

Received July 19, 2020, accepted August 2, 2020, date of publication August 7, 2020, date of current version August 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015043

Differential Evolution Based Manifold Gaussian Process Machine Learning for Microwave Filter's Parameter Extraction

XUEZHI CHEN^{ID}, YUBO TIAN^{ID}, TIANLIANG ZHANG, AND JING GAO^{ID}

School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China

Corresponding author: Yubo Tian (tianyubo@just.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61771225, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province, China, and in part by the Qinglan Project of Jiangsu Higher Education.

ABSTRACT Gaussian process (GP) is a rapidly developing supervised machine learning (ML) method in recent years, which has been widely used in the establishment of surrogate models in the field of electromagnetics. However, it has the problems of large sample demand, high computational complexity and low accuracy when processing high dimensional data. To solve this problem, a manifold Gaussian process (MGP) ML method based on differential evolution (DE) algorithm is proposed in this study. For the proposed method, the DE algorithm is used to get dimension reduction parameters, and the method can work very well with the optimized parameters. Compared with the traditional GP model, the dimensionality reduction method based on Isomap is adopted to simplify the mapping relationship between data pairs. Therefore, the model is more suitable for the problem of insufficient samples and high data dimension. In this study, the proposed DE-based MGP (DE-MGP) is applied to the extraction of coupling coefficients of the fourth-order and sixth-order coupling filters, in which the test error of the fourth-order coupling filter surrogate model can be reduced to 0.84%, and the test error of the sixth-order coupling filter is expected to be reduced to 1.53%, which proves that the proposed method is very effective.

INDEX TERMS Manifold Gaussian process, isometric mapping, differential evolution, parameter extraction.

I. INTRODUCTION

As an important part of wireless communication, microwave filter has a long history. As early as the 1930s, Mason and Sykes used ABCD parameters to derive a large number of useful filter image impedance phase and attenuation functions [1]. In the 1950s, for the design of synchronously tuned cascade resonant filters, the S.B.Cohn designed a direct-coupled cavity filter with transmission zero at wireless distance based on the low-pass filter prototype [2]. In the 1960s, J. D. Rhodes eliminated the redundant coupling parameters by rotating the coupling coefficient matrix of the transformation filter, making the design structure of the filter simple [3]. In the 1970s, Atia and Williams developed a comprehensive general theory that can obtain the topological structure of coupled resonance filters of less than the fourth order by

analytical method [4]. In the 1980s, Cameron developed this theory and generalized the original Chebyshev function into a generalized Chebyshev function and applied it to the synthesis of cross-coupled filters. In Cameron's theory, we can reduce the synthesis work to the calculation coupling matrix by determining the low-pass prototype filter function of the filter [5], [6]. In the 1990s, Cameron proposed a similar transformation method to eliminate the coupling coefficient matrix [7]. In the early 21st century, Amari proposed the circuit optimization method, which is not limited by the coupling topology structure of the filter and can easily obtain the coupling matrix we need [8].

In recent years, there have been many new design methods for microwave filters. In 2013, Tian Yubo *et al.* realized the optimization of the frequency characteristics of the electric band-gap structure of bow unit through particle swarm optimization algorithm [9]. In 2014, Otani H *et al.* carried out high-speed automatic optimization design of the

The associate editor coordinating the review of this manuscript and approving it for publication was Yuhao Liu.

filter through genetic algorithm, and obtained better design results [10]. In 2019, in order to provide a simple and effective solution to the complex tuning problem of microwave filter, J. C. Melgarejo *et al.* proposed the space mapping technology of microwave filter, and verified the method by experimentation of adjusting the six-pole induction waveguide filter [11].

