symmetry

MDPI

Article

Gaussian Process-Based Robust Optimization with Symmetry Modeling and Variable Selection

Zebiao Feng [1,*], Xinyu Ye [1], Weidong Huang [1] and Cuihong Zhai [2]

[1] School of Management, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; b21111503@njupt.edu.cn (X.Y.); huangwd@njupt.edu.cn (W.H.)
[2] School of Economics and Management, Nanchang Hangkong University, Nanchang 330063, China; cuihongzhai@nchu.edu.cn
* Correspondence: zbfeng@njupt.edu.cn

Abstract: Gaussian process (GP)-based robust optimization is an effective tool in product quality improvement. However, most existing variable selection methods are designed for parametric models and are unsuitable for nonparametric GP models. Additionally, improving the prediction accuracy of GP models with limited design points remains a significant challenge in robust optimization. To address these issues, this article proposes a GP-based multi-stage robust parameter optimization method that integrates symmetry modeling, sensitivity analysis (SA), and Markov Chain Monte Carlo (MCMC) techniques. First, a modified expected improvement (EI) criterion is introduced to enhance the utilization efficiency of design points. Second, a nonparametric variable selection technique based on SA is developed for GP models to identify significant variables. This method considers both independent variables and their interactions, improving the interpretability of GP models. Finally, the selected variables are used to construct the robust optimization model, and the genetic algorithm (GA) is employed to search for the optimal solution within the feasible domain. Numerical simulations and real-world experiments demonstrate the effectiveness of the proposed method. Comparative results indicate that the proposed method obtains more robust optimal input parameter settings compared to existing approaches.

Keywords: robust parameter optimization; Gaussian process regression; symmetry modeling; variable selection; nonparametric methods

check for updates

Academic Editor: Hsien-Chung Wu

Received: 19 December 2024
Revised: 9 January 2025
Accepted: 11 January 2025
Published: 13 January 2025

Citation: Feng, Z.; Ye, X.; Zhai, C. Gaussian Process-Based Robust Optimization with Symmetry Modeling and Variable Selection. *Symmetry* **2025**, *17*, 113. https://doi.org/10.3390/sym17010113

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Robust parameter optimization is a powerful tool for enhancing product quality in industrial manufacturing processes [1–3]. This approach seeks to minimize the deviations of the product quality responses from the targets while reducing the response variation [4,5]. A critical component of this process is the response surface model, which estimates the relationships between inputs and outputs, serving as a crucial link between empirical data and process optimization [6,7]. Robust optimization is achieved by identifying optimal parameter settings within the feasible region, which can be transformed into engineering optimization problems [8,9]. Therefore, developing accurate response surfaces with limited training points is crucial for improving the efficiency of robust optimization [10]. Furthermore, variable selection can enhance the interpretability of response surface models, further aiding engineering practitioners in identifying key input variables of manufacturing processes.

In recent years, GP-based robust parameter optimization has been widely applied to improve product quality in engineering [11,12]. Sun et al. [13] derived a sampling distribu-

tion from a fast-fitted Gaussian process based on previously evaluated solutions to balance exploitation and exploration. Costa and Lourenço [14] explored the benefits of GP models as a surrogate model for problems developed in robust optimization. Parnianifard et al. [15] proposed a simulation/optimization method with limited knowledge of the probability distribution of uncertain variables. Wang and Xu [16] developed a GP-based nonparametric functional regression method that handles both scalar and functional responses with mixed multidimensional functional and scalar predictors. Feng et al. [17] proposed a weighted modeling method to improve the flexibility and prediction accuracy of the GP model. Ghosh et al. [18] developed a linear mixed-effect GP modeling framework with B-spline random components to model multivariate profiles, addressing both within- and between-profile correlations for enhanced flexibility and reduced computational cost. Shen et al. [19] employed the multi-task Gaussian process upper confidence bound algorithm for automatic task selection. Ouyang et al. [20] proposed a new stochastic Kriging model to deal with the robust optimization for non-Gaussian responses. Lee [21] introduced a new adaptive Kriging-based optimization framework to improve the robustness of the probability of feasibility. Subsequently, Han et al. [22] presented a GP-based robust optimization method to reduce the computational complexity in the nested-loop structure of the worst-case analysis. Ozdemir and Cho [23] highlighted that the predictive accuracy of the GP response surface model directly impacts the effectiveness of the robust optimization results. Joseph et al. [24] developed a maximum projection (MaxPro) criterion to improve the space-filling of the design points. Krishna et al. [25] introduced an experimental scheme with the D-optimal design to improve the model accuracy. Ur Rehman et al. [26] and Chen et al. [27] adopted the EI technique to enhance the utilization efficiency of design points and the model accuracy. However, these methods with EI often excessively allocate design points at extreme locations. As data dimensionality and model complexity grow, the number of design points is often constrained by experimental limitations and computational resources. Consequently, improving the prediction accuracy of GP models with a limited number of design points has become a critical area of research. Enhancing the space-filling capability of the EI criterion is essential for achieving better prediction accuracy under these constraints. Additionally, insignificant input variables in GP regression models may lead to unnecessary computational costs and damage model prediction accuracy. Therefore, selecting significant variables for the GP model can further improve its prediction accuracy.

The variable selection technique of GP models reduces the computational load and improves the model interpretability [28–30]. Linkletter et al. [31] adopted the inert variable to identify the significant input factors. Chen et al. [32] proposed a Bayesian variable selection method with the MCMC technique to improve interpretability and prediction accuracy. Yi et al. [33] pointed out that the Bayesian variable selection suffers from the large variance of parameter estimation and high predictive errors. Consequently, Jiang et al. [34] adopted the Schwartz method to deal with the minimax-optimal error problem in variable selection. Park et al. [35] developed a forward stagewise approach with embedded gradient descent to select the significant input variables. However, the effectiveness of the above variable selection methods depends heavily on the prediction accuracy of the GP model. Specifically, GP models may exhibit poor prediction accuracy with limited design points, which leads to misleading variable selection results. Moreover, the abovementioned methods lack an adaptive criterion for determining the final variables to construct the GP model. The SA technique can evaluate the influence of input variables on model outputs without relying on specific model assumptions. Compared to other feature selection techniques, such as those based on autocorrelation coefficients or MCMC, SA offers better applicability for high-dimensional input scenarios, particularly when dealing

with complex objective functions and limited data samples. Therefore, it is meaningful to develop the SA-based GP model variable selection method for modern complex industrial products or processes.

This paper proposes a novel multi-stage adaptive GP modeling and variable selection method. First, an adaptive design approach is developed by modifying the EI criterion to improve the prediction accuracy. Second, an adaptive variable selection strategy is developed to identify significant variables for GP models. Compared to existing methods, the proposed symmetry-additive modeling approach separately models each input dimension, significantly reducing the complexity of the covariance structure. Building upon the symmetry-additive covariance structure, we developed a variable selection approach based on SA. The selection approach does not require pre-assumptions about the model and enhances the effectiveness of the selection results through an adaptive process. The proposed MCMC modeling technique mitigates overfitting in scenarios with limited sample data while providing a more flexible model that enhances scalability. The proposed GP-based variable selection method can effectively reduce computational complexity while improving the prediction accuracy and robustness of the model. This makes it suitable for dealing with robust parameter design problems with high-dimensional inputs. Consequently, it has broad application prospects in engineering optimization, quality control, and reliability analysis, including aerospace, automotive engineering, electronic product design, precision machinery manufacturing, and additive manufacturing.

The structure of this paper is organized as follows. Section 2 provides an overview of Gaussian process models, detailing the modeling procedure and the structure for the additive GP model. Section 3 introduces the proposed adaptive experimental design method, which balances exploitation and exploration through the modified EI technique. Section 4 demonstrates the effectiveness of the proposed method using a numerical simulation example and an AM case study. Finally, Section 5 offers concluding remarks and outlines potential directions for future research.

2. GP Modeling with Symmetric Positive Definite Covariance Matrix

The GP model is a flexible nonparametric surrogate modeling technique widely used in computer experiments and machine learning [36]. In practice, the expensive experimental process is always regarded as a black-box function: $f : \mathrm{R}^d \to \mathrm{R}$. To save computation costs, the surrogate model is commonly adopted to estimate the process $\widehat{f}_n : \mathrm{R}^d \to \mathrm{R}$, which approximates f based on a small number of n training points with inputs and outputs. Given an $n \times d$ matrix of training inputs $\mathbf{X}_n$ and the corresponding outputs $\widehat{\mathbf{Y}}_n$, where n is the number of inputs, d is the dimensionality of inputs. The expression of the GP model can be given by

$$\mathbf{Y}_n \sim \mathrm{N}(\widehat{\mathbf{Y}}_n, \sigma^2), \tag{1}$$

where σ^2 is the noise variance parameter. $\widehat{\mathbf{Y}}_n \sim \mathrm{N}(\boldsymbol{\mu}, \alpha_z^2 \mathbf{R}(\mathbf{X}_n))$, where $\boldsymbol{\mu}$ is the mean function, which is always set as 0. α_z^2 is the unknown process variance. $\mathbf{R}(\mathbf{X}_n)$ is the correlation matrix for the training inputs $\mathbf{X}_n$. The squared exponential function is one of the commonly used kernel functions. The correlation between x_i and x_j with the squared exponential kernel can be given by

$$r_{i;j} = \exp(-\sum_{g=1}^{d} l_g^{-2}(x_{i,g} - x_{j,g})^2), \tag{2}$$

where l_g is the length-scale parameter, which measures the importance of the g-th factor. Then, the covariance matrix $\mathbf{\Sigma}_n = (\alpha_z^2 \mathbf{R}(\mathbf{X}_n) + \sigma^2 \mathbf{I})$ can be calculated by $[\mathbf{R}]_{i,j} = r_{i,j}$. For the given training data $\mathbf{D}_n = \{\mathbf{X}_n, \mathbf{Y}_n\}$, the log-likelihood can be given by

$$\log p(\mathbf{Y}_n | \alpha_z, l_g, \sigma) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{\Sigma}| - \frac{1}{2} \mathbf{Y}_n^{\mathrm{T}} \mathbf{\Sigma}^{-1} \mathbf{Y}_n. \tag{3}$$

