

Leveraging Response Surface Methodology, Variance Analysis, and Purple Basil for Process Optimization

Zhixuan Zhao¹, Jinrong He², Xin Yao^{1,*}

1 Guangxi Eco-engineering Vocational and Technical College, Liuzhou, GuangXi, China 450200

2 Dongfeng Liuzhou Motor Co., Ltd

Abstract: Process optimization is a critical approach to improving production efficiency, quality, and stability. Response Surface Methodology (RSM) and Analysis of Variance (ANOVA) are widely used statistical tools in optimizing complex systems and enhancing manufacturing processes. This paper presents a process optimization algorithm based on RSM and variance modeling analysis. First, RSM is used to construct a process model and analyze the relationship between influencing factors and output variables. Then, ANOVA is applied to conduct a detailed statistical analysis of experimental data, identifying and quantifying the significance of factors and their interactions. By adjusting key parameters in the optimization process, this paper demonstrates how to effectively improve process performance, reduce variability, and ultimately achieve process optimization. Experimental results show that the proposed method can effectively improve process control in practical applications, providing valuable insights for process optimization in related fields.

Keywords: Perilla peach cake, Response surface method, Analysis of variance, Process optimization

DOI: 10.63887/jtie.2025.1.3.17

INTRODUCTION

Process optimization plays a fundamental role in enhancing operational efficiency, reducing costs, and improving product quality and consistency across various industries^[1-2]. As industrial systems become more complex, optimizing these processes requires not only an understanding of the technical variables involved but also an ability to navigate the interdependencies and nonlinear relationships that often arise between them. Traditional optimization methods, while useful, often fail to fully account for the interactions between multiple factors, especially in complex, real-world scenarios^[3-4].

Response Surface Methodology (RSM) has emerged as a powerful statistical technique to address these challenges. RSM enables the creation of mathematical models that approximate the relationship between input

factors and system outputs, allowing for the identification of optimal process conditions. This approach is particularly effective in situations where multiple factors interact in a nonlinear manner, which is typical in many industrial processes^[5-6]. By fitting a series of second-order polynomial models to experimental data, RSM provides insights into both the main effects and interaction effects of key variables, guiding decision-making towards optimal process parameters^[7-8].

To enhance the reliability of the optimization results, this paper also incorporates Analysis of Variance (ANOVA). ANOVA is a statistical method used to analyze the variability within the data, helping to determine the significance of each factor and the interactions between them. By performing ANOVA, the study ensures that the identified optimization parameters

are statistically significant and that the conclusions drawn from the RSM analysis are robust and reliable.

The aim of this paper is to integrate RSM and variance modeling analysis to develop a more comprehensive optimization algorithm for industrial processes. This integrated approach will provide a clearer understanding of the factors influencing process performance and offer a more accurate framework for improving efficiency, reducing variability, and ultimately optimizing outcomes. The findings of this research are expected to provide valuable insights and practical tools for industries looking to enhance their process optimization strategies, with potential applications in manufacturing, chemical engineering, and quality control.

1 Research method

1.1 Response surface optimization

Response Surface Optimization (RSO) is a sophisticated statistical technique that is widely used in various scientific and industrial fields to model and analyze the relationships between multiple independent variables (factors) and a dependent variable (response). The primary goal of RSO is to identify the optimal combination of factor levels that maximizes or minimizes the response variable, ultimately improving the performance and efficiency of a process or system. This optimization process is particularly valuable in complex systems where multiple factors interact, and traditional experimental methods may not offer adequate insight into these interactions.

RSO involves designing a set of carefully structured experiments, known as a design of experiments (DOE), where data is collected at different levels of the independent variables. This experimental design typically includes a range of factor combinations to ensure comprehensive coverage of the system's operating conditions. Once the data is collected, a mathematical model—often a quadratic regression model—is fitted to describe the relationship between the factors and the

response variable. This model is essential for understanding the interactions between factors, their individual contributions to the response, and the curvature of the response surface. By fitting the model, researchers can uncover intricate relationships and predict the behavior of the system under various conditions.