To solve these problems, many researchers are now introducing machine learning (ML) techniques into the design of microwave devices. The surrogate model technology has gradually come into our field of vision. By establishing an surrogate model for microwave devices, we can realize its rapid simulation and optimization [12]. At the present stage, functional surrogate modeling techniques for microwave devices mainly include artificial neural network (ANN) [13], support vector machine (SVM) [14], polynomial regression (PR) [15] and kernel extreme learning machine (KELM) [16]. However, with the study of Bayesian neural network, a new training model – Gaussian process (GP) has gradually attracted people's attention [17]. This training model has many advantages such as small sample demand, few training parameters and flexible acquisition, so it is very suitable for solving high dimensional nonlinear problems. At present, GP model has been widely applied in the field of electromagnetism. From 2009 to 2010, Villiers successfully modeled the gap antenna fed by ultra wide band (UWB) dual-frequency coplanar waveguide (CPW) using GP model, which indicated that the GP could be used as an alternative to full-wave analysis in the design of microwave devices [18], [19]. In 2012, Jacobs successfully modeled the microwave filter using the GP of non-standard kernel function [20]. In 2013, Jacobs proposed a two-stage network modeling method for antenna input characteristics [21], compared the antenna input characteristics of single mode and ensemble mode with GP modeling, and the results showed that the GP model of set mode had smaller test errors [22]. In 2015, Jacobs proposed the use of GP to model the resonant frequency of dual-frequency microstrip antenna, which extended the application range of Gaussian process in the electromagnetic field [23]. In the same year, Jacobs proposed a second-stage network modeling method for GP, which was successfully applied to the design of microwave filters [24]. In 2016, Vargas Cardona *et al.* proposed a multi-output GP to enhance diffusion-tensor field resolution [25]. In 2018, Chen Yi *et al.* used GP regression to obtain prior knowledge, and then used knowledge-based neural network (KBNN) to model electromagnetic problems. This method greatly reduced the calculation time and improved the computational efficiency of modeling [26]. In 2019, Xiao-Hong Fan *et al.* proposed the design optimization of GP fast microwave antenna based on particle swarm optimization algorithm, and successfully optimized the substitute model as the fitness function of intelligent optimization algorithm and obtained the design parameters we need [27]. In the same year, Xiao-Hong Fan modeled and optimized the cone-core horn antenna through a coarse grid GP [28]. In 2020, Jing Gao *et al.* proposed a semisupervised learning GP, which combines unlabeled samples to improve

the accuracy of the GP model and reduce the number of labeled training samples required [29].

Thanks to the unremitting efforts of our predecessors, we are also deeply inspired when we do relevant research. In order to further solve the problems of insufficient samples, high dimensions and high computing cost in electromagnetic optimization design, this paper proposes a differential evolution (DE) based manifold GP (DE-MGP) algorithm, and applies it to the inverse model parameter extraction of coupled microwave filters [30]. Compared with the traditional GP model, the DE-MGP has better adaptability to high-dimensional data in the case of the same amount of data. The dimensionality reduction operation is carried out on the training data, i.e. the spatial complexity of the data is reduced, which reduces the dependence of the model on the amount of data. At the same time, it also greatly reduces the calculation amount of the model operation. The main contribution details of this paper are as follows:

- 1) a differential evolution based manifold Gaussian process (DE-MGP) is established. Firstly, different neighborhoods are divided in the data space, and the geodesic distance between sample points in each neighborhood is calculated by the shortest path algorithm [30]. Secondly, multi-dimensional scaling (MDS) [32] is used to obtain the projection of the training data in the low-dimensional space, so as to obtain the isometric mapping (Isomap) of the training data we need in the low-dimensional space. Then the processed data are used as input of the GP to calculate the test error. Finally, the error is taken as the fitness function of the DE algorithm, and the dimensionality reduction parameters in the Isomap are optimized. This means that DE algorithm is used to train dimensionality reduction parameters. After continuous iteration, when the optimization algorithm converges, we can get a set of optimal dimensionality reduction parameters. By applying this set of optimal parameters in the Isomap, our training model can be constructed. Compared with simply combining manifold learning and GP, the advantage of this method is that it can train not only GP model but also manifold parameters through DE algorithm, which we can greatly improve its training efficiency and generalization ability.
- 2) In order to verify the training effect of our proposed algorithm and further improve our theory, we use the proposed DE-MGP algorithm to extract the inverse parameters of a fourth-order coupled filter and a sixth-order coupled filter. In the experiments, we take the S-parameter of the filter as the system input and the coupling coefficient matrix as the system output to establish the model. In the above two filters, we compare the parameter extraction effects of the normal GP model and the DE-MGP model under the conditions of 300 training samples and 600 training samples, respectively. It is found that the DE-MGP has better fitting ability for the two filters under the same

number of samples, and its test error is much smaller than the ordinary GP model, which also shows that the DE-MGP has better adaptability for high-dimensional data, and its training accuracy depends less on the number of training samples. These characteristics are very meaningful for us to deal with the problems in the field of electromagnetics, which have the disadvantages of large amount of calculation, insufficient samples and high cost of obtaining samples.

The rest of this paper is organized as follows. In section two, we will introduce some of the algorithmic theories used in our study. In the third part, we will mainly introduce the training model we proposed in this paper. In the fourth part, we will conduct experiments with two coupled filters in different number of samples, and illustrate the advantages of the proposed model by comparing it with the initial GP model and four classical models. The last part is the summary of this research and the prospect of future work.