The GP model hyperparameters are estimated using the MCMC sampling method introduced by Gelman et al. [37]. The posterior predictive distribution of the test point $\mathbf{x}_{te}$ can be given by

$$y_{te}(\mathbf{x}_{te}) \sim N(f(\mathbf{x}_{te}), V(\mathbf{x}_{te})), \tag{4}$$

where $f(\mathbf{x}_{te}) = \mathbf{\Sigma}(\mathbf{x}_{te}, \mathbf{X}_n) \mathbf{\Sigma}_n^{-1} \mathbf{Y}_n$, $V(\mathbf{x}_{te}) = \mathbf{\Sigma}(\mathbf{x}_{te}, \mathbf{x}_{te}) - \mathbf{\Sigma}(\mathbf{x}_{te}, \mathbf{X}_n) \mathbf{\Sigma}_n^{-1} \mathbf{\Sigma}(\mathbf{X}_n, \mathbf{x}_{te})$, $\mathbf{\Sigma}_n = \mathbf{\Sigma}(\mathbf{X}_n, \mathbf{X}_n)$ is an $n \times n$ symmetric positive definite matrix to capture the covariance relationship between the training points, $\mathbf{\Sigma}(\mathbf{x}_{te}, \mathbf{X}_n)$ is a $1 \times n$ matrix to capture the covariance relationship between $\mathbf{x}_{te}$ and the training points. The additive GP model constructs a separate covariance matrix for each input dimension, which facilitates the application of classical sensitivity analysis techniques for nonparametric variable selection. Therefore, the additive modeling technique is adopted to construct the GP response surface. The posterior predictive distribution is given by an additive function $f = f^{(1)} + \ldots + f^{(J)}$ such that each component $j = 1, \ldots, J$ is a Gaussian process:

$$f^{(j)} \sim GP(0, \alpha_j^2 k_j(\mathbf{x}, \mathbf{x}')), \tag{5}$$

and each component is independent. This means:

$$f \sim GP(0, k(\mathbf{x}, \mathbf{x}')), \tag{6}$$

where

$$k(\mathbf{x}, \mathbf{x}') = \Sigma_{j=1}^{J} \alpha_j^2 k_j(\mathbf{x}, \mathbf{x}'). \tag{7}$$

Consider N observations $\{\mathbf{x}_n\}_{n=1}^{N}$, and the multivariate normal prior of GP model can be given by

$$f^{(j)} = \left[f^{(j)}(\mathbf{x}_1), \ldots, f^{(j)}(\mathbf{x}_N) \right]^{\mathrm{T}} \sim N(0, K^{(j)}), \tag{8}$$

where each vector $f^{(j)}$ corresponds to $j = 1, \ldots, J$. The elements of the matrix $K^{(j)}$ are defined as $\{K^{(j)}\}_{ik} = \alpha_j^2 k_j(\mathbf{x}_i, \mathbf{x}_k)$. The prior distribution for $f = f^{(1)} + \ldots + f^{(J)}$ is

$$f \sim N(0, K), \tag{9}$$

where $K = \Sigma_{j=1}^{J} K^{(j)}$ is the symmetric positive definite matrix. The symmetry modeling approach simplifies the model structure by independently modeling each input, improving the efficiency and reliability of the GP model. Specifically, it offers the following advantages: Firstly, it reduces the complexity of inputs and lowers computational costs. Secondly, it clarifies the influence of each input on the output, facilitating variable selection through the sensitivity analysis technique. Finally, The model can be easily extended without the need to re-model the complex covariance matrix for new inputs, and it also helps reduce the risk of overfitting in situations with limited data. The unknown parameters can be estimated using the Hamiltonian Monte Carlo technique of MCMC. The squared exponential kernel is selected to avoid loss of generality, which is particularly effective for modeling nonlinear functional relationships and capturing correlations within the input space. To ensure the validity of the prior assumptions, the prior distribution assumptions introduced by Gelman et al. [37] are adopted for the hyperparameters of the Gaussian process ker-

nel. The detailed prior settings are as follows: the prior distributions for the process variance α_j^2, length scale l_g, and noise variance σ^2 are $\alpha_j^2 \sim$ Inverse-Gamma(a, b), $l_g \sim$ Uniform$(l_{\min}, l_{\max})$, and $\sigma^2 \sim$ Inverse-Gamma(a', b'), respectively. The covariance matrix $\mathbf{\Sigma}$ is calculated based on the current hyperparameters. Consider the training dataset $\mathbf{D}_n = \{\mathbf{X}_n, \mathbf{Y}_n\}$, where $\mathbf{X}_n$ is an $n \times d$ input matrix and $\mathbf{Y}_n$ is the corresponding output vector. According to the Bayesian theorem, the joint posterior distribution of the hyperparameters is given by:

$$p(\alpha_j^2, \{l_g\}, \sigma^2 | \mathbf{Y}_n) \propto p(\mathbf{Y}_n | \alpha_j^2, \{l_g\}, \sigma^2) \cdot p(\alpha_j^2) \cdot p(\{l_g\}) \cdot p(\sigma^2), \tag{10}$$

where $p(\mathbf{Y}_n | \alpha_j^2, \{l_g\}, \sigma^2)$ is the likelihood function presented in Equation (3). Next, hyperparameters are sampled from their conditional distributions. First, the conditional posterior distribution of process variance α_j^2 is given by the product of the likelihood function and its prior distribution. The likelihood function is expressed as:

$$p(\mathbf{Y}_n | \alpha_j^2, \{l_g\}, \sigma^2) \propto \exp\left(-\frac{1}{2\alpha_j^2} \mathbf{Y}_n^{\mathrm{T}} \mathbf{\Sigma}^{-1} \mathbf{Y}_n \right). \tag{11}$$

Combining the Inverse-Gamma prior distribution $\alpha_j^2 \sim$ Inverse-Gamma(a, b), the conditional posterior distribution is also an Inverse-Gamma distribution:

$$\alpha_j^2 | \mathbf{Y}_n, \{l_g\}, \sigma^2 \sim \text{Inverse-Gamma}\left(a + \frac{n}{2}, b + \frac{1}{2} \mathbf{Y}_n^{\mathrm{T}} \mathbf{\Sigma}^{-1} \mathbf{Y}_n \right). \tag{12}$$

Thus, α_j^2 can be directly sampled from this conditional posterior distribution.

Then, the prior distribution for the length scale l_g is uniform, $l_g \sim$ Uniform$(l_{\min}, l_{\max})$, with prior density given by:

$$p(l_g) = \frac{1}{l_{\max} - l_{\min}}. \tag{13}$$

The conditional posterior distribution is:

$$p(l_g | \mathbf{Y}_n, \alpha_j^2, \sigma^2) \propto p(\mathbf{Y}_n | \alpha_j^2, l_g, \sigma^2) \cdot p(l_g). \tag{14}$$

Since $p(l_g)$ is uniform, the main contribution comes from the likelihood function (3). This conditional posterior distribution does not have an analytical form, thus requiring MCMC methods (such as the Metropolis-Hastings algorithm) for sampling. The specific steps are as follows: 1. Generate a new proposed length scale l_g^{new} from the proposal distribution. 2. Compute the likelihood ratio:

$$r = \frac{p(\mathbf{Y}_n | \alpha_j^2, l_g^{\text{new}}, \sigma^2)}{p(\mathbf{Y}_n | \alpha_j^2, l_g^{\text{current}}, \sigma^2)}. \tag{15}$$

3. Accept the new proposed length scale l_g^{new} based on the acceptance probability $\min(1, r)$.

Finally, the prior distribution for the noise variance σ^2 is given by the Inverse-Gamma distribution $\sigma^2 \sim$ Inverse-Gamma(a', b'). Combining with the likelihood function (3), the conditional posterior distribution of σ^2 is also an Inverse-Gamma distribution:

$$\sigma^2 | \mathbf{Y}_n, \alpha_j^2, \{l_g\} \sim \text{Inverse-Gamma}\left(a' + \frac{n}{2}, b' + \frac{1}{2} \mathbf{Y}_n^{\mathrm{T}} \mathbf{\Sigma}^{-1} \mathbf{Y}_n \right), \tag{16}$$

where $\mathbf{\Sigma} = \alpha_j^2 \mathbf{R}(\mathbf{X}_n) + \sigma^2 \mathbf{I}$ is the symmetric positive definite covariance matrix that includes the noise variance σ^2. Therefore, σ^2 can also be directly sampled from this conditional posterior distribution.

This process continues until a sufficient number of samples are obtained. Ultimately, using the posterior samples, the predictive distribution for new test points $\mathbf{x}_{te}$ can be computed, thereby enabling effective predictions for the model.

