Developing a Geologically-Based $V_S^{30}$ Site-Conditions Model for Portugal: Methodology and Assessment of the Performance of Proxies

Susana P. Vilanova$^1$ and João Narciso, João P. Carvalho, Isabel Lopes, Mário Quinta-Ferreira, Carlos C. Pinto, Rui Moura, José Borges, Eliza S. Nemser

$^1$Corresponding Author; Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisbon, Portugal; susana.vilanova@tecnicoulisboa.pt;

The electronic supplement to this paper contains the $V_S^{30}$ database flat-file developed and used in the context of the paper.

November 8, 2017
Abstract

The inclusion of site-specific conditions is essential to adequately represent the seismic hazard and the seismic risk for a region. We acquired, gathered and organized a near surface shear-wave velocity database for Portugal, and applied a three-step methodological approach for developing a $V_{S30}$ site-conditions map using extrapolation based on surface geology. The methodology includes: 1) defining a preliminary set of geologically defined units; 2) calculating the probability distribution of $\log V_{S30}$ for each unit; and 3) merging the units according to the results of statistical tests. The final model comprises three geologically defined units characterized by $\log V_{S30}$ distributions that are statistically significantly different from each other: F1 - Igneous, metamorphic and old sedimentary rocks; F2 - Neogene and Pleistocene formations; and F3 - Holocene formations. The site conditions for F3 unit may be further refined using correlations with topographic slope based on the SRTM3 dataset. We analysed the performance site-conditions models based on correlations with exogenous data (topographic slope and surface geology analogues). The results show that the residual distributions between $\log V_{S30}$ values measured and estimated from those proxies are strongly biased for some geological units, emphasizing the need for acquiring regional $V_{S}$ data.
Introduction

Earthquake ground motion maps such as seismic hazard maps or instrumental intensity maps provide critical information for a variety of societal applications. They support decision-making processes that include the development of regulatory legislation, the estimation of insurance rates, land-use planning and emergency planning. Since local site conditions strongly affect the characteristics of ground motion, estimating first-order site-conditions at the regional scale is essential for improving the information delivered by such maps.

The importance of the near-surface shear-wave velocity ($V_S$) structure on ground motion amplification is supported by both theoretical considerations (e.g., Aki and Richards, 1980; Stein and Wyssession, 2009) and observational studies (e.g., Joyner et al., 1981; Borcherdt, 1994). Within unbounded and homogeneous media $V_S$ is proportional to the square root of the quotient between the shear modulus and the density of the medium. Both the shear modulus and the density tend to increase with depth and the overall tendency of $V_S$ is to decrease as the waves propagate from depth towards the surface. This reduction of $V_S$ has important implications for the conservation of elastic energy. For vertically propagating SH waves in an elastic medium the energy flux is given by the product of density and $V_S$ (seismic impedance) and the particle velocity squared (e.g., Aki and Richards, 1980). Because the conservation of energy requires the flux to remain constant, the decrease of impedance needs to be compensated by the increase in particle velocity and therefore amplitude of the seismic wave. This effect is however partially counteracted by that of anelastic attenuation, which tends to be greater on soft soils (e.g., Reiter, 1990; Kramer, 1996).

The modification of ground motion by site-conditions, usually referred to as site-effects, includes local ground response, basin effects, and topographic effects (e.g., Stewart et al., 2001; Kamai et al., 2016). Local ground response represents the effects of the variation of the physical properties of the near-surface materials on nearly-vertically propagating SH waves. In local ground response abrupt changes in medium impedance at depth result in large amplification at specific frequency ranges of ground motion (resonance). Both basin
and topographic effects refer to the influence of 2-D or 3-D geometric configurations on the
propagation of seismic waves, and its importance, which can be large, is usually restricted
to specific locations. Although the detailed study of site-effects is essential for site-specific
studies, regional assessments must necessarily rely on simplified approaches.

Joyner et al. (1981) proposed the use of a quantitative $V_S$-based parameter for a simplified
representation of site conditions the $V_S$ corresponding to the depth associated with one
quarter-wavelength at the period of interest. Due to the economical constraints associated
with obtaining data at the required depths this measure has been superseded by the use of
the time averaged $V_S$ to 30 m depth, given by

$$V_{S30} = \frac{30}{\sum t_i}$$  (1)

where $t_i$ is the traveltime of the S wave within each layer up to the depth of 30 m.

Borcherdt (1994), based on previous empirical work, recommended the use of $V_{S30}$ for
classifying sites for building codes and the parameter is included in both in the National
Earthquake Hazard Reduction Program (NEHRP) seismic design provisions (e.g., BSSC,
2004) and in the Eurocode 8 (CEN, 2004).

Because $V_{S30}$ is strongly correlated with deeper velocity structures (Boore, 2004; Boore et al.
2011) it has been shown to describe site-effects at ground motion frequencies corresponding
to wavelengths much longer than 30 m (Stewart et al., 2014).

$V_{S30}$ has increasingly become the reference parameter for classifying site conditions in
several applications. It is currently used for characterizing site-conditions in ground motion
prediction equations, and for modeling ground motion amplification in both seismic hazard
maps and instrumental intensity maps. It has long been known that $V_{S30}$ present several lim-
itations as a site-conditions parameter (e.g. Castellaro et al., 2008). Additional parameters,
such as the natural frequency of the site, are being investigated for improving the regional
site characterization in a variety of applications (e.g. Cadet et al., 2010; Motazedian et al.,
Developing $V_{S30}$ Site-Conditions Maps

Estimating spatially-continuous variables from discrete datasets requires either the correlation with spatially extensive datasets or the application of extrapolation or interpolation techniques.

Surface geology-derived classification schemes have been used to produce regional $V_S$ site conditions maps based on rock type and/or geological age (e.g., Tinsley and Fumal, 1985; Park and Elrick, 1998; Wills and Silva, 1998). The correlation between $V_S$ and geologic units relies on the fact that $V_S$ depends on physical properties of the materials such as density, porosity, cementation, and fracture spacing.

Wills and Silva (1998) correlated $V_{S30}$ data with geologic units in California and extrapolated based on surface geology in order to obtain a statewide map of $V_{S30}$. That approach has been further refined by using depositional environment and geographic criteria as additional constraints (Wills and Clahan, 2006; Wills et al., 2000). Similar approaches have been also used for the state of Utah (e.g., Ashland and McDonald, 2003; McDonald and Ashland, 2008).

Park and Elrick (1998) developed a geologically-based ($V_{S30}$) map for the southern California region. Their approach differs from that of Wills and Silva (1998) in that their goal was to achieve the simplest model supported by the dataset. To attain that objective they used statistical tests (the t-test and the Komolgorov-Smirnov test) to justify the subdivision of an initial set of geological units, if statistically significant.

The most extensively used $V_{S30}$ extrapolation method is however that based on topographic slope. The approach, which has been introduced by Wald and Allen (2007), relies on the correlation of $V_{S30}$ measurements and the topographic slope for both regions of active tectonics and stable tectonics. Although there is no explicit physical relationship connecting $V_{S30}$ and topographic slope, it is expected that the later will relate to different geomorphologic environments and lithology in a broad sense, since more competent high-velocity materials can maintain a steep-slope, whereas fine basin sediments will be deposited in nearly-flat
basins. The main advantage of the method is that since topographic data are globally available, a $V_{S30}$ model can be derived for any region. One of the limitations of the model is that it is not expected to be effective in regions of flat-lying rocks.

