UNIVERSIDADE D
COIMBRA

Alberto João Morgado Ribeiro Malafaia Teixeira

Analysis and optimization of the Drying process of an aerogel for biomedical applications

Thesis submitted to the Faculty of Sciences and Technology of the University of Coimbra for the degree of Master in Biomedical Engineering with the specialization in Biomaterials, supervised by Professor Doutor Marco P. S. dos Reis and Doutora Mara E. M. Braga

October of 2020

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

Alberto João Morgado Ribeiro Malafaia Teixeira

Analysis and optimization of the Drying process of an aerogel for biomedical applications

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

Supervisors:
Professor Doutor Marco P. S. dos Reis
Doutora Mara E. M. Braga

Coimbra, 2020

O tema desta proposta de dissertação de mestrado insere-se num projeto intitulado "Green method to prepare sterilized biopolymers based aerogel" - STERILAEROGEL, referência POCI-01-0145-FEDER-032625, Co-financiado pelo Programa Operacional Competitividade e Internacionalização (FEDER) e pelo Orçamento de Estado.

Esta cópia da tese é fornecida na condição de que quem a consulta reconhece que os direitos de autor são pertença do autor da tese e que nenhuma citação ou informação obtida a partir dela pode ser publicada sem a referência apropriada.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

Marie Curie

Acknowledgements

Acima de tudo, quero começar por agradecer aos meus orientadores, ao Professor Marco Reis com o seu imprescindível conhecimento na área, que sempre foi paciente comigo e me deixou à vontade para desenvolver o trabalho ao meu ritmo, estando sempre disponível para ajudar até ao mais pequeno pormenor e à Professora Mara Braga, pela prontidão e rigor a responder a todas as minhas questões acerca do projeto e pela preocupação em que eu percebesse sempre todo o processo que foi a essência deste trabalho. Sem os Professores e a sua constante disponibilidade, este trabalho não seria possível.

Não podia deixar passar a oportunidade de agradecer à minha família e em especial aos meus pais, que durante estes cinco anos sempre me possibilitaram viver da melhor forma toda a experiência académica, estando lá para absolutamente tudo o que eu precisasse e querendo tanto ou mais que eu, que estes fossem realmente os melhores anos da minha vida. Estarei para sempre grato por isso e por mais.

Aos meus amigos, em especial aqueles que de uma forma mais próxima partilharam estes anos comigo, a minha família em Coimbra. Agradeço só por ter tido a oportunidade de passar estes anos com pessoas tão especiais e que posso realmente dizer que, independentemente das voltas que a vida dê, vocês estarão sempre lá para mim, assim como eu estarei lá para vocês. Todas as manhãs, tardes e noites de estudo, todos os exames e apresentações e, acima de tudo, todas as experiências académicas, nada teria sido o mesmo sem os meus grandes amigos a meu lado.

Queria também deixar um agradecimento especial a mais duas famílias. Primeiro, à Secção de Basquetebol da Associação Académica de Coimbra, que me acolheu ainda nem eu era caloiro e me proporcionou uma das melhores experiências da minha vida, que durante três anos tantas alegrias me deu. Se sei o que é o verdadeiro significado de equipa é graças de vocês. E à família da Junior Enterprise for Science and Technology, que tem um ambiente sem igual e que me deu a conhecer um lado diferente da Faculdade, algo que tanto me ajudou a

desenvolver enquanto pessoa, como enquanto parte duma equipa.

Resumo

A produção de scaffolds para engenharia de tecidos é uma área de investigação, cujo interesse tem aumentado ao longo do tempo. A procura de métodos "verdes" que promovam tanto processos eficientes a nível energético, como a redução de desperdícios, tem requerido metodologias ótimas e que garantam qualidade.

Tomando como inspiração o exemplo de outras indústrias de produção, como a indústria farmacêutica que tem promovido a adoção de abordagens de qualidade pelo design para assim melhorar os seus processos, otimizando-os, este trabalho enaltece a importância de adotar uma abordagem semelhante para o processo de produção de um aerogel para engenharia de tecidos óssea.

Os dados que foram recolhidos para este processo não foram obtidos conforme uma metodologia ótima, como é o caso do design de experiências e, como tal, foi mais difícil extrair a máxima informação a partir de um número limitado de experiências. Contudo, o objetivo principal foi o de encontrar uma forma de estimar os valores dos parâmetros capazes de otimizar o processo. Para isso, em primeiro lugar propôs-se uma divisão sistemática e rigorosa do processo. Assim, aplicou-se um método de seleção de variáveis através de uma regressão passo-a-passo com seleção em frente que levou à obtenção de modelos capazes de representar o processo em estudo. Posteriormente, aplicou-se uma otimização multi-objetivo, com recurso a funções desejáveis.

A abordagem mencionada foi capaz de otimizar várias respostas em simultâneo, sendo estas propriedades específicas do aerogel. Consequentemente, com esta abordagem foi possível não só perceber como os parâmetros do processo influenciam o produto final, mas também como estes podem ser geridos para produzir um aerogel adequado a uma aplicação médica específica. Para além disto, é também possível monitorizar a produção tendo em conta as limitações do processo e/ou dos recursos.

Keywords: Processo de Formação de um Aerogel, Regressão Passo-a-passo, Otimização Multi-objetivo, Seleção de Modelos, Propriedades do Aerogel

Abstract

The production of scaffolds for tissue engineering is a research area that has received increasing interest through the course of time. The search for greener manufacturing methods to promote energy-efficient processes, as well as waste reduction, has demanded for both optimal and quality-assuring methodologies.

Inspired by the example of other manufacturing industries, such as the pharmaceutical industry, which has promoted the adoption of quality by design approaches to its processes in order to optimize them, this work highlights the importance of adopting such an approach to the specific process of aerogel production for bone tissue engineering.

The data collected for this process was not obtained according to an optimal methodology such as the design of experiments, making it more difficult to extract the maximum amount of information from a limited number of experimental runs. Nonetheless, the main goal was to find a way of estimating the best settings of the parameters that optimized this process. Therefore, a rigorous and systematic division of the process was initially proposed. On that account, a variable selection procedure using stepwise regression with forward selection was applied, leading to models able to mimic the process studied. Afterwards, a multi-objective optimization using desirability functions was employed.

This approach was able to simultaneously optimize several responses which were aerogel specific properties. Consequently, it was possible, not only to understand how the process parameters affect the final product, but also how they can be managed to produce aerogels fit for a specific medical application. Moreover, it can also monitor the production with regard to the process and/or resources limitations.

Keywords: Aerogel Formation Process, Stepwise Regression, Multi-Objective Optimization, Model Selection, Aerogel Properties

Contents

List of Figures

List of Tables

List of Abbreviations

AIC Akaike's Information Criterion.
AICc corrected Akaike's Information Criteria.
ANOVA Analysis of Variance.

CCD Central Composite Design.
CMA Critical Material Attribute.
CMC Carboxymethyl Chitosan.
CNTs Carbon Nanotubes.
CPP Critical Process Parameter.
CQA Critical Quality Attribute.

DoE Design of Experiments.
DS Design Space.
DSC Differential Scanning Calorimetry.

EDA Exploratory Data Analysis.
EMA European Medicines Agency.

FDA Food and Drug Administration.

GA Graphene Aerogel.
GCA 3D Composite Graphene-CNTs Aerogel.
GO Graphene Oxide.

LARS Least Angle Regression.
LASSO Least Absolute Shrinkage and Selection Operator.

MOO Multi-Objective Optimization.
MSE Mean Square Error.

OFAT One-Factor-At-a-Time.
OLS Ordinal Least Squares.

QbD Quality by Design.
QTPP Quality Target Product Profile.

RF Resorcinol-Formaldehyde.
RI Refractive Index.
RMSE Root Mean Square Error.
RSM Response Surface Methodology.

SNP Single Nucleotide Polymorphism.

TGA Therapeutics Goods Administration.

1

Introduction

1.1 Scope

Historically, tissue engineering developed into a concept in 1987 that is still what defines it: "Tissue Engineering is the application of the principles and methods of engineering and life sciences toward the fundamental understanding of structure-function relationships in normal and pathologic mammalian tissue and the development of biological substitutes to restore, maintain, or improve function" [1].

Therefore, the goal of tissue engineering is tissue and organ reconstruction using scaffolds, cells and biomolecules [1]. Herein, bone tissue engineering is a widely studied branch, since the replacement or treatment of the damaged bone due to trauma, injury or disease has been one of the main concerns in medical science [2].

Recently, one of the main focus of bone and tissue engineering investigation is on developing greener methods for the synthesis of biomaterials, through the use of safer and biocompatible materials or reducing waste materials, by improving the manufacturing process [2].

In fact, this greener initiative has broader principles, described by Anastas et al (1998) [3], called the 12 principles of green chemistry. The main goal of these and other green principles is to generate energy-efficient processes in order to prevent the formation of wastes, using safer solvents, renewable chemicals and reducing the number of side products, while also economizing matter and energy [4].

Moreover, medical devices have to overcome several obstacles in regulatory agencies such as the Food and Drug Administration (FDA), European Medicines Agency (EMA) or the Therapeutics Goods Administration (TGA) in order to reach the market [5]. Therefore, efforts have been made for monitoring and ensuring the quality of this devices at early stages of their development so that errors will not take place at later stages of the product lifecycle, where they can prove to be fatal

[5]. Herein is where a Quality by Design (QbD) approach might become essential.

The aforementioned approach was firstly named by Joseph M. Juran (1992) [6] and it emphasizes the importance of planning quality into the process of developing a product rather than testing it only when that final product is obtained (quality by testing) [7, 8].

From a regulatory standpoint, only in the beginning of this millennium did the FDA publish the Pharmaceutical cGMPS for the 21st century: a Risk-Based Approach [9], with the FDA finally recognizing that what matters most for product development is the knowledge of the product development process by integrating both science and risk management in development and manufacturing activities [7, 9]. This initiative was embraced by some pharmaceutical companies and by the International Council for Harmonization (ICH), which also published a guideline in 2009 to facilitate the quality control, by introducing the concept of Design Space (DS), which is considered to be the combination and interaction of input variables that have demonstrated to achieve the desired goals [10].

Hence, the QbD methodology was designed to help both the industry and regulatory entities, through continuous improvement [7].

1.2 Motivation

Even though QbD has been increasingly applied in the pharmaceutical industry and the benefits from it are notorious, it has not yet been fully embraced [7]. Furthermore, the scenario is similar for tissue engineering, with the relatively recent implementation of QbD requiring several opinions from researchers, designing engineers and manufacturers to develop a more cemented quality process, since it is also possible to establish a DS for clinical and product quality aspects of a product [5, 11].

Implementing this methodology will translate into better understanding of both the process and the product. It will also enable real-time quality control and reduce the number of tests after the product manufacturing. Moreover, by monitoring every stage of the development process it will be easier to check if the product is meeting the specified regulations and to obtain faster approval times, as the probability of it being rejected at later lifecycle stages is reduced.

The time has come for the tissue engineering industry to follow the recent example of pharmaceutical industry and other industries, who adopted QbD methodologies,

and start making use of the rich data that scaffold manufacturing industries can provide. Besides, the cost and waste reduction that this approach can provide are good enough reason to adopt it.

However, not all the companies and institutions have the resources or funds to develop the necessary QbD statistical tools and procedures. Thus, there is still a considerable gap in the market, especially in the production of aerogels for tissue engineering.

Overall, the work developed at this master thesis aims to clearly define every step of the aerogel formation process (or solvent exchange and supercritical drying) and make use of all the information available to select the most important inputs of the experiment and work within their design space to produce optimized outputs.

1.3 Goals and Contributions

The goal of this project was to optimize the production of an aerogel for bone tissue engineering by building models that could accurately mimic that process. Although the current literature has proved that there are effective ways of optimizing the aerogel formation process, the development of a systematic and reliable methodology remains an on-going challenge.

On that account, a comprehensive analysis of the process and every stage that it involves is required. Additionally, the labeling of every parameter and the measurement of every outcome plays a great role in the fullfilment of the QbD objectives.

Overall, optimization of the process by combining interdisciplinary knowledge of experimenters, researchers and engineers was fundamental, considering that it provided the means to optimize all the steps inherent to this process.

Therefore, this work aimed at providing the best parameter settings for the aerogel solvent exchange and supercritical drying steps. And even though it was not employed a Design of Experiments (DoE) methodology while the experiment was developed, it was still possible to find those optimal settings. For that matter, Exploratory Data Analysis (EDA) was employed to firstly scrutinize the dataset, followed by the use of stepwise regression with forward selection for variable screening and a final Multi-Objective Optimization (MOO) to find the best parameter combination so that all the responses are simultaneously enhanced.

Consequently, the proposed methodology is a significant contribution, since it provides a process optimization regardless of the dataset limitations.

It also presents the perfect opportunity for the implementation of QbD approach with DoE for the product formulation, as both these concepts are thoroughly described, with QbD concepts in particular being introduced to this process.

Moreover, it was also desired to study the efficiency of the solvent exchange process, and therefore it was introduced the possibility of quantifying the predictive relation between the response and a given parameter.

Concluding, this work provides an opportunity for bridging the gap between the work of experimenters and statisticians, by both finding the optimal settings to perform validation experiments and establishing some important concepts and guidelines to implement rigorous quality-based procedures for this process.

1.4 Document Structure

In order to develop the present work, the following structure was adopted: Chapter 2 provides a description of experimental concepts important to understand the process described, as well as optimized ways of developing experiments. It also goes through variable selection procedures that can be used to better achieve the experimental goals. In Chapter 3 the given dataset for the aerogel formation process along with the specifications for building the models to optimize it are mentioned. Furthermore, Chapter 4 describes all the methodology developed to achieve this work's goals and how it was implemented. The results of the aforementioned methodology are described in Chapter 5, as well as the discussion on its strengths and limitations for achieving the optimal settings for the process studied. Finally, Chapter 6 provides an important reflection on all the results achieved, as well as some of the remaning key issues to improve this process in future work.

2

State of the Art

In order for the reader to understand the context of the present study, this chapter comprises a description of QbD's important concepts, which lays the foundation for the optimization of the aerogel formation process that this project aims at. It is important to note that for this QbD methodology to be correctly implemented, the experimental process must be planned in advance with every step of the way being carefully analysed. Therefore, two opposed ways of experimenting are mentioned, with special focus on the experimental design, considering its symbiotic relationship with QbD.

Moreover, some variable selection techniques will be contemplated. Herein it is desired to familiarize the reader with procedures that are able to reduce the number of predictors to build models, especially when the process has more predictors than experimental runs, which was the case of this work. Besides, some insight will be given on how further optimization of manufacturing processes can be induced.

Finally, as the aerogels are the biomaterials demanding property optimization, the historical evolution of this biomaterial is described, with emphasis on the aerogels with similar composition to the one used in this study. Common aerogel properties along with tissue engineering specific properties are also mentioned.

2.1 Quality by Design Methodology

The first step for a QbD methodology is to identify the Quality Target Product Profile (QTPP), which involves describing how the product will be used by the end-user [12]. This is a teamwork effort based on the expertise of scientists, engineers and medical specialists in order to help setting a clear goal for the product, preventing its failure at later development stages, and ultimately accelerating its time to market, as it also facilitates the communication between companies and regulatory agencies [5, 13].

Identifying a QTPP involves selecting a Critical Quality Attribute (CQA) as well, since it is fundamental to guide the product development process [10]. These CQAs are physical, chemical, biological or microbiological characteristics of the product that should be within an appropriate range to ensure product quality and they are identified through risk assessment and experimentation [10].

However, even though it is possible to identify the CQAs at an early stage of the product, the proper definition of their limits requires more knowledge of the developed process [5].

Furthermore, it is essential to identify the process parameters and material attributes that influence the CQAs [10]. A Critical Process Parameter (CPP) is defined by ICH Q8(R2) [10] as a parameter whose variability influences the CQAs. Although there is no definition of Critical Material Attribute (CMA) by the ICH, it can be defined similarly to the CQA, as chemical, physical, biological or microbiological properties of an input material that should be within a certain limit to ensure the quality of the product [14].

