Mixture Design: Development of a Graphical User Interface for Determining Mixture Parameters

Laíres Lima, Ana I. Pereira, Clara Vaz
Research Centre in Digitalization and Intelligent Robotics (CeDRI)
Polytechnic Institute of Bragança, Bragança, Portugal
{laires.lima, apereira, clvaz}@ipb.pt

Olga Ferreira
Mountain Research Center (CIMO)
Polytechnic Institute of Bragança, Bragança, Portugal
oferreira@ipb.pt

Abstract — **Predicting the performance of a mixture is crucial to designing experiences in product development and formulation research. In this work, an application, MDesign, is proposed to construct models in a mixture design with a practical, educational, and intuitive approach. Developed in MATLAB software, the standalone application aims to contribute to the study of mixtures through the definition of multivariate models of different orders, enabling their statistical analysis to verify the robustness of each of those models. Compared to the obtained results from other applications using data experiments published in the literature, the proposed application presents accurate results and good execution. MDesign can be considered an automatic, robust, and valuable tool to support the mixture design in an industrial context.**

Keywords – mixture designs, graphical user interface, predictive methods.

I. Introduction

Mixture experiments are applied in several areas of industry and science, such as chemical and pharmaceutical formulations, product development, and food processing. They address the influence of process variables on specific responses. Thus, the primary aim of mixture design is to reduce the number of experiments, or replications, to improve the quality of the information obtained through the results [1, 2].

Mixture designs belong to a design class of response surfaces, represented by a regular dimension simplex, in which the levels of experimental factors correspond to the proportions of the components in a blend, with values ranging between 0 and 1 and the sum of the proportions of the components is equivalent to 1 (100%). In this case, the response of interest should depend directly on the proportions of the elements, being irrelevant or negligible to the total amount of the mixture [3]. In some cases, the components of the mixture are subject to additional constraints, implying the change of the response surface from a regular simplex to an irregular-shaped polyhedron within the simplex region.

Considering the aspects presented, building a model that can represent the functionality and behavior of the system under study is fundamental for the success of a mixture project. There is a high demand for open software, due to the high cost of acquiring commercial software, commonly used for experimental design and data analysis [4]. Some open free extension packages were developed, mainly in R language: AlgDesign, OptimalDesign, QualityTools, Mixexp, DoseFinding, *inter alia* [5]. This work presents a simple, noise-robust, highly intuitive standalone application with a graphical user interface (GUI), allowing the construction of mathematical models that describe the mixture behaviour and totally developed in the software MATLAB.

The notations and theoretical background about mixture designs and other concepts inherent to the development of this application are described in Sections II to V. Section VI addresses the features and tools of the application MDesign, as well as an overview of its functioning. Sections VII and VIII present the case studies used to illustrate the application and the obtained results using MDesign. Conclusions about the developed work are provided in Section IX.

II. Experimental domain

As mentioned in the previous section, the sum of the proportions of all components must be equal to 1 (or 100%). Therefore, for any mixture of k components, the following equation can be applied [3]:

$$\sum_{i=1}^{k} x_i = 1 \quad (1)$$

where x_i represents the proportion of the *i-th* component. The proportions of the components are dependent on each other, which implies that the ratio value of the last component is always defined by the compositions of k-1 components [3].

Therefore, in the case of mixture designs, there is a reduction in the degree of freedom of the process and the dimension of the experimental domain of the mixture, as shown in Fig. 1. For example, in a four-component system, it is impossible to graphically represent the experimental design for this mixture without fixing the proportion of one of the components or by describing it in abstract k-dimensional spaces. Three components are the maximum number of components that the representation can hold (Fig.1.A). On the other hand, in the mixture design, it is possible to represent the

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

behavior by means of a tetrahedron (Fig. 1.D). Other experimental domains for different numbers of components can be represented by a linear graph for two components (Fig. 1.B) and triangular for three components (Fig. 1.C).