When dense $V_S$ datasets are available and the values are spatially correlated geostatistical interpolation tools can be used to develop spatially-continuous $V_{S30}$ models. Thompson et al. (2007) employed such an approach for mapping $V_{S10}$ in the San Francisco Bay Area, finding spatial horizontal correlations on the order of 4 km.

Thompson et al. (2014) presented a framework that combines surface geology maps with topographic data for developing $V_{S30}$ maps. Their approach is based on identifying trends between surface-geology derived $V_{S30}$ residuals and topographic slope. The results show that both Quaternary alluvium and Pleistocene sedimentary units exhibit trends with topographic gradient. They applied a kriging-with-a-trend technique to obtain a final $V_{S30}$ map for California.

The terrain-based classification is an automatic procedure developed by Iwahashi and Pike (2007) and relies on the development of a set of geomorphic categories based on gradient, convexity and surface texture, using an automatic procedure. This methodology has been applied to characterize $V_{S30}$ in California with promising results (Yong et al., 2012).

Due to the scarceness of shear-wave velocity data in most regions, models developed for data-rich regions have been employed to estimate site conditions elsewhere. In particular, the topographic slope method has become the standard way for incorporating site effects into regional studies worldwide given the convenience provided by the global $V_{S30}$ server (Allen and Wald (2007); see Data and Resources).

Lemoine et al. (2012) evaluated the performance of the topographic slope method for stable and active regions of Europe using the $V_{S30}$ dataset compiled in the context of project SHARE (Seismic Hazard Harmonization in Europe). The results show that while the method provided better estimates than pure randomness for active regions that was not the case for stable continental regions. Lemoine et al. (2012), however, acknowledged the fact that their
analysis for stable continental was based on a very limited dataset. Stewart et al. (2014) compiled and analyzed a large $V_S$ dataset for Greece. They propose a framework for estimating $V_{S30}$ for sites in Greece based on geology and slope. They recommend both the geology-slope approach and the terrain approach of Yong et al. (2012) over the slope approach of Wald and Allen (2007). They nevertheless acknowledge that the latter is probably the only available approach for many regions of the world.

Project SCENE, funded by the Portuguese Foundation for Science and Technology (FCT), aimed at gathering and acquiring shear-wave velocity profiles in diverse lithological and geological formations in Portugal, in order to 1) characterize sites where strong-motion stations are deployed and 2) develop a regional site conditions map to be used in seismic hazard maps.

In this paper we focus on the methodological approach used for developing a statistically robust site-conditions map for Portugal. The database includes 160 $V_S$ profiles and is the largest published for stable continental regions, making it particularly suited to test the applicability of proxies based on exogenous $V_S$ empirical correlation.

**Brief Tectonic and Geological Setting of the Study Region**

The study area, Portugal, is located in the western region of the Iberian Peninsula, within the Eurasian tectonic plate (Figure 1a). It is defined as a stable continental region according to the geological criteria proposed by Johnston (1989), and displays moderate seismicity rates (e.g., Custódio et al., 1996). The seismic record for Portugal includes several intraplate earthquakes with magnitude estimates in the M6.0 – M7.3 range, both historical (Vilanova and Fonseca, 2007; Stucchi et al., 2013) and pre-historical (Rockwell et al., 2009; Canora et al., 2015). Western Iberia is also affected to some extent by the large to great interplate earthquakes nucleating in the Azores-Gibraltar plate boundary, such as the M$_S$7.8
1969 earthquake (Fukao, 1973) and the M8.5−M8.7 1755 Lisbon earthquake (e.g., Johnston, 1996; Vilanova et al., 2003; Martínez Solares and López Arroyo, 2004; Fonseca, 2005).

Portugal displays in general moderate hazard levels ($0.1 \text{ g} \leq \text{PGA} \leq 0.25 \text{ g}$ for 10% exceedance probability in 50 years) according to the 2013 European Seismic Hazard Map (Woessner et al., 2015). This result is consistent with the previous regional study by Vilanova and Fonseca (2007).

The basement of the Iberian Peninsula, known as the Hesperic Massif, or Iberian Massif, is composed of igneous and metamorphic rocks of Paleozoic and Precambian ages, which have been accreted together during the Paleozoic. The Hesperic massif represents the largest continuous exposure of the Variscan Orogen in Europe. Above this cratonic block several basins developed in both the western and southern margins as a consequence of the rifting episodes that, during the Mesozoic, led to the opening of the Atlantic Ocean and the Tethys Ocean. These basins have been further deformed and inverted during subsequent compressive tectonics in the Eocene (Pyrenean Orogeny) and Miocene (Africa-Eurasia collision). Further details on the geology and geological evolution of the region can be found, for instance, in Ribeiro et al. (1979), Pinheiro et al. (1996) and references therein.

Data and Methods

Both invasive and non-invasive methods can be employed to characterize the near-surface structure of the shear-wave velocity. Determining the shear-wave velocity using invasive methods involves directly measuring the wave travel-time to a range of depths. Non-invasive methods involve the acquisition of waves at the surface and require the use of an inversion algorithm and/or forward modeling to resolve the structure at depth.

Although invasive methods are well known and highly reliable, the non-invasive approaches are significantly less costly. The latter also have the advantage of providing a more spatially extensive sample of the subsurface. Comparisons between invasive and non-invasive
data at same sites show that, in general, compatible velocity structures or $V_{S30}$ values are obtained (e.g., Xia et al., 2002; Williams et al., 2003; Scott et al., 2006; Boore and Asten, 2008). Moss (2008) evaluated the intramethod uncertainty in measuring $V_{S30}$ from different techniques both invasive and non-invasive reporting coefficients of variation on the order of 1%-3% for invasive techniques and 5%-6% for non-invasive techniques.

**Seismic Refraction**

We used seismic refraction as the main tool for characterizing the shear-wave subsurface structure at the selected locations. The seismic refraction is a widely known and applied method in geophysics. It uses active seismic sources at the surface and involves measuring the travel times of the seismic waves as they travel from the source towards a set of aligned receivers. Assuming that wave velocity increases with depth, at some distance from the source the direct waves will be overcome by the critically refracted waves at the first layer interface. Likewise, at greater distances the waves refracted at deeper layers will overcome those refracted above. Due to its underling assumptions, the method cannot detect velocity inversions with depth. However, some indications of the presence of a velocity inversion or hidden layer may be obtained using interpretation methods (e.g. Palmer, 1981) or well control.

Within the scope of project SCENE thirty sites where strong motion stations are installed have been characterized using this technique. The surveys were performed, in general, within 200 m from the stations and within the same geologic unit, according to geological maps and field inspections. The SCENE shear-wave database also includes a significant amount of shear-wave refraction data available from FCT project NEFITAG using the same methodological approach. We used as shear-wave source a 3 m long wood beam, coupled to the ground with a four-wheel drive, and stricken on both sides by a sledgehammer, in order to allow data corroboration and to eliminate P-wave contamination (Hasbrouck, 1991). Two shots were performed at both ends (minimum offset distance of 1.75 m) of the array and
three shots within the array. The recording system consisted of a linear array of 24 40-Hz horizontal receivers and 24 50-Hz vertical receivers spaced 3.5 m apart. The overall length of the array, which constrains the depth reached by the survey, was 84 m.