Once the CQAs, CPPs and CMAs are identified, it is helpful to create a process flow diagram, which will provide more accuracy regarding the most important processes, while also providing a way to control and monitor quality [5].

As described earlier, risk assessment and experimentation help understanding the relationship between CPPs (and CMAs) and CQAs and it screens the variable ranges that achieve consistent quality [10]. This relationship between inputs and outputs are described in the DS [10]. Consequently, it is of utter importance to describe a DS.

Pharmaceutical products and medical devices are frequently manufactured by a combination of unit operations [14]. For instance, aerogel formation involves gelation, solvent exchange, supercritical drying and final sterilization. In these cases, the output of the first unit operation becomes the input for the subsequent unit operations and, therefore, the process can be analysed from the perspective of each unit operation or as a combination of the unit operations [14].

2.2 The Importance of Experimentation

Science evolves with experimental observation. However, to actually understand the importance of a certain factor to an experiment, the best way of doing so, is by changing it. It is through observation that one can formulate theories or hypothesis,

but those can only be confirmed if experiments are performed [15].

An experiment can be defined as a series of runs or a test, in which some input factors are deliberately changed in order to evaluate the impact of those changes on a specific response or output [15, 16]. Therefore, the inputs are independent variables and the outputs are dependent variables.

Frequently, the goal of experimentation is to identify which input variables are fundamental to optimize a system or to achieve other specific goal [15].

Furthermore, to study the effect of a factor in a response, it is necessary two or more values of the factor. These values are also called levels or settings [16].

It is also important to note that there are two different types of factors: quantitative factors and qualitative factors. On the one hand, quantitative factors are the ones that can assume any value within a continuous range, which gives a certain flexibility to a factor. On the other hand, qualitative factors assume values between a discrete number of values. Hence, these factors bring less flexibility to an experiment [16].

However, this notion of flexibility can be rather tricky, as choosing the number of values of a factor can also be influenced by cost and some practical constraints. For instance, if reducing the number of levels represents a cheaper experimental plan, or if some factor levels combined at the same time can lead to harmful results, adjustments shall be made in order for the experiment to prosper [16].

Thus, not all the factors impact the experiment in a significant way and the ones with a significant effect on the system's outputs are the active factors. The individual effect of a factor in an experiment is a main effect, whereas the joint effect of N factors is a N factor interaction [17]. Subsequently, whenever the effect of a factor on the response depends on the level of another factor, it means that there is an interaction effect between those factors [15, 18].

2.2.1 One-Factor-At-a-Time Approach

The traditional One-Factor-At-a-Time (OFAT) method is an extensively used approach for experimentation. In this method, a baseline set of levels is defined, establishing a baseline value for each and every factor. After that, each factor is varied within its range of levels, while the others are held constant at their baseline values [15]. However, this method suffers from major setbacks. Firstly, as it depends highly of the baseline or starting point, it does not cover the optimal settings and only gives a local knowledge (where the experiments are performed), as opposed to

the experimental design that promotes knowledge in the whole experiment domain (global knowledge) [18, 19].

Additionally, this approach is not good to distinguish between "noise" and actual improvement unless a significant amount of runs under the same conditions is made [19].

Finally, OFAT is not able to identify any possible interactions between factors [15].

In fact, in comparison to the DoE approach, this method is less efficient, since DoE requires fewer runs to achieve the same statistical power and this difference only gets more pronounced as the number of factors on the design increases [20].

2.2.2 Design of Experiments

The Design of Experiments (DoE), or experimental design, as it may sometimes be named, is the strategy of experimentation used to extract the maximum information from the collected data with the fewest experimental runs possible [21].

The experimental design plays a major role on product realization, as it improves product manufacturing, reduces product cost and shortens both the design and development time. Some areas where DoE is widely implemented besides science and engineering, are marketing, market research, transactions and service operations, and general business operations [15].

According to Wass (2010) [22], the DoE methodology consists of three main stages: Screening; Response Surface Methodology (RSM); Model Validation.

Firstly, the factor screening stage shows what factors influence the response(s) the most, providing more information about the overall process. Then, with RSM, these screened factors are used in order to define the optimal result space. Finally, the model validation stage provides conclusions and recommends a course of action [15, 22].

Moreover, in order for DoE to be applied, three basic principles need to be respected. At first, the design needs to make sure that the allocation of materials and the order to perform the experimental runs are random. This randomization also assures that the observations are independently distributed random variables, which is required for performing statistical methods. Secondly, by replicating an independent run of each factor combination, the experimenter can obtain an estimate of the experimental error. At last, to reduce the variability transmitted by the nuisance factors (factors that can influence the system's response, but are not of interest) blocking

is a valuable tool. A block is a set of homogeneous experimental conditions and running an experiment in blocks, with each level of a nuisance factor corresponding to a block, reducing the "unwanted" variability [15].

Wass (2010) [22] describes how the DoE process starts, describing the *screening designs*. Usually, two levels are considered in such a design, with the representation adopted being "+1" for the high level and "-1" for the low level [22]. This might be often called orthogonal coding or the effects coding [15].

Therefore, these designs require fewer runs and promote a positive relationship between cost and information [22]. However, one of the problems that might be encountered is regarding the linearity assumption. Even though in many cases of the 2^k design (design of k factors with 2 levels) the linearity assumption will hold very well in approximation, the interactions between the factors might introduce some curvature and lack of fit to the model [15].

Consequently, it can be of great interest to test the linearity of the model and, because of that, some points known as "centre points", represented by "0" are introduced, as they detect curvature in a rather simple and inexpensive way [22]. It is also stated that three or four centre points should be included in a design, as repetition allows the determination of confidence intervals [23].

When centre points are added to a design, it is assumed that the k factors are quantitative, but it can also be the case that most of the factors are quantitative and a few are qualitative [15].

All in all, when it comes to selecting designs to kick-start the DoE methodology, there are several options, but the more employed ones are the two-level factorial designs: both the full factorial designs and fractional factorial designs [7].

2.2.3 Full Factorial Designs

In a full factorial design, it is possible to study all experimental factors and interactions between factors and how they all influence the responses. In the case of having only two levels and k factors, the full factorial design will consist of 2^k experiments. Therefore, as the number of factors increases, the number of experiments affordable will be exceeded [23].

2.2.4 Fractional Factorial Designs

This resource limitations might mean that using a full factorial design is impossible. Consequently, investigators will look for experimental designs that require fewer experiments [12]. In fractional factorial designs, the experiments cover as much of the experimental domain as possible, without the necessity of doing an excessive number of experiments [23]. For instance, assuming that some interactions will not have much influence on the outputs and that some factors might be "inert" locally, i.e., inactive in the current space of experimental interest, it is possible to investigate three factors with four experiments, seven factors with eight experiments, fifteen factors with sixteen experiments and so on and so forth [23, 24]. In order to characterize a two-level fractional factorial design, the notation used is 2^{k-p}, being a $\frac{1}{p}$ fraction of a 2^k full factorial design, with 2^{k-p} runs [24].

Figure 2.1 shows an example of the representation of a 2^{3-1} fractional factorial on the left side of the figure. This design uses a 2^2 full factorial model, which is a half fraction of the 2^3 design, to obtain the points represented in the figure. Furthermore, when the design was expanded to a three factor format, the experiments form a tetrahedron, which corresponds to the largest volume occupied by four points in three dimensions. In comparison to the right side of the figure, where a 2^3 full factorial design is represented, it is clear that a fractional factorial design covers as much experimental domain as possible [23].

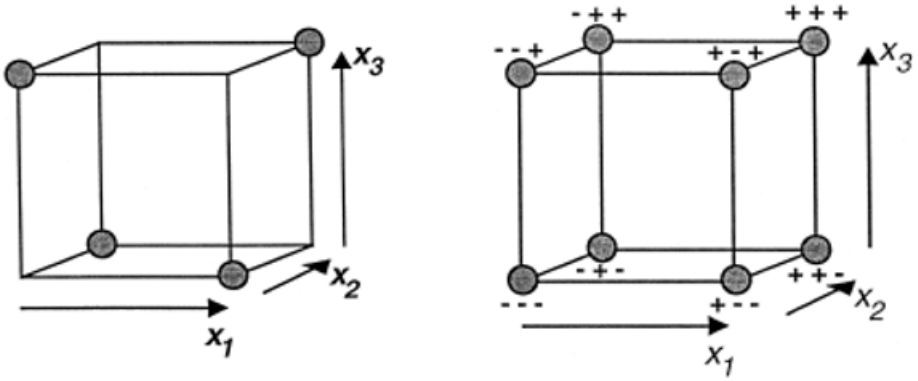


Figure 2.1: Fractional and full factorials. Adapted from [23].

This positive trade-off between information and less experimental runs is only possible by confounding the factors' main effects with interaction effects. In order to know what effects will be confounded, it is necessary to introduce a new term in the design: the generator of a fractional factorial [23].

To that end, the effect coding of a higher-order interaction will be perfectly correlated with a main effect (positively or negatively) [12, 24]. Thus, we cannot separately distinguish the interaction effect from the main effect, i.e., what is obtained is a sum of both effects and one effect is an alias of the other [24].

For instance, if we think of a column of plus signs, I (Identity column), when we multiply another column for I we will have the same coding as in the original column and if we multiply a factor column by itself, we will have I. With a design with three factors such as the one in figure 2.1 and imagining $x_3 = x_1x_2$, then $x_3.x_3 = x_1x_2.x_3$, which means that $I = x_1x_2x_3$. This is the generating relationship or defining relationship for the design. Since the defining relation is composed of the "word" $x_1x_2x_3$, which has the "length" 3, the design is of resolution III [16, 23, 24].

This concept of a design's resolution is of greater importance, because as the resolution increases, the main effects and second-order interactions become less confounded with higher-order interactions. As a matter of fact, for designs of resolution III or higher, no main effects are alias of other main effects [12].

However, if the goal is to have all the second-order interactions not confounded with each other, the resolution of the design needs to be V or higher. The problem with these designs is that as the number of factors increases, it may become unfeasible to perform the amount of runs necessary to develop the design [25].

2.3 Effects Principles

Li *et al.* (2006) [17] identified three key principles to understand the factorial designs: effect sparsity, hierarchical ordering and effect heredity.

According to the sparsity of effects principle or the Pareto Principal of Experimental Design, when there are several factors, the system's responses are likely to be driven primarily by some main effects and low-order interactions [15, 17]. In fact, most investigations assume that third-order interactions or higher are negligible and can be excluded from the model [23].

Furthermore, the sparsity effect promotes the design efficiency through the projection property, since the DoE techniques allow the design to have the desired properties when dimensions of the experimental domain are collapsed [17]. The projection property is well represented in Figure 2.2. This practical use of the design projectivity can be explained for a fractional factorial design of resolution R, as being defined by $P = R - 1$, which means that for any subset of P factors, a complete

factorial can be produced and, therefore, if P or fewer factors are active, there will always be a complete factorial design to study them [24].

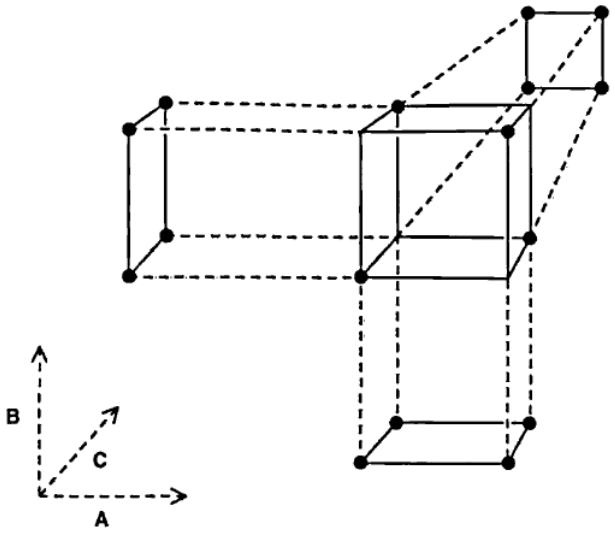


Figure 2.2: Representation of the projectivity property. Taken from [24].

Hierarchical ordering or effect hierarchy states that the main effects tend to be larger than the two factor interactions and these last ones tend to be larger than higher order interactions, suggesting that when resources are limited, the priority is the estimation of lower order effects [16, 17].

Finally, the effect heredity claims that for an interaction to be significant, at least one of its precedent factors needs to be significant as well [16].

2.4 Variable Screening Procedures

When there are many experimental variables, it can be hard to make precise regression models for the outputs [26]. Thus, several variable screening procedures will be described bellow, as they might play a fundamental role on the process being studied in order to extract valuable information about the most significant parameters for the outputs.

2.4.1 Subset Selection Approaches

Recently, a different type of design has been increasingly analysed: the supersaturated design.

This design has the particularity of having at least as much factors as experimental runs [27]. Consequently, as least squares methods cannot be used with more factors

than runs, a subset of the factors must be selected.

Sonoda *et al.* (2007) [26] proposed a stepwise multiple regression method for variable screening in a MOO of electric machinery. It was stated that when there are many independent variables being studied, a screening procedure to reduce the DS is required. Therefore a three-step stepwise multiple regression was employed: firstly, a regression model was chosen; secondly, the significant independent variables were either added or removed according to the F-test; finally the procedure stopped when a specific criterion was met and the best model was found or a pre-defined threshold for the number of iteration steps was reached.

Although stepwise selection methods are efficient for unsaturated and saturated designs, they were found ineffective for supersaturated designs, as they often select many inactive factors as well as missing the active ones [28, 29].

Abraham *et al.* (1999) [27] studied the possibility of using all-subsets for effect screening at supersaturated designs and the results obtained showed that this approach identified better models. Nevertheless, this is a time-consuming approach, since for "k" factors, it will evaluate 2^k models, also making it infeasible when the number of factors is very large [29].

Even so, supersaturated designs should be treated carefully, since even all-subsets regression can be ineffective revealing which factors are active, urging caution and follow-up experimentation [27].

2.4.2 Penalized Least Squares Approaches

These statistical methods were specifically tailored for the analysis of high-dimensional data, i.e. in which the number of factors is bigger than the number of experimental runs [30]. Thus, they are indicated for the aforementioned supersaturated designs.

Hence, it is fundamental to mention the ridge regression which, like OLS includes all the predictors in the model, but with smaller coefficients that minimize a penalized sum of squares [31].

Both the subset selection and ridge regression improve the OLS estimation model. However they have disadvantages: On the one hand, the subset selection is a discrete process and, therefore, small changes in the data can lead to very different models being selected reducing the prediction accuracy. On the other hand, the ridge regression is a continuous process that shrinks coefficients, but does not set any coefficient to zero and consequently, the model interpretation is not easy [32].

Therefore, this last method will not be discussed in detail in this section due to the fact that, as all coefficients become nonzero, it is not suited for variable selection [31]

Herein lies the reasons for the appearance of the different approaches: the Least Absolute Shrinkage and Selection Operator (LASSO) and elastic net.

2.4.2.1 Least Absolute Shrinkage and Selection Operator - LASSO

The LASSO approach shrinks some coefficients, while setting others to zero, combining the best of both the subset selection and ridge regression approaches [32].

Mueller-Using *et al.* (2016) [33] built a predictive model which linked demographical, socio-economical, clinical and immunological factors to a public health response, in this case the co-infection of HIV patients with tuberculosis and used the LASSO method for pre-selection of predictors.

Since classical LASSO treats the candidate variables individually, as it selects an interaction term, it might not select both of its parent factors. Therefore, Choi *et al.* (2010) [34] extended the LASSO method so that it simultaneously fitted the regression model and identified interaction terms obeying to the strong heredity principle. The model produced by this approach proved to be more effective for prediction performance than the classical LASSO method.

Moreover, Usai *et al.* (2009) [35] took advantage of the LASSO's ability to produce a model with only a subset of explanatory variables, each being a Single Nucleotide Polymorphism (SNP). To that effect, LASSO was implemented using a modification of the Least Angle Regression (LARS) [36] and a cross-validation step to define the best size of subset SNPs.

