, 1907–1919.

Hietel, E., Waldhardt, R., Otte, A., 2007.Statistical modeling of land-cover changes based on key socio-economic indicators. Ecological Economics 62, 496-507.

Hodgson, M.E., Jensen, J.R., Tullis, J.A., Riordan, K.D., Archer, C.M., 2003. Synergistic Use of Lidar and Color Aerial Photography for Mapping Urban Parcel Imperviousness. Photogrammetric Engineering & Remote Sensing 69:9, 973–980.

Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004. Development of a 2001 national land-cover database for the United States. Photogramm. Eng. Remote Sens. 70, 829–840.

Hong, Y., Adler, R. F. , Hossain, F., Curtis, S., Huffman, G.J., 2007. A First Approach to Global Runoff Simulation using Satellite Rainfall Estimation, Water Resources Research 43(8), W08502.

Hou, A.Y., kakar, R., Neeck, S., Azabarzin, A., Kummerow, C., Kojima, M., Oki, R., Nakamura, K., Iguchi, T., 2013. The Global Precipitation Measurement (GPM) Mission, Bulletin of the American Meteorological Society, e-View doi: http://dx.doi.org/10.1175/BAMS-D-13-00164.1.

Hou, A.Y., Skofronick-Jacksonm G., Kummerow, C.D., Shepherd, J.M., 2008. Global precipitation measurement, Precipitation: advances in Measurement, Estimation and Prediction, S. C. Michaelides, Ed. Springer-Verlag, 131 - 170.

Houska, T.R. 2012, EarthExplorer, U.S. Geological Survey General Information Product 136, 2 p..

Houska, T.R., Johnson, A.P., 212. GloVis, U.S. Geological Survey General Information Product 137, 2 p..

Hooker, S.B., Rees, N.W., Aiken, J., 2000. An objective methodology for identifying oceanic provinces. Prog. Oceanogr. 45, 313–338.

Horner-Devine, A.R., Fong, D.A., Monismith, S.G., Maxworthy, T., 2006. Laboratory experiments simulating a coastal river inflow. Journal of Fluid Mechanics 555, 203–232.

Horner-Devine, A.R., Jay, D. a., Orton, P.M., Spahn, E.Y., 2009. A conceptual model of the strongly tidal Columbia River plume. J. Mar. Syst. 78, 460–475.

Howarth, R.W., 2008. Coastal nitrogen pollution: A review of sources and trends globally and regionally. Harmful Algae 8, 14–20.

Hovis, W.A., Clark, D.K., Anderson, F., Austin, R.W., Wilson, W.H., Baker, E.T., Ball, D., Gordon, H.R:, Mueller, J.L., El-Sayed, S.Z., Sturm, B., Wrigley, R.C., Yentsch, C.S., 1980. Nimbus-7 coastal zone color scanner - system description and initial imagery, Science 210 (4465), 60-63.

Hu, Z., Pan, D., He, X., Bai, Y., 2016. Diurnal variability of turbidity fronts observed by geostationary satellite ocean color remote sensing. Remote Sens. 8, 1–15.

Hu, C., Feng, L., Lee, Z., Davis, C.O., Mannino, A., McClain, C.R., Franz, B.A., 2012. Dynamic range and sensitivity requirements of satellite ocean color sensors: learning from the past. Appl. Opt. 51, 6045–6062.

Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., Vogelmann, J.E., 2010. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. Remote Sens. Environ. 114, 183–198.

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., 2013. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG), Algorithm Theoretical Basis Document, Version 4.1, NASA, 25 p..

Huffman, G., Adler, R., Bolvin, D., Gu, G., Nelkin, E., Bowman, K., Hong, Y., Stocker, E.,Wolff, D., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. 8, 38–55.

Huffman, G.J., Bolvin, D.T., 2008. TRMM and Other Data Precipitation Data Set Documentation. Laboratory for Atmospheres, NASA Goddard Space Flight Center and

Science Systems and Applications Inc., available at: ftp://precip.gsfc.nasa.gov/pub/trmmdocs/3B42_3B43_doc.pdf.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Nelkin, E.J., 2010. The TRMM Multi-Satellite Precipitation Analysis (TMPA). Chapter 1 in satellite rainfall applications for surface hydrology. *In*: Hossain, F., Gebremichael, M. (Eds.), Springer Verlag. pp. 3–22.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., 2015. Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation, NASA Goddard Space Flight Center and Science Systems and Applications Inc., available at: https://pmm.nasa.gov/sites/default/files/document_files/IMERG_doc.pdf , 48 p.

Hurd, J.D., Civco, D.L., 2004. Temporal Characterization of impervious surfaces for the State of Connecticut, in: ASPRS Annual Conference Proceedings. Denver, Colorado, p. Unpaginated CD ROM.

Huthnance, J.M., 1984. Slope Currents and "JEBAR." J. Phys. Oceanogr. 14, 795-810.

ICIMOD, International Centre for Integrated Mountain Development, 2013. Validation of NOAA CPC_RFE2.0 satellite-based rainfall estimates in the central Himalayas – working paper 2013/5. available at:. http://lib.icimod.org/record/28331.

IDC, 2013. Android growth drives another strong quarter for the worldwide tablet market, Retrieved January 2013, Available at: http://www.idc.com/getdoc.jsp?containerId=prUS24420613.

Immerzeel, W.W., Rutten, M.M., Droogers, P., 2009. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. Remote Sens. Environ. 113, 362–370.

IGP, 2010. Carta de Uso e Ocupação do Solo de Portugal Continental para 2007 (COS2007). Memória descritiva. Instituto Geográfico Português, Lisboa, 77 p.

INE, 2012. Censos 2011. Informação à comunicação social (in Portuguese), Instituto Nacional de Estatística, Accessed online at www.ine.pt, April 2016, 41 p.

INE, 2014. Projeções de população residente 2012-2060 (in Portuguese), Instituto Nacional de Estatística, March 2014, Lisbon, 18 p.

INE, 2015. Digital Census Tables, Instituto Nacional de Estatística, Accessed online at www.ine.pt, May 2015.

IOCCG (International Ocean-Color Coordinating Group), 2009. Partition of the Ocean into Ecological Provinces: Role of Ocean-Color Radiometry. Mark Dowell, Trevor Platt (Eds.), Reports of the International Ocean-Color Coordinating Group, No. 9, IOCCG, Dartmouth, Canada, 98. p..

IOCCG (International Ocean-Color Coordinating Group), 2008. Why Ocean Color? The Societal Benefits of Ocean-Color Technology. Trevor Platt, Nicolas Hoepffner, Venetia Stuart, Christopher Brown (Eds.), Reports of the International Ocean-Color Coordinating Group, No. 7, IOCCG, Dartmouth, Canada, 147 p..

Iorgulescu, I., beven, K.J., 2004. Nonparametric direct mapping of rainfall-runoff relationships: An alternative approach to data analysis and modeling? Water Resources Research, 40, W08403.

IPCC - Intergovernmental Panel on Climate Change, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex , P.M. Midgley (Eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 p..

Irons, J.R., Dwyer, J.L., Barsi, J. a., 2012. The next Landsat satellite: The Landsat Data Continuity Mission. Remote Sens. Environ. 122, 11–21.

Isemer, H., Hasse, L., 1987. The Bunker Climate Atlas of the North Atlantic Ocean. Springer-Verlag Berlin Heidelberg, 255 p..

Jackson, C. R., 2004. An Atlas of Internal Solitary-Like Waves and Their Properties. 2nd ed., Global Ocean Associates, Washington, D. C.

Janssen, L.L.F., van der F. J. M, W., 1994. Accuracy assessment of satellite derived land cover data: a review. Photogramm. Eng. Remote Sensing 60, 419–426.