The data interpretation was performed with commercial software relying on the generalized reciprocal method (Palmer, 1981) and slope intercept method, and the method of Haeni et al. (1987). The latter uses delay-times for constraining a first preliminary velocity model, followed by three iterations of ray tracing and minimization of residuals by least squares. The results are a 2-D $V_S$ cross section. Further details on both the survey and data interpretation can be found on Carvalho et al., unpublished report, 2017, see Data and Resources;

A total of 61 $V_S$ depth sections have been acquired using this methodological approach. In general we are confident to have reached 30 m deep in the seismic sections. In many cases we were actually able to identify interfaces deeper than 30 m, which demonstrates that both the source used and the equipment setup allowed for the 30 m to be reached. However, investigation depth depends on the velocity distribution at each site. Therefore, we compared our interpretations with other available information such as borehole data in the vicinity of the profiles. The vertical resolution in seismic refraction data is usually accepted to vary between 10% to 20% of the reflector’s depth or one quarter of the wavelength (e.g., Briaud (2013), page 155). Therefore using the described procedure we are not expected to detect layers thinner than around 3 m (except the uppermost layer) and around 6 m at 20 – 30 m deep. The lateral resolution is typically around 1/2 to 1 of the spacing between receivers, which corresponds to about 2 – 3 m.

**Multichannel Analysis of Surface Waves**

The use of surface-wave methodologies to estimate the $V_S$ depth structure of a site relies on the dispersive characteristics of Rayleigh-type surface waves traveling through a heterogeneous medium. The velocity of this type of waves depends on the mechanical properties
of the propagation medium. Since lower frequencies and long wavelength waves penetrate deeper into the material than high frequency and short wavelength waves, their velocities will reflect the differences in the mechanical properties of the volumes they travel through. Because more information is available for the upper layers, these are better constrained than the deeper ones.

The dispersion curve is obtained by converting the data to frequency domain and by identifying the different propagation modes. An inversion algorithm is then applied for obtaining a $V_S$ structure compatible with the experimental dispersion curve. The multichannel analysis of surface wave technique (MASW) (Gabriels et al., 1987; Park et al., 1999) uses an array of receivers to record the seismic wave-field produced by an active source. The refraction microtremor technique (ReMi) (Louie, 2001) employs a similar approach but with passive sources.

In spite of the inherent non-uniqueness associated with MASW results several studies have shown that different profiles that fit a particular dispersion curve lead to similar $V_{S30}$ values (e.g. Comina et al., 2011).

Seven sites have been analyzed using MASW methodology. We used a 10 kg sledgehammer striking a metal plate as source, and the acquisition system was composed of a 48 m long line with 24 vertical 4.6 Hz geophones spaced 2 m apart. Ten shots were performed at both ends of the acquisition line with an offset of 2 m. We used 2 s long recording intervals with a sampling rate of 1 ms. Several separate acquisitions were performed in order to evaluate the uncertainty in determining the dispersion curve. In some cases, considering the geological setting and the site-specific characteristics, different line configurations were tried. The data processing and inversion was performed with the SWAN software (Geostudi Astier) although software Dinver (Wathelet et al., 2004; Wathelet, 2008) was used for comparison. The software SWAN uses an automatic inversion algorithm that allows for an iterative trial and error fit to the dispersion curve. The Dinver algorithm searches the space of solutions by minimizing a misfit curve. The final model was built using the best-fit models together with
additional constraints rendered by geological and geotechnical information in the vicinity of the profiles, and some degree of expert judgment with respect to the velocity of the deeper layers.

**Invasive Measurements**

The seismic cone penetrometer test (SCPTu) is an invasive methodology for directly measuring $V_S$ at specific depths. The probe introduced into the soils contains seismic receiver that records the shear wave travel time from a source located at the surface to the recording depth, as in a downhole test. In four sites located within soft sediments we used SCPTu methodology to determine $V_S$ subsurface structure. We used a single receiver seismic cone and the data was interpreted using the cross-correlation method. The seismic signal was acquired and processed using commercial software provided by the manufacturer of the equipment (See Data and Resources). Measurements were performed every meter until reaching stiff material, which occurred within the depth range of 22 – 26 m.

**Other Available Data**

We included in the database $V_S$ information available from both the literature and unpublished technical reports using a variety of methodologies: seismic refraction (24 sites, Carvalho et al. (2008), Carvalho et al. (2009)), MASW (56 sites, Lopes (2005), Lopes et al. (2005), Santos (2011), Fontoura (2013), and unpublished surveys performed in the context of service provisions and scientific projects provided by Rui Moura), and ReMi (8 sites, Carvalho et al. (2016)). In general, the depths of the profiles included in the database range from 20 – 30 m. However, for some seismic refraction sections the deepest mapped interface is shallower than 15 m, raising questions about the actual depth of the models.

The shear-wave database presently consists of 160 profiles or sections from a variety of lithological/geological formations. From these, about 40% have been acquired within the framework of projects SCENE and NEFITAG, and more than 50% have been estimated
using the seismic refraction method. Figure 1b shows the distribution of \( V_{S30} \) values in the database. Figure 1c and Figure 1d show, respectively, the geographic distribution of sites sorted by the geologically defined unit and by the methodology used to characterize the \( V_S \) depth structure.

Developing the Database Flat-File

In this section we describe the procedures employed in the parametrization of the \( V_{S30} \) database. The corresponding flat-file is available as Table S1 in the electronic supplement to this article.

Calculating \( V_{S30} \) and Associated Variability

As discussed previously, for seismic refraction data acquired within the scope of projects SCENE and NEFITAG, we are confident that the \( V_S \) models extend to 30 m deep. However, since the seismic refraction method maps the interfaces between subsurface layers characterized by different seismic velocities, unless an interface has been actually detected below 30 m, one cannot be totally sure that a depth of 30 m has been reached for a particular section. To evaluate the impact of this uncertainty in the \( V_{S30} \) distributions we combined the use of a best-case scenario in which we assumed that all sections reached 30 m (e.g., we assumed constant extrapolation \( V_{30C} \)), with that of a worst-case scenario in which we assumed that the deepest interface roughly corresponds to the maximum depth of the model. In the latter case we use an extrapolation method to obtain \( V_{S30} \) from \( V_{SZ} \) (\( V_{S30z} \)). We used the same approach for profiles whose \( V_S \) model is shallower than 30 m. Overall, only about 11% of the profiles are suspected to have depth models that do not reach 15 m deep.
Assuming constant extrapolation down to 30 m deep ($V_{S30c}$)

To calculate the value of $V_{S30}$ for each site included in the database we proceeded as follows.

For seismic refraction sections acquired in the context of projects SCENE and NEFITAG we used the values of the 2D velocity models at the receivers locations and calculated the $V_{S30}$ according to equation (1) assuming that all $V_S$ models reached 30 m depth.

We then calculated the log-average value for the section. We used the log-average instead of the arithmetic mean because of our underlying assumption that $V_{S30}$ follows a lognormal distribution (see section Developing a geologically-based $V_{S30}$ model for Portugal). Using the arithmetic mean leads, however, to very similar values of $V_{S30}$.

The 2D sections obtained within the scope of previous refraction campaigns (Carvalho et al., 2008, 2009) have been graphically interpolated at five locations within the acquisition line because the original data files have been lost. Then, the same procedure previously described for 2-D sections has been applied. For all available seismic refraction sections we calculated the standard deviation associated with the log $V_{S30}$ value at each site. This provides a measure of the spatial variability of $V_{S30}$ associated with the sites.

For the MASW-based measurements we calculated the $V_{S30}$ using the preferred final profile, using constant extrapolation when required to reach the depth of 30 m. We calculated the standard deviation of log $V_{S30}$ at each site by using a set of automatically inverted best-fit profiles as a measure of the variability associated with the methodology.

The log-average has been calculated for the profiles presented by Carvalho et al. (2016) for each single site analyzed with the refraction microtremor technique. In this case the associated standard deviation provides a measure of the uncertainty associated with the technique.