Nevertheless, the LASSO approach is still not very satisfactory when the number of predictors is much bigger than the number of observations [37].

2.4.2.2 Elastic Net

This procedure is similar to the LASSO, as it does automatic variable selection and coefficient shrinkage and it can select groups of correlated factors [37]. When predictors are highly correlated, the Elastic Net adds a quadratic penalty term to the LASSO penalty that accounts for the correlation between predictors [30].

In the specific case of a microarray data set, with thousands of genes (predictors) and usually less than 100 samples, genes with the same biological "pathway" might

be highly correlated, belonging to the same group. Therefore, the Elastic Net is ideal in this situation since it can select the whole group of genes if one of them is selected [37].

For instance, Cho *et al.* (2009) [38] used a simple stepwise approach with Elastic Net regularization in order to account for the correlation structure of SNPs when selecting disease-causing genes of rheumatoid arthritis. To that effect, firstly a screening step to eliminate noise SNPs using logistic regression is performed and, afterwards, predicted disease-causing SNPs are identified through an Elastic Net based variable selection.

2.5 Optimization

Now that it was mentioned several ways of screening important factors for building significant models in order to improve processes, it is also important to briefly mention how the tissue engineering research has optimized its processes.

2.5.1 Multi-Objective Optimization

Often, in optimization studies, there is the need to reconcile conflicting objectives in MOO problems, whether it is in pharmaceutical processes or other medical applications [39].

In tissue engineering applications there is frequently the need to address different objectives, such as biocompatibility, different mechanical properties and biodegradability. These multiple objectives are often contradictory, as the maximization of one property might be detrimental to another [40]. Therefore, there should be a tradeoff between the objective targets and their importance to the formulation process of a certain product.

Bukzem *et al.* (2016) [41] studied the synthesis of Carboxymethyl Chitosan (CMC), which has numerous applications in cosmetics, foods and pharmaceuticals. They developed their work with a DoE approach, using a Central Composite Design (CCD) with subsequent RSM and creation of a global desirability function that graphically shows the parameter values that can maximize this function, thus optimizing the carboxymethylation process.

More specifically, the formation of collagen elastin-like polypeptide composite scaffolds for bone tissue engineering was investigated in order to find the factor settings that maximize the mechanical properties of the scaffolds [40]. For this purpose,

once again a CCD was adopted to study the best settings of 2 factors to maximize 3 outputs, with RSM being used to identify the optimal composition of the scaffolds. Then, with Analysis of Variance (ANOVA) the significant factors or interaction between factors were identified and the following optimization of the responses was performed with a two factor-three objective optimization giving equal weights to all the three responses.

This last work by Gurumurthy *et al.* (2018) [40] is particularly interesting since it also studies the formation process of a scaffold for bone tissue engineering even though its composition is different from the one studied in this thesis. However, it is worth mentioning that they chose to employ desirability functions to optimize their process, which will be a methodology further described in this work at Chapter 4.

2.6 Aerogels

This section is destined to familiarize the reader with the aerogel evolution and its wide range of applications, so that the transition to the next chapters is smoother.

2.6.1 Historical Perspective

The concept of an aerogel was firstly introduced by Kistler in 1931 [42]. It can be defined as a dry gel with a large pore volume, even though the value depends on the solid's characteristics [43]. Typically, these materials are produced through sol-gel processes, with the liquid part of the gel being replaced by gas without the collapse of its porous nanostructure, using a proper drying technique [44, 45].

Historically, for about 70 years after its invention, aerogel research was limited mainly to silica and other inorganic compounds [45]. It was only in 1989 that Pekala published the discovery that Resorcinol-Formaldehyde (RF) gels were not affected by solvent exchange and could be supercritically dried to produce low density aerogels, which translated onto the first "organic aerogels" [46].

The discovery above was what really paved the way for the "investigation boom" that occurred in the aerogel industry on the 21st century [45]. The technology development enabled the improvement of the former "organic gels". For instance, Graphene Oxide (GO) was used as a nanofiller to prevent the shrinkage and collapse of the nanopores of RF aerogels when dried, mainly at ambient pressure [47]. Even Carbon Nanotubes (CNTs) can be incorporated into a Graphene Aerogel (GA) producing a 3D Composite Graphene-CNTs Aerogel (GCA) that is ideal for oil

adsorption, as it is green, ultralight and has high adsorption and mechanical strength [48].

The aforementioned examples denote the evolution of the aerogel technology all the way to the usage of both CNTs and GAs. Moreover, carbide or even carbonitride aerogels were added to the aerogel portfolio [49, 50].

2.6.2 Applications

Although the silica aerogels are the most broadly studied, their application is limited due to low mechanical strength and to their hygroscopic nature [45].

However, there were recent advances to benefit from the high porosity, low density and low thermal conductivity for thermal insulation in aerospace applications, by reinforcing silica aerogels with polymers, improving both the aerogel's load bearing capacity and elastic response [51].

When it comes to life sciences and biomedical applications (especially when biodegradability is required), polysaccharide aerogels can present a suitable alternative to silica aerogels [44]

For instance, Quignard *et al.* (2008) [52], presented ways of preparing polysaccharide aerogels as catalysts.

Other major applications of these aerogels are as drug delivery systems, with starch and alginate aerogels being successful drug carriers [53]. Moreover, hybrid aerogels composed of both polysaccharide (organic) and inorganic components can combine the better mechanical properties of the inorganic part with the biodegradability of the organic [54]. Hence, these ideal properties, allied to the nanostructured porosity in the aerogel and the interconnectivity of the pores that allow the diffusion of nutrients and oxygen, makes these aerogels ideal for tissue engineering, especially for developing scaffolds for bone tissue repairment [44, 54, 55].

Furthermore, organic aerogels are also used for blood compatible implants and exhibit properties that make them suitable for several cardiovascular applications [56].

Finally, polysaccharide aerogels have showed promise regarding being cost-effective and providing a biocompatible and biodegradable solution for food preservation and food processing technologies [57].

3

Process Definition and Dataset

3.1 Process Definition

3.1.1 Aerogel Formation

Considering that the aim of this work is to produce an aerogel suitable for tissue engineering applications, the path chosen was to produce a polysaccharide-based aerogel composed of both alginate, which is a polysaccharide widely studied due to its biocompatibility and relatively low cost [58] and gelatin, which is a collagen-derived material suitable for scaffolds for cartilage tissue engineering [59].

In order for an aerogel to become a scaffold for tissue engineering, it has to go through the following steps: gelation, which induces the formation of a hydrogel, i.e., a gel swollen by water or other aqueous solution; solvent exchange, which leads to the formation of an alcogel and is essential, since the water must be replaced with a solvent with high affinity for the supercritical fluid (Ethanol is the one used in this specific case); and supercritical drying to arrive at the aerogel end product [54].

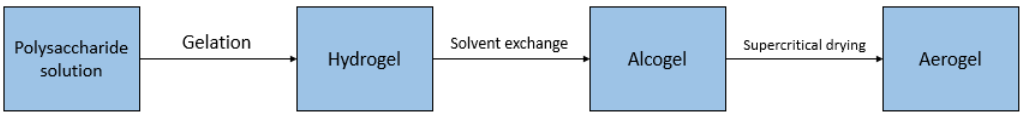


Figure 3.1: Aerogel formation process.

The steps of the aerogel formation that will be further described were performed by Agostinho (2019) [60] and a sample of the final product can be seen in Appendices. Thus, the data available for this study and that was subject of a thorough analysis, was obtained according to the aforementioned work.

Consequently, the steps of the aerogel formation process explored in depth will be

the solvent exchange and supercritical drying.

3.1.2 Solvent Exchange

Within the solvent exchange step, two different approaches are established and both were performed for comparison purposes [60].

Firstly, the solvent exchange step was based on Baldino *et al.* (2015) [61] work, being performed at ambient temperature and pressure, while varying the solvent concentration.

Then, the solvent exchange step was modified as it was done under high pressure and temperature conditions by combining it with the drying step. Some variations of both temperature and pressure were also introduced as the solvent concentration changed. Moreover, the solvent flow had dynamic periods where it had fixed values rather than just being static. This methodology was based on Raman *et al.* (2015) [62] work.

For analysis purposes, this step was clearly divided into 5 stages, with each of these stages being characterized by 5 variables: Pressure; Temperature; Time; EtOH (Solvent) flow; [EtOH] (Solvent concentration). However, it is important to bear in mind that not all the experimental runs went through these 5 stages, since in some experimental runs there were only 4 stages and only in 2 experimental runs were there 5 stages.

Nevertheless, all the experimental runs went through at least 3 stages of solvent exchange, henceforth being the necessary number of stages to arrive at a functional alcogel.

The process flow diagram in Figure 3.2 clearly represents the scenario described above.

3.1.2.1 Alcogel Formation Evaluation

Furthermore, in order to evaluate the alcogel formation process, the analysis already proposed by Agostinho (2019) [60] of calculating the water presence in the alcogel was complemented. In the stated work, it was verified that the Refractive Index (RI) is a simple and reliable method for estimating the ethanol concentration in the sample and thus providing a measure of the amount of water that left the system. Moreover, the Differential Scanning Calorimetry (DSC) was used to measure the mass of water in the sample, hence establishing a relation between the mass of

water and the ethanol concentration.

To validate these relations and further study of the solvent exchange efficiency, an attempt of predicting the mass of water in the sample from the RI was made.

3.1.3 Supercritical Drying

The supercritical drying technique is an alternative to the conventional drying techniques, which is assisted by the use of supercritical fluids, $scCO_2$ in this case, since it has the advantage of preserving the pore structure and textural properties of the gel [63].

Generally, supercritical drying can be classified into two types regarding the contact between the gel and the supercritical fluid: static, if the gel contacts with the supercritical fluid in different cycles or dynamic, if the gel is continuously in contact with the supercritical fluid [54]. Agostinho (2019) [60] developed her work with the dynamic procedure.

Thus, CO_2 Flow, which is one of the variables of the supercritical drying step, does not vary within experimental runs. There are 3 more variables related to this step, which are: Drying Time; Drying Pressure; and Drying Temperature. Figure 3.2 also provides the representation of this step and its role on the overall aerogel formation process.

3.1.4 Process Flow Diagram

For better process understanding, the process flow diagram is represent in Figure 3.2.

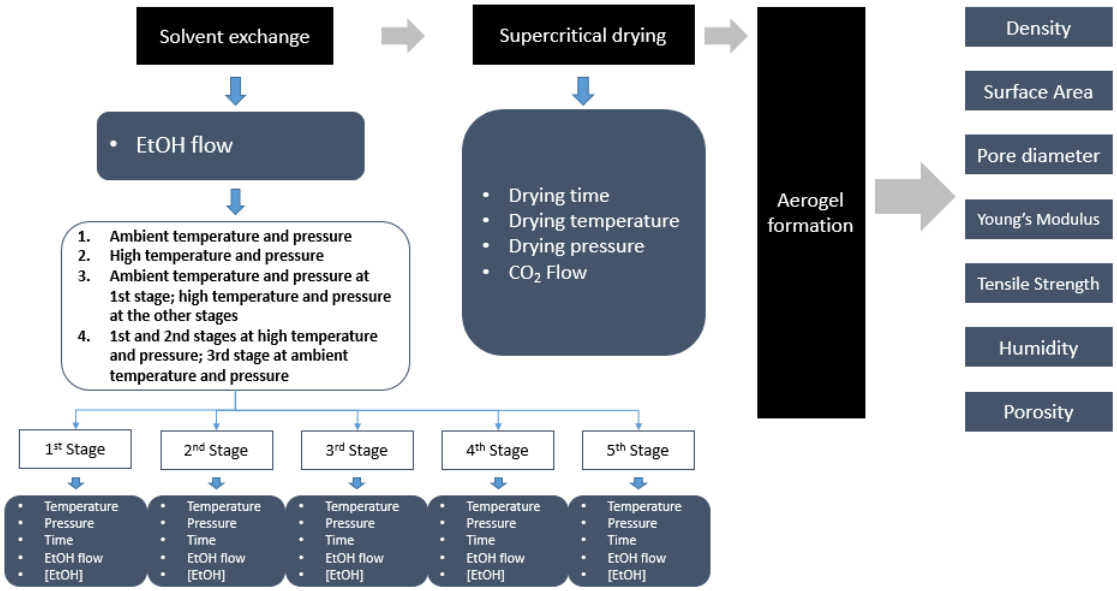


Figure 3.2: Solvent exchange and supercritical drying steps of the aerogel formation process.

3.1.5 Dataset

3.1.5.1 Description of the Dataset

Two steps of the aerogel formation process will be analysed as described before: the solvent exchange and the supercritical drying. From a "step" standpoint, the solvent exchange can be divided into 5 unit operations, which will be further named as stages. Whereas for the supercritical drying, which can be considered as only 1 unit operation, it will continue to be mentioned as a single step.

However, none of the unit operations produces unit-specific responses or CQAs. The CQAs are from now on primarily named as responses or outputs, being only measured at the end of the process (after both steps were performed).

Process Inputs

These steps of the aerogel formation are characterized by 29 process parameters or input factors. As previously described in sections 3.1.2 and 3.1.3, a total of 25 factors were measured in the 5 stages of the solvent exchange and 4 factors in the supercritical drying step. The inputs for this process are clearly represented in Figure 3.3.

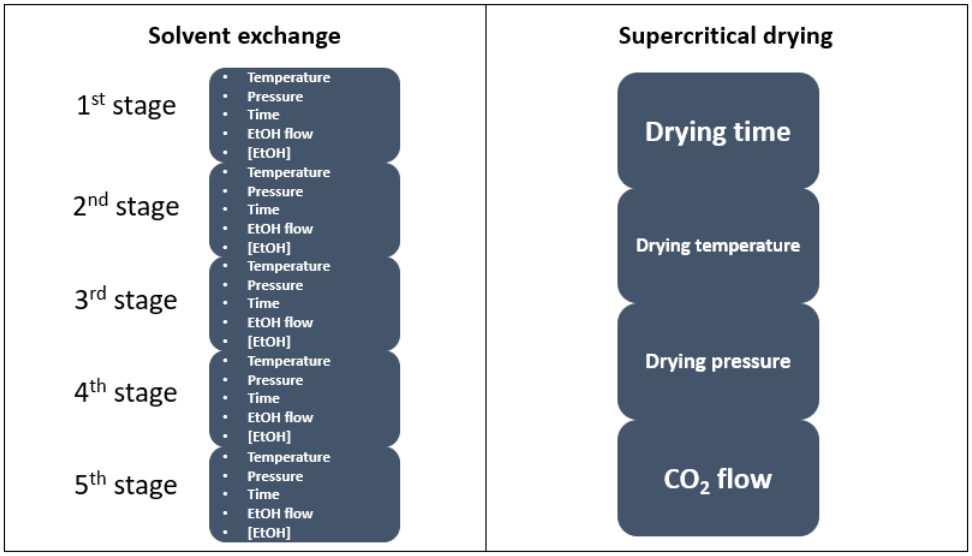


Figure 3.3: Representation of all the process inputs.

It was interesting to see how a QbD methodology could be adapted to the parameters of the process studied:

- **CMAs**: Concentration from all the 5 stages of the solvent exchange step.
- **CPPs**: Temperature, Pressure, Time and Solvent flow from all the 5 stages of the solvent exchange step; Time, Temperature, Pressure and CO_2 flow from the supercritical drying step.

Nevertheless, as mentioned before the solvent exchange step combined static solvent flow for some samples, whilst for others the flow was dynamic, and therefore, it is not quantifiable. The inputs were regrouped accordingly, omitting from now on the solvent flow factor from the list presented in Figure 3.3.

Process Outputs

When it comes to the process outputs or CQAs, none of the unit operations produces unit-specific responses, being only measured at the end of the process (after both steps were performed).

- **CQAs**: Density; Surface Area; Pore Diameter; Young's Modulus; Tensile Strength; Humidity; Porosity.

These responses are essentially aerogel properties and all of them are continuous variables. Table 3.1 provides all the important information about the specific outputs of the aerogel, in order to make the data analysis.