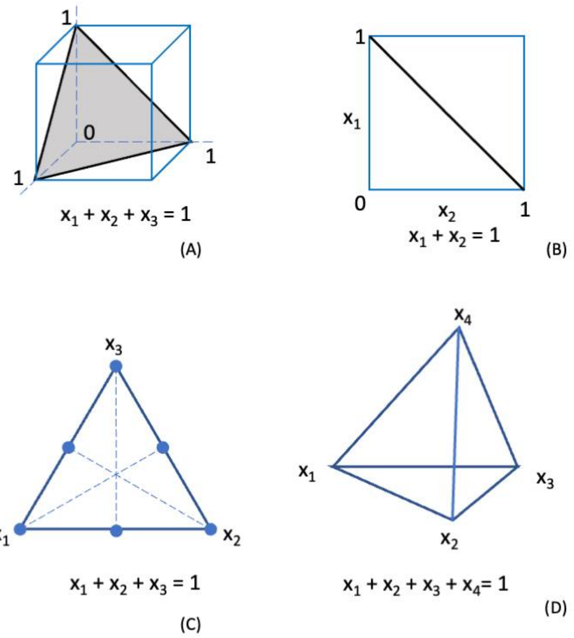

Figure 1. Graphical representations of factorial experimental design for three components (A) and mixing designs with reduced dimensionality for two (B), three (C) and four components (D) (Adapted from [1]).

For the two-component mixture in Fig. 1.B, the linear representation corresponds to the possible ratios of the binary mixture, and the extreme points represent the pure component. In Fig. 1.C, corresponding to the tertiary mixture, the vertices represent the pure components, the edges correspond to the binary mixtures, and the internal region to the tertiary mixtures; similar reasoning can be applied to Fig. 1.D (4-component mixture).

III. Types of Mixture Designs

In mixture designs, the experimental region can be constrained and, as described previously, it is not possible to change the levels of each component independently. Therefore, classical design of experiments (DOE), such as factorial designs, are not able to analyze these cases correctly. Instead, conditioned approaches to mixture design are able to overcome the limitations of these classic DOE. Selecting a design method should take into account the number of independent variables, the type of interactions, the target responses, feasibility and process costs [6]. The most used designs are lattice simplex, centroid-simplex, augmented simplex-centroid, polynomial design, optimal design, and space filling design [6].

IV. Constraints in Mixture Designs

In many cases, mixture studies considering all composition ranges can be time-consuming and unnecessary. Many engineering problems demand minimum, or maximum, proportions of a specific component to achieve an expected result. Thus, limiting the experimental domain through upper and lower constraints can be an alternative. These constraints build a new, more practical experimental space, being a fraction of the complete environment, and the original components are parameterized based on the original fraction, being called pseudocomponents [3].

In mixtures where the components have a lower limit for their proportion, the original components composition x_i, and the pseudocomponents composition x_i' can be mathematically described as:

$$x_i' = \frac{x_i - L_i}{R_L}, 0 < L_i \leq x_i \leq 1 \quad (2)$$

where L_i is the lower limit for the i component and R_L is the range of the region between lower limits [3]:

$$R_L = 1 - \sum_{i=1}^{k} L_i, \text{ where } \sum_{i=1}^{k} L_i < 1. \quad (3)$$

In cases where the upper constraints delimit the experimental region, the relation of the composition of pseudocomponents and components is given by [3]:

$$x_i' = \frac{U_i - x_i}{R_U}, 0 \leq x_i \leq U_i < 1 \quad (4)$$

where U_i is the upper limit for the i component and R_U is the range of the region between upper limits [3]:

$$R_U = \left(\sum_{i=1}^{k} U_i\right) - 1, \text{ where } \sum_{i=1}^{k} U_i > 1 \quad (5)$$

In cases where lower and upper constraints (R_L and R_U, respectively) of each component delimit the domain, intersection bands construct the space (R_i), represented by [3]:

$$R_i = U_i - L_i, \qquad i = 1,2, \ldots, k \quad (6)$$

V. Mathematical modeling and Statistical Validation

Mixing experiments consist of blending different ratios of at least two components k and verifying the influence of proportions on response variables. Modeling tools are commonly adopted to perform this type of study. For the application development, lattice simplex design applied for mixture models were adopted, in which the linear model (7) was

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

the simplest approach. Models of the second, third and fourth order are described in (8) and (9) [1, 3].

$$y = \beta_0 + \left(\sum_{i=1}^{k} \beta_i x_i \right) + \varepsilon \quad (7)$$

$$y = \beta_0 + \left(\sum_{i=1}^{k} \beta_i x_i + \sum_{\substack{i=j \\ j>1}}^{k-1} \sum_{j=2}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 \right) + \varepsilon \quad (8)$$

$$y = \beta_0 + \left(\sum_{i=1}^{k} \beta_i x_i + \sum_{\substack{i=j \\ j>1}}^{k-1} \sum_{j=2}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq i}}^{k} \beta_{ijj} x_i x_j^2 + \beta_{123} x_1 x_2 x_3 \right) + \varepsilon \quad (9)$$

where x_i represents the proportion of the *i-th* component. In the model above, y represents the response variable under study, β corresponds to the coefficients and ε represents the associated random error.