II. RELATED WORKS

In this section, we will introduce some related theories, including differential evolution algorithm, Gaussian process and manifold learning theory.

A. DIFFERENTIAL EVOLUTION (DE) ALGORITHM

DE is a kind of intelligent optimization algorithm which can perform in the way of parallel search. It uses floating-point vector encoding. The optimal fitness value in solution space is searched through the continuous evolution of the population. In the non-linear and non-differentiable continuous space, the DE algorithm has a good ability of optimization.

In each iteration of DE, the population needs to go through three steps: variation, crossover and selection. The individual x_i in the initial population $\{x_1, \dots, x_n\}$ will continue to evolve towards the optimal solution, and each individual searches in the m dimensional space. This process can be described by the following formulas.

Initialization:

$$x_{ij} = x_{ij}^L + rand(0, 1)(x_{ij}^U - x_{ij}^L) \quad (1)$$

Variation:

$$v_i = x_{r1} + F(x_{r2} - x_{r3}) \quad (2)$$

Crossover:

$$u_{ij}(g) = \begin{cases} v_{ij}(g), & rand(0, 1) \leq CR \text{ or } j = j_{rand} \\ x_{ij}(g), & \text{others} \end{cases} \quad (3)$$

Selection:

$$x_i(g+1) = \begin{cases} u_i(g), & f(u_i(g)) \leq f(x_i(g)) \\ x_i(g), & \text{others} \end{cases} \quad (4)$$

where x_{ij} represents the j th element of the i th individual, x_{ij}^L and x_{ij}^U , respectively, represents its lower limit and upper limit. x_{r1} , x_{r2} and x_{r3} are three different individuals randomly selected from the current population, and F is the scaling

factor. $u_{ij}(g)$ represents the crossover result of the g generation, CR is the crossover rate, and j_{rand} is a random integer from 1 to m , so as to ensure that at least one characteristic of the crossed individuals comes from the mutation result v_i . $x_i(g+1)$ is an evolutionary progeny.

B. GAUSSIAN PROCESS (GP) MODELING

GP is a random process composed of an infinite number of random variables that obey the Gaussian distribution, and any subset of these random variables obeys joint Gaussian distribution. Its mean value function and covariance function are respectively expressed as follows:

$$\begin{cases} m(\mathbf{x}) = E[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') = E\{[f(\mathbf{x}) - m(\mathbf{x})][f(\mathbf{x}') - m(\mathbf{x}')]\} \end{cases} \quad (5)$$

where $\mathbf{x}, \mathbf{x}' \in R^d$ is any random variable, $m(\mathbf{x})$ is the mean value function of the random variable, and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function. Therefore, we define the GP model as:

$$f(x) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (6)$$

In the case that system noise is considered, the training model is represented as:

$$y = f(x) + \varepsilon \quad (7)$$

where ε is the additive noise following a normal distribution, with a mean value of 0 and a variance of σ_n^2 , i.e. $\varepsilon \sim N(0, \sigma_n^2)$.

x is the input vector, y is the observed value polluted by noise, and the prior distribution of y is:

$$y \sim N(0, \mathbf{K} + \sigma_n^2 \mathbf{I}) \quad (8)$$

where, $\mathbf{K} = K(X, X)$ is a positive definite covariance matrix of $n \times n$ order symmetry, and matrix elements are used to measure the correlation between x_i and x_j . The joint Gaussian prior distribution consisting of the output y with n training samples and the output f^* with n^* test samples is

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim N\left(0, \begin{pmatrix} K(X, X) + \sigma_n^2 \mathbf{I} & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{pmatrix}\right) \quad (9)$$

where, $K(X, X^*)$ is the covariance matrix with $n \times n^*$ order, and $K(X^*, X^*)$ is the covariance matrix of test output samples with $n^* \times n^*$ order.

The covariance function of the Gaussian process model must satisfy the Mercer condition, that is, any set of its points must be guaranteed to produce a non-negative definite covariance matrix. Usually we use the square exponent (SE) function as the covariance function.

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp[-0.5(\mathbf{x} - \mathbf{x}')^T M (\mathbf{x} - \mathbf{x}')] \quad (10)$$

where the matrix $M = \text{diag}(l)$ and l are the positive feature length scale parameters corresponding to the input variable elements, and σ_f^2 is the signal covariance. The hyperparameters of Gaussian process determine the properties of the model, which can be obtained by the maximum likelihood

estimation method based on the prior data. The form of the negative logarithmic likelihood function is:

$$l = \log p(y|X) = -\frac{1}{2}y^T K^{-1}y - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi \quad (11)$$

According to the Bayesian principle, given the new input x^* , training input X and training output y , the most probable predictive posterior distribution of y^* is inferred to be:

$$y^*|x^*, X, y \sim N(\mathbf{m}, \Sigma) \quad (12)$$

Its mean and covariance are as follows:

$$\mathbf{m} = K(X^*, X)K(X, X)^{-1}y \quad (13)$$

$$\Sigma = K(X^*, X^*) - K(X^*, X)K(X, X)^{-1}K(X, X^*) \quad (14)$$

\mathbf{m} contains the most likely values of the test output related to the test input vector in x^* , and the corresponding prediction variance is given by the covariance matrix Σ .