Taking a look at the desired characteristics for polysaccharide-based aerogels and

the specifications of the responses available in the dataset, a brief guide on how the responses were optimized is provided:

- Density: All the densities reported in the dataset have a low value (bellow 1,45 g/cm^{-3}) which is in complience with the statement that aerogels are typically low density materials [44]. Furthermore, the measured density is the skeletal density or true density, which is the mass of the particle, divided by its volume, excluding the porosity [64]. Consequently, the goal was to minimize the density values.

- Surface Area: The surface area is intimately related to the porous structure of the aerogels, since a high surface area (surface to volume ratio) promotes protein adsorption as well as cell adherence and migration through the porous network [1]. Polysaccharide-based aerogels are characterized by having high surface areas, between 70 and 680 m^2/g [45]. As expected, the values in the dataset are mainly within this threshold and the objective was to maximize them.

- Pore Diameter: The application of a scaffold can vary greatly depending on its pore size. For pore widths not above 2 nm, the pores are considered micropores, whereas pores between 2 and 50 nm are designated as mesopores and lastly, macropores have their pore size above 50 nm [65]. This dataset comprises only mesopores, hence the primary goal was inducing macroporosity in the aerogel.

  Larger pores and pore interconnectivity are essential for cell attachment and adhesion, as well as subsequent nutrition and proliferation. Nevertheless, for medical applications such as the developed aerogel, there should be a balance of different pore sizes as larger pores also affect the scaffold stability and, therefore, there should be a tradeoff between cellular and mechanical needs [1].

- Young's Modulus: Defined by the ratio between tensile stress and tensile strain [1]. For instance, the Young's Modulus for the cortical bone is between 3000 and 30000 MPa [66], but the target value for this response depends on the scaffold exact application. Nevertheless, as the dataset comprises values between 3 and 13 MPa, the goal for this response was maximization.

- Tensile Strength: As it corresponds to the maximum tensile stress handled by the aerogel, its increase implies capability of supporting higher tensions [1]. It was reported that the maximum tensile strength for the bone was 200 MPa

[67]. As an aerogel for bone tissue engineering, the goal was to maximize the tensile strength.

- Humidity: Essentially measures the efficiency of the solvent exchange step and the experimenters established a desired value of less than 5% for it, whereas the values obtained for the dataset (bellow 19%) might imply incomplete alcogel formation [60]. Therefore, minimization for the humidity response was proposed.

- Porosity: The aerogels are differentiated by their highly porous structure and open porosity, with higher mesoporosity than the native extracelular matrix [68]. This statement is supported by the around 80% porosity observations in the dataset, which is similar to the tabulated value for the trabecular bone porosity (79,3%) [69]. However, similarly to the Pore Diameter response, macroporosity is required and therefore porosity's maximization was desired.

Table 3.1: Summary of the optimization goals for the process outputs.

Response	**Goal**
Density	Minimize
Surface Area	Maximize
Pore Diameter	Maximize
Young's Modulus	Maximize
Tensile Strength	Maximize
Humidity	Minimize
Porosity	Maximize

Even though the data was not extracted according to a DoE methodology, it is appropriate to mention that in order to produce a full factorial design, it would require 2^{29} runs. Besides, this scenario would only be possible with each factor having 2 levels, which does not occur in the context of the work developed.

Nevertheless, with only 16 experimental runs, the dataset needed cutting on every redundant factor for the design in order to understand which effects are significant for the process and to even learn what changes and adjustments would be needed to be made for a future adoption of a QbD methodology.

Moreover, as reported in Chapter 2, in experiments such as the one performed, interactions between factors are of utmost importance. Specially the two-way interactions.

3.2 Model Preparation

In order to overcome all the setbacks of having a dataset with the described conditions, some efforts were made to simplify it.

Firstly, the dataset was separated by solvent exchange stage to analyze the influence of each stage combined with the supercritical drying step. Therefore, since there were 5 stages for the solvent exchange process and the supercritical drying was considered a single step, 5 different model proposals could be analysed.

However, for the 4^{th} and 5^{th} stages there are not enough runs to evaluate the effect of these stages on the process responses. Notwithstanding this, the first 3 stages were individually analyzed and the models were created considering the above specifications, overall summarized in Figure 3.4.

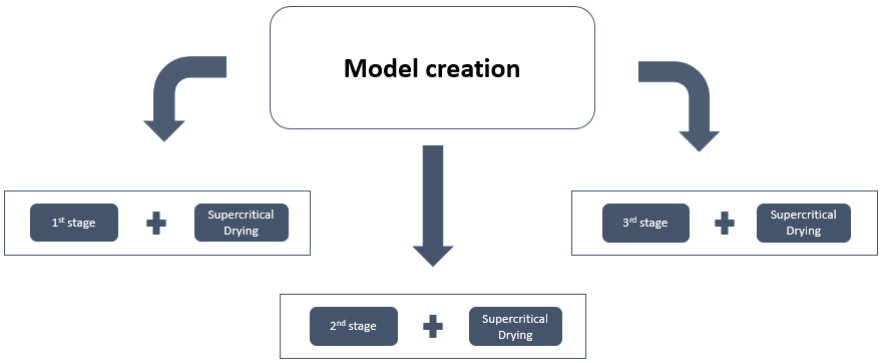


Figure 3.4: Representation of the stage-wise model creation process, in which the responses are analyzed based on the values of each stage of the solvent exchange step individually, plus the supercritical drying step.

Then, since the goal was to study and optimize the entire process of solvent exchange and supercritical drying of the aerogel, the first 3 stages that were analyzed separately before, were all combined for evaluating their effect together on the responses, making a global model. Furthermore, as the supercritical drying was a single step it could be analyzed with the combination of the 3 stages of solvent exchange like the Figure 3.5 demonstrates.

For this combination of stages, notwithstanding the fact that the responses remained the same, the inputs had to change for the solvent exchange step. Since the first 3 solvent exchange stages were combined, the variables were built differently.

Hence, the inputs changed so that all the parameters related to the solvent exchange steps became an average of all the 3 stages of that process except for the Total Time,

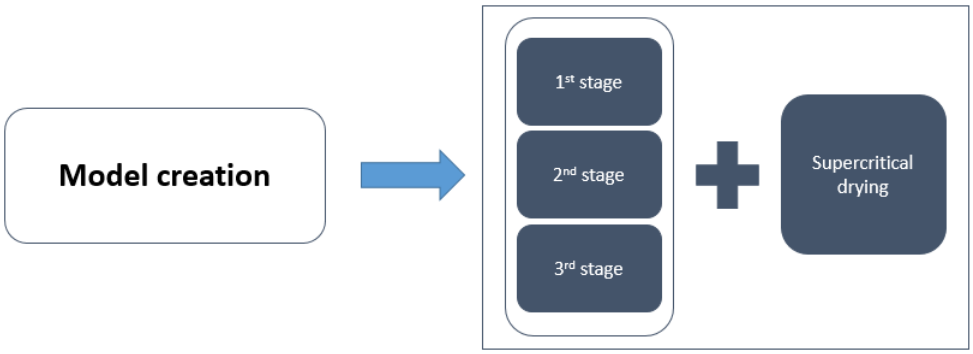


Figure 3.5: Representation of the global model creation when all the first 3 stages of the solvent exchange are combined, plus the supercritical drying.

which corresponds to the sum of the time for each stage. Since the supercritical drying was a single step, it did not need to be further adapted.

- Inputs: Average temperature; Average pressure; Average Concentration; Total time; Drying time; Drying temperature; Drying pressure; CO_2 flow
- Outputs: Density; Surface Area; Pore Diameter; Young's Modulus; Tensile Strength; Humidity; Porosity

By opting for this approach, the goal was to verify if the models selected for both the stages individually and all of them combined were characterized by the same significant factors. Furthermore, the selected values of those factors to optimize the process also had to make sense for both the approaches aforementioned.

Concluding, it was tested whether the global approach for the combination of stages was a generalization of the stage-wise approach. With this generalization, it was possible to obtain the best and simplest models that achieved the optimum operating conditions for the experimental parameters, given the studied dataset.

4

Methods

4.1 Exploratory Data Analysis (EDA)

Before going deeper into the data analysis methodology, it is essential to have an overview of the dataset and its main variation patterns and relevant features for the analysis goals.

Hence an EDA stage was considered since it uses techniques, such as summary statistics and graphical tools, to find patterns in the data, check assumptions and find anomalies [70].

Statistically speaking, the term correlation is related to whether an association of two variables exists or not. This association is measured by a unit-free correlation coefficient, which can assume values between -1 and +1 [71]. For perfectly positive correlations, the coefficient assumes the +1 value, whereas -1 implies a perfectly negative correlation between the variables. When there is no correlation, the coefficient value is 0. There are two main types of correlation coefficients: the Pearson's product moment correlation coefficient and the Spearman's rank correlation coefficient [72]. The Pearson's coefficient is suited for the assessment of normally distributed variables, linearly related. On the other hand, when the relationship in non-linear, but monotone (increasing or decreasing), the Spearman's coefficient is the most appropriate [72, 73].

The Pearson's coefficient is computed with the formula:

$$r = \frac{\sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{[\sum_{i=1}^{n}(x_i - \bar{x})^2][\sum_{i=1}^{n}(y_i - \bar{y})^2]}} \tag{4.1}$$

The Spearman's rank correlation coefficient can be obtained by the following ex-

pression:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{4.2}$$

With d_i being the difference between ranks for x and y.

Nonetheless, correlation does not necessarily mean causation and caution must be taken when interpreting correlations [71].

4.2 Multiple Linear Regression

Multiple linear regression is the extension of simple linear regression to multiple predictors (or input variables) [74].

Consequently, the model for a response is now represented as:

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \varepsilon_i \tag{4.3}$$

For each observation represented by i, $i = 1, ..., n$ and n being the number of observations. p represents the number of independent variables. ε_i represents an error term, which is a random variable that introduces *noise* in the relationship between the dependent variable and the predictors.

4.3 Parameter Estimation through Ordinary Least Squares (OLS)

The method used for estimating the model regression coefficients was OLS. The OLS finds the best fit to data points so that the sum of squared deviations from the line (squared residuals) is minimal [75].

Based on equation 4.3, it was possible to estimate the regression function:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + ... + \hat{\beta}_p x_{ip} + \hat{\varepsilon}_i = \hat{y}_i + \hat{\varepsilon}_i \tag{4.4}$$

Where $\hat{y}_i$ is the estimated value of y_i and $\hat{\varepsilon}_i$ is the difference between the actual and the estimated values, from now on mentioned as residuals. Furthermore, OLS

adopts the *least squares criterion*, which states that the sum of squared residuals should be as small as possible [75].

Hence the sum of squared residuals is a function of the parameter estimates and was estimated as follows:

$$\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = f(\hat{\beta}_0, \hat{\beta}_p) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.5}$$

The optimal values for the parameter estimates are the ones that minimized the sum of squared residuals.

To clearly demonstrate how the parameter estimates are obtained, matrix notation is more convenient, with equation 4.5 assuming the following form:

$$\sum_{n=1}^{i} (y_i - x_i^T b)^2 = (y - Xb)^T (y - Xb) \tag{4.6}$$

Where b is a candidate coefficient vector for the OLS parameters, $(\beta_0, \beta_1, ..., \beta_p)^T$ while X is the $n \times (p + 1)$ matrix with each row being an input vector and y being the n-vector of the responses [76].

Consequently, the unique global minimum at b was obtained by:

$$\hat{b} = \underset{\text{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \right\} = (X^T X)^{-1} X^T y \tag{4.7}$$

However, in order for the model parameters to be estimated by OLS, certain statistical assumptions regarding multiple linear regression had to hold no matter how the data was generated [75]. More specifically, the errors should have zero mean, as well as being independent (no autocorrelation) and homoscedastic. The last means that the error term has constant variance in each observation. Finally, these errors are assumed to be normally distributed [77].

4.3.1 Model Fitting Evaluation

4.3.1.1 Coefficient of Determination

The coefficient of determination (R^2) is a dimensionless goodness-of-fit measure that assumes values between 0 and 1 and corresponds to the variance in the data that

can be explained by the model, with higher values indicating a better model fit [78]. It was formulated as:

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.8}$$

In which the $SS_{residuals}$ is the residual sum of squares and SS_{total} is the total sum of squares. According to equation 4.8 and the aforementioned definition of R^2, it can be deducted that the difference between SS_{total} and $SS_{residuals}$ is the variation explained by the model [79].

However, when several predictors are added to a model, the R^2 will increase accordingly. That being said, adjusted R^2 allows the comparison of models with different numbers of predictors, introducing a penalization to each additional predictor introduced [79]:

$$Adj.R^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1} = 1 - \frac{SS_{residual}/DF_{residual}}{SS_{total}/DF_{total}} \tag{4.9}$$

So that n is the number of observations and k the number of predictors. Hence, the adjusted R^2 can be negative and will never be bigger than R^2.

4.3.1.2 Root Mean Square Error

The Root Mean Square Error (RMSE) is one of the most common methods for evaluating model performance and the smaller its value, the better the model's performance [80]. It can be formulated as:

$$RMSE = \sqrt{\frac{SS_{residual}}{DF}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{DF}} \tag{4.10}$$

From 4.10 it can be seen that RMSE is the square root of the Mean Square Error (MSE), in which DF represents the degrees of freedom of the error.

4.4 Stepwise Regression

The stepwise regression method aims at selecting a subset of input factors that explains the response efficiently [26].

This method combines two variable selection approaches: the forward selection and the backward elimination [81].

4.4.1 Forward Stepwise Selection

In this work, the forward stepwise selection was implemented. It started with a model with only the intercept. Then, the first variable entered in the model was the one with the smallest p-value (the most significant value) [82]. Afterwards, the most significant terms to improve the model are added, with only one term being added at each step. In stepwise methods, if a predictor is found to be no longer significant after the addition of others, it can be removed. At last, the best model according to a stopping rule is the chosen one. For this procedure, the chosen stopping rule was the minimum corrected Akaike's Information Criteria (AICc), as the best model was the one with lowest AICc.

Besides, in a process such as the one analyzed, interactions between factors play a fundamental role. Still, when an interaction term is added to a model, it is important to note that its parent factors should also be included in the model, since they are also significant terms according to the effect hierarchy principle.

Thus, the *Combine* option of JMP Pro [82] was employed. This option calculates 2 p-values for separate tests. The first p-value, $p1$, groups the interaction term with its parent factors, calculating a joint F-test. The second value, $p2$, reflects the significance of the interaction term for entering the model, after its parent factors are already in the model. Finally, the final significance probability to the entry of the interaction term is $max(p1,p2)$, thereby avoiding the inclusion of non-significant interaction terms.

4.5 Model Selection

In a process like the one addressed in this work, where there are many factors to be considered, a model selection from a set of competing models with different numbers of factors derived from the dataset, is required. This approach has received greater attention, since the best model is the least complex and that can provide most information about the process [83]. For model selection, an important concept is the likelihood function, $L(\theta)$, which is the product of the probability density functions $f(x|\theta)$ at the obtained data values [82]:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{4.11}$$

For the data values obtained, the maximum likelihood estimation finds the values for the parameters that optimize $L(\theta)$ [83].

4.5.1 Akaike's Information Criteria

The Akaike's Information Criterion (AIC) [84] provides an estimate of the distance between the fitted model and the unknown true mechanism of the process that is responsible for the collected data [85]. The aforementioned criterion is based on the maximum likelihood concept since it is formulated as:

$$AIC = -2logL(\theta) + 2k \tag{4.12}$$

With the $log(L(\theta))$ being the loglikelihood and the goal the minimization of the negative loglikelihood [83].

However, if the number of factors (k) is large relatively to the number of replicates n with a ratio $n/k < 40$, then a AICc fits the models better [85]. Therefore, the AICc was used at this work:

$$AICc = -2logL(\theta) + 2k(\frac{n}{n-k-1}) \tag{4.13}$$

Consequently, this value was used to compare several models within the dataset to determine the best fitting one. And although the model with the smallest AIC was often the preferred one, it were the AIC differences that mattered most in selecting the most significant model.