At last, for sites analyzed with different methodological approaches, $V_{S30}$ was calculated for each method as described previously and the final $V_{S30}$ value for the site was log-averaged.

In a few sites we had reservations regarding the results of some measurements due to specific difficulties faced during acquisition or analysis. This was the case for seismic re-
fraction section at SC-VFX and multichannel analysis of surface waves profile at SC-BEJ. The corresponding $V_S$ structures were not further considered in the analysis. The standard deviation of $\log V_{S30}$ at sites characterized by multiple techniques provides an estimate of the inter-method variability.

**Using Extrapolations Based on $V_{SZ}$ ($V_{S30z}$)**

We checked the applicability of the relationship proposed by Boore (2004) and Boore et al. (2011) to extrapolate $V_{S30}$ from shallower velocity structures, based on data from California and Japan, respectively, to our data. The relationships are based on the parameter $V_{SZ}$, which represents the time averaged $V_S$ to the depth $z$, and is given by

$$V_{SZ} = \frac{z}{\sum t_{iz}}$$

where $t_{iz}$ is the travel time within each layer up to the depth $z$.

We calculated $V_{SZ}$ for profiles reaching 30 m deep for $z = 10, 15, 20,$ and $25$ m. In the seismic refraction data we assumed that profiles exhibiting interfaces at depth $z > 25$ m did reach 30 m depth. Refraction profiles that reached high values of $V_S$, typical of bedrock, have also been included. Profiles obtained with other methodologies were included only if the model explicitly reached a depth of 30 m. $V_{S30}$ is plotted as a function of $V_{SZ}$ in Figure 2, together with the relationships proposed for California (Boore, 2004) and for Japan (Boore et al., 2011). The results indicate that the relationships developed for California are more suited to represent the trends of regional data than those for Japan, in particular in what concerns the shallower depths considered ($z = 10,$ and $15$ m). We therefore consider the functional forms proposed by Boore (2004) to extrapolate the profiles that may have not reached 30 m depth.
Classifying the Surface Geology

The site classification was performed using the 1:50,000 scale geological maps published by Serviços Geológicos de Portugal. If that scale was not available we used the 1:200,000 scale geological maps. In few locations the only available geological map was at the 1:500,000 scale.

The consistency of the maps was a problem in particular for the southernmost region of Portugal. In some cases the same unit was attributed a different geological age in adjacent maps. We corrected the units according to the 1:200,000 scale geological map, which was consistent throughout the region, and a comment was introduced in the flat-file. This type of inconsistency has been also reported by Stewart et al. (2014) for Greece.

Calculating the Topographic Slope

Following Wald and Allen (2007) we calculated the topographic-slope associated with each site in the database using the Generic Mapping Tools slope function `grdgradient`; Wessel and Smith (1991). We used freely available topographic data sets from the Shuttle Radar Topography Mission at 30 arcsec resolution (SRTM30) and at 3 arcsec resolution (SRTM3)(see Data and Resources). The topographic-slope value for each site was calculated using the nearest neighbor interpolation.

Developing a geologically-based $V_{S30}$ model for Portugal

In this section we describe the methodological approach used for deriving a $V_{S30}$ site condition model for Portugal using extrapolation based on surface geology. Our objective was to estimate the most accurate model statistically supported by the dataset. To accomplish this goal we followed an iterative three-step procedure which consisted of 1) defining a preliminary set of geologically defined units based on the literature; 2) estimating the probability distribution of log $V_{S30}$ for each of those units; and 3) performing statistical tests in order
to estimate the statistical significance of the difference in the \( \log V_{S30} \) distribution characteristics between the units. The units were merged according to the results of the statistical tests and the procedure was repeated.

It has been debated whether \( V_{S30} \) or the (decimal) logarithm of \( V_{S30} \) (\( \log V_{S30} \)) should be used as the variable for deriving \( V_{S30} \) predicting models. (e.g., Lemoine et al., 2012). The use of \( \log V_{S30} \) as a variable assumes that \( V_{S30} \) observations follow a lognormal distribution (e.g., Park and Elrick, 1998; Ashland and McDonald, 2003). Boore et al. (2011) show that, unlike \( V_{S30} \), \( \log V_{S30} \) values in their database follow a normal distribution. This is also the case for our dataset as can be graphically illustrated by the quantile-quantile plot in Figure 3. Since most statistical tests require that data are normally distributed we used \( \log V_{S30} \) as the dependent variable in our model.

The preliminary model consisted of six preliminary geologically defined units that are summarized in Table II. We used as variables \( V_{S30_c} \), in which we assumed constant extrapolation for all profiles, and \( V_{S30_z} \), in which we used the functional forms proposed by Boore (2004) to extrapolate the profiles that did not or may have not reached 30 m depth. Figure 4 shows histograms for \( \log V_{S30_c} \) together with fitted normal distributions for both \( \log V_{S30_c} \) and \( \log V_{S30_z} \). The dataset is not, in general, significantly affected by uncertainty regarding the extrapolation method for profiles that may have not reached 30 m depth. The unit most affected by this type of uncertainty is P4 (Pliocene formations). The dispersion of data is similar and around 0.2 for every geologically defined unit except P4 (\( \sigma = 0.1 \)). This issue may be related with the relatively lower lithological variety associated with the Pliocene age in the region. However, more data is necessary in order to confirm this hypothesis. In general the \( \log V_{S30} \) distributions for each geologically defined unit do not show systematic trends with the data’s geographical region.
Declustering

Attributes measured in clustered datasets may not be representative of those of the population since closely spaced observations may exhibit strong spatial autocorrelation (e.g., De Smith et al., 2015). This issue is particularly relevant for cases where preferential sampling applies. In preferential sampling a large number of observations are spatially aggregated in regions of interest, where the variable to be analyzed is expected to take consistently high or low values. In these cases population attributes such as the mean values, standard deviation will probably be substantially biased.

The database developed for this study includes shear-wave profiles acquired in the context of research projects with different aims. For project SCENE, the adopted strategy of acquiring data in the vicinity of sites where strong-motion instruments were installed led to a dataset that is spatially disperse. That is also the case for project NEFITAG, and data from Carvalho et al. (2008, 2009, 2016) whose data-acquisition policy aimed at sampling different geological units within relatively large regions. However, in other studies, a relatively small region was extensively sampled, producing datasets that exhibit strong spatial clustering. In particular, data from Santos (2011) are probably affected by preferential sampling since the aim of that study was to map the thickness of altered rocks.

Declustering methods are based on the weighting of the sample data in order to account for spatial representativity. Closely spaced observations receive a reduced weight because of its redundancy. Cell declustering and polygonal declustering are the most widely used declustering methods (e.g., Olea, 2007). In polygonal declustering the domain is divided into polygons that define the area of influence of each observation and the attributed weights are proportional to that area. This method has the disadvantage of being extremely sensitive to the location of the domain boundaries (e.g., Olea, 2007; De Smith et al., 2015). In cell declustering (Journel, 1983; Deutsch, 1989; Deutsch and Journel, 1992) a regular grid of cells is superimposed over the data domain and the attributed weights are inversely proportional to the number of observations per cell. Deutsch (2015) discusses the parametrization for cell
declustering and proposes that the cell size should be related to the spacing of data in sparse sampled areas. A set of randomly selected locations is usually used for the origin of the cell grid.