Jeans, D., Sherwin, T., 2001. The variability of the strongly non-linear solitary internal waves observed during an upwelling season on the Portuguese shelf. Continental Shelf Research 21, 1855–1878.

Jensen, J., Cowen, D., 1999. Remote sensing of urban suburban infrastructure and socio-economic attributes. Photogramm. Eng. Remote Sensing 65, 611–622.

Jia, K., Liang, S., Wei, X., Zhang, L., Yao, Y., Gao, S., 2014. Automatic land- cover update approach integrating iterative training sample selection and a Markov Random Field model. Remote Sens. Lett. 5, 148–156.

Jiang, L., Liao, M., Lin, H., Yang, L., 2009. Synergistic use of optical and InSAR data for urban impervious surface mapping: a case study in Hong Kong. Int. J. Remote Sens. 30, 2781–2796. doi:10.1080/01431160802555838

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., Xian, G., 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sens. Environ. 132, 159–175.

Kahru, M., Di Lorenzo, E., Manzano-Sarabia, M., Mitchell, B.G., 2012. Spatial and temporal statistics of sea surface temperature and chlorophyll fronts in the California Current. J. Plankton Res. 34, 749–760.

Kelley, D., 2014. oce: Analysis of Oceanographic data. R package version 0.9-14. Available at: http://CRAN.R-project.org/package=oce

Kelly, W.R., Panno, S. V., Keith Hackley, 2012. The Sources, Distribution, and Trends of Chloride in the Waters of Illinois. Illinois State Water Survey Bulletin B-74, Champaign, Illinois, 59 p..

Keogh, E., Ratanamahatana, C.A., 2004. Exact indexing of dynamic time warping. Knowledge and Information Systems 7(3), 358-386.

Kidd, C., Levizzani, V., 2011. Status of satellite precipitation retrievals. Hydrol. Earth Syst. Sci. 15, 1109–1116.

Kidd, C., Bauer, P., Turk, J., Huffman, G.J., Joyce, R., Hsu, K.-L., Braithwaite, D., 2012. Intercomparison of high-resolution precipitation products over Nortwest Europe. J. Hydrometeorol. 13, 67–83.

Kilpatrick, K. a., Podestá, G., Walsh, S., Williams, E., Halliwell, V., Szczodrak, M., Brown, O.B., Minnett, P.J., Evans, R., 2015. A decade of sea surface temperature from MODIS. Remote Sens. Environ. 165, 27–41.

Kim, K., Thompson, A.M., Botter, G., 2008. Modeling of thermal runoff response from an asphalt-paved plot in the framework of the mass response functions. Water Resour. Res. 44, 1–13.

Klein, R. D. (1979). Urbanization and Stream Quality Impairment. Water Resources Bulletin 15(4), 948-963.

Kosro, P.M., Huyer, A., Ramp, S.R., Smith, R.L., Chavez, F.P., Cowles, T.J., Abbott, M.R., Strub, P.T., Barber, R.T., Jessen, P., Small, L.F., 1991. The Structure of the Transition Zone Between Coastal Waters and the Open Ocean off Northern California, Winter and Spring 1987. J. Geophys. Res. 96 (C8), 14707-14730.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–263.

Kraas, F., 2007. Megacities and global change: Key priorities, in: Geographical Journal. pp. 79–82.

Krellenberg, K., Link, F., Welz, J., Harris, J., Barth, K., Irarrazaval, F., 2014. Supporting local adaptation : The contribution of socio-environmental fragmentation to urban vulnerability. Appl. Geogr. 55, 61–70.

Lamigueiro, O.P., Hijmans, R., 2014. rasterVis. R package version 0.31. Available at: doi:10.5281/zenodo.12394

Lande, R., Yentsch, C., 1988. Internal waves, primary production and the compensation depth of marine phytoplankton. Journal of Plankton Research 10 (3), 565–571.

Langanke, T., Büttner, G., Dufourmont, H., Iasillo, D., Probeck, M., M.Rosengren, Sousa, A., Strobl, P., Weichselbaum, J., 2013. GIO land (GMES/Copernicus initial operations land) High Resolution Layers (HRLs) – summary of product specifications.

Legeckis, R., Brown, C.W., Chang, P.S., 2002. Geostationary satellites reveal motions of ocean surface fronts. J. Mar. Syst. 37, 3–15.

Legeckis, R., 1978. A survey of worldwide sea surface temperature fronts detected by environmental satellites. Journal of Geophysical Research 83 (C9), 4501–4522.

Le Fèvre, J., 1986. Aspects of the biology of frontal systems. Adv. Mar. Biol. 23, 163– 299.

Loveland, T.R., Sohl, T.L., Stehman, S. V, Gallant, a L., Sayler, K.L., Napton, D.E., 2002. A Strategy for Estimating the Rates of Recent United States Land-Cover Changes. Photogramm. Eng. Remote Sensing 68, 1091–1099.

Li, M., Zang, S., Wu, C., Deng, Y., 2015. Segmentation-based and rule-based spectral mixture analysis for estimating urban imperviousness. Adv. Sp. Res. 55, 1307–1315.

Li, X., Zhang, Q., Xu, C., 2014. Assessing the performance of satellite-based precipitation products and its dependence on topography over Poyang Lake basin. Theor. Appl. Climatol. 115, 713–729.

Lillebø, A.I., Pardal, M.Â., Marques, J.C., 1999. Population structure, dynamics and production of Hydrobia ulvae (Pennant) (Mollusca: Prosobranchia) along an eutrophication gradient in the Mondego estuary (Portugal). Acta Oecologica 20, 289–304.

Lima, F.P., Wethey, D.S., 2012. Three decades of high-resolution coastal sea surface temperatures reveal more than warming. Nat. Commun. 3, 704.

Liu, J.W., Xie, S.P., Yang, S., Zhang, S.P., 2016. Low-cloud transitions across the Kuroshio front in the East China Sea. J. Clim. 29, 4429–4443.

Liu, Z., Ostrenga, D., Vollmer, B., Deshong, B., MacRitchie, K., Greene, M., Kempler, S., 2016b. Global Precipitation Measurement (GPM) Mission Products and Services at the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC). Bull. Am. Meteorol. Soc. BAMS–D–16–0023.1.

Liu, Z., Ostrenga, D., Teng, W., Kempler, S., Milich, L., 2014. Developing GIOVANNI-based online prototypes to intercompare TRMM-related global gridded-precipitation products, Computers and Geosciences 66,168 - 181.

Liu, Z., Ostrenga, D., Teng, W., and Kempler, S., 2012. Tropical Rainfall Measuring Mission (TRMM) Precipitation Data and Services for Research and Applications, Bulletin of the American Meteorological Society, doi: http://dx.doi.org/10.1175/BAMS-D-11-00152.1

Liu, Z., Rui, H., Teng, W.L.,Chiu, L.S.,Leptoukh, G.G., Kempler, S., 2009, Developing an Online Information System Prototype for Global Satellite Precipitation Algorithm Validation and Intercomparison. Journal of Applied Meteorology and Climatology-IPWG Special Issue, 48(12) (December 2009), 2581-2589.

Liu, Z., Rui, H., Teng, W., Chiu, L., Leptoukh, G., Vicente G., 2007, Online visualization and analysis: A new avenue to use satellite data for weather, climate and interdisciplinary research and applications. Measuring Precipitation from Space - EURAINSAT and the future, Advances in Global Change Research 28, 549-558.

Longhurst, A.R., 2007. Ecological Geography of the Sea (2nd Edition). Academic Press, Burlington, USA, 542 p..

Longhurst, A. (1995), Seasonal cycles of pelagic production and consumption, Prog. Oceanogr. 36(2), 77–167.