This AIC difference can be computed as Δ_i [85]:

$$\Delta_i = AIC_i - AIC_{min} \tag{4.14}$$

In order to evaluate a model, it is argued that models with $\Delta_i \leq 2$ have substantial evidence of being a good fit for the process [85].

4.6 Multiple Response Optimization

In manufacturing processes, it is common to have several responses. Although each response requires specific levels of the input factors in order to optimize it, relationships between the different responses may lead to trade-offs that lead to sub-optimal individual performances, but a better overall behaviour [86]. Therefore, the set of

experimental conditions that lead to the best overall response is called the optimum condition for the process [86].

To obtain the optimum condition for the process, one can use the multiple response optimization techniques.

Generally, in DoE the optimization of the process/product is an inherent phase of the RSM [87]. However, in this work RSMs are not followed from the beginning. Therefore, the path chosen for the multiple response optimization was by using desirability functions.

4.6.1 Desirability Functions

The desirability function was introduced by Harrington (1965) [88] and is a transformation of the estimated response models ($\hat{y}$) to a 0 to 1 scale [87]. These desirability functions (d) can have different shapes and 0 represents a completely undesirable response, as opposed to 1 which represents a perfectly desirable response [86].

These individual desirability functions were modified by Derringer and Suich (1980) [89], which promoted their use in this optimization, since they are smooth piecewise functions with three defining points. For instance, when the goal was to maximize the response, the individual desirability function one-sided transformation was:

$$d_i = \begin{cases} 0, & \hat{y} < L \\ \frac{\hat{y}-L}{T-L}, & L \leqslant \hat{y} \leqslant T \\ 1, & \hat{y} > T \end{cases} \tag{4.15}$$

On the contrary, to minimize the response, the one-sided transformation corresponded to:

$$d_i = \begin{cases} 0, & \hat{y} > U \\ \frac{\hat{y}-U}{T-U}, & T \leqslant \hat{y} \leqslant U \\ 1, & \hat{y} < T \end{cases} \tag{4.16}$$

For the cases of maximization and minimization above, L stands for lower bound and U for upper bound, while T is the target values desired so that $L \leqslant T \leqslant U$. Besides, the curve for the minimization function will be the maximization curve flipped along an horizontal line that passes through a center point in the plot [90].

In order to optimize multiple responses, an overall desirability function (D) was created. This function corresponds to the geometric mean of the desirability functions for the individual responses [90].

Representing the k responses' individual desirability functions by $d_1, ..., d_k$, the overall desirability function will take the form:

$$D = \sqrt[k]{d_1}...\sqrt[k]{d_k} \tag{4.17}$$

In this work, all the responses being optimized were assumed to have the same importance, thereby being equally weighted.

Therefore, by choosing the optimum condition for the process, the overall desirability function is maximized, with the individual desirability functions all being optimized simultaneously.

Furthermore, since all the responses studied were continuous, the optimization algorithm used for maximizing the overall desirability function was a gradient descent algorithm [90].

4.7 Significance Level

For this work, a significance level (α) of 0,05 was considered. However, as the number of experiments is limited, p-values close to this significance value were also considered as relevant for the analysis.

4.8 Software

Data analysis was entirely performed with JMP-Pro 15, including EDA [91], variable selection using stepwise forward selection [82] and MOO [90].

5

Results and Discussion

In this chapter, the results of the exploratory data analysis will first be presented, followed by the results of the methodology proposed in Chapter 4. Finally, the global optimization of the process parameters selected for the best models will be presented and all the required tradeoffs between responses for this procedure will be analysed.

However, since this aerogel formation process was not performed following a systematic DoE approach, the models produced in order to achieve this project's main goal were not obvious and data analysis was adapted in order to take the most out of the existing experiments.

Furthermore, it is important to mention that, throughout the remaining of this work, the input factors were also named as inputs, factors or predictors, whereas the outputs were also mentioned as responses.

5.1 Exploratory Data Analysis

As an overview of the entire process, it was important to analyse both the inputs (independent variables) and the outputs (dependent variables).

5.1.1 Inputs

Due to the fact that 29 factors was a large number of factors to analyse, even so when they were not collected according to a DoE methodology, the overview provided by the EDA shed some light regarding possible connections and patterns in the dataset and, consequently aided in its simplification and comprehension.

Through correlation analysis it was identified that in each and every stage of the 5 that compose the solvent exchange step, the temperature and pressure are perfectly correlated, since their correlation coefficient (r) was 1 within every stage.

Thus, it was verified that the temperature and pressure are confounded, i.e. one is an alias of the other, hence indistinguishable. Consequently, after deciding with the experimenters, for further analysis only the pressure was used.

Furthermore, the factor values for Drying Temperature, Drying Pressure and CO_2 Flow of the supercritical drying step did not vary throughout the experiments and, consequently their contribution to the regression models could not be assessed. Therefore, they were omitted from the model building for the rest of this thesis.

5.1.2 Outputs

Table 5.1: Spearman's rank correlation coefficients for the dataset's responses.

Responses	Density	Surface Area	Pore Diameter	Young's Modulus	Tensile Strength	Humidity	Porosity
Density	1						
Surface Area	0,6990	1					
Pore Diameter	0,0162	0,2324	1				
Young's Modulus	0,0107	-0,0500	0,0214	1			
Tensile Strength	-0,0321	-0,1179	0,1393	0,6786	1		
Humidity	-0,1854	0,0706	0,2382	0,1143	-0,1393	1	
Porosity	0,4091	0,6471	-0,3559	-0,1929	-0,3071	-0,1441	1

In Table 5.1, the Spearman's rank correlations between the responses are represented. Visualizing the values obtained, it is important to mention that the correlations between responses that were significant were the following:

- Correlation between Density and Surface Area: $r_s = 0{,}6990$; $p{-}value = 0{,}0026$
- Correlation between Surface Area and Porosity: $r_s = 0{,}6471$; $p - value = 0{,}0067$
- Correlation between Young's Modulus and Tensile Strength: $r_s = 0{,}6786$; $p - value = 0{,}0054$

Therefore there is a significantly positive correlation between the responses mentioned above. On that account, when one of the responses changed according to certain process parameters, the correlated response followed the same pattern. However, it is important to remember that the displayed results do not support the theory of the variation of a response being caused by the variation of the other.

Finally, it is important to mention that an outlier analysis was also performed in the next section (see also the Graphics presented in the Appendices).

5.2 Model Selection

Taking into consideration the limitations of the dataset, it was tested if the global model approach was indeed consistent with stage-wise models. Therefore, for the stage-wise approach, the 1st, 2nd and 3rd stages' best models and subsequent estimated parameter coefficients were obtained. Additionally, the analysis of the best model and subsequent parameter coefficient estimation was also developed for the global approach, which combined the 3 stages.

It is worth mentioning that even though the supercritical drying step is not always mentioned in model descriptions since it did not vary, it was always taken into consideration in the model building process.

Moreover, for both the stage-wise and global approaches, the variable selection was performed by stepwise regression with forward selection, using the *Combine* option to enforce effect heredity. For this procedure, both the main effects and interaction effects were considered.

Besides, the best models for each response, which are the ones presented in the next section, were the ones with the lowest AICc, whereas the model parameters' coefficent estimation was carried out by OLS. In this model analysis, out of the 7 responses, for only 4 of them were the models obtained composed by at least one relevant main effect, which was the case of Density, Young's Modulus, Tensile Strength and Humidity.

5.2.1 1st Stage

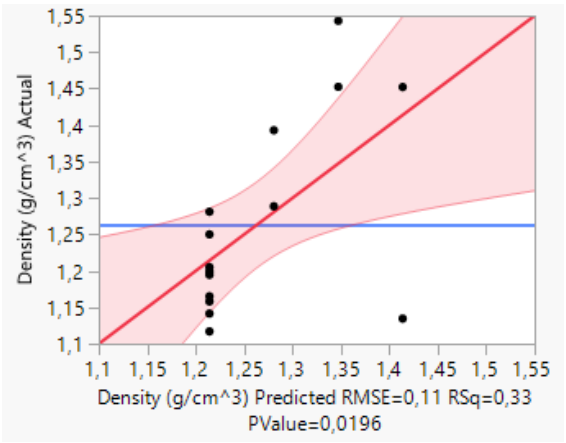
At first, only the 1st stage of the solvent exchange along with the supercritical drying step was analysed. The analysis consisted of an initial variable selection performed by the stepwise regression with forward selection, followed by an OLS parameter estimation.

Table 5.2: Best model obtained for each response using forward stepwise selection with OLS parameter estimation, when only the 1st stage of the solvent exchange and the subsequent supercritical drying were considered.
* relevant term

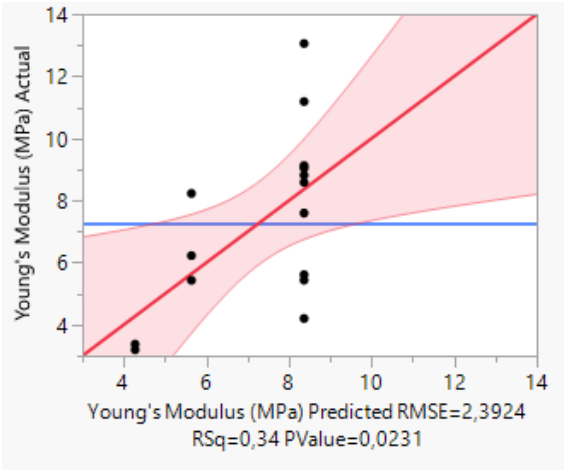
Response	Significant Terms	Estimates	*p-value*	R^2	Adj. R^2	RMSE
Density	Drying Time	0,066	0,020	0,33	0,28	0,11
Young's Modulus	Time 1st Stage	0,045	0,023	0,34	0,29	2,39
Tensile Strength	[EtOH] 1st Stage*	0,782	0,060	0,25	0,19	10,01
Humidity	Drying Time	-0,996	0,024	0,31	0,26	1,71

Table 5.2 shows the responses that were able to produce models with relevant terms with its parameter estimations and significance value, as well as the models' R^2, adjusted R^2 and RMSE.

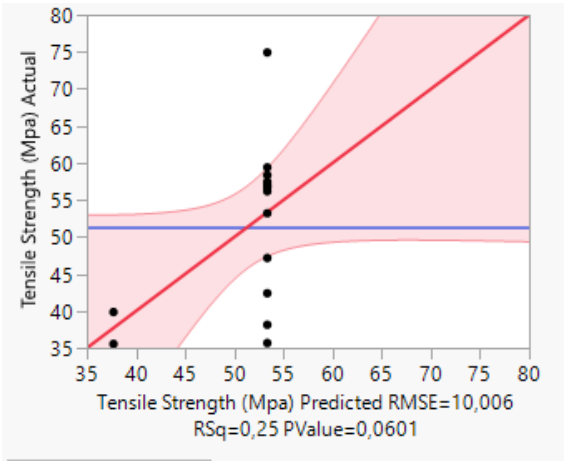
Furthermore, only one relevant term for each response was obtained: for the Density, Drying Time was the significant term; in the case of Young's Modulus it was the Time 1st stage; in the Tensile Strength case, the [EtOH] 1st stage was not significant, but was considered relevant ($p-value = 0{,}060$), being fitted into a regression model to better understand this response; and finally, for Humidity the significant term was again the Drying Time.



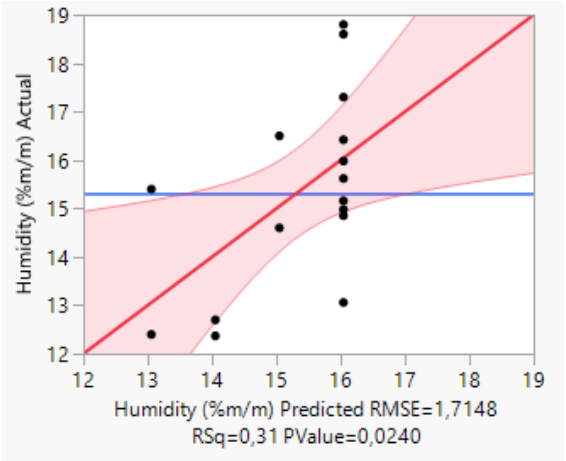
(a) Actual by predicted plot for the Density response.



(b) Actual by predicted plot for the Young's Modulus.



(c) Actual by predicted plot for the Tensile Strength.



(d) Actual by predicted plot for the Humidity.

Figure 5.1: Actual by predicted plots of the responses for the 1st stage alone with the supercritical drying step.

In Figure 5.1 it was displayed that the portion of predictors related to the 1st stage produced models with low R^2 and Adj. R^2. Moreover, in Figure 5.1c the RMSE was high since the predictor was not significant.

5.2.2 2nd Stage

As described above, the 2nd stage of the solvent exchange step was also analysed separately, along with the subsequent supercritical drying. Therefore, stepwise regression with forward selection for variable selection and OLS parameter estimation were performed.

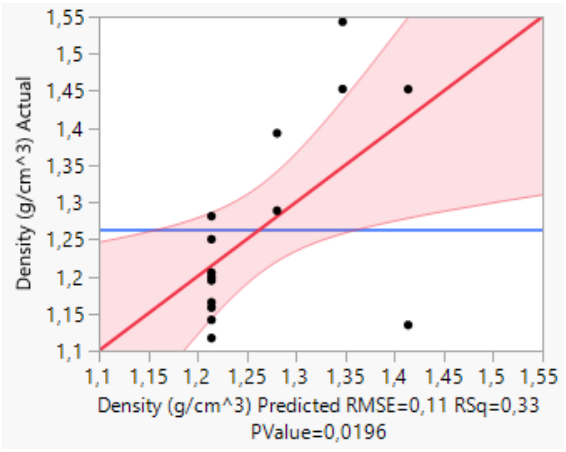
Table 5.3: Best model obtained for each response using forward stepwise selection with OLS parameter estimation, when only the 2nd stage of the solvent exchange and the subsequent supercritical drying were considered.

Response	Significant Terms	Estimates	*p-value*	R^2	Adj. R^2	RMSE
Density	Drying Time	0,066	0,020	0,33	0,28	0,11
Young's Modulus	Pressure 2nd Stage	-0,057	0,003	0,51	0,47	2,05
Tensile Strength	Pressure 2nd Stage	-0,175	0,029	0,32	0,26	9,53
Humidity	Time 2nd Stage	0,026	0,040	0,51	0,43	1,50
	Drying Time	-1,742	0,003			

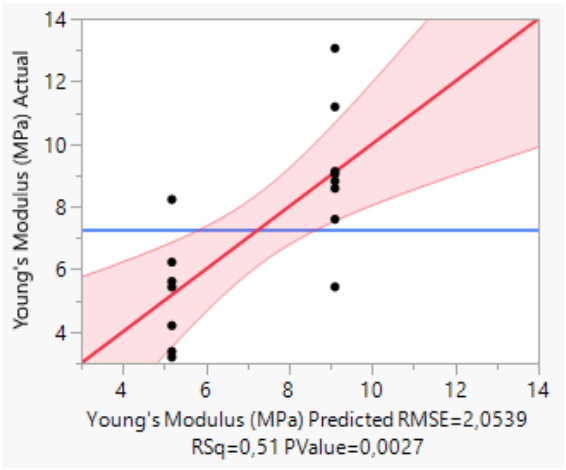
From Table 5.3, the significant terms that provided the best model fits for the responses are displayed as well as their R^2, adjusted R^2 and RMSE.

At this stage, the best model fits were obtained with the terms: for Density, Drying Time was once again significant; for both Young's Modulus and Tensile Strength it was the Pressure 2nd Stage; and for Humidity, there were two significant main effects, the Time 2nd Stage and Drying Time.