The next step in RSO is to analyze the fitted model using various statistical tools, including contour plots, response surface plots, and umbrella diagrams, to visualize how changes in the factor levels affect the response variable. These graphical representations provide a clear, intuitive way to identify regions of the factor space that lead to optimal responses, thus facilitating the determination of ideal operating conditions. Researchers can adjust the factors to achieve desired outcomes, whether the objective is to maximize production yield, minimize waste, or optimize efficiency.

One of the unique strengths of RSO is its ability to handle complex systems where factors are not simply additive but interact in nonlinear ways. This is particularly true in real-world applications, where factors such as temperature, pressure, concentration, and time often influence the outcome in a synergistic or antagonistic manner. RSO offers a robust approach to analyzing these interactions, providing deeper insights than traditional methods like one-factor-at-a-time (OFAT) experimentation, which may overlook important cross-factor effects.

In practical applications, RSO is widely used in a variety of industries, such as food production, pharmaceuticals, chemicals, and manufacturing, to fine-tune processes, enhance product quality, and improve operational efficiency. In food production, for example, RSO is used to optimize processing conditions to enhance flavor, texture, and nutritional content. In pharmaceuticals, RSO helps refine drug formulation processes, ensuring that the final product meets strict regulatory standards for consistency and efficacy.

Manufacturing industries leverage RSO to streamline production processes, reduce defects, and enhance productivity, all while minimizing resource usage and waste.

Moreover, the flexibility of RSO allows it to be integrated with other optimization techniques, such as Genetic Algorithms, Simulated Annealing, or Particle Swarm Optimization, for more sophisticated and automated optimization in high-dimensional problems. The continued advancement of computational tools has made RSO even more powerful, allowing researchers and engineers to explore more complex models, analyze larger datasets, and achieve higher levels of precision in optimization.

In conclusion, Response Surface Optimization is a valuable tool for improving process design and system performance. Its ability to model complex interactions between multiple variables and identify optimal conditions for a system makes it an indispensable technique in many industries striving to enhance product quality, efficiency, and innovation.

1.2 Model variance

Model variance refers to the variability observed in the predicted values of a model, and it is a key indicator of how well the model fits the actual data. In regression and statistical modeling, understanding model variance is critical for assessing the accuracy and reliability of a model's predictions. Variance essentially measures how much the model's predictions fluctuate based on the input data, and it helps in determining whether the model is generalizing well to new data or overfitting to the training set.

A model with high variance suggests that the model is highly sensitive to the data it was trained on, meaning that it is fitting not only the genuine underlying patterns but also the noise or random fluctuations present in the data. This typically leads to overfitting, where the model performs well on the training data but struggles to

generalize to unseen data. Overfitting is a common problem, especially when a model becomes excessively complex, fitting even small idiosyncrasies in the data that do not represent real-world patterns.

On the other hand, a model with low variance may indicate that the model is too simplistic, failing to capture important relationships and patterns in the data. This condition is known as underfitting, where the model's predictions are imprecise because it cannot adequately represent the underlying complexities of the system. In underfitting, the model may exhibit poor performance both on the training data and on new data, as it lacks the complexity needed to model the real system.

Therefore, a balance between high variance and low variance is essential for creating models that generalize well and provide reliable predictions. Achieving this balance often involves selecting an appropriate model complexity and using techniques such as regularization or cross-validation to avoid overfitting or underfitting.

1.3 The Role of R^2 and Adjusted R^2

One of the most common metrics used to assess how well a model explains the variance in the data is the coefficient of determination (R^2). R^2 quantifies the proportion of the total variability in the response variable that is explained by the model. R^2 values range from 0 to 1, with values closer to 1 indicating a better fit of the model to the data. A high R^2 value implies that the model is able to account for most of the variability in the data, which is a positive indicator of the model's explanatory power.

However, R^2 alone does not always provide a complete picture, especially when comparing models with different numbers of predictors. In such cases, the adjusted R^2 is a more accurate measure of fit. The adjusted R^2 adjusts the R^2 value for the number of predictors in the model, penalizing models with unnecessary or irrelevant variables. This adjustment ensures that adding more predictors to a model only

improves the R^2 value if those predictors meaningfully contribute to explaining the response. Thus, adjusted R^2 helps to prevent overfitting by providing a more reliable evaluation of model fit when comparing models with varying complexities.