Łuczak, M., 2016. Hierarchical clustering of time series data with parametric derivative dynamic time warping. Expert Syst. Appl. 62, 116–130.

Macary, F., Morin, S., Probst, J.-L., Saudubray, F., 2014. A multi-scale method to assess pesticide contamination risks in agricultural watersheds. Ecol. Indic. 36, 624–639.

Malila, W.A. Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat. In *Proceedings of 6th Annual Symposium on Machine Processing of Remotely Sensed Data Soil Information Systems and Remote Sensing and Soil Survey*, West Lafayette, IN, USA, 3–6 June 1980; pp. 326-335.

Mandel, P., Maurel, M., Chenu, D., 2015.Better understanding of water quality evolution in water distribution networks using data clustering., Water Research 87, 69-78.

Mann, K.H., Lazier, J.R.N. , 1991. Dynamics of Marine Ecosystems: Biological-physical Interactions in the Oceans. Blackwell Scientific Publications, Oxford, 466 p..

Mantas, V.M., Liu, Z., Caro, C., Pereira, A. J.S.C., 2015. Validation of TRMM multi-satellite precipitation analysis (TMPA) products in the Peruvian Andes. Atmos. Res. 163, 132–145.

Mantas, V.M., Pereira, A.J.S.C., Neto, J., Patrício, J., Marques, J.C., 2013. Monitoring estuarine water quality using satellite imagery. The Mondego river estuary (Portugal) as a case study. Ocean Coast. Manag. 72, 13–21.

Mantas, V.M., Pereira, A.J.S.C., Liu, Z., 2013b. Development of a Web Service and Android 'App' for the distribution of satellite rainfall estimates.Proceedings of the ESA Living Planet Symposium, ESA SP-722, 9–13 September 2013, Edinburgh in United Kingdom.

Marchesiello, P., Estrade,P. , 2007. Eddy activity and mixing in upwelling systems: a comparative study of Northwest Africa and California regions. International Journal of Earth Sciences 98 (2), 299 - 308.

Margalef, R., 1997. Our biosphere. Volume 10 of Excellence in Ecology. Inter-Research, Oldendorf, Germany, 176 p..

Marinoni, O., Higgins, A., Coad, P., Navarro Garcia, J., 2013. Directing urban development to the right places: Assessing the impact of urban development on water quality in an estuarine environment. Landsc. Urban Plan. 113, 62–77.

Marques, J.C., Graça, M.A., Pardal, M.A., 2002. Introducing the Mondego river basin. *In*: A.M. Pardal, J.C. Marques, M.A.S. Graça (Eds.) Aquatic ecology of the Mondego river basin. Global importance of local experience. Imprensa da Universidade de Coimbra, Portugal, 7-12.

Marques, J.C., Nielsen S.N., Pardal M.A., Jørgensen S.E., 2003. Impact of eutrophication and river management within a framework of ecosystem theories. Ecol Model 166, 147–168.

Marques, J.C., Basset A, Brey T, Elliott M., 2009. The ecological sustainability trigon — a proposed conceptual framework for creating and testing management scenarios. ar Pollut Bull 58, 1773–1779.

Marta-Almeida, M., Dubert, J., Peliz, A., 2002. Simulations of extreme shelf current along the North-Western Iberian Shelf forced by wind and river runoff. *In*: Proceedings of the 3rd Iberian meeting on Geophysics and Geodesy, Valencia, February 4–8, pp. 1555–1559.

Martínez-Fernández, J., Ruiz-Benito, P., Zavala, M. a., 2015. Recent land cover changes in Spain across biogeographical regions and protection levels: Implications for conservation policies. Land use policy 44, 62–75.

Martins, C., Hamann, M., Fiú za, A., 2002. Surface circulation in the eastern North Atlantic, from drifters and altimetry. Journal of Geophysical Research 107 (C12), 3217.

Maryland General Assembly, 2012. House Bill 987, Stormwater Management - Watershed Protection and Restoration Program.

Masek, J.G., E.F. Vermote, N. Saleous, R. Wolfe, F.G. Hall, F. Huemmrich, F. Gao, J. Kutler, and T.K. Lim. 2013. LEDAPS Calibration, Reflectance, Atmospheric Correction Preprocessing

Code, Version 2. Model product. Available at: http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

Mason, E., Coombs, S., Oliveira, P.B., 2006. An overview of the literature concerning the oceanography of the eastern North Atlantic region. Relatórios Científicos e Técnicos IPIMAR, Série Digit. 33, 58 p..

Matzler, C., Standley, A., 2000. Relief effects for passive microwave remote sensing. Int. J. Remote Sens. 21, 2403–2412.

Mavor, T.P., Bisagni, J.J., 2001. Seasonal variability of sea surface temperature fronts on Georges Bank. Deep-Sea Research II 48, 215–243.

May, D. A., Parmeter, M. M. , Olszewski, D. S. , McKenzie, B. D., 1998. Operational processing of satellite sea surface temperature retrievals at the naval oceanographic office. Bulletin of the American Meteorological Society 79, 397-407.

McClain, C.R., 2009. A decade of satellite ocean color observations. Ann. Rev. Mar. Sci. 1, 19–42.

McClure, C.J.W., Korte,A.C., Heath,J.A., Barber,J.R., 2015.Pavement and riparian forest shape the bird community along an urban river corridor 4, 291-310.

McGranahan, D., Cromartie, J., Wojan, T., 2010. Nonmetropolitan outmigration counties: some are poor, many are prosperous. United States Department of Agriculture (USDA), Economic Research Service, ERR - 107, 29 pp.

McIntosh, J., Yuan, M., 2005. Assessing similarity of geographic processes and events. Trans. GIS 9 (2), 223–245.

Mehaffey, M.H., Nash, M.S., Wade, T.G., Ebert, D.W., Jones, K.B., Rage, R.A., 2005. Linking land cover and water quality in New York City‘s water supply watersheds. Environ. Monit. Assess., 107, 29–44.

Meng, J., Li, L., Hao, Z., Wang, J., Shao, Q., 2014. Suitability of TRMM satellite rainfall in driving a distributed hydrological model in the source region of Yellow River. J. Hydrol. 509, 320–332.

Miller, P.I., 2004. Multispectral front maps for automatic detection of ocean color features from SeaWiFS. International Journal of Remote Sensing 25(7-8), 1437-1442.

Miller, P., 2009. Composite front maps for improved visibility of dynamic sea-surface features on cloudy SeaWiFS and AVHRR data. J. Mar. Syst. 78, 327–336.

Miller, P., 2011. Case Study 16: Detection and Visualization of Oceanic Fronts. *In*: J. Morales, V. Stuart, T. Platt, S. Sathyendranath (Eds.) Handbook of Satellite Remote Sensing Image Interpretation: Applications for Marine Living Resources Conservation and Management. Dartmouth, EU PRESPO & IOCCG, 229-239.

Miltner, R., D. White and C. Yoder (2004). The biotic integrity of Streams in urban and suburbanizing landscapes. Landscape and Urban Planning 69(1), 87-100. doi:

Mobley, C.D., Stramski, D., Bissett, W.P., Boss, E., 2004. Optical modeling of ocean waters: Is the Case 1 - Case 2 classification still useful? Oceanography 17, 60–67.

Moita, T., Oliveira, P.B., Mendes, J.C., Palma, A.S., 2003. Distribution of chlorophyll-a and *Gymnodinium catenatum* associated with coastal upwelling plumes off central Portugal. Acta Oecologica 24, 125-132.

