Response Surface Method combined with Data Analysis to Optimize Extraction Process Problem

L. A. Lima[1,2], A. I. Pereira[1], C. B. Vaz[1], O. Ferreira[2], and M. Carocho[2]

[1] Research Center in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Bragança, Portugal
[2] Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Bragança, Portugal
limalaires1@gmail.com,{apereira,clvaz,oferreira,mcarocho}@ipb.pt

Abstract. Find and develop an appropriate optimization approach is directly associated with the reduction of the time and labor employed in a given chemical process and could be decisive for quality management. In this context, this work presents an approach to implement the Response Surface Method. This technique combines Response Surface Method with Genetic Algorithm and data mining. The main objective is to develop in MATLAB® a method able to optimize the surface function based on three variables using Hybrid Genetic Algorithms combining with Cluster Analysis to reduce the number of experiments and to find the closest value to the optimum within the established restrictions. The results are in accordance with those reported in a previous study. The proposed method has proven to be a promising alternative strategy since the optimal value was achieved without going through derivability unlike conventional methods, and fewer experiments were required to find the optimal solution in comparison to the previous work using the traditional Response Surface Method.

Keywords: Optimization · Genetic algorithm · Cluster analysis.

1 Introduction

The search and optimization methods have several principles, the most relevant being: the search space, where the possibilities of solving the problem in question are considered; the objective function (or cost function) and the problem coding, that is, a way to evaluate an objective in the search space [1]. Conventional optimization techniques, such as traditional RSM, start from an initial value or vector that is iteratively manipulated through some heuristic or deterministic process directly associated with the problem to be solved [3]. The great difficulty in solving a problem using the stochastic method is the number of possible solutions growing with a factorial speed, making it impossible to list all possible solutions to the problem [4]. Evolutionary computing techniques, as Genetic Algorithm, operate on a population that changes with each iteration. Thus, they can search in different regions of the viable space, allocating an adequate number of members to search in different areas [4].

Thus, considering the importance of predicting the behavior of analytical processes and avoiding costly procedures, this study aims to propose an alternative for optimizing multivariate problems, *e.g.* yield from the extraction of plant matrices. This study presents a comparative analysis between two optimization methodologies (traditional RSM and dynamic RSM), both developed in MATLAB® software (version R2019a 9.6), which aim to maximize the yield of a extraction process.

1.1 Case Study

This work is based on the published experimental data regarding the extraction conditions of male chestnut flowers using heat-assisted extraction [2]. The study parameters: time (t, in min), temperature (T, in °C), solvent (S, organic solvent in %v/v of ethanol) were optimized based on the Circumscribed Central Composite Design (CCCD) and the result was based on the extraction yield (Y, expressed as a percentage of dry extract). The CCCD design selected by the authors is based on a cube circumscribed to a sphere in which the vertices are at an α distance from the center and have 5 levels for each factor (t, T, and S). In this case, the α values vary between -1.68 and 1.68, and correspond to each factor level.

2 Dynamic Response Surface Method

For the newly proposed optimization method, briefly described in the flowchart shown in Fig. 1, the structure of the design of the experience was maintained and the imposed constraints on the responses and variables were adopted for dynamic RSM to elude awkward solutions.

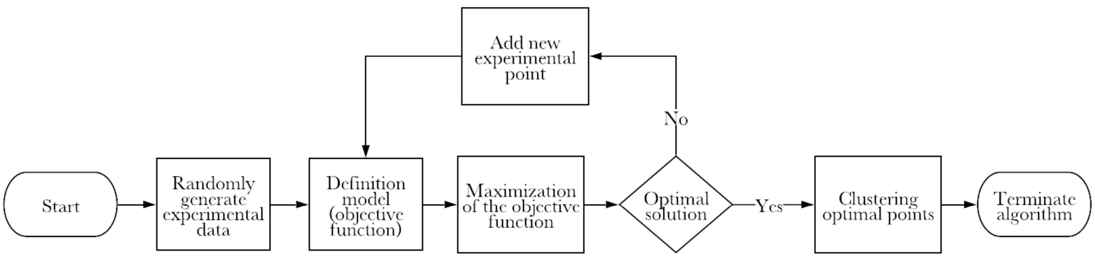


Fig. 1: Flowchart of dynamic RSM integrating Genetic Algorithm and cluster analysis to the process.

The algorithm was developed in MATLAB® software and starts by generating a set of 15 random combinations between the levels of combinatorial analysis. From this initial experimental data, a multivariate regression model is calculated, being this model the objective function of the problem. The Genetic Algorithm was used to solve the optimization problem. The optimal combination is identified and it is used to define the objective function. The process stops when no new optimal solution is identified

For the implementation of dynamic RSM in this case study, one hundred tests were performed to assess the effectiveness of the method. Considering the stochastic nature of the problem, cluster analysis is used to evaluate the outputs of the tests and identify the best candidate ideal solution. To handle the variability of the optimal solution achieved, the bootstrap method is used to estimate the confidence interval at 95%, by re-sampling data (1000 re-samples).

3 Numerical Results

According to the previous work using traditional RSM, the optimal conditions for maximum yield were: 120.0 min, 85.0 °C, and 44.5% of ethanol in the solvent, producing 48.87% of dry extract [2]. Using dynamic RSM, the estimated optimal conditions for the same response variable were: 118.5 min, 84.1 °C, and 46.1% of ethanol in the solvent, producing 45.87% of dry extract. In this case, the obtained optimal conditions for time and temperature were in accordance with approximately 80% of the tests.

The clustering analysis for each response variable was performed considering the means and the medoids for the output population (optimal responses from all tests) and the *bootstrap* analysis enables to make the inference concerning the results achieved, as shown in Figs. 2 and 3.