In a matrix form, the models [1] can be expressed as follows:

$$y = X\beta + \varepsilon \quad (10)$$

where $y = (y_1, y_2,..., y_n)$ is the column vector of n observed responses, X is the $(n \times p)$ matrix with p model elements, β is the column vector corresponding to the model coefficients, and $\varepsilon = (\varepsilon_1, \varepsilon_2,..., \varepsilon_n)$ the column of random errors, normally distributed, and common variance.

Assuming that the matrix of the errors (ε) has mean equals to zero and has a variance matrix given by $\sigma^2 I_n$, being I_n the identity matrix with dimension n, the matrix of the estimates of the coefficients can be defined as:

$$\mathrm{b} = (X^T X)^{-1} X^T y \quad (11)$$

where X^T represents the transpose matrix of X. The variance property of the least-squares coefficients (b) is expressed in terms of diagonal elements of the matrix $(X^T X)^{-1}$. Once multiplied by σ^2, the covariance matrix of b is obtained (12).

$$cov(b) = \sigma^2 (X^T X)^{-1} \quad (12)$$

As previously mentioned, the compositions can be limited to lower or higher constraints, that is, the proportions of the components can vary only in a sub-region of the original simplex. In these conditions, the parametrization of the original proportions is performed in terms of pseudo-components, maintaining the interval between 0 and 1 [6].

Besides the existence of constraints, it is important to recognize the precision and the statistical significance of the coefficients since the choice of more compact and robust models leads to the success of the mixture study. For this, the coefficients are analyzed using the standard error for each b_i. The confidence interval (13) can be estimated based on the standard error and *t*-student distribution for each of the model coefficients [1].

$$b_i - t_{\left(\frac{\alpha}{2}\right),(n-p)} SE(b_i) \leq \beta_i \leq b_i + t_{\left(\frac{\alpha}{2}\right),(n-p)} SE(b_i) \quad (13)$$

Alternatively, the *t*-test at the statistical significance level of α can be performed for each regression coefficient β_i to assess its statistical significance, using the *t*-statistic (14) which is contrasted with the t-student distribution with (n-p) degrees of freedom.

$$t_i = \frac{b_i - \beta_i}{SE(b_i)} \quad (14)$$

VI. APPLICATION DESIGN AND FEATURES

The overall functioning of the MDesign app is described in the flowchart in Fig. 2. In terms of code structure, the application has 44 initialization components, 19 callbacks and utility functions, and some dynamic initialization controls. These built-in features allow the user to design mixtures in a simple and efficient platform.

The algorithms for calculating the regression models employed the mixture design methodology combined with matrix calculations. The aspects and statistical methods mentioned in the previous section were considered. For a better understanding of the behavior of the mixture and the interaction between components, a graphical user interface (GUI) was developed in this work, in standalone application format, using the App Designer toolbox of the MATLAB software (MathWorks, Inc., R2021b 9.11 version). The application MDesign has an initial dialog box in which the user can enter data from the experimental runs and their respective responses and should select the option of multivariate model in terms of linear (7), quadratic (8) or cubic (9), being the first two models the most frequently used by researchers.

The data is displayed in a table format, along with the coefficients of each designated experimental model, following the order of the terms displayed on the App.