3. Adaptive Robust Optimization Process

3.1. Adaptive Modeling

There are many popular experimental design strategies for GP modeling, such as the LHS, pairwise distances, and mixture design methods. However, most of these methods are not sequential, which may reduce the efficiency of the training data. The accuracy of the model prediction is crucial for the effectiveness of the obtained optimal solution in robust optimization. An adaptive experimental design strategy based on the EI algorithm and active learning technique is developed to improve the prediction accuracy to construct the GP model with the fixed n experiment points. The response of input $\mathbf{x}$ can be regarded as a random value $Y(\mathbf{x})$, and then the improvement beyond the current observed minimum value $y_{min} = \min(y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_n))$ with $\mathbf{y} = [y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_n)]$ can be derived as

$$\mathrm{I}(\mathbf{x}) = \max(y_{\min} - Y(\mathbf{x}), 0). \tag{17}$$

The EI function is conditional on the observations and derived in the closed form:

$$\begin{aligned} \mathrm{EI}(\mathbf{x}) &= \mathrm{E}(\mathrm{I}(\mathbf{x}) | Y(\mathbf{x}) = \mathbf{y}) \\ &= (\min(\mathbf{y}) - f(\mathbf{x}))\Phi\left(\frac{\min(\mathbf{y}) - f(\mathbf{x})}{v(\mathbf{x})}\right) + V(\mathbf{x})\phi\left(\frac{\min(\mathbf{y}) - f(\mathbf{x})}{v(\mathbf{x})}\right), \end{aligned} \tag{18}$$

where $f(\mathbf{x})$ and $v(\mathbf{x})$ are the predictive response and the predictive standard deviation at input $\mathbf{x}$, respectively. Φ and ϕ are the cumulative distribution function (CDF) and the probability density function (PDF) of the standard Gaussian distribution. The EI algorithm can find the minimum value using the following expression:

$$\mathbf{x}^{n+1} = \text{argmax} \quad \mathrm{EI}(\mathbf{x}^{n+1} | \mathbf{X}), \tag{19}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ is the observations. The points generated by the EI algorithm are mostly near the extremums, which may lead to the waste of the computational cost. Then, the MaxPro criterion is adopted to measure the distance between points:

$$\max_{\mathbf{x} \in \Omega} \mathrm{M}(\mathbf{x}_i, \mathbf{x}_j | i \neq j) = \left\{ \prod_{g=1}^{d} (x_{i,g} - x_{j,g})^2 \right\}^{1/d}. \tag{20}$$

Equation (20) can be used to measure the average distance between the new point $\mathbf{x}_{new}$ and observations $\mathbf{X}$:

$$\begin{aligned} \mathrm{ME}(\mathbf{x}_{new}, \mathbf{X}) &= \mathrm{mean}(\mathrm{M}(\mathbf{x}_{new}, \mathbf{x}_1) + \mathrm{M}(\mathbf{x}_{new}, \mathbf{x}_2) + \ldots + \mathrm{M}(\mathbf{x}_{new}, \mathbf{x}_n)) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{g=1}^{d} (x_{new,g} - x_{i,g})^2 \right\}^{1/d} \end{aligned}. \tag{21}$$

Then, the EI design criterion is modified (MoEI) using Equation (21):

$$\mathbf{x}^{n+1} = \text{argmax} \quad \mathrm{MoEI}(\mathbf{x}^{n+1} | \mathbf{X}) = \mathrm{EI}(\mathbf{x}^{n+1}) \cdot (1 - \mathrm{ME}(\mathbf{x}^{n+1}, \mathbf{X})), \tag{22}$$

where $\mathbf{x}^{n+1}$ represents the new point based on the n previous runs. Equation (22) can be used to generate the training points sequentially. The adaptive design criterion is developed based on Equations (19)–(22). The adaptive design process is given in Algorithm 1. The total number of training points is set as N. The initial n training points are generated by the LHS technique with midpoints, which can avoid points clustering, $\mathbf{D}_{in} = \{\mathbf{X}_{in}, \mathbf{Y}_{in}\}$. Consequently, construct the GP model with the training data $\mathbf{D}_{in}$. Finally, the MoEI algorithm is employed to select the new points in the design region. In Gaussian process modeling, prediction uncertainty in extreme regions is typically high, representing a major source of response surface errors. Therefore, the MoEI algorithm is employed to explore these extreme regions, aiming to reduce uncertainties and improve the effectiveness of the response surface model. Additionally, the distance constraint of MoEI can effectively avoid local clustering and resource wastage. Consequently, the proposed MoEI algorithm not only effectively explores extreme regions but also improves the efficiency of design point utilization, thereby enhancing the overall performance of the GP response surface model.

Algorithm 1 Pseudocode of adaptive design process with MoEI

Input: Maximum number of experimental runs N, initial training data $\mathbf{D}_{in} = \{\mathbf{X}_{in}, \mathbf{Y}_{in}\}$ from LHS technique.
Construct the GP model using $\mathbf{D}_{in}$.
for $p = 1$ to $(N - n)$ **do**
    Generate the new point $\mathbf{x}^p$ using MoEI in Equation (22).
    Calculate ME($\mathbf{x}^p$,$\mathbf{X}_{in}$) using Equation (21).
    if MoEI($\mathbf{x}^p$,$\mathbf{X}_{in}$) $>\kappa_{md}$ **then**
        Update $\mathbf{D}_{in}$ with new points: $\mathbf{X}_{in} = [\mathbf{X}_{in}, \mathbf{x}^{new}]$, $\mathbf{Y}_{in} = [\mathbf{Y}_{in}, \mathbf{y}^{new}]$.
        Set $\mathbf{D}_1 = \mathbf{D}_{in}$.
        if Length($\mathbf{D}_1$) $\leq N$ **then**
            Generate additional $N -$ Length($\mathbf{D}_{in}$) new points via MoEI.
            Update $\mathbf{D}_{in}$.
        else
            Break.
        end if
    end if
end for
Output: Final training dataset $\mathbf{D}_{tr} = \mathbf{D}_1$.

3.2. Adaptive Variable Selection

The variable selection technique can be adopted to save computational costs and improve model interpretability. The length-scale parameter l_g of the correlation function can be used to measure the importance of the input variable, which is called automatic relevance determination (ARD). The drawback of the ARD method is that it heavily depends on the accuracy of the parameter estimation. The variable selection technique based on sensitivity analysis is adopted to measure the predictive relevance unaffected by parameter estimation. The proposed selection method measures the ranking of input variables by analyzing the changes in the predicted mean along each variable direction. When the value of an individual input variable changes, large variations in the latent mean suggest that the variable is important for predicting the target variable. Only the latent mean is considered to save computational costs, and the prediction uncertainty is ignored during the selection process. Another advantage of this selection method is that it utilizes information from a larger region of the input space.

The distribution of input variables is approximated to estimate the variance of the latent function mean, which is achieved by calculating the sample mean $\boldsymbol{\mu}$ and sample covariance $\boldsymbol{\Sigma}$ for the n training inputs $\mathbf{X} = (x^{(1)}, \ldots, x^{(n)})$. For any given point, $\mathbf{x}_i$, $p(x_g|\mathbf{x}_i)$ represents the conditional distribution of variable g. The predictive response can be estimated by

$$\begin{aligned} &x_g|\mathbf{x}_{-g} \sim N(m_g, s_g^2), \\ &m_g = \mu_g + \boldsymbol{\sigma}_{g,-g}\boldsymbol{\Sigma}_{-g,-g}^{-1}(\mathbf{x}_{-g} - \boldsymbol{\mu}_{-g}), \\ &s_g^2 = \sigma_{g,g} - \boldsymbol{\sigma}_{g,-g}\boldsymbol{\Sigma}_{-g,-g}^{-1}\boldsymbol{\sigma}_{-g,g}. \end{aligned} \tag{23}$$

where the subscript g refers to selecting the g-th row or column, and $-g$ represents removing them. The variance of the posterior mean $f_{i,g} = \mathrm{E}[f(x_g)|\mathbf{x}_{i,-g}]$ can be given by integrating over the conditional distribution $y(x_g|\mathbf{x}_{i,-g})$:

$$V_{f_{i,g}} = \int f_{i,g}^2(x_g)N(x_g|m_g, v_g)dx_g - (\int f_{i,g}(x_g)N(x_g|m_g, v_g)dx_g)^2. \tag{24}$$

For the change in variables, $k = \frac{x_g - m_g}{\sqrt{2v_g}}$. Under the new variable, the form of the integral can be given by

$$I = \int f(x_g)N(x_g|m_g, v_g)dx_g = \int f(k\sqrt{2v_g} + m_g)\frac{e^{-k^2}}{\sqrt{\pi}}dk. \tag{25}$$

By the Gaussian-Hermite quadrature (GHQ), Equation (25) can be approximated as

$$I \approx \Sigma_{i=1}^{N} w_i f(k_i\sqrt{2v_g} + m_g). \tag{26}$$

Then, Equation (26) can be numerically approximated by the GHQ:

$$\begin{aligned} V_{f_{i,g}} &= \int f_{i,g}^2(x_g)(\sqrt{2}ks_g + m_g)\tfrac{e^{-k^2}}{\sqrt{\pi}}dk - \left(\int f_{i,g}(x_g)(\sqrt{2}ks_g + m_g)\tfrac{e^{-k^2}}{\sqrt{\pi}}dk\right)^2 \\ &\approx \pi^{-1/2}\Sigma_{n_k=1}^{N_k} w_i f_{i,g}^2(x_g)(\sqrt{2}k_i s_g + m_g) - \pi^{-1}\left(\Sigma_{n_k=1}^{N_k} w_i f_{i,g}^2(x_g)(\sqrt{2}k_i s_g + m_g)\right), \end{aligned} \tag{27}$$

where N_k represents the number of weights and evaluation points k_i of the GHQ approximation. The k_i is given by the roots of the Hermite polynomial $H_{N_k}(k)$ and the weight w_i is given by

$$w_i = \frac{2^{N_k-1}N_k!\sqrt{\pi}}{N_k^2[H_{N_k-1}(k_i)]^2}, \tag{28}$$

Repeating Equations (23)–(27) for all the training points $\mathbf{X}_n$ can obtain the predictive variance mean of any dimension x_g

$$\mathrm{PV}_g = \frac{1}{n}\sum_{i=1}^{n} V_{i,g}, \tag{29}$$

where $\mathrm{PV_g}$ can be used to estimate the predictive relevance of variable g. Then, the adaptive variable selection method is developed to determine the significant variables. The pseudocode of this adaptive algorithm is summarized in Algorithm 2.