Monteiro, M., Tavares, A.O., 2015. What is the Influence of the Planning Framework on the Land Use Change Trajectories? Photointerpretation Analysis in the 1958–2011 Period for a Medium/Small Sized City. Sustainability 7, 11727-11755.

Mooers, C. N. K., Flagg, C.N., Boicourt, W.C., 1978. Prograde and retrograde fronts, *In*: M. Bowman, W. Esias (Eds.) Oceanic Fronts in Coastal Processes, Springer, New York, 43–58.

Moore, P. G., 1977. Inorganic particulate suspensions in the sea and their effects on marine animals. Oceanogr. Mar. Biol. Ann. Rev. 15, 225-363.

Morgan, R.P., Kline, K.M., Kline, M.J., Cushman, S.F., Sell, M.T., Weitzell,R.E., Churchill,J.B., 2012.Stream Conductivity: Relationships to Land Use, Chloride, and Fishes in Maryland Streams, North American Journal of Fisheries Management 32(5), 941-952.

Morisawa, Marie, and Ernest LaFlure. 1979. Hydraulic geometry, stream equilibrium and urbanization. *In*: D. D. Rhodes, G. P. Williams (Eds.) Adjustments of the fluvial systems—Proceedings of the 10th annual Geomorphology Symposium Series. Binghamton, NY.

Moutin, T., Karl, D., Duhamel, S., Rimmelin, P., Raimbault, P., Van Mooy, B.A.S., Claustre, H., 2008. Phosphate availability and the ultimate control of new nitrogen input by nitrogen fixation in the tropical Pacific Ocean. Biogeosciences 5, 95−109.

Muacho, S., Silva, J.C.B., Brotas, V., Oliveira, P.B., 2013. Deep-Sea Research I Effect of internal waves on near-surface chlorophyll concentration and primary production in the Nazaré Canyon (west of the Iberian Peninsula). Deep. Res. Part I 81, 89–96.

Mueller, J. L., 2000. SeaWiFS algorithm for the diffuse attenuation coefficient, K(490), using water-leaving radiances at 490 and 555 nm, *In*: S.B. Hooker, E.R. Firestone (Eds.) SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3, SeaWiFS Postlaunch Tech. Rep. Ser., vol. 11, NASA Goddard Space Flight Cent., Greenbelt, Md., 24 – 27.

Mullaney, J.R., Lorenz, D.L., Arntson, A.D., 2009. Chloride in groundwater and surface water in areas underlain by the glacial aquifer system, northern United States [electronic resource] / by John R. Mullaney, David L. Lorenz, and Alan D. Arntson. U.S. Geological Survey Reston, Va.

NASA, Retrieved December 2015, Available at: http://oceancolor.gsfc.nasa.gov/cms/atbd/chlor_a#sec_2

NASA, Retrieved April 2016, available at: https://modis.gsfc.nasa.gov/about/specifications.php

Nevitt, G., 1999. Olfactory foraging in Antarctic seabirds: a species- specific attraction to krill odours. Mar. Ecol., Prog. Ser. 177, 235– 241.

Nieblas, A.E., Drushka, K., Reygondeau, G., Rossi, V., Demarcq, H., Dubroca, L., Bonhommeau, S., 2014. Defining Mediterranean and Black Sea biogeochemical subprovinces and synthetic ocean indicators using mesoscale oceanographic features. PLoS One 9 (10), e111251.

NOAA, 2004. Report on the Delineation of Regional Ecosystems. Regional Ecosystem Delineation Workshop, Charleston, SC, 31 August - 1 September 2004. Washington, D.C..

Nykjær, L., Van Camp, L., 1994. Seasonal and interannual variability of coastal upwelling along northwest Africa and Portugal from 1981 to 1991. Journal of Geophysical research 99, C7, 197 - 207.

Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC), 2011. MODIS subsetted land products, Collection 5. Retrieved January 2014, Available at: http://daac.ornl.gov/MODIS/modis.html from ORNL DAAC, Oak Ridge, Tennessee, U.S.A.,

Ocean Biology Processing Group (NASA), OBPG, Retrieved September 2016, available at: http:// oceancolor.gsfc.nasa.gov .

Oliveira, M.R.L., Monteiro, M.T. (1992). "Blooms" de Cyanophyceae na Albufeira da Aguieira: Efeitos na qualidade da água e no zooplâncton. Relatório Técnico Científico INIP (61), 57p.

Oliveira, P.B., Nolasco, R., Dubert, J., Moita, T., Peliz, Á., 2009. Surface temperature, chlorophyll and advection patterns during a summer upwelling event off central Portugal. Cont. Shelf Res. 29, 759–774.

Oliphant, T.E., 2007. Python for scientific computing, Comput. Sci. Eng. 9, 10-20.

Olson, D.B., Backus, R.H., 1985. The concentrating of organisms at fronts: a coldwater fish and a warm-core Gulf Stream ring. J. Mar. Res. 43, 113– 137.

Olson, D.B., 2002. Biophysical dynamics of ocean fronts. Biological –Physical Interactions in the Sea. *In*: Allan R. Robinson, James J. McCarthy, Brian J. Rothschild (Eds.) The Sea, vol. 12. Wiley, New York, USA.

O'Reilly, Tom, 2007. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, Communications & Strategies 65, 17 - 37.

O'Reilly, J. E., et al. (2000). In S. B. Hooker, & E. R. Firestone (Eds.), SeaWiFS postlaunch calibration and validation analyses, part 3. NASA Tech. Memo. 2000-206892, 11. NASA Goddard Space Flight Center, 49 pp.

Oregon Coastal Atlas , Retrieved September 2016, available at: http://www.coastalatlas.net/.

OSM, Open Street Map, Retrieved January 2015, Available at: http://www.openstreetmap.org.

Otero, P., Ruiz-Villarreal, M., Peliz, Á., 2009. River plume fronts off NW Iberia from satellite observations and model data. ICES J. Mar. Sci. 66, 1853–1864.

Owen, Robert W., Jr., 1981. Fronts and eddies in the sea: Mechanisms, interactions and biological effects. *In*: A. R. Longhurst (Ed.) Analysis of marine ecosystems. Academic Press, London, 197-233.

Pacifici, F., 2014.The Importance of Physical Quantities for the Analysis of Multitemporal and Multiangular Optical Very High Spatial Resolution Images.IEEE Transactions on Geoscience and Remote Sensing, 52, 10, 6241 – 6256.

Panagopoulos, T., Barreira, A.P., 2013. Understanding the Shrinkage Phenomenon in Portugal. WSEAS Trans. Environ. Dev. 9, 1 – 12.

Pardal, M. A., Marques, J.C., Metelo, I., Lillebø, A. I., Flindt, M.R., 2000. Impact of eutrophication on the life cycle, population dynamics and production of *Ampithoe valida* (Amphipoda) along an estuarine spatial gradient (Mondego estuary, Portugal). Mar. Ecol. Prog. Ser. 196, 207–219.

Parece, T.E., Campbell, J.B., 2013.Comparing Urban Impervious Surface Identification Using Landsat and High Resolution Aerial Photography. Remote Sensing, 5, 4942-4960.

Passow, M., 2010. TRMM - Tropical Rainfall Measuring Mission: Bringing remote sensing of precipitation into your classroom, Terrae didat. 6(1), 3 - 8.

Patino, J.E., Duque, J.C., Pardo-Pascual, J.E., Ruiz, L. a., 2014. Using remote sensing to assess the relationship between crime and the urban layout. Appl. Geogr. 55, 48–60.

Paul, M., J. Meyer and C. Couch (2006). Leaf breakdown in streams differing in catchment land use. Freshwater Biology 51(9), 1684-1695.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space, Philosophical Magazine 6(2), 559-572.