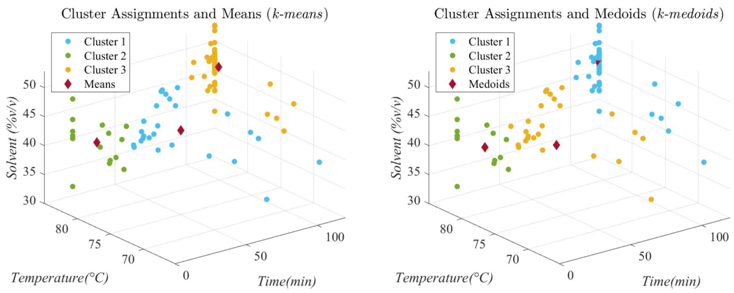


Fig. 2: Clustering analysis of the outputs from yield optimization using dynamic RSM: *k-means* and *k-medoids*, from left to right.

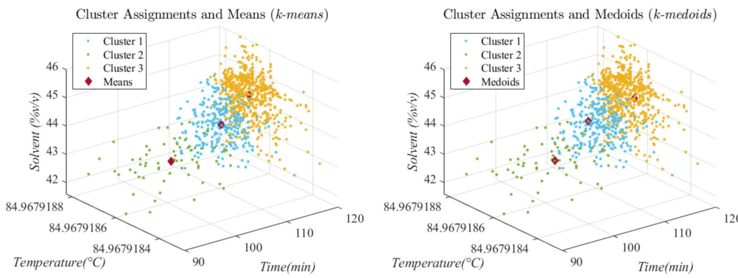
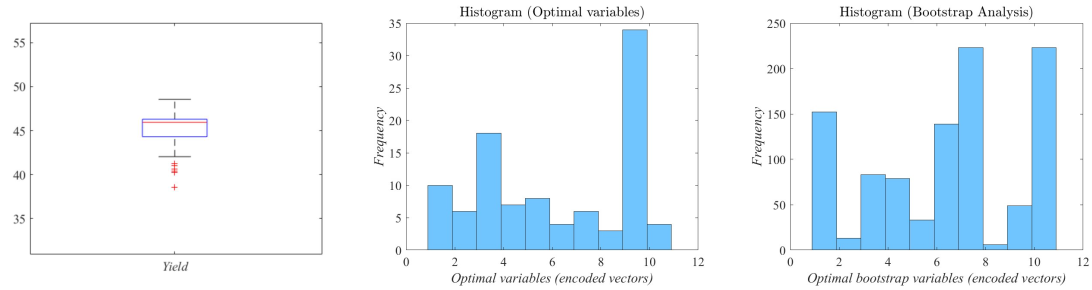


Fig. 3: Clustering analysis of the bootstrap points from yield optimization using dynamic RSM: *k-means* and *k-medoids*, from left to right.

The box plot corresponding to the group of optimal responses (yield) from dynamic RSM shows that the variance within each group is small, given that the difference between the set of responses is narrow. The histograms concerning the set of dynamic RSM responses and the bootstrap distribution of the mean (1000 re-samples) are shown in Fig. 4.



(a) Box plot of the outputs. (b) Histogram of the outputs. (c) Histogram of the bootstrap means.

Fig. 4: Box plot and histograms for the outputs of the testes using dynamic RSM.

The results obtained in this work are satisfactory, since the dynamic RSM returned values sufficiently close to those obtained using the traditional RSM [2]. Dynamic RSM took from 15 to 18 experimental points to find the optimal coordinates while some authors use the traditional RSM approach containing 20 combinations, including repetitions at the centroid. In addition to cluster analysis, bootstrapping was also applied, in which the sampling distribution of the statistic of interest is simulated through the use of several re-samplings with replacement of the original sample, enabling statistical inference. Bootstrapping was used to calculate confidence intervals to obtain unbiased estimates of the proposed method. In this case, the confidence interval was calculated at the 95% level (two-tailed), analogous to the previous study [2]. It was observed that the Dynamic RSM approach also allows the construction of confidence intervals with a smaller margin of error than the Traditional RSM approach, leading to a more precise definition of the optimal conditions for the experiment.

4 Conclusion and Future Work

For the presented case study, applying dynamic RSM using Genetic Algorithm coupled with clustering analysis returned positive results, in accordance with previous published data [2]. Both methods proved to be applicable to the resolution of this particular case regarding the optimization of the extraction of target compounds from plant matrices. However, although the dynamic RSM method proposes an alternative and different approach, the number of experimental points is slightly lower in comparison to the traditional method. Further studies are suggested to improve the algorithm and evaluate its performance using other case studies. In conclusion, dynamic RSM can be used as an alternative to traditional optimization methods, being effective and robust in estimating optimal conditions.

Acknowledgments

This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope: UIDB/05757/2020.

References

1. Beasley, D., Bull, D. R., Martin, R. R.: An overview of genetic algorithms: Part 1, fundamentals. University Computing, vol. 2(15), pp. 1–16 (1993)
2. Caleja C., Barros L., Prieto M. A., Bento A., Oliveira M.B.P., Ferreira, I C.F.R.: Development of a natural preservative obtained from male chestnut flowers: Optimization of a heat-assisted extraction technique. In: Food and Function, p. 1352-1363, vol. 10 (2019)
3. Geiger, E.: Statistical Methods for Fermentation Optimization. In: Vogel H. C., Todaro C. M. Fermentation and Biochemical Engineering Handbook: Principles, Process Design, and Equipment, 3rd edn. pp. 415-422, Elsevier Inc. (2014)
4. Schneider, J., Kirkpatrick, S.: Genetic Algorithms and Evolution Strategies. In: Stochastic Optimization, vol. 1, pp. 157-168, Springer-Verlag Berlin Heidelberg (2006)