ISBN: 978-989-33-3436-2

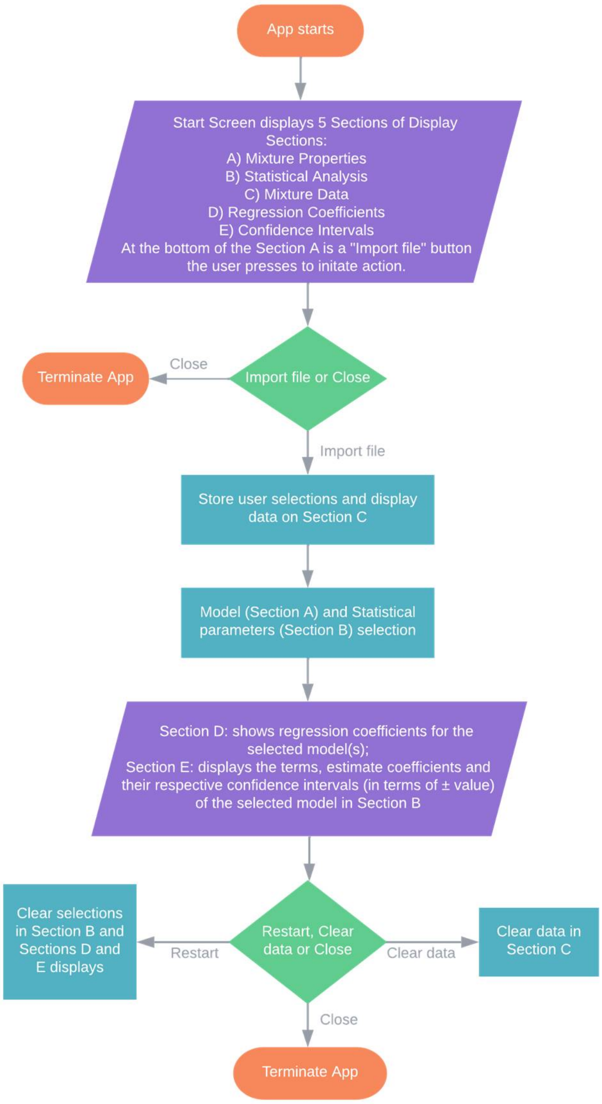

Figure 2. Flowchart of the app.

Furthermore, the application has a section for residual analysis and confidence interval of the coefficients. Statistical results are based on analysis of variance (ANOVA), ordinary and adjusted coefficients of determination (R^2 and R^2_{Adj}) and residual analysis (Standard, Studentized and Cook's distance), which are displayed in a modal box. The user can further count on a statistical significance test for each coefficient of the designated regression model, and verified by the *t*-statistic, both based on hypothesis tests at the statistical significance level of 5% (default option of α). Coefficients that do not show statistical significance for the model can be eliminated. Thereby, it is possible to obtain a multivariate model suitable for the process, but more compact, with only the significant terms.

VII. CASE STUDIES

To test the efficiency and accuracy of the application developed in this work, three experimental case studies retrieved from the book "Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data", written by J. Cornell [3] were used. The choice of these three case studies was based on the premise that they have differences among them, as the number of components and constraints. Concerning the topic of the case studies, mixture design is an important tool in analytical and chemometric studies related to the solubility of organic compounds in multi-solvent systems.

A. *Vapor pressure for organic solvents/water mixture*

Any liquid can vaporize until the pressure reached by its vapor meets the equilibrium. Once reached, the pressure in the vapor phase is called vapor pressure (VP) of the liquid [7]. Vapor pressure is a colligative property of solvent mixtures and is directly linked to the amount of solute present in a solution, which is a fundamental aspect in the development of new chemical, biochemical and pharmaceutical formulations. Vapor pressure readings (y) for each of the 12 different mixtures of ethanol (x_1), propylene glycol (x_2) and water (x_3) were evaluated in this case study. The proportions in this experiment follow the following constraints $0.0463 \leq x_1 \leq 0.7188$; $0.0272 \leq x_2 \leq 0.5776$; $0.2272 \leq x_3 \leq 0.9265$ [3, p. 560].

B. *Study of the solubility of phenobarbital*

The study of the solubility of medicinal agents is widely regarded as a fundamental task in the development of pharmaceutical products and studies of active components in drugs [8]. In this case study, several mixtures of glycerol (x_1), propylene glycol (x_2) and water (x_3) were evaluated in order to study the solubility (y) of phenobarbital as a function of the composition of the solvent [3, p. 564].

C. *Study of the solubility of butoconazole nitrate*

Polaxomer 407 was blended with a solution of polyethylene glycol (x_1), glycerin (x_2), polysorbate 60 (x_3) and water (x_4) to form an antifungal agent: butoconazole nitrate. The amount of Polaxomer 407 was kept constant, while the remaining portion varied between different proportions of each component, according to the constraints $0.1 \leq x_1 \leq 0.4$; $0.1 \leq x_2 \leq 0.4$; $0.0 \leq x_3 \leq 0.08$; $0.3 \leq x_4 \leq 0.7$ [3, p. 567]. For this case study, the solubility of butoconazole nitrate was evaluated in terms of proportions of the components.