Peliz, Á., Rosa, T.L., Santos, a. M., Pissarra, J.L., 2002. Fronts, jets, and counter-flows in the Western Iberian upwelling system. J. Mar. Syst. 35, 61–77.

Peliz, Á., Santos, A. M.P., Oliveira, P.B., Dubert, J., 2004. Extreme cross-shelf transport induced by eddy interactions southwest of Iberia in winter 2001. Geophys. Res. Lett. 31, 2–5.

Peliz, Á., Dubert, J., Santos, A. M.P., Oliveira, P.B., Le Cann, B., 2005. Winter upper ocean circulation in the Western Iberian Basin - Fronts, Eddies and Poleward Flows: An overview. Deep. Res. Part I Oceanogr. Res. Pap. 52, 621–646.

Pelleg, D., Moore, A., 2000. X-means: Extending K-means with efficient estimation of the number of clusters ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning, 727 - 734.

Perez, F. F., Castro, C. G. , Alvarez-Salgado, X.A., Rios, A.F., 2001. Coupling between the Iberian basin-scale circulation and the Portugal boundary current system: a chemical study, Deep-Sea Research I 48, 1519-1533.

Petitjean, F., Inglada, J., Gançarski, P., 2012. Satellite image time series analysis under time warping. IEEE Trans. Geosci. Remote Sens. 50 (8), 3081–3095.

Petus, C., Chust, G., Gohin, F., Doxaran, D., Froidefond, J., Sagarminaga, Y., 2010. Estimating turbidity and total suspended matter in the Adour River plume (South Bay of Biscay) using MODIS 250-m imagery, Continental Shelf research 30(5), 379-392.

Picado, A., Alvarez, I., Vaz, N., Dias, J.M., 2013. Chlorophyll concentration along the northwestern coast of the Iberian Peninsula vs. atmosphere-ocean- land conditions. J. Coast. Res. 2047–2052.

Picart, S., Sathyendranath, S., Dowell, M., Moore, T., Platt, T., 2013. Remote sensing of assimilation number for marine phytoplankton. Remote Sens. Environ. 146, 87–96.

Pickett, S.T. A., Cadenasso, M.L., Grove, J.M., Groffman, P.M., Band, L.E., Boone, C.G., Burch, W.R., Grimmond, C.S.B., Hom, J., Jenkins, J.C., Law, N.L., Nilon, C.H., Pouyat, R. V., Szlavecz, K., Warren, P.S., Wilson, M. A., 2008. Beyond Urban Legends: An Emerging Framework of Urban Ecology, as Illustrated by the Baltimore Ecosystem Study. Bioscience 58, 139.

Pilli, R., 2012. Calibrating CORINE land cover 2000 on forest inventories and climatic data: An example for Italy. Int. J. Appl. Earth Obs. Geoinf. 19, 59–71.

Pingree, R., Le Cann, B., 1990. Structure, strength and seasonality of the slope currents in the Bay of Biscay region. Journal of the Marine Biology Association of UK 70, 857–885.

Polovina, J.J., Howell, E., Kobayashi, D.R., Seki, M.P., 2001. The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. Progress in Oceanography 49, 469–483.

Prodran, R., Sperk, M., 2013. Scientific computing with Google App Engine, Future Generation Computer Systems 29, 1851 - 1859.

Polovina, J.J., Howell, E. a., Kobayashi, D.R., Seki, M.P., 2015. The Transition Zone Chlorophyll Front updated: Advances from a decade of research. J. Math. Anal. Appl. 1–7. (Corrected Proof).

Polovina, J.J., Howell, E., Kobayashi, D.R., Seki, M.P., 2001. The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. Progress in Oceanography 49, 469–48.

Porcù, F., Milani, L., Petracca, M., 2014. On the uncertainties in validating satellite instantaneous rainfall estimates with raingauge operational network. Atmos. Res. 144, 73–81.

Platt, T., Sathyendranath, S., 1988. Oceanic primary production: Estimation by remote sensing at local and regional scales. Science 241, 1613-1620

Portugal, Decreto-Lei N.º 52/2012 – Diário da República N.º 48/2012, Série I de 2012-03-07

Quaresma, L.S., Pichon, A., 2013. Modelling the barotropic tide along the West-Iberian margin, Journal of Marine Systems 109-110 (Supplement), S3-S25.

Quinlan, J.R., 1992. Learning with Continuous Classes. *In*: 5th Australian Joint Conference on Artificial Intelligence, World Scientific, Singapore, 343-348.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

Rabalais, N. N., Turner, R. E., Díaz, R. J., Justić, D., 2009. Global change and eutrophication of coastal waters. – ICES Journal of Marine Science 66, 1528–1537.

Raciti, S.M., Hutyra, L.R., Newell, J.D., 2014. Mapping carbon storage in urban trees with multi-source remote sensing data: Relationships between biomass, land use, and demographics in Boston neighborhoods. Sci. Total Environ. 500-501, 72–83.

Reese, D.C., O'Malley, R.T., Brodeur, R.D., Churnside, J.H., 2011. Epipelagic fish distributions in relation to thermal fronts in a coastal upwelling system using high-resolution remote-sensing techniques. ICES Journal of Marine Science 68(9), 1865-1874.

Relvas, P., Luís, J., Santos, A. M.P., 2009. Importance of the mesoscale in the decadal changes observed in the northern Canary upwelling system. Geophys. Res. Lett. 36, 1–4.

Relvas, P., Barton, E.D., Dubert, J., Oliveira, P.B., Peliz, Á., da Silva, J.C.B., Santos, A. M.P., 2007. Physical oceanography of the western Iberia ecosystem: Latest views and challenges. Prog. Oceanogr. 74, 149–173.

Ribeiro, A., Peliz, A., Santos, A., 2005. A study of the response of chlorophyll-a biomass to a winter upwelling event off Western Iberia using SeaWiFS and in situ data. Journal of Marine Systems 53, 87–107.

Ridd, M.K., 1995. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. Int. J. Remote Sens. 16, 2165–2185.

Roberts, J.J., Best, B.D., Dunn, D.C., Treml, E.A., Halpin, P.N., 2010. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. Environmental Modelling & Software 25: 1197-1207.

Robinson, I.S., 2010. Discovering the oceans from space: the unique applications of satellite oceanography. Springer-Praxis, Chichester. 638 p..

Rodwell, L.D., Roberts, C.M., 2004. Fishing and the impact of marine reserves in a variable environment. Canadian Journal of Fisheries and Aquatic Sciences 61, 2053-2068.

Roden, G., 1991. Subarctic-subtropical transition zone of the North Pacific: largescale aspects and mesoscale structure. *In*: J.A. Wetherall (Ed.), Biology, Oceanography, and Fisheries of the North Pacific Transition Zone and Subarctic Frontal Zone. Hon. Lab, SWFSC, Nat. Mar. Fish. Ser. Tech. Rpt. NMFS 105, 1–38.

Roed, L.P., Shi, X.B., 1999. A numerical study of the dynamics and energetics of cool filaments, jets and eddies off the Iberian peninsula. Journal of Geophysical Research 104, 29817–29841.

Roff, J.C., Taylor, M.E., 2000. National frameworks for marine conservation —a hierarchical geophysical approach. Aquatic Conservation: Marine and Freshwater Ecosystems 10, 209–23.

Roff, J.C., Taylor, M.E., Laughren, J., 2003. Geophysical approaches to the classification, delineation and monitoring of marine habitats and their communities. Aquatic Conservation: Marine and Freshwater Ecosystems 13, 77–90.