We evaluated the extent to which spatial clustering or preferential sampling affects the mean value of \( \log V_{S30} \). We compared the histograms calculated from both the full dataset and a dataset obtained by using the cell declustering technique. The cell size was chosen on the basis of the average nearest neighbor distance for the sparse areas of the dataset. Cell sizes of 10 km, 15 km and 20 km have been tested with similar results. For each grid size a randomly selected set of 10 grid origins have been used. A square grid size of 10 km was retained for the final analysis.

Figure 4 shows the fitted normal distributions for the declustered dataset. The normal distribution fitted to generalized geologic unit P1 shows a significant degree of bias that may be attributed to spatial clustering or preferential sampling. However, the attributes of the remaining generalized geological units are not significantly affected by declustering. For the subsequent analysis we used a declustered version of the dataset assigned to geologic unit P1.

**Statistical Tests**

The analysis of variance (ANOVA) is a statistical test that is used for assessing whether there are statistical significant differences between the means of a set of independent groups. The method relies on computing the \( F \) value, which is the ratio between the variances within and between groups, and determining the corresponding \( F \)-distribution under the null hypothesis that data from all groups belong to a common distribution function. A \( p \)-value determined from the \( F \)-distribution reflects the probability that the calculated \( F \) value has occurred by chance. The ANOVA test assumes that 1) the samples are independent, 2) the underlying populations are normally distributed, and 3) the variance of data in groups are homogeneous. Unlike using multiple t-tests, the ANOVA procedure ensures that the final significance level
is achieved. If the null hypothesis is rejected one can proceed the analysis using a post-hoc test such as the Tukey-HSD (Honestly Significant Difference). The Tukey-HSD approach uses the Studentized Range distribution to evaluate which group’s means are significantly different from each other. The test computes the value $q$, which is the difference between the means divided by the standard deviation, for all pairwise comparisons. The corresponding $p$-value is obtained by comparing that value with Studentized Range distribution for the null hypothesis.

**Results**

We tested the null hypothesis for the independent variable $\log V_{S30}$ distributed by six groups, corresponding to the preliminary set of geologically defined units. The resulting $F$ value of 22.9467 corresponds to a $P$-value of $2.5 \times 10^{-15}$, much below the common significance level of 5%. Therefore we reject the null hypothesis and proceed the post-hoc analysis using the Tukey-HSD method. The results for the Tukey-HSD test for the preliminary model are summarized in Table 2.

The Tukey-HSD post-hoc test results indicate that there is no statistically significant difference between groups P1 and P2 and between groups P3, P4 and P5, and to a lesser extent between groups P2 and P3 and P2 and P5. We merged the groups exhibiting higher values for $p$-value and repeated the procedure. The resulting set of groups of geological units defined by the tests – F1, F2 and F3 – are summarized in Table 3. The results of the statistical tests for F1, F2 and F3 are presented in Table 2. The ANOVA test produced a $F$ value of 57.4279 which corresponds to a $p$-value of $10^{-16}$. The ensuing Tukey-HSD test indicates that the difference in the means of the groups is statistically significant.

The final model is illustrated in Figures 5 and 6. The median $V_{S30}$ values for the geologically defined units F1, F2 and F3 are 829 m/s, 470 m/s and 237 m/s, respectively. However, the 68% confidence interval for $V_{S30}$ overlaps for F1 and F2 and for F2 and F3, i.e., the lower limit for F1 is lower than the upper limit for F2, and the lower limit for F2 is lower than the
upper limit for F3. This is partially related to the inherent limitations of surface geology as a predictor for $V_{S30}$. A geologically defined unit includes different rock types, lithologies and layer thicknesses, which influence the VS depth structure and consequently the corresponding $V_{S30}$ value. Other geologically-based $V_{S30}$ models display similar dispersion values despite the fact of presenting more specific geologically defined units (Wills and Clahan, 2006). Nevertheless, some dispersion could be related to the limited size of the database. A larger dataset that would allow the definition of more specific geologically defined units or the inclusion of other geographic criteria (e.g., Holocene in narrow valleys, small basins, etc.), might eventually decrease the dispersion within some units.

**Relationship with Topographic-Slope**

We investigated the relationship between topographic slope and log $V_{S30}$ in order to evaluate the extent to which that variable could be used to refine the $V_{S30}$ model. Figure 7 shows log $V_{S30}$ as a function of topographic slope for both the SRTM30 and SRTM3 elevation datasets, sorted by the final set of generalized geological units F1, F1 and F3. The relationship between $V_{S30}$ and topographic slope is in general extremely poor for the SRTM30 dataset, regardless of the geological unit. There is however a slight tendency for some F3 sites to concentrate in the lower-left part of the graphic (lower $V_{S30}$ corresponding to lower slope). The SRTM3 dataset shows a much clearer correlation between those two variables for F3 sites only. Tentative $V_{S30}$-slope classes for F3 sites are outlined in Figure 7b). A t-test run indicates that the differences in $V_{S30}$ distributions pertaining to the topographic slope classes $0.002m/m < slope < 0.016m/m$ $0.016m/m < slope < 0.100m/m$ are statistically significant at a 5% confidence level (pvalue=0.08). However, since the sample sizes are fairly small (less than 15) the statistical power of the result is low, which means that there is a reduced likelihood that a statistically significant result reflects a true effect.
Evaluation of Proxies Based on Exogenous Data

In the absence of local $V_{S30}$ data, it is common practice to estimate that variable using proxies derived from data pertaining to other regions. We used the database developed in this study to evaluate the performance of $V_{S30}$ proxies proposed in the literature. The topographic-slope is the most widely used $V_{S30}$ proxy (Wald and Allen, 2007). The model relies on correlations between the topographic-slope calculated for the SRTM30 elevation data set and $V_{S30}$ data from California, Utah, Central United States, Taiwan, Italy, and Australia. We use the model as implemented in the global $V_{S30}$ server (Allen and Wald, 2007; see Data and Resources).

In the geologic analogue proxy approach, local geologic units are correlated with geologic categories developed in a different geographic context, which are characterized by $V_{S30}$ distributions. Vilanova et al. (2012) used the geologically-based $V_{S30}$ model developed by Wills and Clahan (2006) for California as a proxy for estimating $V_{S30}$ at sites in Portugal, where ground motion stations were deployed. It has been shown by Stewart et al. (2008) that Wills and Clahan’s (2006) model had no significant bias with respect to $V_{S30}$ distributions for geologic units in Italy. Vilanova et al. (2012) used Stewart et al. (2008) $V_{S30}$ distributions for estimating $V_{S30}$ for geologic conditions that do not have geological analogue in California.

Silva et al. (2015) likewise used both the geologically-based $V_{S30}$ models of Wills and Clahan (2006) and Stewart et al. (2008) to estimate the site conditions for Portugal. They selected, however, dissimilar geological analogues with respect to Vilanova et al. (2012) for the local geological units. For instance, Silva et al. (2015) used “Quaternary (Pleistocene) sand deposits” ($V_{S30} = 302 \pm 46 \text{ m/s}$) as the geological analogue for “Sandstones, more or less argillaceous limestone, sands, gravels, clays, from Miocene and Pliocene”, while Vilanova et al. (2012) correlated that unit with “Tertiary sandstone units” ($V_{S30} = 515 \pm 215 \text{ m/s}$). This example illustrates the difficulties associated with implementing $V_{S30}$ proxies based on the geological analogue methodological approach.

Figure 8 shows the residual distributions for log $V_{S30}$, for both the proxy model based on
slope and that based on the geological analogues. We used the Silva et al. (2015) geologically-based model because since it applies to the most extensively representative geological units of Portugal, the exercise corresponds to a truly blind comparison.