Algorithm 2 Pseudocode of adaptive variable selection procedure

Input: Training data $\mathbf{D}_{tr} = \{\mathbf{X}_{tr}, \mathbf{Y}_{tr}\}$, testing data $\mathbf{D}_{te} = \{\mathbf{X}_{te}, \mathbf{Y}_{te}\}$.
Construct GP model with $\mathbf{D}_{tr}$.
Compute variable importance PV_g for each input variable $g = 1, \ldots, d$, set threshold ε_x.
for each $g = 1$ to d **do**
    Select variables with $\mathrm{PV}_g > \varepsilon_x$ and form new input set $\mathbf{x}_{new}$.
    Define removed variables set $\mathbf{x}_{removed}$ for $\mathrm{PV}_g < \varepsilon_x$.
end for
Update data: $\mathbf{D}_{tr}^{new}$, $\mathbf{D}_{te}^{new}$.
Build GP model with updated training data and compute $RMSE_1$ on testing data.
Select the variable $\mathbf{x}_{removed}^{l}$ with highest PV_g from $\mathbf{x}_{removed}$, form $\mathbf{x}_{nr} = \{\mathbf{x}_{new}, \mathbf{x}_{removed}^{l}\}$.
Update data: $\mathbf{D}_{tr}^{nr}$, $\mathbf{D}_{te}^{nr}$.
Build GP model and compute $RMSE_2$.
if $(RMSE_1/RMSE_2) > \varepsilon_R$ **then**
    Update ε_x, go to Step 3.
else
    Break.
end if
Output: Final selected variables $\mathbf{x}_{new}$.

The detailed iteration procedure is as follows: (1) Design the experimental scheme and collect the experimental data. (2) Construct the GP model and variable selection model with the experimental data. (3) Calculate predictive relevance for input variables, and set the variable relevance threshold ε_x. (4) Construct the GP model using the variables with relevance values greater than εx, and calculate the root mean squared error (RMSE) with the testing data to assess the prediction accuracy. The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}{N}}, \tag{30}$$

where N is the number of test points. $\widehat{y}_i$ is the predictive response. The RMSE calculated in this step is denoted as $RMSE_1$. (5) Select the variable with the largest relevance from those removed in the previous step and combine it with the variables selected in Step 4. Then, construct the GP model using the new variable combination and calculate the $RMSE$, denoted as $RMSE_2$. (6) Calculate $w = RMSE_2/RMSE_1$. If $w < \varepsilon_R$, terminate the iteration process and output the final input variables. If $w > \varepsilon_R$, return to Step (3), reset the threshold, and continue the iteration.

3.3. Robust Optimization Model

Taguchi et al. [38] first introduced the quality loss function to find input parameter settings that make the response value close to the target with minimal variance. Tan and Wu [39,40] extended Taguchi's loss function to a generalized expected loss function for GP models. The expected quality loss function is expressed as

$$E(L(y)) = \gamma^2 + (E(y) - T)^2, \tag{31}$$

where T is the target of the response. y and γ^2 are the predictive response and variance of the GP model, respectively. As shown in Equation (31), both the variance and deviation between the response from the target contribute to quality loss. The quality loss function is adopted to construct the robust parameter optimization model.

3.4. Steps of the Proposed Approach

This section outlines the modeling process of the proposed adaptive robust optimization method, which comprises three phases: GP modeling using an adaptive experimental design strategy, adaptive variable selection with the SA technique, and robust optimization with the quality loss function. The flowchart of the proposed procedure is depicted in Figure 1, and the procedure is detailed as follows.

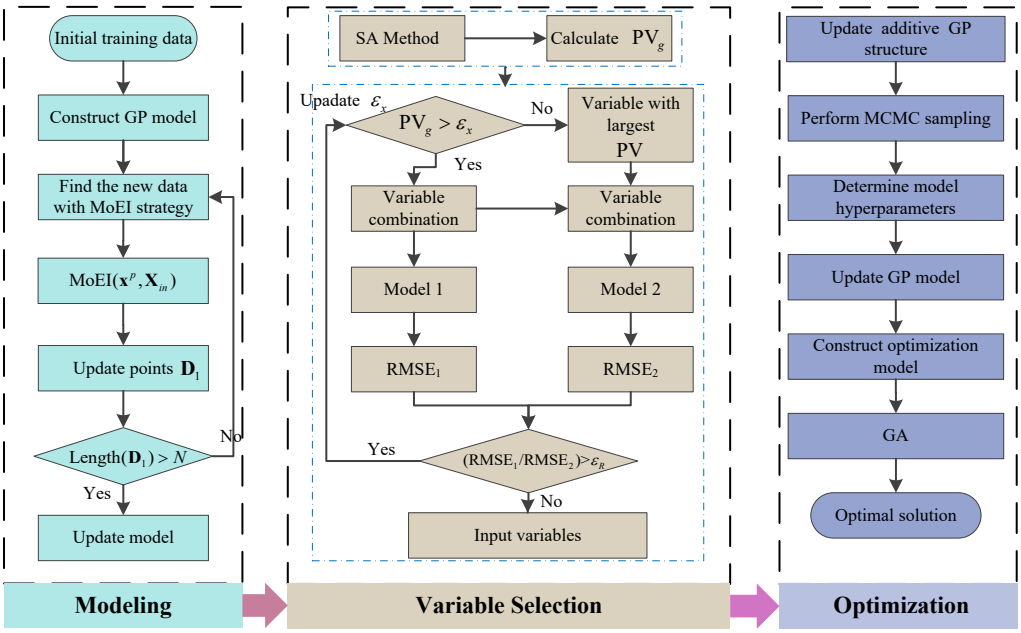


Figure 1. Flowchart of the proposed optimization approach.

In Phase I, the LHS technique is adopted to generate the initial experimental data for constructing the GP model. The adaptive experimental design is then applied using the MoEI algorithm to generate additional data and update the GP model. The detailed process is described in Algorithm 1.

In Phase II, the significant variables are identified using the adaptive variable selection technique based on the SA technique. The GP model is subsequently updated with the selected variables, as outlined in Algorithm 2.

In Phase III, the robust optimization model is constructed using the quality loss function. Then, the 'GA' algorithm is employed to find the optimal variable settings identified in Phase II within the feasible region.

4. Case Study

4.1. Numerical Example

This section illustrates the effectiveness of the proposed method with a numerical simulation example introduced by Friedman [41]. To highlight the advantages of the proposed variable selection method, the test function has been modified by incorporating an unimportant term, $0.0001x_6^2$. The modified test function is given by

$$y = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5^2 + 0.0001x_6^2 + x_4 x_5, \tag{32}$$

where $x_i \in [0,1], i = 1,2,3,4,5$. It is important to note that x_6^2 is the insignificant variable in Equation (32). The variable selection process considers both first-order and second-order effects: $x_1, x_2, x_3, x_4, x_5, x_6, x_1^2, x_2^2, x_3^2, x_4^2, x_5^2, x_6^2,\ x_1x_2, x_1x_3, x_1x_4, x_1x_5, x_1x_6, x_2x_3, x_2x_4, x_2x_5,$

$x_2x_6, x_3x_4, x_3x_5, x_3x_6, x_4x_5, x_4x_6, x_5x_6$. The proposed variable selection method is then employed to identify significant variables among these 27 variables.

4.1.1. Adaptive Modeling

In this example, a total of 70 training points are utilized, with 54 initial training points generated using the LHS technique based on Equation (32). The remaining 16 design points are selected through the adaptive learning design process outlined in Algorithm 1. Subsequently, the GP model is constructed using the 70 design points. The RMSE and mean absolute error (MAE) are used to measure the prediction accuracy. To illustrate the effectiveness of the proposed method (PM), a comparative analysis is conducted using the CoEI approach, which performs robust optimization and GP modeling via the widely adopted EI technique [26]. To ensure the robustness of the comparison, the adaptive modeling process is repeated 20 times, regenerating both the initial 54 training points and the 16 additional points in each iteration. The RMSE and AE results for both methods are presented in Figure 2.

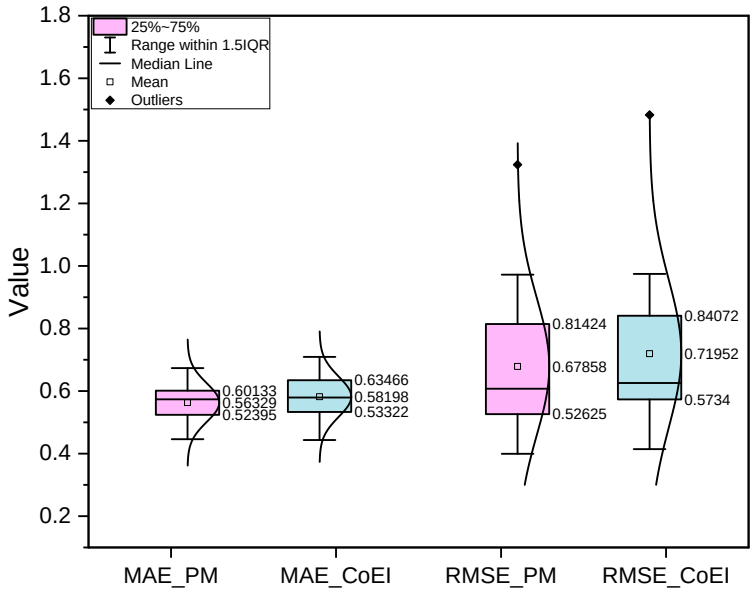


Figure 2. The MAE and RMSE of the two methods.