Roffer's, Retrieved September 2016, available at: https://www.roffs.com/

Rose, S., and N. E. Peters. 2001. Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. Hydrological Processes 15, 1441–1457.

van Rossum, G.,Drake, F.L. (Eds.), 2001. Python Reference Manual, PythonLabs, Virginia, USA, 2001. Available at http://www.python.org

Roy, A. H., A. D. Rosemond, M. J. Paul, D. S. Leigh and J. B. Wallace (2003). Stream macroinvertebrate response to catchment urbanisation (Georgia, USA). Freshwater Biology 48(2), 329-346.

Rubel, F., Kottek, M., 2010: Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. Meteorol. Z. 19, 135-141.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithmoptimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. ASSP-26, 43–49.

Santos, A. M.P., Peliz, A., Dubert, J., Oliveira, P.B., Angélico, M.M., Ré, P., 2004. Impact of a winter upwelling event on the distribution and transport of sardine (Sardina pilchardus) eggs and larvae off western Iberia: A retention mechanism. Cont. Shelf Res. 24, 149–165.

Santos, A.M.P., Kazmin, A.S., Peliz, A., 2005. Decadal changes in the Canary upwelling system as revealed by satellite observations: their impact on productivity. Journal of Marine Research 63, 359–379.

Santos, A. M.P., Chícharo, A., Dos Santos, A., Moita, T., Oliveira, P.B., Peliz, Á., Ré, P., 2007. Physical-biological interactions in the life history of small pelagic fish in the Western Iberia Upwelling Ecosystem. Prog. Oceanogr. 74, 192–209.

Santos, F., Gomez Gesteira, M., deCastro, M., 2011. Coastal and oceanic SST variability along the western Iberian Peninsula. Cont. Shelf Res. 31, 2012–2017.

Sapiano, M., Arkin, P., 2009. An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. J. Hydrometeorol. 10 (1), 149–166.

Saraceno, M., Provost, C., Piola, A.R., 2005. On the relationship between satellite-retrieved surface temperature fronts and chlorophyll a in the western South Atlantic. J. Geophys. Res. Ocean. 110, 1–16.

Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., et al. (2009). Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. Marine Ecology Progress Series, 383, 73–84.

Scheel, M.L.M., Rohrer, M., Huggel, C., Santos Villar, D., Silvestre, E., Huffman, G.J., 2011. Evaluation of TRMM Multi-Satellite Precipitation Analysis (TMPA) performance in the Central Andes region and its dependency on spatial and temporal resolution. Hydrol. Earth Syst. Sci. 15, 2649–2663.

Schluessel, P., Emery, W.J., Grassl, H., Mannen, T., 1990. On the Bulk-Skin Temperature Difference and Its Impact on Satellite Remote Sensing of Sea Surface Temperature. J. Geophys. Res. 95, 13,341–13,356.

Schueler, T., Fraley-McNeal, L., Cappiella, K., 2009. Is Impervious Cover Still Important? Review of Recent Research. J. Hydrol. Eng. 14, 309–315.

Schwarz, N., 2010. Urban form revisited—Selecting indicators for characterising European cities. Landsc. Urban Plan. 96, 29–47.

Scofield, R.A., Kuligowski, R.J., 2003. Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. Weather Forecast. 18, 1037–1051.

Serra, N., Ambar, I., 2002. Eddy generation in the mediterranean undercurrent. Deep-Sea Research II 49, 4225–4243.

Sexton, J.O., Song, X.P., Huang, C., Channan, S., Baker, M.E., Townshend, J.R., 2013. Urban growth of the Washington, D.C.-Baltimore, MD metropolitan region from 1984 to 2010 by annual, Landsat-based estimates of impervious cover. Remote Sens. Environ. 129, 42–53.

Shang, S., Lee, Z., Shi, L., Lin, G., Wei, G., Li, X., 2016. Changes in water clarity of the Bohai Sea: Observations from MODIS. Remote Sens. Environ. 186, 22–31.

Sharif, S.M., Kusin, F.M., Asha'ari, Z.H., Aris, A.Z., 2015. Characterization of Water Quality Conditions in the Klang River Basin, Malaysia Using Self Organizing Map and K-means Algorithm. Procedia Environmental Sciences 30, 73-78.

Sherman, K., Alexander, L.M., Gold, B.D. (Eds.), 1990. Large Marine Ecosystems: Patterns, Processes and Yields. Amer. Assoc. Adv. Sci. Washington D.C. 242p.

Sherman, K., 2005. A modular strategy for recovery and management of biomass yields in large marine ecosystems. *In*: E. Levner, I. Linkov, J. Proth (Eds), Strategic Management of Marine Ecosystems. NATO Science Series. Series IV: Earth and Environmental Series – Vol. 50. Springer. 65-80.

Sherman, K., Celone, P., Adams, S., 2004. NOAA Fisheries Service's Large Marine Ecosystems Program: Status Report U. S., NOAA Technical Memorandum NMFS-NE-183, NOAA, Woods Hole, Massachusetts, 30 p..

Sherman, K., Aquarone, M.C., Adams, S. (Eds.) 2009. Sustaining the World's Large Marine Ecosystems. Gland, Switzerland, IUCN., 142p.

Shi, W., Wang, M., 2010. Satellite observations of the seasonal sediment plume in central East China Sea. J. Mar. Syst. 82, 280–285.

Silva, A., Palma, S., Moita, M.T., 2008. Coccolithophores in the upwelling waters of Portugal: Four years of weekly distribution in Lisbon bay. Cont. Shelf Res. 28, 2601–2613.

da Silva, J.C.B., Ermakov, S.A., Robinson, I.S., Jeans, D.R.G., Kijashko, S.V., 1998. Role of surface films in ERS SAR signatures of internal waves on the shelf. 1. Short period internal waves. Journal of Geophysical Research: Oceans 103 (C4), 8009–8031.

da Silva, A.J., 1992. Dependance of upwelling related circulation on wind forcing and stratification over the Portuguese northern shelf. ICES C.M./C:17, Hydrography Committee.

Slonecker, E.T., Jennings, D.B., Garofalo, D., 2001. Remote sensing of impervious surfaces: A review. Remote Sens. Rev. 20, 227–255.

Small, J., 2002. Internal tide transformation across a continental slope off Cape Sines, Portugal. Journal of Marine Systems 32, 43-69.

Smale, D. a., Wernberg, T., 2009. Satellite-derived SST data as a proxy for water temperature in nearshore benthic ecology. Mar. Ecol. Prog. Ser. 387, 27–37.

Smyth, T.J., Miller, P.I., Groom, S.B., Lavender, S.J., 2001. Remote sensing of sea surface temperature and chlorophyll during Lagrangian experiments at the Iberian margin. Progress in Oceanography 51, 269-281.

SNIRH – Sistema Nacional de Informação de Recursos Hídricos, Retrieved July 2016, Available at: http://www.snirh.pt.

Song, Y., Li, F., Wang, X, Xu, C., Zhang, J., Liu, X., Zhang, H., 2015. The effects of urban impervious surfaces on eco-physiological characteristics of Ginkgo biloba: A case study from Beijing, China. Urban Forestry & Urban Greening, 14:4, 1102-1109.

Sordo, I., Barton, E.D., Cotos, J.M., Pazos, Y., 2001. An Inshore Poleward Current in the NW of the Iberian Peninsula Detected from Satellite Images, and its Relation with *G. catenatum* and *D. acuminata* Blooms in the Galician Rias. Estuarine, Coastal and Shelf Science 53, 787–799.

Spalding, M.D., Agostini, V.N., Rice, J., Grant, S.M., 2012. Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. Ocean Coast. Manag. 60, 19–30.

Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z. A., Finlayson, M., Halpern, B.S., Jorge, M. A., Lombana, A., Lourie, S. A., Martin, K.D., Mcmanus, E., Molnar, J., Recchia, C. A., Robertson, J., 2007. Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. Bioscience 57, 573.