1.4 Model Variance in Optimization and Experimental Design

In the context of optimization and experimental design, understanding model variance is crucial for evaluating how changes in the input variables (factors) influence the output response. Optimization often aims to find the input values that maximize or minimize the response variable, and a robust model is key to achieving this goal. A model with low unexplained variance (the variance that is not explained by the model) and high explained variance (the variance accounted for by the model) suggests that the model has successfully captured the system's behavior and can provide reliable predictions under different conditions.

For example, in process optimization, a model with high explained variance implies that the factors being manipulated have a significant effect on the outcome, making it easier to identify optimal conditions. Conversely, a model with high unexplained variance may suggest that important factors or interactions have not been accounted for, making it harder to pinpoint the optimal settings. Therefore, a model's variance plays a central role in the interpretation and reliability of optimization results.

2 Experimental results and process

2.1 Experiment preparation

A preliminary single-factor experiment was conducted using a basic formula consisting of low-gluten flour, lard, sucrose, egg yolk liquid, baking soda, baking powder, and oven settings, excluding baking soda and salt. The goal was to investigate the effects of different factors on the outcome. Perilla seed meal was incorporated into

the formula at five varying levels: 5%, 10%, 15%, 20%, and 25%.

Following the single-factor experiment, additional ingredients, including xylitol and maltitol, were mixed at a 1:1 ratio, with the following dosage levels: 40g, 30g, 20g, 10g, and 0g. Similarly, lard was tested at five different levels: 45g, 50g, 55g, 60g, and 65g, to determine its influence on the final product. Additionally, egg yolk liquid was varied at levels of 10g, 12.5g, 15g, 17.5g, and 20g.

To explore the interactions between these factors, an orthogonal experiment was carried out using the L9 (3⁴) design. This experimental design was chosen because it allows for the evaluation of multiple factors at different levels while minimizing experimental trials, making it efficient for identifying the most significant factors influencing the product.

2.2 Response surface optimization

In this study, we employed Design-Expert V8.0.6 software to conduct a comprehensive three-level optimization process on three key influencing factors: soybean oil (A), white sugar (B), and the other variable (C), as detailed in Table 2. The optimization process was based on the principles of the Box-Behnken experimental design, which is widely used for its efficient exploration of quadratic response surfaces and factor interactions. The coding of optimization factors and the central combination optimization experiment setup are outlined in Table 1 and Table 2, respectively.

Through this optimization procedure, a regression equation was derived that quantitatively describes the relationship between the independent variables and the sensory evaluation (Y value). The regression model, derived from the experimental data, represents the combined effects of these factors and their interactions.

Further, a variance analysis of the model data was performed to identify the significant interactions and main effects. The results from this analysis indicate that

there are significant interaction effects between the three influencing factors (A, B, and C), as well as their corresponding interaction terms (AB, AC, BC) and quadratic terms (A^2 , B^2 , C^2). These interactions and terms contribute differently to the sensory evaluation outcomes, with varying levels of influence observed across the factors.

Specifically, the main effects of the individual factors on the sensory evaluation (Y value) were ranked as follows: B (white sugar) had the highest influence, followed by C (the other variable), and A (soybean oil) had the least impact. This finding suggests that white sugar (B) plays a dominant role in determining the sensory quality of the product, whereas soybean oil (A) has a relatively smaller effect.

When examining the combined effects of factors, the most significant interactions were found between A and C (AC), followed by the interaction between A and B (AB), and then the interaction between B and C (BC). This suggests that the combined effects of soybean oil and the other variable (A and C) are more pronounced in

affecting the sensory evaluation than the interactions involving white sugar (B).

In terms of the quadratic effects, the influence of the squared terms was also significant, with A^2 (soybean oil) having the strongest influence, followed by B^2 (white sugar) and C^2 (the other variable). This indicates that the non-linear effects of soybean oil and white sugar are more impactful than those of the other variable (C).

Overall, among the three types of combined factors (interaction terms, quadratic terms, and individual factors), the most significant influences were attributed to white sugar (B), the interaction between soybean oil and the other variable (AC), and the quadratic term for soybean oil (A^2). These findings highlight the importance of both individual factors and their interactions in shaping the sensory qualities of the product and provide valuable insights into the optimization process.