The predicted mean and variance of GP model describe a Gaussian distribution that the predicted output may follow. If the predicted mean is regarded as the predicted output value of a general nonlinear fitting tool, the predicted variance can be regarded as an evaluation of the uncertainty of the predicted mean. If a training input vector in the training set is close to the predicted output vector, a small prediction variance will be obtained, and the predicted mean value will be close to the actual output. In other words, the size of the prediction variance reflects the model accuracy of the GP model at that point.

C. ISOMETRIC MAPPING (Isomap)

Manifold learning is a dimensionality reduction method that draws on the concept of topological manifolds. 'manifold' is a space that is homogenous with Euclidean space locally. In other words, it has the property of Euclidean space locally and can use Euclidean distance to calculate the distance.

Isomap is a classical manifold learning algorithm based on multi-dimensional scaling (MDS). It is misleading to directly calculate Euclidean distance in high-dimensional space after embedding low-dimensional manifold into high-dimensional space, because the linear distance in high-dimensional space cannot be expressed in low-dimensional space. In this case, we need to express the distance between two points on the low-dimensional embedded manifold by 'geodesic' distance. The detailed process of isometric mapping is shown in algorithm 1.

When using Isomap algorithm to process data, it is very important to select a suitable neighbor parameter. If the selected neighborhood is too large, the algorithm will mistake the distant points as near neighbors, which will lead to 'short circuit'; if the neighborhood we selected is very small, many points on the manifold surface may not have close neighbor relationship with other points, which will lead to 'open circuit'.

Algorithm 1 Isometric Mapping

Input:

Data set $D = \{x_1, x_2, \dots, x_m\}$;
neighbor parameter K ;
dimension of low dimensional space d' .

Process:

- 1: for $i = 1, 2, \dots, m$ do
 - 2: Partition k -nearest neighbor of x_i ;
 - 3: The distance between x_i and its k -nearest neighbor samples is set to Euclidean distance, and the distance between X_i and other points is set to infinity;
 - 4: end for
 - 5: Calculate the distance $dist(x_i, x_j)$ between any two sample points (Dijkstra, Floyd) using the shortest path algorithm;
 - 6: Use $dist(x_i, x_j)$ as the input of MDS algorithm;
- Output:** The projection of the sample in the low-dimensional space is the output of the MDS algorithm recorded as $Z = \{z_1, z_2, \dots, z_m\}$.

III. PROPOSED ALGORITHM MODEL

In order to obtain the optimal data dimensionality reduction parameters scientifically, improve the generalization ability of the manifold GP model, and reduce the computational complexity and time consumption, a differential evolution based manifold Gaussian process (DE-MGP) is proposed in this study. A new manifold parameter training method is provided for the proposed surrogate model through DE algorithm. We take the manifold parameters as the optimization object, and the output of GP model as the optimization guidance of the DE algorithm. With the continuous iteration of DE, we train the dimension reduction parameters of manifold learning method while training the GP model. Firstly, we use the method of Isomap to process the collected initial data, and obtain a set of dimensional-reduction data with less information loss, and then use it to train GP. The mean absolute percentage error (MAPE) of the model is calculated and applied to the DE algorithm as the fitness function. The DE algorithm is used for iterative optimization, and the dimensionality reduction parameters can be solved after reaching the output conditions. Finally, the optimal parameters obtained by the algorithm are fed back to the Isomap, so that our optimal parameter model is established. The main algorithm flow is shown in algorithm 2, and its flow chart is shown in figure 1.

In the algorithm 2, there are several key points that need to be explained in detail:

- 1) In the process of using DE algorithm, we choose the mean absolute percentage error (MAPE) as the fitness function, and its calculation formula is as follows:

$$MAPE = \frac{1}{n} \sum \left| \frac{observed - predicted}{observed} \right| \times 100\% \quad (15)$$

where, n is the number of samples, $observed$ is the label, and $predicted$ is the output of the model.

Algorithm 2 Main Framework**Input:**

$Traindata = \{x_1, x_2, \dots, x_m\}$ includes training input $TrainX$ and training output $TrainY$.