The use of proxies based either on geological analogues or on correlations with the topographic slope shows fairly unbiased total residual distributions of log $V_{S30}$. However, the performance of the methods varies significantly with the generalized geological unit analyzed. The topographic-slope proxy is biased towards lower values of $V_{S30}$ for F1 sites (Igneous, metamorphic and old sedimentary rocks) and it is biased towards higher values of $V_{S30}$ for F3 sites (Holocene formations). It is unbiased for the most extensive dataset which pertains to F2 sites (Neogene and Pleistocene formations). The residual’s distribution shows clear linear trends with the independent variable (topographic-slope) for all geological categories. The residuals are positive for lower values of log $V_{S30}$ slope and negative for higher values of topographic slope, indicating that the relationship between $V_{S30}$ and topographic-slope assumed by the model doesn’t apply to the three subsets.

The geological analogue model of Silva et al. (2015) is slightly biased for sites located both on F1 and F3 geological units (Igneous, metamorphic and old sedimentary rocks, and Holocene formations, respectively). It is however strongly biased towards lower values of $V_{S30}$ for F2 sites (Neogene and Pleistocene formations). This is probably related to the fact that Pleistocene formations in our dataset display log $V_{S30}$ distributions similar to those of Neogene, with a mean value higher than what would be expected from the model of Wills and Clahan (2006). In addition, the fact that Silva et al. (2015) considered some Miocene and Pliocene formations correlated with Pleistocene formations in California, also contributed to exacerbating the bias.
Discussion and Conclusions

We developed a $V_{S30}$ database for Portugal, by acquiring and gathering of $V_s$ profiles using different techniques. Most of the sites in the database have been characterized in terms of $V_s$ depth profile using the seismic refraction technique. Other techniques used include MASW, seismic cone penetrometer and ReMi. Few sites tested using different techniques showed in general compatible $V_s$ depth profiles and corresponding $V_{S30}$ values.

We present a geologically-based $V_{S30}$ model for Portugal, which includes three geological categories: F1 - Igneous, metamorphic, and sedimentary rocks of Mesozoic or Paleogene age; F2 - Neogene and Pleistocene Formations; and F3 - Holocene Formations. The log $V_{S30}$ distributions pertaining to each geologic category are statistically significantly different from each other.

The methodological approach used for developing this model involves an iterative three-step procedure which consists of: 1) Defining a preliminary set of geologically defined units based on the literature; 2) calculating the log $V_{S30}$ distribution for each geologically defined unit; and 3) merging the units according to the results of statistical testing.

We investigated the correlation between log $V_{S30}$ and topographic-slope in order to evaluate the extent to which the last could be used as a variable for refining the model. The topographic-slope has been successfully used for this purpose by Thompson et al. (2014) and Stewart et al. (2014).

We find that, in general, and in what concerns our dataset, the correlation between slope and $V_{S30}$ is poor. The relationship is similar to that reported by Lemoine et al. (2012) for stable continental regions within Europe, with slope values ranging between 0.05 – 0.10m/m, regardless of the log $V_{S30}$ value. Part of the F3 sites (Holocene formations) in our dataset tends, however, to display lower topographic-slope values than the remaining geological defined units. This becomes more evident using SRTM3 elevation dataset for calculating the topographic-slope than using the SRTM30 dataset. In this case the topographic slope can be used to refine the model for F3 sites. However, because the sample sizes are relatively
small, this issue needs to be further investigated whenever a larger dataset pertaining to F3 is available. The correlation between topographic-slope and $V_{S30}$ for Holocene formations is probably related to the relationship between the sedimentation environment and grain size. 

Stewart et al. (2014) also reported that the topographic slope calculated using the SRTM3 dataset revealed better correlation with $V_{S30}$ than either higher or lower resolution digital elevation models. The decrease in performance with higher than 3 arc seconds resolution elevation models has been attributed to canopy effects (Allen and Wald, 2009; Stewart et al., 2014).

We believe that our final model, although relatively broad, is the best that can be achieved with the currently available dataset. Whenever a larger dataset is available, it may be possible to develop a better model, both in terms of accuracy and precision, without compromising the corresponding statistical robustness.

The underlying dataset presents several important limitations. In particular, some data are heavily clustered, some geological units are poorly sampled, and some geographical regions are underrepresented. This database will be used to assist the selection of future sites to be characterized in terms of $V_S$ depth distribution. For instance Holocene formations along the western coast and small basins need to be better sampled in the future. The log $V_{S30}$ distributions show, however, no evidence of systematic trends with geographic location, suggesting that those limitations in the dataset do not significantly affect the results.

We evaluated the performance of models for $V_{S30}$ developed from proxies, such as topographic-slope or surface geology, with data coming from exogenous regions. We used the model based on topographic-slope as implemented by Allen and Wald (2007) and the model based on geologic analogues with the model by Wills and Clahan (2006) for California as implemented by Silva et al. (2015). Both models display overall unbiased residuals between estimated and measured $V_{S30}$ values. However, their performance relative to data pertaining to each geologically defined unit is highly irregular. The model based on topographic slope is unbiased for F2 sites, but strongly biased for both F1 and F3 sites. We find that the model based
on topographic-slope presents, overall, spurious spatial variations of $V_{S30}$. A positive point about this methodology is that it seems to effortlessly be able to partially identify F3 sites, which are in general characterized by lower values of $V_{S30}$ with respect to F1 and F2 sites.

The model based on geological analogues is fairly unbiased for both F1 and F3 sites, but is severely biased for F2 sites. This bias is in part related to the challenges associated with correlating geological units from regions with different geological and lithological conditions. Due to these difficulties, the usefulness of this model in estimating $V_{S30}$ should be regarded in a qualitative sense only.

We conclude that in the absence of endogenous data the method based on analogue surface geology units should be preferred to that based on topographic-slope. We stress however that topographic-slope may be useful in identifying Holocene basins in the absence of more pertinent data. Both proxies should be regarded as supplying qualitative information on the distributions of $V_{S30}$, emphasizing the need for acquiring regional $V_S$ data.

### Data and Resources

The database (flat-file) used in this paper is available as Table S1 in the electronic supplement to the paper. The digital elevation models used, the 3-arcsec resolution (SRTM3) and the 30-arcsec resolution (SRTM30) datasets, were obtained respectively at [http://srtm.csi.cgiar.org](http://srtm.csi.cgiar.org) last accessed June 2017, and [https://dds.cr.usgs.gov/srtm/version2_1/SRTM30/](https://dds.cr.usgs.gov/srtm/version2_1/SRTM30/) last accessed June 2017.

The global $V_{S30}$ model based on slope is available at [https://earthquake.usgs.gov/data/vs30](https://earthquake.usgs.gov/data/vs30) last accessed June 2017.


The figures were plotted using the Generic Mapping Tools package developed by Paul
The SCPTu data was acquired and interpreted using the software provided by the manufacturer (Pagani Geotechnical Equipment, http://www.pagani-geotechnical.com, last accessed October 2017).

The following reference is in the process of publication: “Near surface characterization of the Lisbon and Lower Tagus Valley area, Portugal, for seismic hazard assessment: $V_{S30}$ and soil classification maps” by J. Carvalho, R. Dias, R. Ghose, J. Borges, J. Narciso, C. Pinto, and J. Leote.

Acknowledgements

The Portuguese Foundation for Science and Technology (FCT) funded this work through research projects SCENE (PTDC-CTE/GIX/103032/2008), NEFITAG (PTDC-CTE/GIX/102245/2008), SEICHE (EXCL/GEO-FIQ/0411/2012) and SHARPE (IF-EXPLOR). S.P.V. acknowledges FCT for her contract IF/01561/2014/CP1214/CT0006 under IF2014 Program. CERENA research unit is funded by FCT through strategic project UID/ECI/04028/2013.