As shown in Figure 2, the PM method achieves mean MAE and RMSE values of 0.56329 and 0.58198, respectively, which are lower than those obtained by the CoEI method (0.58198 and 0.71952, respectively). The PM method develops the MoEI design strategy to modify the adaptive design strategy with the standard EI strategy, as given in Equation (22). This adjustment enables a more balanced exploration and exploitation of the function extrema and the overall response surface, resulting in superior prediction accuracy compared to the CoEI method.

4.1.2. Adaptive Variable Selection

In this section, variable selection is performed using the GP model developed in Section 4.1.1. The proposed adaptive variable selection method identifies significant variables based on the SA technique, as shown in Algorithm 2 and Figure 1. The selection results are given in Table 1, where the top 10 significant variables, ranked by relevance, are presented for analysis.

Table 1. Variable selection results of the two methods.

Method	Type	Variable Ranking									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
PM	Input	x_1x_2	x_3	x_3^2	x_4^2	x_5^2	x_4	x_5	x_2x_4	x_2	x_1^2
	RMSE	/	/	/	1.674	0.369	0.165	0.16	0.17	0.171	0.17
	ε_R	/	/	/	0.22	0.447	0.970	1.063	1.006	0.994	/
CoEI	Input	x_1x_2	x_3	x_3^2	x_4	x_4^2	x_5^2	x_5	x_2	x_1^2	x_2x_4
	RMSE	/	/	/	1.703	1.667	0.191	0.184	0.19	0.188	0.19
	ε_R	/	/	/	0.979	0.115	0.96	1.033	0.989	1.011	/

The ranking of variables in Table 1 is based on the SA technique introduced in Section 3.2. Input variables with a relevance greater than 0.02 ($PV_g > 0.02$) were selected to construct the initial GP model. Subsequently, variables were adaptively and progressively selected based on their contribution to the RMSE. If adding a new variable results in insufficient improvement for the RMSE (i.e., the value of $RMSE_2/RMSE_1$ exceeds the stopping criterion), the new variable will not be selected. The adaptive variable selection process, described in Algorithm 2, is then performed with a stopping criterion set to $\varepsilon_R = 0.95$ ($RMSE_2/RMSE_1 > 0.95$). This criterion implies that the variable selection process is halted if adding a variable does not significantly enhance the model's predictive accuracy. Conversely, if $RMSE_2/RMSE_1 \leq 0.95$, the selection process continues. As shown in Table 1, the stop condition of the PM method is met after selecting the sixth variable, with an RMSE of 0.165 and a convergence function value of 0.970, which exceeds the stopping criterion of 0.95. Similarly, the CoEI method stops after selecting the fourth variable, yielding an RMSE of 0.191 and a convergence function value of 0.979. Consequently, the significant variables of the PM method are x_1x_2, x_3, x_3^2, x_4^2, x_5^2 and x_4, while those selected by the CoEI method are x_1x_2, x_3, x_3^2, and x_4. As shown in Table 1, the variable selection results of the PM and CoEI methods exhibit notable differences. Furthermore, the results derived from the PM method demonstrate a closer alignment with the true importance of the input variables in Equation (32). This is because the PM method adopted the MoEI technique to improve the accuracy of the response surface, which leads to a more accurate variable selection process. The selected variables are then utilized to construct the GP response surface models. The 'GA' algorithm of R software is adopted to minimize Equation (32), with the target T set to 0. The tuning parameters are set as follows: popSize = 200, maxiter = 200, run = 100. The optimization results for both methods are provided in Table 2.

Table 2. Comparison results of the two methods.

Method	Type				GA
	$RMSE_b$	MAE_b	$RMSE_a$	MAE_a	
PM	0.237	0.189	0.165	0.130	0.054
CoEI	0.310	0.239	0.191	0.148	0.096

As shown in Table 2, the PM method achieves a smaller objective function value (GA = 0.054) compared to the CoEI method (GA = 0.096). By incorporating the proposed adaptive experimental design strategy, the PM method considers both the prediction accuracy of the entire response surface and the function extrema. As a result, the PM method finds a more precise minimum value than the CoEI method. In contrast, the CoEI method tends to update points primarily around the function extremum, potentially compromising the global accuracy of the surface model. This reduced global accuracy may affect variable selection precision, leading to overestimated optimal solutions. In terms of model prediction accuracy, the $RMSE_b$ (0.237) and AE_b (0.189) of the PM method before

variable selection are smaller than those of the CoEI method (0.310 and 0.239, respectively). Moreover, the $RMSE_a$ (0.165) and AE_a (0.130) of the PM method after variable selection are also smaller than those of the CoEI method (0.191 and 0.148, respectively). The PM method improved the utilization efficiency of design points using the MoEI technique, thereby enhancing the predictive accuracy and resulting in a more precise response surface model than the CoEI method before variable selection. Then, the PM method identified more accurate significant input variables, as shown in Table 1, which further strengthened its advantage in predictive accuracy after variable selection. These findings underscore the superior predictive accuracy of the GP model developed using the PM method, both for extrema and the entire response surface. Compared with the CoEI method, the PM method offers a more effective response surface model for robust optimization in practice, thereby enhancing the robustness and reliability of the optimization results.

4.1.3. Optimization

This section further investigates the robust optimization process by repeating it 20 times to comprehensively compare the performance of the two methods. Based on the variable selection results in Section 4.1.2, the insignificant input variable x_6 (x_5 and x_6 for the CoEI method) is fixed at 0.5 during the optimization process without loss of generality. The 'GA' algorithm is employed for robust optimization, with the results of the different methods presented in Table 3 and Figure 3.

Table 3. The optimization results of the two methods.

No.	Method	Input						
		x_1	x_2	x_3	x_4	x_5	x_6	y
1	CoEI	0.094	0.013	0.491	0.002	0.5	0.5	1.315
	PM	0.02	0.019	0.503	0.017	0.017	0.5	0.203
2	CoEI	0.055	0.112	0.497	0.015	0.05	0.5	0.36
	PM	0.085	0.013	0.503	0.002	0.068	0.5	0.078
3	CoEI	0.05	0.017	0.504	0.006	0.5	0.5	0.138
	PM	0.045	0.015	0.497	0.005	0.063	0.5	0.096
4	CoEI	0.082	0.027	0.498	0.002	0.5	0.5	0.096
	PM	0.086	0.01	0.501	0.005	0.05	0.5	0.093
5	CoEI	0.021	0.04	0.498	0.013	0.05	0.5	0.172
	PM	0.039	0.042	0.503	0.009	0.048	0.5	0.149
6	CoEI	0.009	0.128	0.494	0.024	0.5	0.5	0.303
	PM	0.043	0.047	0.491	0.01	0.046	0.5	0.177
7	CoEI	0.021	0.046	0.509	0.011	0.5	0.5	0.192
	PM	0.03	0.007	0.505	0.007	0.021	0.5	0.077
8	CoEI	0.016	0.07	0.498	0.016	0.5	0.5	0.23
	PM	0.001	0.039	0.498	0.009	0.062	0.5	0.113
9	CoEI	0.029	0.036	0.521	0.01	0.5	0.5	0.142
	PM	0.015	0.033	0.5	0.009	0.017	0.5	0.108
10	CoEI	0.025	0.113	0.492	0.006	0.5	0.5	0.183
	PM	0.022	0.019	0.498	0.012	0.034	0.5	0.136
11	CoEI	0.051	0.053	0.508	0.007	0.5	0.5	0.179
	PM	0.006	0.05	0.5	0.011	0.035	0.5	0.121
12	CoEI	0.055	0.039	0.511	0.01	0.055	0.5	0.183
	PM	0.028	0.021	0.505	0008	0.037	0.5	0.102
13	CoEI	0.006	0.061	0.49	0.011	0.5	0.5	0.161
	PM	0.021	0.07	0.495	0.001	0.07	0.5	0.079
14	CoEI	0.01	0.02	0.487	0.014	0.032	0.5	0.152
	PM	0.032	0.025	0.518	0.01	0.048	0.5	0.141
15	CoEI	0.033	0.085	0.501	0.003	0.5	0.5	0.132
	PM	0.169	0.004	0.492	0.004	0.091	0.5	0.102

Table 3. *Cont.*

No.	Method	Input						y
		x_1	x_2	x_3	x_4	x_5	x_6	
16	CoEI	0.024	0.071	0.507	0.008	0.5	0.5	0.133
	PM	0.01	0.051	0.515	0.003	0.073	0.5	0.076
17	CoEI	0.028	0.064	0.505	0.011	0.5	0.5	0.197
	PM	0.075	0.029	0.474	0.001	0.028	0.5	0.094
18	CoEI	0.015	0.077	0.498	0.009	0.5	0.5	0.146
	PM	0.045	0.044	0.511	0.002	0.067	0.5	0.111
19	CoEI	0.023	0.061	0.51	0.014	0.5	0.5	0.188
	PM	0.012	0.154	0.524	0.003	0.019	0.5	0.103
20	CoEI	0.055	0.039	0.511	0.01	0.5	0.5	0.183
	PM	0.028	0.021	0.505	0.008	0.037	0.5	0.102