Stathopoulou, M., Cartalis, C., 2007. Daytime urban heat islands from Landsat ETM+ and CORINE land cover data: An application to major cities in Greece. Sol. Energy 81, 358–368.

Steele, M.K., Aitkenhead-Peterson, J. A., 2011. Long-term sodium and chloride surface water exports from the Dallas/Fort Worth region. Sci. Total Environ. 409, 3021–3032.

Stramma, L., 1984: Geostrophic transport in the warm water sphere of the eastern subtropical North Atlantic. Journal of Marine Research 42, 537-558.

Stramma, L., 2001. Current systems in the Atlantic Ocean. *In*: J.H. Steele, K.K. Turekian, S.A. Thorpe (Eds.), Encyclopaedia of Ocean Sciences, Academic Press, London, 589-598.

Stravs, L., Brilly, M., Sraj, M., 2008. Precipitation Interception Modelling Using Machine Learning Methods – The Dragonja River Basin Case Study. Practical Hydroinformatics 68, 347-358.

Su, F., Hong, Y., Lettenmaier, D., 2007. Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin. Journal of Hydrometeorology 9, 622 – 640.

Suau-Sanchez, P., Burghouwt, G., Pallares-Barbera, M., 2014. An appraisal of the CORINE land cover database in airport catchment area analysis using a GIS approach. J. Air Transp. Manag. 34, 12–16.

Sun, R., Chen, L., Chen, W., Ji, Y., 2013. Effect of Land-Use Patterns on Total Nitrogen Concentration in the Upstream Regions of the Haihe River Basin, China. Environ. Manag. 51, 45–58.

Szekielda, K.H., 1976. Spacecraft oceanography, In: Harold Barnes (Ed.) Oceanography and Marine Biology Annual Review, 14, Aberdeen University Press, Aberdeen, SD, 99-166.

Tapia, F.J., Navarrete, S.A., Castillo, M., Menge, B.A., Castilla, J.C., Largier, J., Wieters, E.A., Broitman, B.L., Barth, J.A., 2009. Thermal indices of upwelling effects on inner-shelf habitats. Prog. Oceanogr. 83, 278–287.

Tapiador, F., Turk, F.J., Petersen, W., Hou, A., García-Ortega, E., Machado, L., Angelia, C., Salio, P., Kidd, C., Huffman, G., de Castro,M., 2012. Global precipitationmeasurement: Methods, datasets and applications. Atmos. Res. 104–105, 70–97.

Tavares, A.O., Pato, R.L., Magalhães, M.C., 2012. Spatial and temporal land use change and occupation over the last half century in a peri-urban area. Applied Geography 34, 432-444.

Teixeira, Z., Teixeira, H., Marques, J.C., 2014. Systematic processes of land use/land cover change to identify relevant driving forces: Implications on water quality. Sci. Total Environ. 470-471, 1320–1335.

Tenore, K. R., Alonso-Noval, M., Alvarez-Ossorio, M., Atkinson, L. P., Cabanas, J. M., Cal, R. M., Campos, H. J., et al., 1995. Fisheries and oceanography off Galicia, NW Spain (FOG): mesoscale spatial and temporal changes in physical processes and resultant patterns of biological productivity. Journal of Geophysical Research 100, 10943–10966.

Thomas, A., Byrne, D., Weatherbee, R., 2002. Coastal sea surface temperature variability from Landsat infrared data. Remote Sens. Environ. 81, 262–272.

Thomas, N., Hendrix, C., Congalton,R.G., 2003.A Comparison of Urban Mapping Methods Using High-Resolution Digital Imagery. Photogrammetric Engineering & Remote Sensing, 69:9, 963-972.

Thomas, A.C., Townsend, D.W., Weatherbee, R., 2003. Satellite-measured phytoplankton variability in the Gulf of Maine. Cont. Shelf Res. 23, 971–989.

Tiner, R.W., 2004. Remotely-sensed indicators for monitoring the general condition of "natural habitat" in watersheds: an application for Delaware's Nanticoke River watershed. Ecol. Indic. 4, 227–243.

Tittensor, D., Mora, C., Jetz, W., Lotze, H., Ricard, D., Vanden Berghe, E., Worm, B., 2010. Global patterns and predictors of marine biodiversity across taxa. Nature 466, 1098–U107.

Torres, R., Barton, E.D., Miller, P., Fanjul, E., 2003. Spatial patterns of wind and sea surface temperature in the galician upwelling region. Journal of Geophysical Research 108 (C4), 3130.

Turbak, F., 2012. Blocks languages for creating tangible artifacts, Visual Languages and Human-Centric Computing (VL/HCC), 2012 IEEE Symposium, Sept. 30 2012-Oct. 4 2012.

Turner, B.L., Lambin, E.F., Reenberg, A., 2007. The emergence of land change science for global environmental change and sustainability. Proc. Natl. Acad. Sci. 104, 20666–20671.

UCS - Union of Concerned Scientists Satellite Database. Retrieved January 2016, Available at: http://www.ucsusa.org/nuclear-weapons/space-weapons/satellite-database#.WA840PkrLIU

Uda, M., 1938. Researches on "siome" or current rip in the seas and oceans. Geophys. Mag. 11(4), 306-372.

Ullman, D. S., Cornillon, P.C., 1999. Satellite-derived sea surface temperature fronts on the continental shelf off the northeast U.S. coast. J. Geophys. Res. 104, 23459 - 23478.

United Nations, 2014. World Urbanization Prospects, the 2014 Revision (ST/ESA/SER.A/352). United Nations, Department of Economic and Social Affairs, Population Division, New York, 27 p.

United Nations, 2012. New York: United Nations. The Millennium Development Goals report 2012. 68 p..

United Nations Environment Programme, 2012. Early warning systems: A state of the art analysis and future directions, Environmental Development 4, 136 - 171.

UNEP-WCMC, 2006. Global Marine Assessments: A Survey of Global and Regional Assessments and Related Scientific Activities of the Marine Environment. *In*: UNEP-WCMC Biodiversity Series, vol. 27.

United States Department of Agriculture, 2013. Summary Report: 2010 National Resources Inventory. Natural Resources Conservation Service and Center for Survey Statistics and Methodology, Iowa State University, Washington, DC.

United States Geological Survey, 2015. Provisional Landsat 8 Surface reflectance Product.Product Guide.Version 1.3. Department of the Interior, United States Geological Survey, 27 p.

Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K., Aonashi, K., Inoue, T., Takahashi, N., Iguchi, T., Morimoto, T., Kawasaki, Z., 2009. A Kalman Filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined Passive Microwave and Infrared Radiometric Data, Journal of the Meteorological Society of Japan 87A, 137 – 151.

Van Aken, H.M., 2001. The hydrography of the mid-latitude Northeast Atlantic Ocean—Part III: the subducted thermocline water mass. Deep Sea Research I 48, 237–267.

Van der Linden, S., Hostert, P., 2009. The influence of urban structures on impervious surface maps from airborne hyperspectral data. Remote Sens. Environ. 113, 2298–2305.

Vanhellemont, Q., Ruddick, K., 2015. Advantages of high quality SWIR bands for ocean color processing: Examples from Landsat-8. Remote Sens. Environ. 161, 89–106.

Vanhellemont, Q., Ruddick, K., 2016. Acolite for Sentinel-2 : Aquatic Applications of Msi Imagery. ESA Living Planet Symp. 9–13.

Veach, A.M., Bernot, M.J., 2011. Temporal variation of pharmaceuticals in an urban and agriculturally influenced stream. Sci. Total Environ. 409, 4553–4563.