Joana Carvalho kindly supplied the refraction microtremor depth profiles from the Carvalho et al (2016) study, and Ana Paula Falco provided assistance in GIS-related issues.

Finally, we would like to acknowledge Editor Thomas Pratt, Associate Editor Mark Stirling and two anonymous reviews for their valuable comments that helped us improving and clarifying the manuscript.
References


Full mailing address for each author

Susana P. Vilanova: Centro de Recursos Naturais e Ambiente, Instituto Superior Tcnico, Universidade de Lisboa, Lisbon, Portugal.

João Narciso: Centro de Recursos Naturais e Ambiente, Instituto Superior Tcnico, Universidade de Lisboa, Lisbon, Portugal.


Isabel Lopes: TPF Planege Cenor and CERIS, Instituto Superior Tcnico, Universidade de Lisboa, Lisbon, Portugal.

Mário Quinta Ferreira: Geosciences Center, Department of Earth Sciences, University of Coimbra, Rua Slvio Lima, 3030-790 Coimbra, Portugal.

Carlos C. Pinto: Halliburton Landmark, 97 Park Dr, Milton, Abingdon OX14 4RY, United Kingdom.

Rui Moura: Instituto de Cincias da Terra (ICT), Faculdade de Cincias da Universidade do Porto, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal.

José Borges: Institute of Earth Sciences, University of vora, Colgio Luis Antnio Verney, Rua Romo Ramalho, 59, 7002-554 vora Portugal.

Eliza S. Nemser: Seismic Hazards Group, AECOM, 300 Lakeside, Suite 400, Oakland, CA 94612, USA.
Table captions

Table 1: Preliminary set of geologically defined units

Table 2: Results of the Tukey-HSD post-hoc tests

Table 3: Statistics for the final model

Figure captions

Figure 1: a) Tectonic setting of the study area. Seismicity $M \geq 3.0$, all magnitude scales, according to ISC (2014) (see Data and Resources) is represented for the period 2000-2014; b) Distribution of $V_{S30}$ values in the database. NEHRP site classes (A - $V_{S30} > 1500$ m/s; B - $760 < V_{S30} \leq 1500$ m/s; C - $360 < V_{S30} \leq 760$ m/s; D - $180 < V_{S30} \leq 360$ m/s; E - $V_{S30} < 180$ m/s) are represented in the background; c) Geographic distribution of $V_S$ depth profiles in the database. The geological units represented are simplified from the 1:500,000 scale geological map of Portugal (Serviços Geológicos de Portugal, 1992); d) Geographic distribution of $V_S$ depth profiles sorted by the characterization method;

Figure 2: $V_{S30}$ as a function of $V_{SZ}$ for $z = 10, 15, 20, \text{ and } 25$ m. The relationships proposed by Boore (2004) for California are represented by solid lines and their 95% confidence limits by dotted lines. The relationship proposed by Boore et al. (2011) for Japan is represented by dashed lines.

Figure 3: Quantiles derived for the normalized $V_{S30}$ and log $V_{S30}$ data distributions as a function of the theoretical quantiles for the normal distribution. The solid line represents the reference 1:1 line. $V_{S30c}$ and $V_{S30z}$ represent, respectively, the datasets derived using constant extrapolation and extrapolation based on $V_{SZ}$.
Figure 4: Normalized frequency distribution for log $V_{S30_{c}}$, sorted by the preliminary set of geologically defined units. The solid line shows the corresponding fitted normal distributions with mean $\mu$ and standard deviation $\sigma$. The dotted lines correspond to the fitted normal distributions for $V_{S30_{c}}$. The fitted normal distribution for the declustered P1 dataset ($\mu = 2.9$ and $\sigma = 0.2$) is represented by a dashed line. NEHRP site classes are represented in the background.

Figure 5: Normalized frequency distribution for log $V_{S30_{c}}$, sorted by the final set of geologically defined units. The solid line shows the corresponding fitted normal distributions with mean $\mu$ and standard deviation $\sigma$. The dotted lines correspond to the fitted normal distributions for $V_{S30_{c}}$. NEHRP site classes are represented in the background.

Figure 6: Geographic distribution for the final $V_{S30}$ model; a) log-averaged $V_{S30}$ value, b) upper limit of the 68% confidence interval for the $V_{S30}$ distribution, and c) lower limit of the 68% confidence interval for the $V_{S30}$ distribution.

Figure 7: $V_{S30}$ as a function of slope sorted by the final set of geologically defined units (F1 - Igneous, metamorphic and old sedimentary rocks, F2 - Neogene and Pleistocene formations, F3 - Holocene formations). The boxes outlined in gray represent the $V_{S30}$-slope class correlations proposed by Wald and Allen (2007) for stable continental regions. The boxes outlined with dashed lines represent tentative $V_{S30}$-slope class correlations for Holocene data in this study.

Figure 8: Residual distributions of log $V_{S30}$ with log $V_{S30}$ values predicted by a) the topographic slope model (see text for details), and b) from the geological analogue method as implemented by Silva et al. (2015).
# Table 1: Preliminary set of geologically defined units

<table>
<thead>
<tr>
<th>Name</th>
<th>Geological Unit</th>
<th>General Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Igneous and metamorphic rocks</td>
<td>Granites, basalts, schists, gabbros, marbles, quartz, turbidites, etc. Includes other formations of Palaeozoic age or older.</td>
</tr>
<tr>
<td>P2</td>
<td>Old Sedimentary rocks (Mesozoic or Paleogene age)</td>
<td>Limestones, marly limestones, dolomites, conglomerates and sandstones</td>
</tr>
<tr>
<td>P3</td>
<td>Miocene formations</td>
<td>Sands, sandstones, clays and conglomerates</td>
</tr>
<tr>
<td>P4</td>
<td>Pliocene formations</td>
<td>Sandstones, gravels, sands and clays</td>
</tr>
<tr>
<td>P5</td>
<td>Pleistocene formations</td>
<td>Sand and clays, terrace deposits</td>
</tr>
<tr>
<td>P6</td>
<td>Holocene formations</td>
<td>Alluvium, mud, sands, clay, silt and sand dunes</td>
</tr>
</tbody>
</table>
Table 2: Results of the Tukey-HSD post-hoc tests

<table>
<thead>
<tr>
<th>Group pairs</th>
<th>q</th>
<th>p-value</th>
<th>Null hypothesis*</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-P2</td>
<td>1.308</td>
<td>0.900</td>
<td>accepted</td>
</tr>
<tr>
<td>P1-P3</td>
<td>5.827</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>P1-P4</td>
<td>6.480</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>P1-P5</td>
<td>5.114</td>
<td>0.006</td>
<td>rejected</td>
</tr>
<tr>
<td>P1-P6</td>
<td>13.164</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>P2-P3</td>
<td>3.902</td>
<td>0.073</td>
<td>accepted</td>
</tr>
<tr>
<td>P2-P4</td>
<td>4.562</td>
<td>0.158</td>
<td>accepted</td>
</tr>
<tr>
<td>P2-P5</td>
<td>3.428</td>
<td>0.158</td>
<td>accepted</td>
</tr>
<tr>
<td>P2-P6</td>
<td>10.481</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>P3-P4</td>
<td>0.939</td>
<td>0.900</td>
<td>accepted</td>
</tr>
<tr>
<td>P3-P5</td>
<td>0.201</td>
<td>0.900</td>
<td>accepted</td>
</tr>
<tr>
<td>P3-P6</td>
<td>8.331</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>P4-P5</td>
<td>1.030</td>
<td>0.900</td>
<td>accepted</td>
</tr>
<tr>
<td>P4-P6</td>
<td>6.984</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>P5-P6</td>
<td>7.449</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>F1-F2</td>
<td>7.639</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>F1-F3</td>
<td>14.987</td>
<td>0.001</td>
<td>rejected</td>
</tr>
<tr>
<td>F2-F3</td>
<td>9.969</td>
<td>0.001</td>
<td>rejected</td>
</tr>
</tbody>
</table>