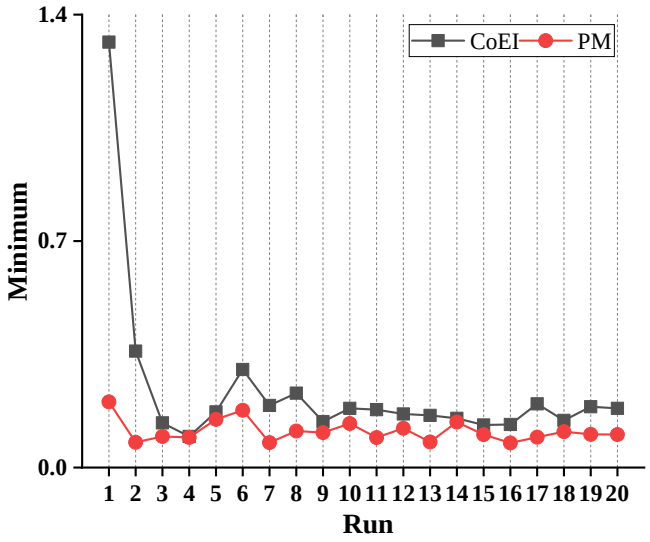


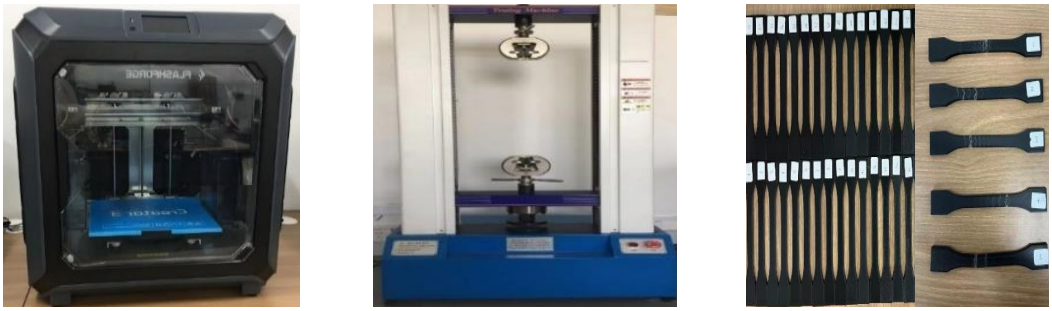
Figure 3. The minimum values of the two methods.

As illustrated in Table 3 and Figure 3, the PM method achieves relatively smaller GA values within the feasible region compared to the CoEI method. The CoEI method tends to excessively concentrate updated points around these extremes, which can diminish its global prediction accuracy and potentially result in overestimated optimal solutions. In contrast, the PM method balances global accuracy and extreme value consideration, thereby enhancing the efficiency of experimental design in response surface modeling.

4.2. Additive Manufacturing Case Study

In this section, a real-world additive manufacturing (or 3D printing) case study is presented to demonstrate the effectiveness of the proposed method. The 3D printer used in this study is Flashforge Creator 3, as shown in Figure 4a. The tensile testing machine utilized is WDW-4204, as shown in Figure 4b, and the experimental products are depicted in Figure 4c. The response of interest is the product deformation y (mm). The objective is to maximize the response y under a fixed force, as it represents a "larger-the-better" characteristic. The study considers three input factors: infill x_1 (10%∼30%), layer height x_2 (0.1 mm∼0.3 mm), and nozzle temperature x_3 (205 °C∼225 °C). The coded levels of these inputs are provided in Table 4. The goal of robust parameter optimization in this case study is to determine a set of input settings within the feasible region that maximizes the response y to satisfy the specified quality requirements. Both main effects and second-order interactions (e.g., $x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$) are considered during the variable selection process. The Latin Hypercube Sampling (LHS) technique is employed to generate

the 30 training and 30 testing points. The GP model is then constructed using the adaptive modeling process described in Algorithm 1, with the experimental data presented in Appendix A.



(**a**) (**b**) (**c**)

Figure 4. Additive manufacturing experiment platform: (**a**) 3D printer. (**b**) Tensile machine. (**c**) Experimental products.

Table 4. Input variables and code levels.

Input	Symbol	Unit	Values		
Infill	x_1	%	10%	20%	30%
Layer Height	x_2	mm	0.1	0.2	0.3
Nozzle Temperature	x_3	°C	205	215	225
Coded Levels			0	0.5	1

Then, the GP model is constructed to simulate the relationship between input variables and output response. Subsequently, variable selection is carried out using the method specified in Algorithm 2, with the results presented in Table 5. In this case study, the stop criterion is set as $\varepsilon_R^s = 1$, indicating that any variable contributing to improved prediction accuracy is utilized in constructing the response surface model. As shown in Table 5, the PM method and the CoEI methods produce similar variable rankings. The PM method reaches its stopping condition at the third variable ($\varepsilon_R = 1.1286$). At the same time, the CoEI method also meets its stopping criterion at the third variable ($\varepsilon_R = 1.1095$). Consequently, the significant variables selected by the PM method are x_2, x_2x_3, and x_2^2, while those selected by the CoEI method are x_2, x_2x_3, and x_1^2. Based on the results of the PM method, inputs x_2, x_2x_3, and x_2^2 are identified as having a significant impact on the response deformation. As shown in Table 5, both methods reached the stopping criteria after adding the third variable, so the importance rankings for the remaining less significant input variables (x_3, x_3^2, x_1x_2, x_1x_3) are not included in this case study. Both main effects and second-order interactions (x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3) were considered in the sensitivity analysis. The proposed method identified the variables x_1 (infill), x_1^2, x_1x_2, and x_1x_3 as insignificant, implying that the effect of x_1 (infill in the 10–30% range) on the output response (product deformation) can be neglected. Other infill-related input combinations were also considered insignificant. Notably, the variable selection results differ between the two methods due to their distinct modeling strategies. This finding enhances the interpretability of the GP model, particularly in capturing the interaction and quadratic effects on the output.

Table 5. Variable selection results of the two methods.

Method	Type	Input Variable Importance Ranking			
		1	**2**	**3**	**4**
PM	Input	x_2	x_2x_3	x_2^2	x_1
	ε_R	/	0.7875	1.1286	1.4344
CoEI	Input	x_2	x_2x_3	x_1^2	x_2^2
	ε_R	/	0.9236	1.1095	1.0083

After obtaining the variable selection results, the GP model is updated using the selected input variables. The model hyperparameters are then estimated via the MCMC method. Trace plots and density plots of parameters (e.g., α_1, l_2, and σ) are employed to verify convergence. As shown in Figure 5, all chains from the sampling process eventually conform to the same distribution, demonstrating the effectiveness and convergence of the parameter estimation process. The quality loss function introduced in Section 3.3 is then constructed for robust optimization to achieve the optimal parameter settings:

$$Q(\mathbf{x}) = \left(\frac{1}{y(\mathbf{x})} - T\right)^2 + \sigma^2, \tag{33}$$

where T is the target and is set to 0. σ^2 is the predictive variance. The 'GA' algorithm is adopted to minimize Equation (33) to obtain the optimal solution. The optimization results of different methods are given in Table 6.

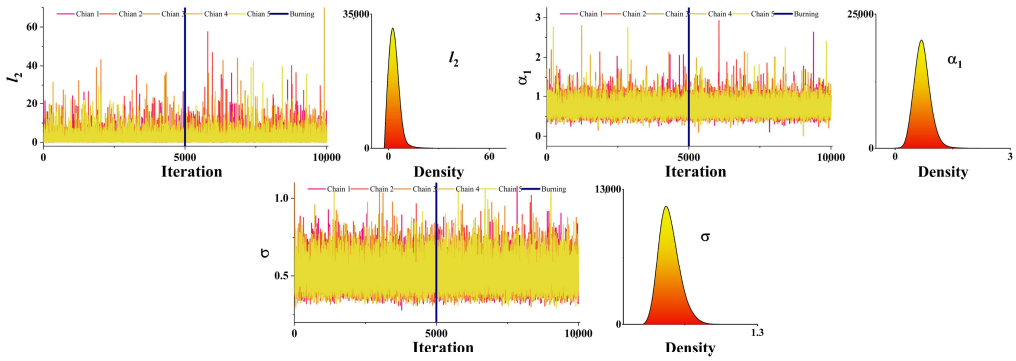


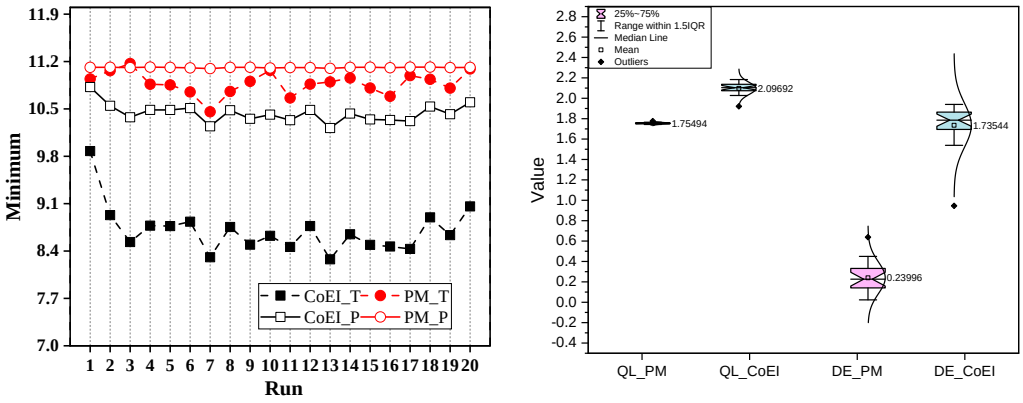
Figure 5. Convergence diagnostic of model hyperparameters (e.g., α_1, l_2, and σ).

Table 6. The optimization results of the two methods.