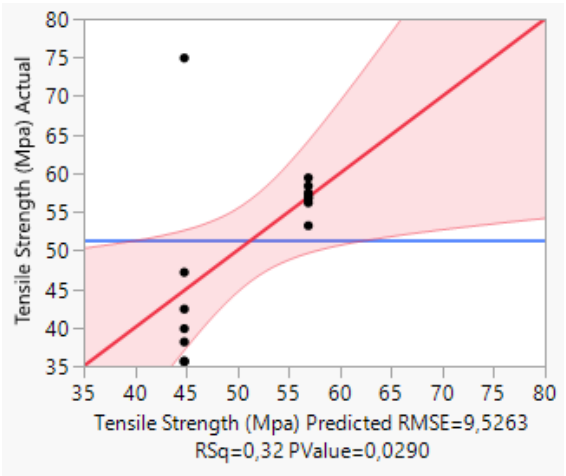
From Figure 5.2 and the values of R^2 and Adj. R^2, it was clear that both Humidity and Young's Modulus were fitted good models. The Density response was fitted the same model from the 1st Stage. Moreover, the Tensile Strength had a potential outlier, motivating further residual graphical analysis and Cook's D represented at Appendices. Nevertheless, after debating with experimenters, this value was considered a legitimate part of the data and in cases such as this one, the researcher with its training and experience in the matter has the final say [92]. Therefore, the observation was not removed from the dataset for further analysis.



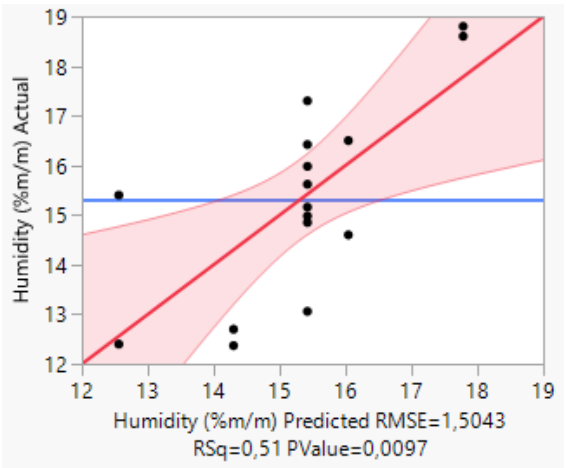
(a) Actual by predicted plot for Density.



(b) Actual by predicted plot for Young's Modulus.



(c) Actual by predicted plot for Tensile Strength.



(d) Actual by predicted plot for Humidity.

Figure 5.2: Actual by predicted plots of the responses for the 2nd stage alone with the supercritical drying step.

5.2.3 3rd Stage

Similarly to the two approaches above, when it came to the last considered stage of the solvent exchange step (3rd stage), the analysis was also performed individually, followed by the supercritical drying step.

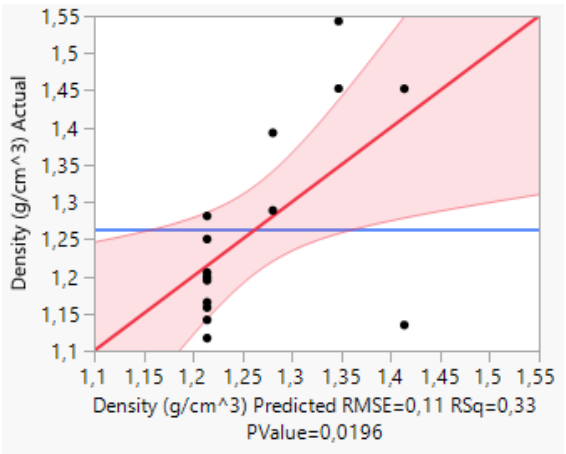
Afterwards, for data analysis, stepwise regression with forward selection was implemented for variable selection and OLS for parameter estimation.

Table 5.4: Best model obtained for each response using forward stepwise selection with OLS parameter estimation, when only the 3rd stage of the solvent exchange and the subsequent supercritical drying were considered.
* relevant term

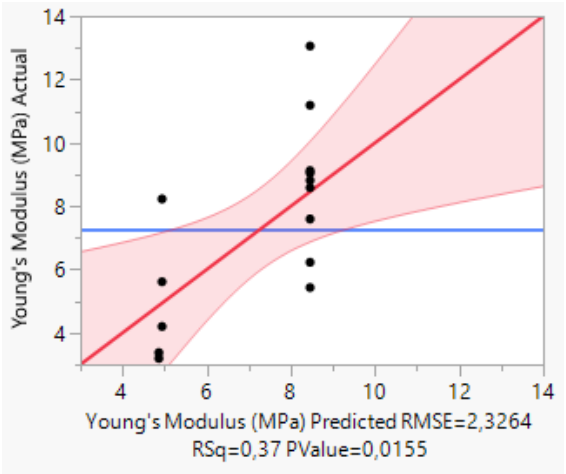
Response	Significant Terms	Estimates	*p-value*	R^2	Adj. R^2	RMSE
Density	Drying Time	0,066	0,020	0,33	0,28	0,11
Young's Modulus	Time 3rd Stage	0,003	0,015	0,37	0,33	2,33
Tensile Strength	[EtOH] 3rd Stage*	0,522	0,060	0,25	0,19	10,01
Humidity	Drying Time	-0,996	0,024	0,31	0,26	1,71

Table 5.4 sums up the performed analysis, with the relevant parameters estimates, as well as the its p-values. It was also described the R^2, adjusted R^2 and RMSE values for the best models fitted.

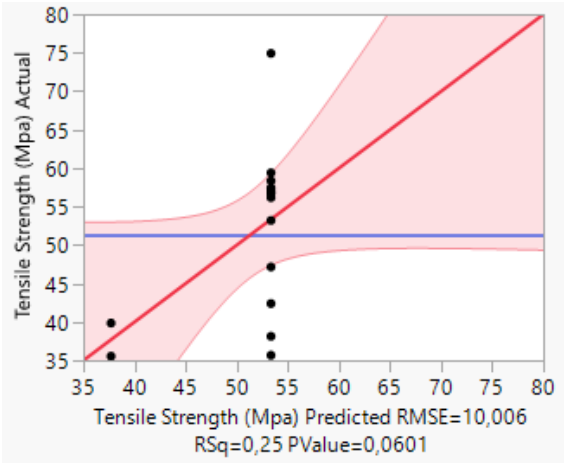
Once again, the models for the responses were obtained with only one significant term: for Density, it was the Drying Time; for Young's Modulus it was the Time 3rd stage; [EtOH] 3rd stage was included in the model for Tensile Strength, but was again not significant; finally, for Humidity, it was the Drying Time.



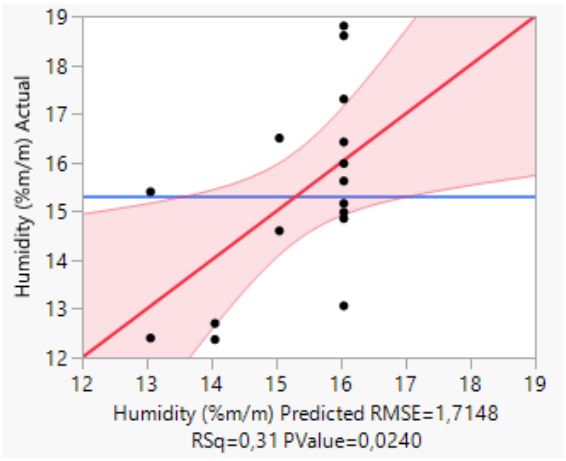
(a) Actual by predicted plot for Density.



(b) Actual by predicted plot for Young's Modulus.



(c) Actual by predicted plot for Tensile Strength.



(d) Actual by predicted plot for Humidity.

Figure 5.3: Actual by predicted plots of the responses for the 3rd stage alone with the supercritical drying step.

From Figure 5.3, it can be seen that the models for Density, Young's Modulus and Humidity are models with R^2 in the range of the [0,30 - 0,40]. Oppositely, the Tensile Strength model has a high RMSE once again, due to the fact that [EtOH] 3rd stage is not significant.

Final remarks for the models obtained in the stage-wise approach:

- For the Density in specific, the best models were the same for every individual stage. This was mainly due to the fact that Drying Time was the only significant parameter and it is independent of the solvent exchange step, since it is a CPP of the supercritical drying step. However, as Figures 5.1a, 5.2a and 5.3a demonstrate, there is also a potential outlier. Once again, after debating with the experimenters, it was concluded that the particular value of Density was a legitimate part of the data and, based on their intuition and process understanding, the decision was to maintain it [92].

- For the response models in general, as each stage-wise model for the responses was produced with different portions of the dataset, it was not possible to compare the models for the same responses across individual stages, regarding their R^2, Adj. R^2 or RMSE.

5.2.4 3 Stages Combined

For both comparison and generalization purposes, the combination of the 3 mentioned stages for the solvent exchange was analysed along with the supercritical drying step. This comprised the global approach for model building.

It is important to note that since the temperature at each stage was not further considered for analysis purposes, the average temperature of the global approach was also not considered for model building.

Similarly to the stage-wise analysis, the model building for the 3 stages combined, plus the supercritical drying step consisted of variable screening by stepwise regression with forward selection and OLS parameter estimation.

Table 5.5: Best models obtained for the global approach using forward stepwise selection with OLS parameter estimation for the global approach.

Response	Significant Terms	Estimates	*p-value*	R^2	Adj. R^2	RMSE
Density	Drying Time	0,0665	0,0196	0,33	0,28	0,11
Young's Modulus	Average Pressure	-0,0665	0,0034	0,50	0,46	2,09
Tensile Strength	Average Pressure	0,3462	0,0007	0,94	0,92	3,09
	Average Concentration	1,4530	<0,0001			
	Total Time	0,0271	<0,0001			
	Average Pressure*Total Time	-0,0012	<0,0001			
Humidity	Drying Time	-0,9955	0,0240	0,31	0,26	1,71

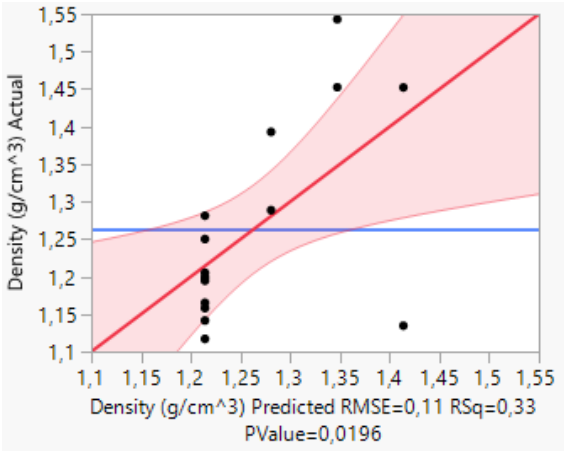
Table 5.5 shows the significant parameters estimates and p-values, as well as the models' R^2, adjusted R^2 and RMSE.

Although the results for the model fitting measures could not be directly compared between the stage-wise approach and the global one, since the set of predictors is different, the significant parameters for each model were compared.

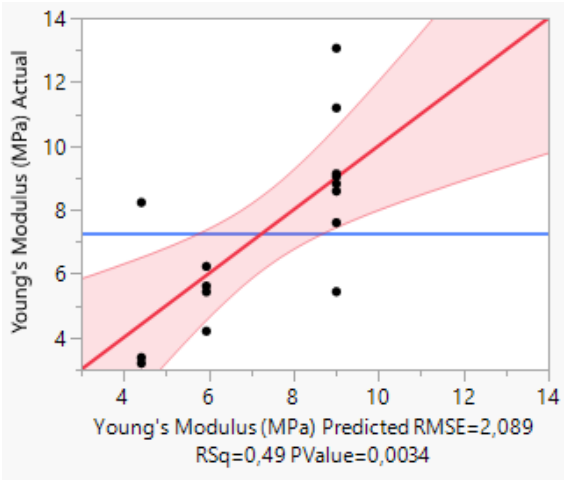
The results presented provided important generalizations: similarly to the stage-wise approach, the global models obtained for Density and Humidity were produced with the Drying Time parameter; for the Young's Modulus, the Average Pressure was the only significant term and, even though in the 1st and 3rd stages it was the Time of each stage the significant parameter, in 2nd stage, the Pressure 2nd stage produced a fairly good model; besides, for the Tensile Strength there were three significant main effects, which were the Average Pressure, Average Concentration, Total Time and one two-factor interaction between the Average Pressure and the Total Time. For the last, the pressure and concentration were significant factors in the stage-wise models, which supported that in order for Total Time to be significant,

the interaction between Average Pressure and Total Time was entered before its precedent factors.

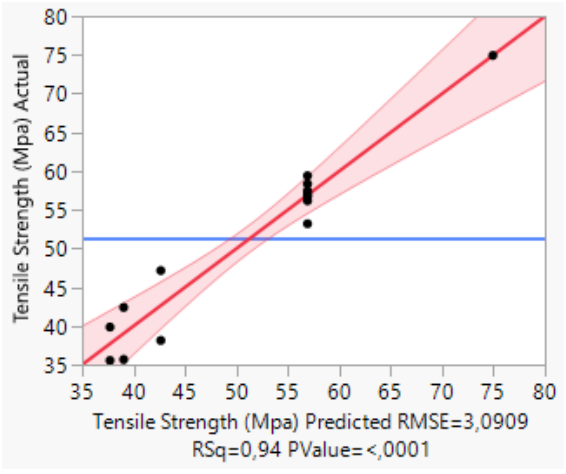
All in all, it can be verified that the global significant effects for all the responses are consistent with the significant parameters for the singular stages of the solvent exchange step, along with the supercritical drying.



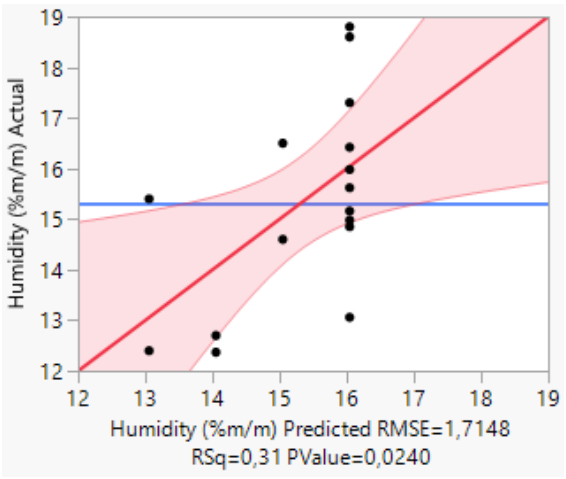
(a) Actual by predicted plot for Density.



(b) Actual by predicted plot for Young's Modulus.



(c) Actual by predicted plot for Tensile Strength.



(d) Actual by predicted plot for Humidity.

Figure 5.4: Actual by predicted plots of the responses for the 3 stages combined, plus the supercritical drying step.

The Actual by predicted plots in figure 5.4 showed a similar fit for the Density compared to the stage-wise approach, which makes sense since the significant parameter for the model is part of the supercritical drying step and therefore doesn't vary with the solvent exchange stage. Besides, for Young's Modulus the fit had a solid R^2

value, whereas for the Tensile Strength, it was fitted a model with both very high R^2 and Adj. R^2, indicating the advantages of the global approach. Finally, the Humidity produced a model similar to the one of the 1st and 3rd stages of solvent exchange.

5.3 Multiple Response Optimization

Although the significant parameters obtained for the global approach were similar for the stage-wise one, it remained to be seen whether the values of the parameters also fitted accordingly.

It was with MOO through desirability functions that the multiple response optimization was achieved. In order to do it, the limits in Table 5.6 were established.

Table 5.6: Established limits for the outputs of the dataset.

Response	Lower Bound	Upper Bound
Density	1,2	1,4
Young's Modulus	4	9,5
Tensile Strength	35	80
Humidity	13	16,5

Therefore, the values for the parameters obtained for the best models were described, as well as the overall desirability for further evaluation on whether the global approach could be in fact a generalization of the stage-wise approach.

1st Stage

Table 5.7: Best values for each significant factor at the 1st stage, so that the overall desirability function's value is maximum.

Factor	Value	Overall Desirability (D)
Drying Time (h)	3,02	
Time 1st Stage (min)	120	0,51
[EtOH] 1st Stage (%)	40	

2nd Stage

Table 5.8: Best values for each significant factor at the 2nd stage, so that the overall desirability function's value is maximum.