This analysis provides a deeper understanding of how each factor and their interactions contribute to the overall outcome and lays the foundation for fine-tuning the formulation to achieve the desired sensory evaluation results.

Table 1 Code of optimization factors

Factor	Level		
	-1	0	1
Perilla seed meal	5	10	15
Soybean oil	35	40	45
Fine white sugar	20	30	40

2.3 Simulated analysis of variance

From the model's P value ($P_{\text{Model}} < 0.0001$) and the P value of the lack of fit ($0.1537 > 0.05$) shown in Table 2, it can be concluded that the model established during the optimization process exhibits significant differences overall, while the lack-of-fit term is not

significant. This indicates that the model has a high degree of fit.

Conclusion

This study conducted a three-level optimization of three influencing factors—soybean oil, white sugar, and another variable—using Box-Behnken experimental design principles and Design-Expert V8.0.6 software.

Through regression analysis and variance analysis, significant interactions were identified between the influencing factors, along with their interaction terms (AB, AC, BC) and quadratic terms (A^2 , B^2 , C^2), which all exhibited varying degrees of influence on the sensory evaluation outcomes.

In terms of individual factor effects, white sugar (B) had the most significant influence on the sensory evaluation, followed by the other variable (C), while soybean oil (A) had the least impact. Regarding combined factor effects, the most significant interaction was found between soybean oil and the other variable (AC), followed by the interaction between soybean oil and white sugar (AB), and the interaction between white sugar and the other variable (BC), which had a relatively smaller effect. Among the quadratic terms, soybean oil

(A^2) had the strongest influence, followed by white sugar (B^2) and the other variable (C^2).

Overall, the most significant influences on sensory evaluation were attributed to white sugar (B) and the interaction between soybean oil and the other variable (AC). This highlights the importance of optimizing white sugar levels and the interactions between soybean oil and the other variable to achieve the desired sensory quality. Additionally, the quadratic term for soybean oil (A^2) also played a crucial role in the formulation adjustments.

These findings provide a theoretical basis for further optimizing the product formulation and offer valuable data support for sensory evaluation improvements. The study also paves the way for future research in this area.

References

- [1]Li, M., Zhou, G. & Lu, C. Peach surface defect identification of complex background based on IDCNN and GWOABC-KM.[J].Multimed Tools Appl.2022,81, 16309–16334.
- [2]Kurtulbaş, E., Şahin, S. Phytochemical properties in the waste by-products of peach: optimization and storage stability studies.[J].Biomass Conv. Bioref.2024,14, 11053–11062.
- [3]Xu, B.; Cui, Q.; Guo, L.; Hao, L. Design and Parameter Optimization of a Combined Rotor and Lining Plate Crushing Organic Fertilizer Spreader.[J].Agronomy2024,14,1732.
- [4]Saha, S.; Mondal, A.K.; Čep, R.; Joardar, H.; Haldar, B.; Kumar, A.; Alsalah, N.A.; Ataya, S. Multi-Response Optimization of Electrochemical Machining Parameters for Inconel 718 via RSM and MOGA-ANN.[J].Machines.2024,12, 335.
- [5]Du, F.; He, L.; Huang, H.; Zhou, T.; Wu, J. Analysis and Multi-Objective Optimization for Reducing Energy Consumption and Improving Surface Quality during Dry Machining of 304 Stainless Steel.[J].Materials.2020,13, 4693.
- [6]Feng Zhunruo, Li Yunhong, Chen Weizhong, et al. Infrared and Visible Image Fusion Based on Improved Latent Low-Rank and Unsharp Masks.[J] SPECTROSCOPY AND SPECTRAL ANALYSIS, 2025, 45(7): 2034-2044.
- [7]Bose, S., Nandi, T. Experimental investigation of WEDM on titanium hybrid composite reinforced with boron powder: a novel approach.[J]Eur. Phys. J. Plus.2020,135,:914.
- [8]Sachin, B., Rao, C.M., Naik, G.M.et al.Influence of slide burnishing process on the surface characteristics of precipitation hardenable steel.[J] SN Appl. Sci.2021,3:223.

ACKNOWLEDGE: This work was financially supported by the analysis of nutritional components of water extract from Perilla leaves and development of peach cake products (2021KY01)