Method	Input			Response		SD	QL
	x_1	x_2	x_3	True	Prediction		
CoEI	0.7396	0.8626	0.9409	9.88	10.82	1.38	1.92
PM	0.5	0.8732	0.9864	10.94	11.11	1.32	1.75

As shown in Tables 5 and 6, the significant variables selected by the PM method do not include x_1. Consequently, x_1 is maintained at an average level of 0.5 during the robust optimization process for the PM method. In Table 6, the true response of the PM method (10.94) and the predictive response (11.11) are both larger than those of the CoEI method (9.88 and 10.82, respectively), indicating that the proposed method more effectively satisfies the quality characteristic requirements. Furthermore, the predictive standard deviation (SD) of the PM method is 1.32, smaller than that of the CoEI method (1.38), suggesting that the proposed method has smaller response fluctuations and higher predictive accuracy. The quality loss (QL) of the PM method is 1.75, also smaller than that of the CoEI method

(1.92), indicating that the optimal solution obtained by the PM method exhibits superior overall performance compared to the CoEI method. As shown in Tables 5 and 6, the proposed adaptive variable selection method accurately identifies the input variables that significantly impact the output response (product deformation). Within the MoEI modeling framework, the proposed method constructs a more robust response surface model using the identified significant input variables. Therefore, the proposed approach not only enhances model interpretability but also improves the prediction accuracy of the response surface model. Compared to the CoEI method, the proposed method achieves smaller prediction standard deviations and reduced quality loss, especially for larger true responses, which better meets the product quality requirements. To further evaluate the performance of the two methods, the optimization process described above was repeated 20 times. The predictive responses (PM_P and CoEI_P), true responses (PM_T and CoEI_T), quality loss (QL_PM and QL_CoEI), and the deviation between predictions and true responses (DE_PM and DE_CoEI) for these iterations are presented in Figure 6 and Table 7.



(**a**) (**b**)

Figure 6. The 20 optimization results of the two methods: (**a**) Predictive and true responses. (**b**) QL and DE.

Table 7. The quality loss of the two methods.

Method	QL			
	Minimum	**Mean**	**Median**	**Maximum**
CoEI	1.921094	2.096921	2.103658	2.183544
PM	1.743114	1.754944	1.754057	1.776462

As illustrated in Figure 6a, the true responses (PM_T) and predictive responses (PM_P) from the PM method consistently outperform those from the CoEI method (CoEI_T and CoEI_P, respectively). This outcome suggests that the PM method achieves superior optimal values and more effectively meets the quality characteristics requirements compared to the CoEI method. Figure 6b further illustrates that the mean quality loss obtained via the PM method is significantly smaller than that of the CoEI method, indicating more robust optimization results. Additionally, the deviation between prediction and true value (DE_PM) from the PM method is smaller than that from the CoEI method (DE_CoEI), demonstrating higher prediction accuracy of the PM method. Table 7 summarizes the minimum, mean, median, and maximum values of the quality loss across the 20 optimization iterations. All these metrics for the PM method are smaller than those for the CoEI method, reinforcing the superior robustness of the proposed method. Moreover, the difference between the maximum and minimum values in Table 7 for the PM method is 0.033, significantly smaller

than the 0.262 observed for the CoEI method, further underscoring the enhanced robustness of the PM method.

In this case study, the proposed adaptive variable selection strategy effectively identifies the significant variables as x_2, x_2x_3, and x_2^2, while excluding x_1. This exclusion suggests that variations in x_1 within the 10–30% range have minimal effect on the response y. Consequently, the proposed variable selection method enhances the interpretability of the GP model by identifying the significant factors. Furthermore, the proposed MoEI-based GP modeling approach accounts for prediction accuracy at both local extremum and global mapping levels, thereby reducing model uncertainty and yielding more robust optimal input parameter settings. As a result, the proposed adaptive robust optimization method offers engineering practitioners a novel approach to tackling quality improvement challenges in additive manufacturing.

5. Conclusions

In this article, a multi-stage GP-based symmetric modeling approach is developed to address the experimental design, variable selection, and robust optimization problems in quality improvement. The proposed method improves the GP response surface predictive accuracy and effectively identifies significant input variables. The developed multi-stage model achieves a more robust optimal solution with minimal quality loss and enhances computational efficiency. Furthermore, the impact of uncertainty on the optimization results is assessed by the quality loss function, which further enhances the robustness and reliability of the optimal solution.

The effectiveness of the proposed method was demonstrated through numerical simulations and additive manufacturing experiments. The proposed approach improves the model prediction accuracy achieved by the modified EI-based variable selection techniques. Comparative results in Table 2 show an average improvement rate of 22.8%. Meanwhile, the proposed robust optimization approach combined with the MCMC technique can enhance the effectiveness of the optimal input settings. Comparative results in Table 7 show an average improvement rate of 15.21%. The proposed GP-based robust optimization approach demonstrates broad applicability, particularly in addressing parameter optimization challenges in scenarios with limited experimental data and high-dimensional input spaces. This method is especially advantageous for production practices involving high experimental costs and long cycles, making it an excellent choice for applications such as parameter optimization in aerospace (e.g., aircraft, rockets, satellites), nuclear fuel development, and reactor design.

The proposed model may have limitations which can serve as fruitful future research areas. First, engineers may need to deal with robust optimization problems for nonstationary responses in real-life situations. GP models need to be expanded into nonstationary responses using Bayesian treed modeling techniques. Second, the GA used in the proposed optimization model has some limitations, such as premature convergence, high computational cost, and parameter sensitivity, which may lead to instability in the optimization process. Other search techniques (e.g., fuzzy evolutionary algorithms [42] and wavelet-oriented evolutionary algorithms [43]) could further improve the robustness of the proposed method. Incorporating such ideas into adaptive learning criteria may further improve the predictive accuracy of the GP response surface model.

Author Contributions: Z.F.: Writing—original draft, Methodology, Visualization; X.Y.: Software; W.H.: Writing—review and editing, Supervision; C.Z.: Investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China [grant numbers 72301146]; the China Postdoctoral Science Foundation [grant number 2023M741802, 2024T170430]; the China Statistical Science Research Foundation [grant number 2023LZ010]; Science and Technology Research Project of Jiangxi Provincial Department of Education [grant number GJJ2401014]; the Natural Science Foundation of the Jiangsu Higher Education Institutions of China [grant number 23KJB630012].

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, [Z. Feng], upon reasonable request.

Conflicts of Interest: The authors have no conflicts of interest to declare that are relevant to the content of this article.

Appendix A

As shown in Table A1, the total number of design points is 44, where runs 1–30 are initial training points, and runs 31–44 are new points generated using the adaptive design method in Algorithm 1.

Table A1. Experimental data.

No.	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3	y
1	0.025	0.466	0.582	0.001	0.218	0.339	0.012	0.015	0.272	5.253
2	0.136	0.420	0.684	0.019	0.176	0.468	0.057	0.093	0.287	5.351
3	0.089	0.305	0.173	0.008	0.093	0.030	0.027	0.015	0.053	4.390
4	0.111	0.814	0.732	0.012	0.662	0.536	0.090	0.081	0.595	7.921
5	0.052	0.214	0.119	0.003	0.046	0.014	0.011	0.006	0.026	4.570
6	0.607	0.652	0.859	0.368	0.425	0.738	0.396	0.521	0.560	5.056
7	0.771	0.592	0.269	0.595	0.351	0.072	0.457	0.207	0.159	5.840
8	0.890	0.501	0.008	0.792	0.251	0.000	0.446	0.007	0.004	6.581
9	0.461	0.069	0.543	0.213	0.005	0.295	0.032	0.250	0.037	5.907
10	0.653	0.495	0.918	0.427	0.245	0.842	0.323	0.599	0.454	5.881
11	0.940	0.368	0.816	0.884	0.135	0.665	0.346	0.767	0.300	6.302
12	0.574	0.916	0.392	0.330	0.839	0.154	0.526	0.225	0.359	5.543
13	0.977	0.984	0.989	0.954	0.969	0.978	0.962	0.966	0.973	6.897
14	0.672	0.271	0.742	0.451	0.074	0.551	0.182	0.498	0.201	5.332
15	0.548	0.128	0.513	0.300	0.016	0.263	0.070	0.281	0.066	6.721
16	0.832	0.861	0.657	0.692	0.741	0.432	0.716	0.547	0.566	8.682
17	0.194	0.152	0.481	0.037	0.023	0.231	0.029	0.093	0.073	7.029
18	0.313	0.033	0.333	0.098	0.001	0.111	0.010	0.104	0.011	8.702
19	0.759	0.338	0.769	0.576	0.114	0.592	0.256	0.584	0.260	6.193
20	0.382	0.632	0.239	0.146	0.400	0.057	0.241	0.091	0.151	6.027
21	0.509	0.758	0.419	0.260	0.575	0.175	0.386	0.213	0.318	5.774
22	0.422	0.677	0.877	0.178	0.458	0.769	0.285	0.370	0.593	6.336
23	0.274	0.254	0.063	0.075	0.065	0.004	0.070	0.017	0.016	4.248
24	0.493	0.784	0.447	0.243	0.615	0.199	0.387	0.220	0.350	7.279
25	0.911	0.175	0.074	0.830	0.031	0.005	0.160	0.067	0.013	4.247
26	0.347	0.540	0.164	0.120	0.292	0.027	0.187	0.057	0.089	4.393
27	0.847	0.702	0.618	0.717	0.492	0.381	0.594	0.523	0.433	6.394
28	0.243	0.948	0.365	0.059	0.899	0.133	0.231	0.089	0.346	6.290
29	0.713	0.877	0.934	0.508	0.768	0.872	0.625	0.665	0.818	12.436
30	0.205	0.058	0.215	0.042	0.003	0.046	0.012	0.044	0.013	6.796
31	0	0.938	0.920	0	0.880	0.846	0	0	0.863	4.526
32	0.999	0.971	0	0.999	0.943	0	0.971	0	0	8.504
33	0.979	0.002	1	0.959	0	1	0.001	0.979	0.002	7.471
34	0.012	0.016	1	0	0	1	0	0.012	0.016	7.499
35	0.647	0.880	1	0.418	0.775	1	0.569	0.647	0.880	12.316
36	0.262	0.877	1	0.069	0.769	1	0.230	0.262	0.877	9.533
37	0.787	0.879	1	0.619	0.773	1	0.692	0.787	0.879	12.054
38	0.668	0.878	0.591	0.447	0.770	0.349	0.587	0.395	0.519	10.681