Veit, R.R., 1995. Pelagic communities of seabirds in the South Atlantic Ocean. Ibis 137, 1 – 10.

Vietz, G.J., Sammonds, M.J., Walsh, C.J., Fletcher, T.D., Rutherfurd, I.D., Stewardson, M.J., 2014. Ecologically relevant geomorphic attributes of streams are impaired by even low levels of watershed effective imperviousness. Geomorphology 206, 67–78.

Villéon, L.P., Carval, T., Blouch, P., Duporte, E., 2003. CORIOLIS: Providing a data management infrastructure for operational oceanography. Elsevier Oceanography Series 69, 611 - 614.

Walsh, C.J., Roy, A.H., Feminella, J.W., Cottingham, P.D., Groffman, P.M., Morgan, R.P., 2005. The urban stream syndrome: current knowledge and the search for a cure. J. North Am. Benthol. Soc. 24, 706–723.

Wang, Z., Li, L., Chen, D., Xu, K., Wei, T., Gao, J., Zhao, Y., Chen, Z., Masabate, W., 2007. Plume front and suspended sediment dispersal off the Yangtze (Changjiang) River mouth, China during non-flood season. Estuar. Coast. Shelf Sci. 71, 60–67.

Wang, J., Wolff, D., 2009. Evaluation of TRMM ground-validation radar-rain errors using rain gauge measurements. J. Appl. Meteorol. Climatol. 49, 310–324.

Ward, E., Buytaert, W., Peaver, L.,Wheater, H., 2011. Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. Adv. Water Resour. 34, 1222–1231.

Wardah, T., Abu Bakar, S.H., Bardossy, A., Maznorizan, M., 2008.Use of geostationary meteorological satellite images in convective rain estimation for flash-flood forecasting. Journal of Hydrology 356(3-4), 283–298.

Weng, Q., 2012. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. Remote Sens. Environ. 117, 34–49.

Weng, Y., Sun, F., Grigsby, J., 2012. GeoTools: An android phone application in geology, Computers & Geosciences 44, 24 - 30.

Weng, Q., 2001. Modeling Urban Growth Effects on Surface Runoff with the Integration of Remote Sensing and GIS. Environ. Manage. 28, 737–748.

Werdell, P.J., Bailey, S.W., Fargion, G.S., Pietras, C., Knobelspiesse, K.D., Feldman, G.C., McClain, C.R., 2003. Unique data repository facilitates ocean color satellite validation, EOS Trans. AGU 84(38), 377-392.

White, M. and K. Grier (2006). The Effects of Watershed Urbanization on the Stream Hydrology and Riparian Vegetation of Los Peñasquitos Creek, California. Landscape and Urban Planning 74, 125-138.

Williamson, R. A., 1997. The Landsat legacy: remote sensing policy and the development of commercial remote sensing. Photogramm. Eng. Remote Sens. 20052, 877–885.

Wolter, K., and M.S. Timlin, 1993. Monitoring ENSO in COADS with a seasonally adjusted principal component index. Proc. of the 17th Climate Diagnostics Workshop, Norman, OK, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, 52-57.

Wood, L.J., 2007. MPA global: a database of the worlds marine protected areas. Sea Around Us Project, UNEP-WCMC & WWF. Retrieved April 2016, Available at: http://www.mpaglobal.org.

Woodson, C.B., McManus, M. A., Tyburczy, J. A., Barth, J. A., Washburn, L., Caselle, J.E., Carr, M.H., Malone, D.P., Raimondi, P.T., Menge, B. A., Palumbi, S.R., 2012. Coastal fronts set recruitment and connectivity patterns across multiple taxa. Limnol. Oceanogr. 57, 582–596.

Wooster,W.S.,Bakun,A.,McLain,D.,1976. The seasonal upwelling cycle along the eastern boundary of the North Atlantic. J. Mar. Res. 34 (2),131–141.

Wu, C., 2004. Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. Remote Sens. Environ. 93, 480–492.

Wulder, M. A., Masek, J.G., 2012. Preface to Landsat Legacy Special Issue: Continuing the Landsat Legacy. Remote Sens. Environ. 122, 1.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat, Remote Sensing of Environment 122, 2 -10.

Xian, G., Homer, C., 2010. Updating the 2001 National Land Cover Database Impervious Surface Products to 2006 using Landsat Imagery Change Detection Methods. Remote Sens. Environ. 114, 1676–1686.

Xian, G., Homer, C., Fry, J., 2009. Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods. Remote Sens. Environ. 113, 1133–1147.

Xian, G., 2008. Mapping Impervious Surfaces Using Classification and Regression Tree Algorithm, in: Weng, Q. (Ed.), Remote Sensing of Impervious Surfaces. CRC / Taylor & Francis, pp. 39 – 58.

Xian, G., 2008b.Satellite remotely-sensed land surface parameters and their climatic effects for three metropolitan regions.Adv. Sp. Res. 41, 1861–1869.

Xie, S.P., Liu, W.T., Liu, Q., Nonaka, M., 2001. Far-reaching effects of the Hawaiian Islands on the Pacific Ocean-atmosphere system. Science 292, 2057–2060.

Xu, M., Xu, H., 2015. Atmospheric responses to Kuroshio SST front in the East China Sea under different prevailing winds in winter and spring. J. Clim. 28, 3191–3211. doi:10.1175/JCLI-D-13-00675.1

Yanagi, T., 1987. Classification of "Siome", Streaks and Fronts. Journal of the Oceanographical Society of Japan, 45, 149-158.

Yang, L., Xian, G., Klaver, J.M., Deal, B., 2003. Urban Land-Cover Change Detection through Sub-Pixel Imperviousness Mapping Using Remotely Sensed Data. Photogramm. Eng. Remote Sens. 69, 1003–1010.

Young, D., Afoa, E., Meijer, K., Wagenhoff, A., Utech, C., 2013. Temperature as a contaminant in streams in the Auckland region, stormwater issues and management options.Prepared by Morphum Environmental Ltd for Auckland Council. Auckland Council technical report, TR2013/044.

Yuan, D., Zhu, J., Li, C., Hu, D., 2008. Cross-shelf circulation in the Yellow and East China Seas indicated by MODIS satellite observations. J. Mar. Syst. 70, 134–149.

Yuan, F., Wu, C.,Bauer, ME., 2008. Comparison of spectral analysis techniques for impervious surface estimation using landsat imagery. Photogrammetric Engineering and Remote Sensing 74:8, 1045-1055.

Yuan, F., Bauer, M.E., 2007. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. Remote Sens. Environ. 106, 375–386. doi:http://dx.doi.org/10.1016/j.rse.2006.09.003

Yue, P., Di, L., Wei, Y., Han, W., 2013. Intelligent services for discovery of complex geospatial features from remote sensing imagery, ISPRS Journal of Photogrammetry and Remote Sensing 83, 151 – 164.

Zeileis, A., Grothendieck, G., 2005. zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27.

Zhang, J., He, C., Zhou, Y., Zhu, S., Shuai, G., 2014. Prior-knowledge-based spectral mixture analysis for impervious surface mapping. Int. J. Appl. Earth Obs. Geoinf. 28, 201–210.

Zhang, J., Li, P., Wang, J., 2014b. Urban Built-Up Area Extraction from Landsat TM/ETM+ Images Using Spectral Information and Multivariate Texture. Remote Sens. 6, 7339–7359.

Zhang, Y., Zhang, H., Lin, H., 2014c.Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. Remote Sensing of Environment, 141, 155-167.

Zhang, H., Zhang, Y., Lin, H, 2012.A comparison study of impervious surfaces estimation using optical and SAR remote sensing images. International Journal of Applied Earth Observation and Geoinformation 18, 148 - 156.

List of Figures

List of Tables