* The null hypothesis is rejected at a 5% significance level.
Table 3: Statistics for the final model

<table>
<thead>
<tr>
<th>Name</th>
<th>Geological Unit</th>
<th>N</th>
<th>$\mu_{\log V_{S30}}$</th>
<th>$\sigma_{\log V_{S30}}$</th>
<th>$V_{S30}$ (m/s)</th>
<th>$V_{S30} \pm 68% CI$ (m/s)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>Igneous, metamorphic and old sedimentary rocks</td>
<td>23</td>
<td>2.91</td>
<td>0.20</td>
<td>829</td>
<td>[523, 1315]</td>
</tr>
<tr>
<td>F2</td>
<td>Neogene and Pleistocene formations</td>
<td>55</td>
<td>2.67</td>
<td>0.15</td>
<td>470</td>
<td>[329, 672]</td>
</tr>
<tr>
<td>F3</td>
<td>Holocene Formations</td>
<td>29</td>
<td>2.38</td>
<td>0.22</td>
<td>237</td>
<td>[144, 392]</td>
</tr>
</tbody>
</table>

* Represents the log-averaged $V_{S30}$ value
† Represents the 68% confidence interval.
Figure 2
Figure 3: Normalized data quantiles vs. Normal theoretical quantiles. The graph shows the following data points:

- $V_{S30c}$
- $V_{S30z}$
- $\log V_{S30c}$
- $\log V_{S30z}$

The data points are plotted on the graph, with the normalized data quantiles on the y-axis and the normal theoretical quantiles on the x-axis. The graph includes a linear trend line for reference.
Figure 7

Comparison of the model predictions with the observed data for three different scenarios (F1, F2, F3) for SRTM at 30 arcsec and 3 arcsec.
Figure 8a
Figure 8b
Figure 6  composed example

(a)

(b)

(c)

Vs30 (m/s)

- 100 - 200
- 200 - 300
- 300 - 400
- 400 - 500
- 500 - 600
- 600 - 700
- 700 - 800
- 800 - 900
- 900 - 1000
- >1000

Click here to download Figure fig_6_composed_example.eps
Figure 8  composed example

(a)

(b)

All: N=105
μ=−0.02 σ=0.29

F1: N=21
μ=0.17 σ=0.25

F2: N=55
μ=0.02 σ=0.22

F3: N=29
μ=−0.25 σ=0.29

All: N=107
μ=0.08 σ=0.21

F1: N=23
μ=0.05 σ=0.2

F2: N=55
μ=0.17 σ=0.2

F3: N=29
μ=−0.08 σ=0.2
Electronic Supplement to

Developing a Geologically-Based $V_{S30}$ Site-Conditions Model for Portugal: Methodology and Assessment of the Performance of Proxies

by

Susana P. Vilanova, João Narciso, João P. Carvalho, Isabel Lopes, Mário Quinta-Ferreira, Carlos C. Pinto, Rui Moura, José Borges, and Eliza S. Nemser

This electronic supplement includes a spreadsheet file containing the database used in this paper (Table S1).

Table S1 – Plain text comma separated values (csv) file including location of the profiles, methodology used to estimate the $V_S$ depth structure, elevation and topographic slope calculated using both the SRTM30 and the SRTM30 datasets, geological classification using maps at 1:500.000, 1:200:000 and 1:50:000 scales, depth value $z$ for which $V_S$ information is available, $V_{SZ}$ calculated, $V_{S30}$ calculated from correlations with $V_{SZ}$, $V_{S30}$ calculated using constant extrapolation, variability of log$V_{S30}$, and $V_{S30}$ calculated from topographic slope, and associated notes and remarks.
Click here to access/download

**Supplemental Material (All Other Files, i.e. Movie, Zip, tar)**

**Vs30_database_flatfile_revision.csv**
BSSA-D- ______________________ [leave blank for new submissions]

Title:  Developing a Geologically-Based VS30 Site-Conditions Model for Portugal: Methodology and Assessment of the Performance of Proxies

Authors:  Susana P. Vilanova, João Narciso, João P. Carvalho, Isabel Lopes, Mário Quinta-Ferreira, Carlos C. Pinto, Rui Moura, José Borges, Eliza S. Nemser

COPYRIGHT
In accordance with Public Law 94-533, copyright to the article listed above is hereby transferred to the Seismological Society of America (for U.S. Government employees, to the extent transferable) effective if and when the article is accepted for publication in the Bulletin of the Seismological Society of America. The authors reserve the right to use all or part of the article in future works of their own. In addition, the authors affirm that the article has not been copyrighted and that it is not being submitted for publication elsewhere.

To be signed by at least one of the authors (who agrees to inform the others, if any) or, in the case of "work made for hire," by the employer.

Signature

Authorized Signature for Copyright

Print Name (and title, if not author)

Date

PUBLICATION CHARGES
The Seismological Society of America requests that institutions supporting research share in the cost of publicizing the results of that research. The Editor has discretion of waiving publication charges for authors who do not have institutional support. If pages are paid for by SSA, then no further page charge waivers can be requested for two years by any author listed on the paper. Page charges for waived papers cannot exceed 12 printed pages. Rejected papers in which a page waiver was requested will be considered toward the limit of one request per two years. In addition to regular publication charges there is a nominal fee for publishing electronic supplements, which will not be waived. Current rates are available at http://www.seismosoc.org/publications/journal-publication-charge.

Color options:  Color figures can be published (1) in color both in the online journal and in the printed journal, or (2) in color online and gray scale in print. Online color is free; authors will be charged for color in print. You must choose one option for all of the color figures within a paper; that is, you cannot choose option (1) for one color figure and option (2) for another color figure. You cannot submit two versions of the same figure, one for color and one for gray scale. You are responsible for ensuring that color figures are understandable when converted to gray scale, and that text references and captions are appropriate for both online and print versions. Color figures must be submitted before the paper is accepted for publication.

Art guidelines are at http://www.seismosoc.org/publications/bssa/bssa-art-submission-guidelines/

Will publication charges be paid? Check one:

___ BOTH PUBLICATION CHARGES AND COLOR CHARGES WILL BE PAID, and all color figures for this paper will be color both online and in print. This option requires full payment of publication & color charges.

X ___ ONLY PUBLICATION CHARGES WILL BE PAID, and all figures for this paper will be gray scale in print. Color figures, if any, will be color online.

___ REQUEST A REDUCTION IN PUBLICATION CHARGES. Send a letter of request and explanation to the Editor-in-Chief at BSSAeditor@seismosoc.org. Color figures, if any, will be color online but gray scale in print.

Send Invoice to:  IST-ID, VAT: 509 830 072; Address: IST, CERENA Av. Rovisco Pais, 1, 1049-003 Lisboa

If your paper is accepted for publication, SSA requires that you fill out and submit your final files.

Questions regarding billing should be directed to the SSA Business Office,
400 Evelyn Avenue, Suite 201 Albany, CA 94706 USA  Phone 510 525-5474  Fax 510 525-7204

Rev. 2016-06-01