Table A1. *Cont.*

No.	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3	y
39	0.423	0.004	0.181	0.179	0	0.033	0.002	0.077	0.001	8.860
40	0.859	0.023	0.046	0.738	0.001	0.002	0.020	0.039	0.001	10.451
41	0.515	0.995	0	0.266	0.991	0	0.513	0	0	7.404
42	1	0.029	0.454	1	0.001	0.206	0.029	0.454	0.013	9.970
43	0.664	0.875	1	0.441	0.766	1	0.581	0.664	0.875	12.059
44	0.677	0.881	0.888	0.459	0.776	0.789	0.597	0.602	0.782	12.644

References

1. Wu, C.F.J.; Hamada, M.S. *Experiments: Planning, Analysis and Parameter Design Optimization*; John Wiley & Sons: New York, NY, USA, 2009.
2. Costa, N.; Lourenço, J.; Pereira, Z. Responses modeling and optimization criteria impact on the optimization of multiple quality characteristics. *Comput. Ind. Eng.* **2012**, *62*, 927–935. [CrossRef]
3. Huang, Q. An impulse response formulation for small-sample learning and control of additive manufacturing quality. *IISE Trans.* **2022**, *55*, 926–939. [CrossRef]
4. Wang, J.; Ma, Y.; Ouyang, L.; Tu, Y. A new Bayesian approach to multi-response surface optimization integrating loss function with posterior probability, *Eur. J. Oper. Res.* **2016**, *249*, 231–237. [CrossRef]
5. Ouyang, L.; Ma, Y.; Wang, J.; Tu, Y. A new loss function for multi-response optimization with model parameter uncertainty and implementation errors. *Eur. J. Oper. Res.* **2017**, *258*, 552–563. [CrossRef]
6. Myers, R.H.; Montgomery, D.C.; Vining, G.G.; Borror, C.M.; Kowalski, S.M. Response surface methodology: A retrospective and literature survey. *J. Qual. Technol.* **2004**, *36*, 53–77. [CrossRef]
7. Grbić, R.; Kurtagić, D.; Slišković, D. Stream water temperature prediction based on Gaussian process regression. *Expert Syst. Appl.* **2013**, *40*, 7407–7414. [CrossRef]
8. Fan, S.-K.S.; Zahara, E. A hybrid simplex search and particle swarm optimization for unconstrained optimization. *Eur. J. Oper. Res.* **2007**, *181*, 527–548. [CrossRef]
9. Ouyang, L.; Zhu, S.; Ye, K.; Park, C.; Wang, M. Robust Bayesian hierarchical modeling and inference using scale mixtures of normal distributions. *IISE Trans.* **2022**, *54*, 659–671. [CrossRef]
10. He, Z.; Zhu, P. F.; Park, S.H. A robust desirability function method for multi-response surface optimization considering model uncertainty. *Eur. J. Oper. Res.* **2012**, *221*, 241–247. [CrossRef]
11. Kleijnen, J.P.C. Regression and Kriging metamodels with their experimental designs in simulation: A review. *Eur. J. Oper. Res.* **2017**, *256*, 1–16. [CrossRef]
12. Feng, Z.; Wang, J.; Zhou, X.; Zhai, C.; Ma, Y. Robust optimization for functional multiresponse in 3D printing process. *Simul. Model. Pract. Theory* **2023**, *126*, 1–13. [CrossRef]
13. Sun, L.; Hong, L.J.; Hu, Z. Balancing exploitation and exploration in discrete optimization via simulation through a Gaussian process-based search. *Oper. Res.* **2014**, *62*, 1416–1438. [CrossRef]
14. Costa N.; Lourenço, J. Gaussian process model-an exploratory study in the response surface methodology. *Qual. Reliab. Eng. Int.* **2016**, *32*, 2367–2380. [CrossRef]
15. Parnianifard, A.; Azfanizam, A.; Ariffin, M.; Ismail, M. Crossing weighted uncertainty scenarios assisted distribution-free metamodel-based robust simulation optimization. *Eng. Comput.* **2019**, *36*, 139–150. [CrossRef]
16. Wang, B.; Xu, A. Gaussian process methods for nonparametric functional regression with mixed predictors. *Comput. Stat. Data Anal.* **2019**, *131*, 80–90. [CrossRef]
17. Feng, Z.; Wang, J.; Ma, Y.; Tu, Y. Robust parameter design based on Gaussian process with model uncertainty. *Int. J. Prod. Res.* **2021**, *59*, 2772–2788. [CrossRef]
18. Ghosh, M.; Li, Y.; Zeng, L.; Zhang, Z.; Zhou, Q. Modeling multivariate profiles using gaussian process-controlled b-splines. *IISE Trans.* **2020**, *53*, 787–798. [CrossRef]
19. Shen, B.; Gnanasambandam, R.; Wang, R.; Kong, Z.J. Multi-task Gaussian process upper confidence bound for hyperparameter tuning and its application for simulation studies of additive manufacturing. *IISE Trans.* **2022**, *55*, 496–508. [CrossRef]
20. Ouyang, L.; Han, M.; Ma, Y.; Wang, M.; Park, C. Simulation optimization using stochastic kriging with robust statistics. *J. Oper. Soc.* **2023**, *74*, 623–6360. [CrossRef]
21. Lee U. A new adaptive Kriging-based optimization (AKBO) framework for constrained optimization problems: A case study on shared autonomous electric vehicle system design. *Expert Syst. Appl.* **2024**, *252*, 124147. [CrossRef]
22. Han, J.; Zheng, Y.; Wang, K.; Yang, C.; Yuan, X. Worst-case robust optimization based on an adaptive incremental Kriging metamodel. *Expert Syst. Appl.* **2025**, *260*, 125372. [CrossRef]

23. Ozdemir, A.; Cho, B.R. Response surface-based robust parameter design optimization with both qualitative and quantitative variables. *Eng. Optim.* **2017**, *49*, 1796–1812. [CrossRef]
24. Joseph, V.R.; Gul, E.; Ba, S. Maximum projection designs for computer experiments. *Biometrika* **2015**, *102*, 371–380. [CrossRef]
25. Krishna, A.; Joseph, V.R.; Ba, S.; Brenneman, W.A.; Myers, W.R. Robust experimental designs for model calibration. *J. Qual. Technol.* **2021**, *54*, 441–452. [CrossRef]
26. Ur Rehman, S.; Langelaar, M. Efficient infill sampling for unconstrained robust optimization problems. *Eng. Optim.* **2015**, *48*, 1313–1332. [CrossRef]
27. Chen, J.H.; Kang, L.L.; Lin, G. Gaussian process assisted active learning of physical laws. *Technometrics* **2021**, *63*, 329–342. [CrossRef]
28. Santner, T.; Williams, B.; Notz, W.I. *Space-Filling Designs for Computer Experiments*; Springer: New York, NY, USA, 2018.
29. Paananen, T.; Piironen, J.; Andersen, M.R.; Vehtari, A. Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, PMLR, Naha, Japan, 16–18 April 2019; pp. 1743–1752.
30. Pavone, F.; Piironen, J.; Bürkner, P.-C.; Vehtari, A. Using reference models in variable selection. *Comput. Stat.* **2022**, *38*, 349–371. [CrossRef]
31. Linkletter, C.; Bingham, D.; Hengartner, N.; Higdon, D.; Ye, K.Q. Variable selection for Gaussian process models in computer experiments. *Technometrics* **2006**, *48*, 478–490. [CrossRef]
32. Chen, T.; Wang, B. Bayesian variable selection for Gaussian process regression: Application to chemometric calibration of spectrometers. *Neurocomputing* **2010**, *73*, 2718–2726. [CrossRef]
33. Yi, G.; Shi, J.Q.; Choi, T. Penalized Gaussian process regression and classification for high-dimensional nonlinear data. *Biometrics* **2011**, *67*, 1285–1294. [CrossRef]
34. Jiang, S.; Tokdar, S.T. Variable selection consistency of Gaussian process regression. *Ann. Stat.* **2021**, *49*, 2491–2505. [CrossRef]
35. Park, C.; Borth, D.J.; Wilson, N.S.; Hunter, C.N. Variable selection for Gaussian process regression through a sparse projection. *IISE Trans.* **2022**, *54*, 699–712.
36. Williams, C.K.I.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, UK, 2006.
37. Gelman, A.; Stern, H.; Carlin, J.; Dunson, D.; Vehtari, A.; Rubin, D. *Bayesian Data Analysis*; Chapman and Hall/CRC: New York, NY, USA, 2013.
38. Taguchi, G.; Elsayed, E.; Hsiang, T.C. *Quality Engineering in Production System*; McGraw-Hill: New York, NY, USA, 1989.
39. Tan, M.; Wu, C. Robust design optimization with quadratic loss derived from Gaussian process model. *Technometrics* **2012**, *54*, 51–63. [CrossRef]
40. Tan, M.; Wu, C. A Bayesian approach for model selection in fractionated split plot experiments with applications in robust parameter design. *Technometrics* **2013**, *55*, 359–372. [CrossRef]
41. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]
42. Naderi, E.; Mirzaei, L.; Pourakbari-Kasamaei, M.; Cerna, F.V.; Lehtonen, M. Optimization of active power dispatch considering unified power flow controller: Application of evolutionary algorithms in a fuzzy framework. *Evol. Intell.* **2023**, *17*, 1357–1387. [CrossRef]
43. Naderi, E.; Mirzaei, L.; Trimble, J.P.; Cantrell, D.A. Multi-objective optimal power flow incorporating flexible alternating current transmission systems: Application of a wavelet-oriented evolutionary algorithm. *Electr. Power Compon. Syst.* **2023**, *52*, 766–795. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.