Factor	Value	Overall Desirability (D)
Drying Time (h)	3,63	
Time 2nd Stage (min)	30	0,59
Pressure 2nd Stage (bar)	1	

3rd Stage

Table 5.9: Best values for each significant factor at the 3rd stage, so that the overall desirability function's value is maximum.

Factor	Value	Overall Desirability (D)
Drying Time (h)	3,02	
Time 3rd Stage (min)	1440	0,51
[EtOH] 3rd Stage (%)	100	

3 Stages Combined

Table 5.10: Best values for each significant factor at the stage-combined approach, so that the overall desirability function's value is maximum.

Factor	Value	Overall Desirability (D)
Drying Time (h)	3,02	
Average Pressure (bar)	1	0,54
Average Concentration (%)	70	
Total Time (min)	1680	

As it can be seen, from the Table 5.10 in comparison to the Tables 5.7 to 5.9, it was concluded that for:

- Drying Time: the value for the global approach is between the interval of values obtained by the stage-wise approach.
- Average Pressure: the value for the global approach is the exact same seen for the 2nd stage, as it was the only situation where the pressure was a significant term modeling a response.
- Average Concentration: the value for the global approach is the average of the values obtained for the stage-wise approach.

- Total Time: the value for the global approach exceeds the sum of the time of the solvent exchange step obtained for each of the individual stages. However, it corresponds to the situation where the experimental runs only had 3 stages of solvent exchange.

This last conclusion about the Total time only highlighted the fact that it was only possible to come to conclusions regarding the influence of the first 3 stages of the solvent exchange step, plus the supercritical drying one, even though there were 4 and 5 stages at times.

Furthermore, Figure 5.5 represents the individual desirability functions (d_i) for each response in the last column of the figure. These were set according to the specifications of the aerogel applications and the values were represented at Table 5.6. Thus, it can be seen that Density and Humidity were set to be minimized as opposed to Young's Modulus and Tensile Strength which were set to be maximized, by looking at the inverted slopes of the functions. Consequently, the overall desirability function, which value is in the last row of the figure, was the result of the tradeoff between the responses' specifications mentioned above.

Moreover, this optimization was developed under the assumption that all the responses were equally important and thereby equal weights were attributed. Hence, if the optimization goal was a specific one, such as improving a certain aerogel response or minimizing for instance a certain parameter, the adjustments in the weights of the responses could be arranged so that the desired specifications would be met.

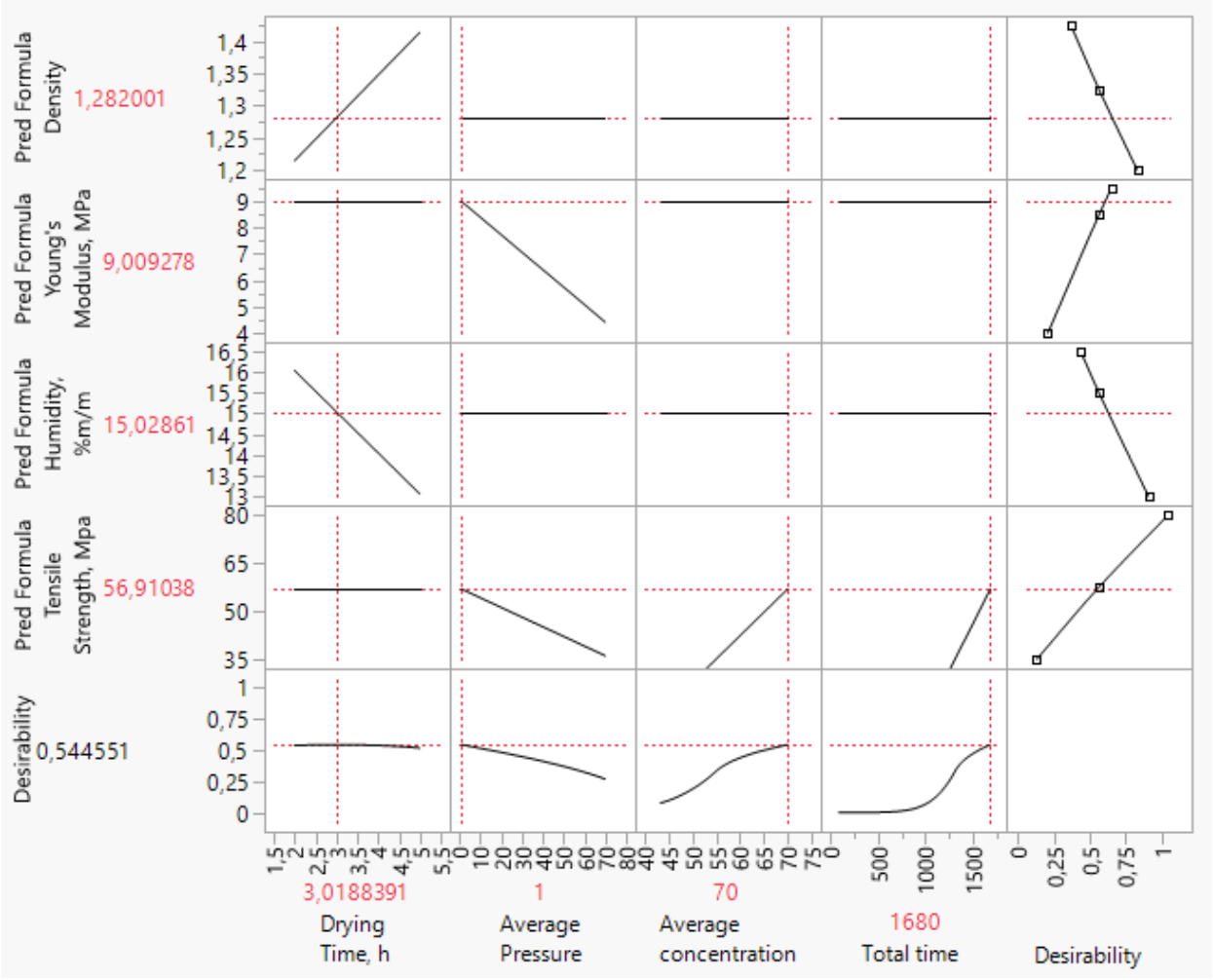


Figure 5.5: Representation of the responses individual desirability functions and the parameters that optimize them simultaneously. Subsquently, the overall desirability function is also displayed.

6

Conclusion

The main goal of this work was to optimize an aerogel formation process, specifically its solvent exchange and supercritical drying steps, by building models that could accurately find the significant parameters for the process and, subsequently the values that would optimize it.

By looking at the dataset available for this process, which was not collected according to a DoE methodology, the main goal was harder to achieve. Some efforts were made to understand how the data could be arranged so that the different steps of the process were clearly defined and, for that reason a process flow chart was the first step to properly organize all the important factors and responses.

Afterwards, it was decided to opt for a variable selection procedure to make the most out of our data. Besides, as the dataset was not collected according to an optimal design, a simple way of producing models to represent the process was required.

Since the process was hard to represent, a variable selection method was required and hence was the stepwise method for variable selection with forward regression employed. The fact that this method also had the property of enforcing heredity was fundamental, since interactions between factors are of utmost importance in processes like the one studied.

However, to perform the analysis described above, some stages of the solvent exchange step had to be removed from the analysis, since they did not have enough observations to be studied. Therefore, the 4^{th} and 5^{th} stages of the solvent exchange were not further analyzed for the purpose of this research. Moreover, not only was the dataset a non-optimal one, but also had limited observations, with only 16 experimental runs. All of these specifications led to a decision of first testing a stage-wise approach where the dataset would be divided by stages and each stage parameters would be tested along with the supercritical drying step parameters to evaluate the models produced. Then, a global approach was developed combining the 3 stages

of the solvent exchange plus the supercritical drying step for a combined analysis.

The aforementioned analysis of both the stage-wise approach and the global one led to the conclusion that the same parameters that were relevant to the responses' models for the individual stages, were also relevant for the same responses' models for the global approach. Consequently, by proving that the generalization of the 3 stages produced models with the same relevant parameters, it was easier to represent the process being studied and obtaining the optimal settings for the parameters that optimized the given responses.

Nevertheless, out of the 7 responses studied, only 4 were able to produce significant models and these were the Density, Young's Modulus and Humidity. And even though such an important property as the Porosity was not possible to study, it was proved that if in further experiments the Surface Area produces significant models, it might also affect the Porosity values, as these two responses are positively correlated.

Finally, a MOO was employed to simultaneously optimize the 4 responses aforementioned, by assigning them desirability functions which were combined in an overall desirability function. This overall desirability function took into consideration the individual optimization objectives for each response and promoted the necessary tradeoffs to optimize all of them at the same time.

Subsequently, not only was the main goal achieved, but also was the ground set for a more optimized methodology to develop this aerogel formation experiment. In order to promote the development of such a methodology, an introduction of the QbD methodology to this process was employed and it is expected to be further developed in future studies.

6.1 Future Work

Bearing in mind the limitations of this project, there is much that can be improved so that this process becomes an optimal one. However, it is expected that these analysis will promote validation experiments, so that more data can be gathered and the aerogel efficiently produced.

It can only be beneficial to implement a QbD approach that promotes building quality into the product. Therefore, carefully establishing the process CQAs, CPPs and CMAs is already a good start. Additionally, risk assessment diagrams might be something to explore and, most importantly the need to employ a DoE methodology to optimally build designs, so that the best combination of factors according to the

number of available runs is explored. This is also a way of saving resources and of allocating them to the desired process in a rigorous manner. This approach perfectly fits the Juran's Trilogy: Quality planning; Quality control; Quality improvement [93].

Nonetheless, it was not possible to develop the prediction of the the mass of water in the alcogel, using the RI measures as a predictor. Future studies will need to address this issue, as it can be an effective way of assessing the alcogel formation process.

Moreover, a connection between the remaining steps of the aerogel formation process, namely the gelation and sterilization can be something interesting to do in order to have a global view of the process and to be able to incorporate the results of this study in the overall process.

Bibliography

[1] U. Meyer, T. Meyer, J. Handschel, and H. P. Wiesmann, *Fundamentals of tissue engineering and regenerative medicine*. Springer, 2009.

[2] H. Jahangirian, E. G. Lemraski, R. Rafiee-Moghaddam, and T. J. Webster, "A review of using green chemistry methods for biomaterials in tissue engineering," *International journal of nanomedicine*, vol. 13, p. 5953, 2018.

[3] P. T. Anastas and J. C. Warner, "Principles of green chemistry," *Green chemistry: Theory and practice*, pp. 29–56, 1998.

[4] O. V. Kharissova, B. I. Kharisov, C. M. Oliva González, Y. P. Méndez, and I. López, "Greener synthesis of chemical compounds and materials," *Royal Society open science*, vol. 6, no. 11, p. 191378, 2019.

[5] D. Martinez-Marquez, A. Mirnajafizadeh, C. P. Carty, and R. A. Stewart, "Application of quality by design for 3d printed bone prostheses and scaffolds," *PloS one*, vol. 13, no. 4, p. e0195291, 2018.

[6] J. M. Juran *et al.*, *Juran on quality by design: the new steps for planning quality into goods and services*. Simon and Schuster, 1992.

[7] H. B. Grangeia, C. Silva, S. P. Simões, and M. S. Reis, "Quality by design in pharmaceutical manufacturing: A systematic review of current status, challenges and future perspectives," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 147, pp. 19–37, 2020.

[8] S. N. Politis, P. Colombo, G. Colombo, and D. M. Rekkas, "Design of experiments (doe) in pharmaceutical development," *Drug development and industrial pharmacy*, vol. 43, no. 6, pp. 889–901, 2017.

[9] U. Food, D. Administration, *et al.*, "Pharmaceutical cgmps for the 21st century—a risk-based approach," *Final report. Rockville, MD*, 2004.

[10] I. Harmonisation, "Pharmaceutical development q8 (r2)," in *International Conference on Harmonisation*, 2009.

[11] A. S. Rathore, "Roadmap for implementation of quality by design (qbd) for biotechnology products," *Trends in biotechnology*, vol. 27, no. 9, pp. 546–553, 2009.

[12] L. M. Collins, J. J. Dziak, and R. Li, "Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs.," *Psychological methods*, vol. 14, no. 3, p. 202, 2009.

[13] U. Food, D. Administration, *et al.*, "Guidance for industry and review staff target product profile—a strategic development process tool," *FDA website*, 2007.

[14] X. Y. Lawrence, G. Amidon, M. A. Khan, S. W. Hoag, J. Polli, G. Raju, and J. Woodcock, "Understanding pharmaceutical quality by design," *The AAPS journal*, vol. 16, no. 4, pp. 771–783, 2014.

[15] D. C. Montgomery, *Design and analysis of experiments*. John wiley & sons, 2017.

[16] C. J. Wu and M. S. Hamada, *Experiments: planning, analysis, and optimization*, vol. 552. John Wiley & Sons, 2009.

[17] X. Li, N. Sudarsanam, and D. D. Frey, "Regularities in data from factorial experiments," *Complexity*, vol. 11, no. 5, pp. 32–45, 2006.

[18] R. Leardi, "Experimental design in chemistry: A tutorial," *Analytica chimica acta*, vol. 652, no. 1-2, pp. 161–172, 2009.

[19] S. A. Weissman and N. G. Anderson, "Design of experiments (doe) and process optimization. a review of recent publications," *Organic Process Research & Development*, vol. 19, no. 11, pp. 1605–1633, 2015.

[20] M. J. Anderson and P. J. Whitcomb, "Design of experiments," *Kirk-Othmer Encyclopedia of Chemical Technology*, pp. 1–22, 2000.

[21] J. Gabrielsson, N.-O. Lindberg, and T. Lundstedt, "Multivariate methods in pharmaceutical applications," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 16, no. 3, pp. 141–160, 2002.

[22] J. A. Wass, "First steps in experimental design-the screening experiment," *Journal of validation technology*, vol. 16, no. 2, p. 49, 2010.

[23] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, Å. Nyström, J. Pettersen, and R. Bergman, "Experimental design and optimization," *Chemometrics and intelligent laboratory systems*, vol. 42, no. 1-2, pp. 3–40, 1998.

[24] G. E. Box, W. G. Hunter, J. S. Hunter, *et al.*, *Statistics for experimenters*, vol. 664. John Wiley and sons New York, 1978.

[25] M. Uy and J. K. Telford, "Optimization by design of experiment techniques," in *2009 IEEE Aerospace conference*, pp. 1–10, IEEE, 2009.

[26] S. Sonoda, Y. Takahashi, K. Kawagishi, N. Nishida, and S. Wakao, "Application of stepwise multiple regression to design optimization of electric machine," *IEEE transactions on magnetics*, vol. 43, no. 4, pp. 1609–1612, 2007.

[27] B. Abraham, H. Chipman, and K. Vijayan, "Some risks in the construction and analysis of supersaturated designs," *Technometrics*, vol. 41, no. 2, pp. 135–141, 1999.

[28] P. H. Westfall, S. S. Young, and D. K. Lin, "Forward selection error control in the analysis of supersaturated designs," *Statistica Sinica*, pp. 101–117, 1998.

[29] X. Lu and X. Wu, "A strategy of searching active factors in supersaturated screening experiments," *Journal of Quality Technology*, vol. 36, no. 4, pp. 392–399, 2004.

[30] F. Bunea, Y. She, H. Ombao, A. Gongvatana, K. Devlin, and R. Cohen, "Penalized least squares regression methods and applications to neuroimaging," *NeuroImage*, vol. 55, no. 4, pp. 1519–1527, 2011.

[31] T. Hesterberg, N. H. Choi, L. Meier, C. Fraley, *et al.*, "Least angle and 1 penalized regression: A review," *Statistics Surveys*, vol. 2, pp. 61–93, 2008.

[32] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[33] S. Mueller-Using, T. Feldt, F. S. Sarfo, and K. A. Eberhardt, "Factors associated with performing tuberculosis screening of hiv-positive patients in ghana: Lasso-based predictor selection in a large public health data set," *BMC Public Health*, vol. 16, no. 1, p. 563, 2016.