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

VIII. RESULTS AND DISCUSSION

To obtain direct comparisons between the regressions previously described in the literature and the ones generated by the application, models of the same order were adopted for each case study. The results for the three case studies were very similar for both sources, literature, and application since the models are equivalent to two decimal places and the same significant β coefficients are identified. Considering the fitness, the models presented adequate F-ratio, R^2 and adjusted R^2, according to the obtained ANOVA tables, shown in Fig. 3 for Case Study A. The coefficients of determination were equal to those previously published in the literature [3].

	Degrees of Freedom	Sum_squares	Mean_square	F_ratio
Regression	2	1.0788e+03	539.4140	46.3964
Residual	9	104.6357	11.6262	0
Total	11	1.1835e+03	0	0
R-squared	0.9116	0	0	0
Adj. R-squared	0.8919	0	0	0

Figure 3. ANOVA table and correlation coefficients for the linear model of Case Study A.

As previously mentioned, the models were statistically evaluated according to their accuracy, to assure that they provide an adequate approximation to the real response values and to certify that the component constraints were not violated. For this verification, in addition to the analysis of variance, residual diagnosis was applied to the regression models. Fig. 4 illustrates the residuals calculated for Case Study A as an example.

	Observation	Response	Estimated respo...	Residuals	Studentized Res...	Cooks Distance
1	1	18.4400	19.6546	-1.2146	-1.0468	0.4063
2	2	25.8100	27.9143	-2.1043	-0.7927	0.0435
3	3	30.7100	30.0067	0.7033	0.2453	0.0033
4	4	30.7300	35.8352	-5.1052	-3.6900	2.7300
5	5	33.2700	32.7444	0.5256	0.2094	0.0040
6	6	35.0300	34.1942	0.8358	0.2930	0.0048
7	7	37.0300	36.7011	0.3289	0.1410	0.0022
8	8	42	39.7419	2.2581	0.7451	0.0193
9	9	43.7000	43.7560	-0.0560	-0.0220	4.2593e-05
10	10	48.4200	41.2812	7.1388	3.8551	0.8745
11	11	48.7900	48.2602	0.5298	0.2042	0.0034
12	12	53.7000	57.5401	-3.8401	-3.8058	5.2409

Figure 4. Residual analysis for the linear model of Case Study A.

Regarding the performance of MDesign app, there were no reports of crash error and startup or terminate errors. The initial layout of all components was executed and displayed as expected. Data outputs and internal functions were evaluated to guarantee the high performance of the app, starting with the most basic, and gradually proceeding to the more complex ones. The screen layout of the app was designed to be simple and didactic (Fig. 5), being directly intended for academic use as a digital research tool for chemical engineering and related areas.