[34] N. H. Choi, W. Li, and J. Zhu, "Variable selection with the strong heredity constraint and its oracle property," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 354–364, 2010.

[35] M. G. Usai, M. E. Goddard, and B. J. Hayes, "Lasso with cross-validation for genomic selection," *Genetics research*, vol. 91, no. 6, pp. 427–436, 2009.

[36] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, *et al.*, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[37] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[38] S. Cho, H. Kim, S. Oh, K. Kim, and T. Park, "Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis," in *BMC proceedings*, vol. 3, pp. 1–6, BioMed Central, 2009.

[39] M. Mehrian and L. Geris, "Optimizing neotissue growth inside perfusion bioreactors with respect to culture and labor cost: a multi-objective optimization study using evolutionary algorithms," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 23, no. 7, pp. 285–294, 2020.

[40] B. Gurumurthy, J. A. Griggs, and A. V. Janorkar, "Optimization of collagen-elastin-like polypeptide composite tissue engineering scaffolds using response surface methodology," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 84, pp. 116–125, 2018.

[41] A. L. Bukzem, R. Signini, D. M. dos Santos, L. M. Lião, and D. P. R. Ascheri, "Optimization of carboxymethyl chitosan synthesis using response surface methodology and desirability function," *International journal of biological macromolecules*, vol. 85, pp. 615–624, 2016.

[42] S. S. Kistler, "Coherent expanded aerogels and jellies," *Nature*, vol. 127, no. 3211, pp. 741–741, 1931.

[43] W. Yin and D. A. Rubenstein, "Biomedical applications of aerogels," in *Aerogels Handbook*, pp. 683–694, Springer, 2011.

[44] H. Maleki, L. Durães, C. A. García-González, P. del Gaudio, A. Portugal, and M. Mahmoudi, "Synthesis and biomedical applications of aerogels: Possibilities and challenges," *Advances in colloid and interface science*, vol. 236, pp. 1–27, 2016.

[45] L. Zuo, Y. Zhang, L. Zhang, Y.-E. Miao, W. Fan, and T. Liu, "Polymer/carbon-based hybrid aerogels: preparation, properties and applications," *Materials*, vol. 8, no. 10, pp. 6806–6848, 2015.

[46] R. Pekala, "Organic aerogels from the polycondensation of resorcinol with formaldehyde," *Journal of materials science*, vol. 24, no. 9, pp. 3221–3227, 1989.

[47] K. Guo, H. Song, X. Chen, X. Du, and L. Zhong, "Graphene oxide as an anti-shrinkage additive for resorcinol–formaldehyde composite aerogels," *Physical Chemistry Chemical Physics*, vol. 16, no. 23, pp. 11603–11608, 2014.

[48] W. Wan, R. Zhang, W. Li, H. Liu, Y. Lin, L. Li, and Y. Zhou, "Graphene–carbon nanotube aerogel as an ultra-light, compressible and recyclable highly efficient absorbent for oil and dyes," *Environmental Science: Nano*, vol. 3, no. 1, pp. 107–113, 2016.

[49] M. A. Worsley, J. D. Kuntz, O. Cervantes, T. Y.-J. Han, A. E. Gash, J. H. Satcher Jr, and T. F. Baumann, "Route to high surface area tio 2/c and ticn/c composites," *Journal of Materials Chemistry*, vol. 19, no. 38, pp. 7146–7150, 2009.

[50] M. A. Worsley, J. D. Kuntz, P. J. Pauzauskie, O. Cervantes, J. M. Zaug, A. E. Gash, J. H. Satcher, and T. F. Baumann, "High surface area carbon nanotube-supported titanium carbonitride aerogels," *Journal of Materials Chemistry*, vol. 19, no. 31, pp. 5503–5506, 2009.

[51] J. P. Randall, M. A. B. Meador, and S. C. Jana, "Tailoring mechanical properties of aerogels for aerospace applications," *ACS applied materials & interfaces*, vol. 3, no. 3, pp. 613–626, 2011.

[52] F. Quignard, R. Valentin, and F. Di Renzo, "Aerogel materials from marine polysaccharides," *New journal of chemistry*, vol. 32, no. 8, pp. 1300–1310, 2008.

[53] T. Mehling, I. Smirnova, U. Guenther, and R. Neubert, "Polysaccharide-based aerogels as drug carriers," *Journal of Non-Crystalline Solids*, vol. 355, no. 50-51, pp. 2472–2479, 2009.

[54] C. García-González, M. Alnaief, and I. Smirnova, "Polysaccharide-based aerogels—promising biodegradable carriers for drug delivery systems," *Carbohydrate Polymers*, vol. 86, no. 4, pp. 1425–1438, 2011.

[55] C. A. Garcia-Gonzalez, A. Concheiro, and C. Alvarez-Lorenzo, "Processing of materials for regenerative medicine using supercritical fluid technology," *Bioconjugate chemistry*, vol. 26, no. 7, pp. 1159–1171, 2015.

[56] W. Yin, H. Lu, N. Leventis, and D. A. Rubenstein, "Characterization of the

biocompatibility and mechanical properties of polyurea organic aerogels with the vascular system: Potential as a blood implantable material," *International Journal of Polymeric Materials*, vol. 62, no. 2, pp. 109–118, 2013.

[57] K. S. Mikkonen, K. Parikka, A. Ghafar, and M. Tenkanen, "Prospects of polysaccharide aerogels as modern advanced food materials," *Trends in food science & technology*, vol. 34, no. 2, pp. 124–136, 2013.

[58] K. Y. Lee and D. J. Mooney, "Alginate: properties and biomedical applications," *Progress in polymer science*, vol. 37, no. 1, pp. 106–126, 2012.

[59] H. A. Awad, M. Q. Wickham, H. A. Leddy, J. M. Gimble, and F. Guilak, "Chondrogenic differentiation of adipose-derived adult stem cells in agarose, alginate, and gelatin scaffolds," *Biomaterials*, vol. 25, no. 16, pp. 3211–3222, 2004.

[60] B. S. Agostinho, "Secagem supercrítica de aerogéis compostos por polímeros naturais para aplicações biomédicas," Master's thesis, Universidade de Coimbra, 2019.

[61] L. Baldino, S. Cardea, and E. Reverchon, "Natural aerogels production by supercritical gel drying," *Chemical Engineering Transactions*, vol. 43, pp. 739–744, 2015.

[62] S. Raman, P. Gurikov, and I. Smirnova, "Hybrid alginate based aerogels by carbon dioxide induced gelation: Novel technique for multiple applications," *The Journal of Supercritical Fluids*, vol. 106, pp. 23–33, 2015.

[63] C. García-González, M. Camino-Rey, M. Alnaief, C. Zetzl, and I. Smirnova, "Supercritical drying of aerogels using co2: Effect of extraction time on the end material textural properties," *The Journal of Supercritical Fluids*, vol. 66, pp. 297–306, 2012.

[64] P. A. Webb, "Volume and density determinations for particle technologists," *Micromeritics Instrument Corp*, vol. 2, no. 16, p. 01, 2001.

[65] K. S. Sing, "Characterization of porous solids: An introductory survey," in *Studies in Surface Science and Catalysis*, vol. 62, pp. 1–9, Elsevier, 1991.

[66] S. Yang, K.-F. Leong, Z. Du, and C.-K. Chua, "The design of scaffolds for use in tissue engineering. part i. traditional factors," *Tissue engineering*, vol. 7, no. 6, pp. 679–689, 2001.

[67] F. H. Silver and D. L. Christiansen, "Mechanical properties of tissues," in *Biomaterials Science and Biocompatibility*, pp. 187–212, Springer, 1999.

[68] M. Martins, A. A. Barros, S. Quraishi, P. Gurikov, S. Raman, I. Smirnova, A. R. C. Duarte, and R. L. Reis, "Preparation of macroporous alginate-based aerogels for biomedical applications," *The Journal of Supercritical Fluids*, vol. 106, pp. 152–159, 2015.

[69] G. Renders, L. Mulder, L. Van Ruijven, and T. Van Eijden, "Porosity of human mandibular condylar bone," *Journal of Anatomy*, vol. 210, no. 3, pp. 239–248, 2007.

[70] P. Patil, "What is exploratory data analysis?." `https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15`, 2018. Online; accessed 3 October 2020.

[71] T. D. V. Swinscow, M. J. Campbell, *et al.*, *Statistics at square one*. Bmj London, 2002.

[72] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi medical journal*, vol. 24, no. 3, pp. 69–71, 2012.

[73] University of the West of England - UWE Bristol, "Data analysis." `http://learntech.uwe.ac.uk/da/Default.aspx?pageid=1442`, 2020. Online; accessed 4 October 2020.

[74] A. J. Duleba and D. L. Olive, "Regression analysis and multivariate analysis," in *Seminars in reproductive endocrinology*, vol. 14, pp. 139–153, Copyright© 1996 by Thieme Medical Publishers, Inc., 1996.

[75] D. N. Gujarati and D. C. Porter, "Basic econometrics (ed.)," *Singapore: McGrew Hill Book Co*, 2003.

[76] T. Hastie, R. Tibshirani, and J. Friedman, "Linear methods for regression," in *The elements of statistical learning*, pp. 43–99, Springer, 2009.

[77] M. N. Williams, C. A. G. Grajales, and D. Kurkiewicz, "Assumptions of multiple regression: Correcting two misconceptions," *Practical Assessment, Research, and Evaluation*, vol. 18, no. 1, p. 11, 2013.

[78] D. R. Legates and G. J. McCabe Jr, "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation," *Water resources research*, vol. 35, no. 1, pp. 233–241, 1999.

[79] J. Miles, "R squared, adjusted r squared," *Wiley StatsRef: Statistics Reference Online*, 2014.

[80] C. Jacovides and H. Kontoyiannis, "Statistical procedures for the evaluation of evapotranspiration computing models," *Agricultural water management*, vol. 27, no. 3-4, pp. 365–371, 1995.

[81] A. C. Rencher, *Methods of multivariate analysis*, vol. 492. John Wiley & Sons, 2003.

[82] "JMP®15 Fitting Linear Models," 2020. SAS Institute Inc., Cary, NC.

[83] H. Bozdogan, "Model selection and akaike's information criterion (aic): The general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.

[84] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*, pp. 199–213, Springer, 1998.

[85] K. P. Burnham and D. R. Anderson, "A practical information-theoretic approach," *Model selection and multimodel inference, 2nd ed. Springer, New York*, vol. 2, 2002.

[86] D. J. Obermiller, "Multiple response optimization using jmp," *The Dow Chemical Company, Midland, MI*.

[87] N. R. Costa, J. Lourenço, and Z. L. Pereira, "Desirability function approach: a review and performance evaluation in adverse conditions," *Chemometrics and Intelligent Laboratory Systems*, vol. 107, no. 2, pp. 234–244, 2011.

[88] E. C. Harrington, "The desirability function," *Industrial quality control*, vol. 21, no. 10, pp. 494–498, 1965.

[89] G. Derringer and R. Suich, "Simultaneous optimization of several response variables," *Journal of quality technology*, vol. 12, no. 4, pp. 214–219, 1980.

[90] "JMP®15 Profilers," 2020. SAS Institute Inc., Cary, NC.

[91] "JMP®15 Basic Analysis," 2020. SAS Institute Inc., Cary, NC.

[92] J. W. Osborne and A. Overbay, "The power of outliers (and why researchers should always check for them)," *Practical Assessment, Research, and Evaluation*, vol. 9, no. 1, p. 6, 2004.

[93] J. Juran and A. B. Godfrey, "Quality handbook," *Republished McGraw-Hill*, vol. 173, no. 8, 1999.

Appendices

A

Aerogel

A.1 Aerogel Sample

Figure A.1: Picture of a sample of an aerogel, obtained by Agostinho (2019) [60].

A.2 Microscopic Section

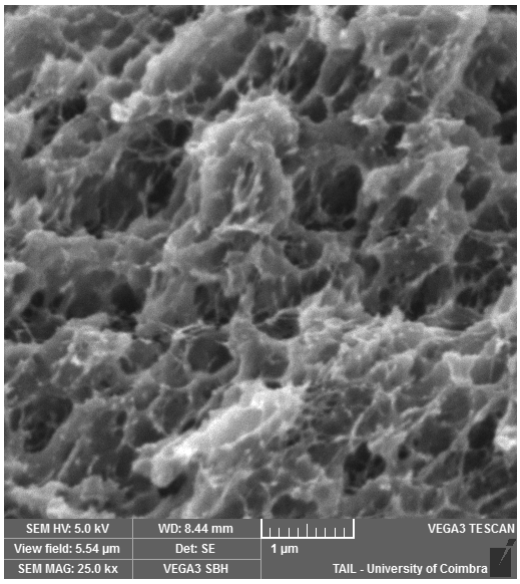


Figure A.2: Microscopic representation of the monolith (SEM), obtained by Agostinho (2019) [60].

B

Outlier Analysis

B.1 Density Model - 1st Stage; 2nd Stage; 3rd Stage; 3 Stages Combined

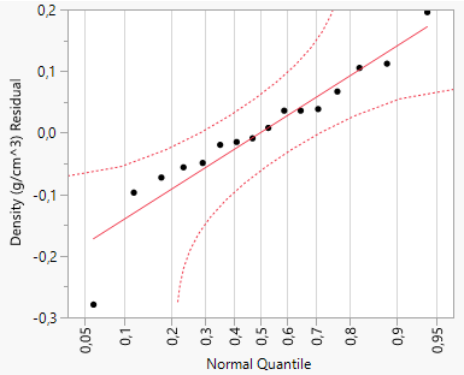


Figure B.1: Residual Normal Quantile Plot for the Density response.

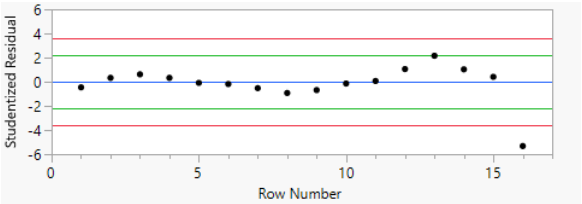


Figure B.2: Studentized residuals for the Density response.

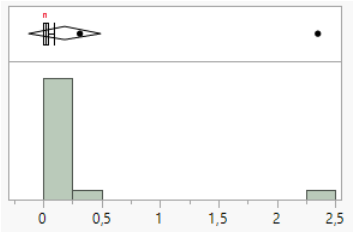


Figure B.3: Cook's Distance Influence for the Density response.

B.2 Tensile Strength Model - 2nd Stage

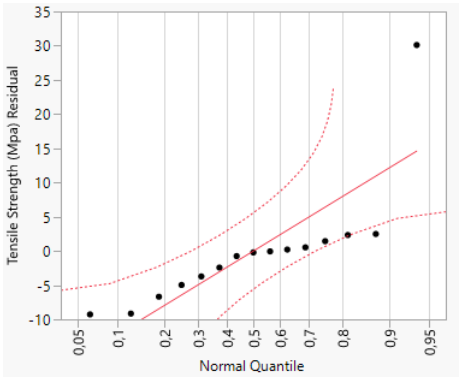


Figure B.4: Residual Normal Quantile Plot for the Tensile Strength response.

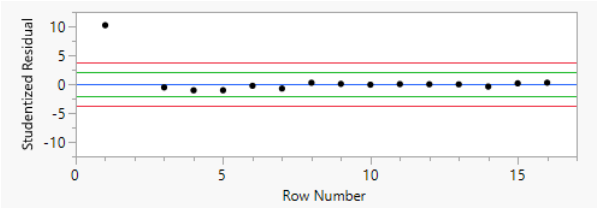


Figure B.5: Studentized residuals for the Tensile Strength response.

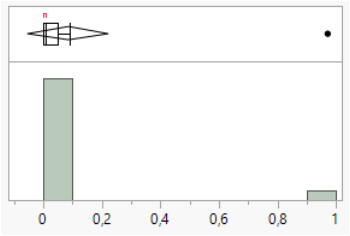


Figure B.6: Cook's Distance Influence for the Tensile Strength response.