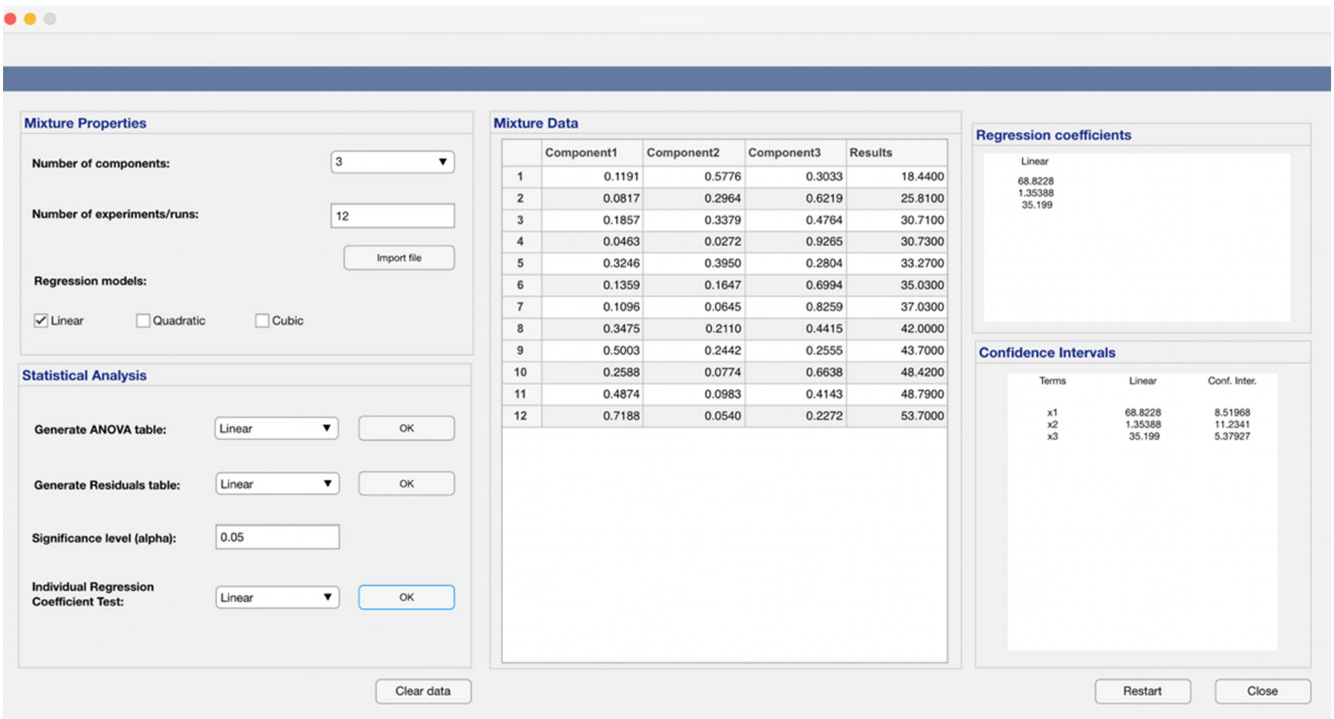

Figure 5. MDesign Layout.

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

IX. CONCLUSIONS

The mixture design methodology implemented as MDesign app can be successfully used for the estimation of parameters in mixture systems through regression models of different orders, combined with statistical significance. On the other hand, the use of this kind of interactive tools can support researchers to monitor the properties of mixtures in a practical way, eliminating the burden of numerous experiments and complex mathematical calculations. The application MDesign was illustrated using data experiments from three case studies published in the literature, in which satisfactory results were obtained. Although further studies should be conducted to consolidate its use for general application in mixture studies, the MDesign application was efficient in predicting the behavior of mixtures and can be a useful and didactic tool for several areas of science and technology.

ACKNOWLEDGMENT

This work has been supported by FCT — Fundação para a Ciência e Tecnologia within the projects CeDRI (UIDB/05757/2020) and CIMO (UIDB/00690/2020), and AllNat – (POCI-01- 0145-FEDER-030463, PTDC/EQU-EPQ/30463/2017), funded by FEDER funds through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI), Portugal 2020 and by national funds FCT/MCTES.

REFERENCES

[1] D. C. Montgomery, *Design and analysis of experiments*, 8th ed. John Wiley & Sons, Inc, 2012.

[2] B. Barros Neto, I. S. Scarminio, and R. E. Bruns, *Como Fazer Experimentos: Pesquisa e Desenvolvimento na Ciência e na Indústria.*, 4th ed. Campinas, SP: Ed. Unicamp, 2010.

[3] J. A. Cornell, *Experiments with Mixtures*, 3rd ed. New York: John Wiley & Sons, Inc., 1992.

[4] M. A. Bezerra *et al.*, "Application of mixture design in analytical chemistry," *Microchemical Journal*, vol. 152, p. 104336, Jan. 2020, doi: 10.1016/j.microc.2019.104336.

[5] "CRAN Task View: Design of Experiments (DoE) & Analysis of Experimental Data." https://cran.r-project.org/web/views/ExperimentalDesign.html (accessed Feb. 15, 2022).

[6] B. Dejaegher and Y. vander Heyden, "Experimental designs and their recent advances in set-up, data interpretation, and analytical applications," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 56, no. 2, pp. 141–158, Sep. 2011, doi: 10.1016/j.jpba.2011.04.023.

[7] W. Limmun, J. J. Borkowski, and B. Chomtee, "Weighted A-optimality criterion for generating robust mixture designs," *Computers & Industrial Engineering*, vol. 125, pp. 348–356, Nov. 2018, doi: 10.1016/j.cie.2018.09.002.

[8] E. Stauffer, J. A. Dolan, and R. Newman, "Chemistry and Physics of Fire and Liquid Fuels," in *Fire Debris Analysis*, Elsevier, 2008, pp. 85–129. doi: 10.1016/B978-012663971-1.50008-7.

[9] Y. Ma *et al.*, "Machine learning-based solubility prediction and methodology evaluation of active pharmaceutical ingredients in industrial crystallization," *Frontiers of Chemical Science and Engineering*, Oct. 2021, doi: 10.1007/s11705-021-2083-5.