$Testdata = \{x'_1, x'_2, \dots, x'_n\}$ includes test input $TestX$ and test output $TestY$.

Initialize:

Generation = 0. Set the control parameters of the algorithm as follows: variation factor $F = 0.9$, crossover rate $CR = 0.9$, maximum number of iterations $IterationMax = 50$, population size $NP = 30$, and population characteristic dimension $D = 2$. The initial individual $Pop(0)$ is set according to formula (1).

Process:

1: while Generation < $IterationMax$

2: For $i = 1: NP$

3: The mutant corresponding to the i th individual is calculated according to the equation (2).

4: According to the method in formula (3), calculate the individual after the i th individual crosses its mutant individual.

5: Select according to the equation (4) to get the next generation individuals.

6: End for

7: Generation = Generation + 1;

8: End while

9: Calculate the fitness value of the best individual: $Bestfitness = fitness(pop(Generation))$, $Pop(Generation, 1)$ is the optimal neighbor parameter, denoted as K , and $pop(Generation, 2)$ is the optimal low-dimensional space, denoted as d' . Among them, MAPE of Gaussian process prediction model is used as the fitness function.

10: For $i = 1, 2, \dots, (m + n)$ do

11: Take K samples as a neighborhood space and determine the neighborhood of each sample point.

12: The distance between x_i and the points in its neighborhood is set to Euclidean distance, and the distance from other points is set to infinity.

13: End for

14: The shortest path algorithm is used to calculate the distance $dist(x_i, x_j)$ between any two sample points.

15: By MDS operation, the low dimensional mapping GPR_{train} and GPR_{test} of model input samples are obtained.

16: $(GPR_{train}, TrainY)$ is taken into the above GP model as training data.

17: $(GPR_{test}, Testy)$ is the calculated mean absolute percentage error MAPE of the test input.

Output:

Return the predicted values $Pred = (p_1, p_2, \dots, p_n)$ and MAPE for the test sample.

- 2) In the manifold neighborhood division from step 10 to part 13, the first step is to partition the neighborhood of all samples including training samples and test

samples together. The problem of 'short circuit' and 'open circuit' in isometric mapping caused by improper neighborhood partition can be effectively avoided by optimizing the dimensionality reduction parameters.

IV. CASES STUDY

In the process of modeling and optimizing of microwave components, there are many problems about training data, such as too many characteristics for training data and time consuming for getting them. Therefore, it is very significant for us to reduce the feature dimension and complexity of samples. In this part, the mapping inverse model between the S-parameter of microwave filter and its coupling coefficient matrix will be established by the proposed DE-MGP, in which the given S-parameter is used as the input to predict the corresponding coupling coefficient of the filter, so as to achieve the purpose of designing the required microwave filter. In order to test the regression ability of the proposed training model in small samples and high-dimensional problems, this paper uses a MGP to extract the parameters of the two filters in the case of 300 groups of training samples and 600 groups of training samples respectively, and compared the experimental results with the ordinary GP model. According to our experimental research, when there are only 300 samples, we can get a surrogate model with high accuracy through the proposed method. On this basis, we continue to increase the training samples. When the number of samples reaches 600, our model accuracy will fully meet our design requirements.

A. THE FOURTH-ORDER COUPLING FILTER

In this case, the proposed DE-MGP will be used to establish the parameter extraction model of a fourth-order coupling filter with a central frequency of 4GHz and a bandwidth of 40MHZ. The initial coupling coefficient matrix [33] is as follows:

$$M = \begin{bmatrix} 0 & 0.84135 & 0 & -0.22423 \\ 0.84135 & 0 & 0.78212 & 0 \\ 0 & 0.78212 & 0 & 0.84135 \\ -0.22423 & 0 & 0.84135 & 0 \end{bmatrix} \quad (16)$$

Four different non-zero terms in the initial coupling coefficient matrix are randomly generated into a group of coupling coefficients within the tolerance of ± 0.3 , and then the S-parameters corresponding to each group of coupling coefficients are calculated respectively, so that the training data needed to establish the dimension reduction fitting model is obtained. The S_{11} curve is sampled at 35 frequency points, and input the 35 points as 35 dimensions of training data:

$$X = [dB(S_{11}(f_1)) dB(S_{11}(f_2)) \dots dB(S_{11}(f_{35}))] \quad (17)$$

According to the previous experiments, these 35 frequency points can better express the characteristics of the whole S_{11} curve, which enables us to obtain the highest surrogate model with the lowest frequency sampling. Four non-zero parameters of the coupling matrix M are taken as model outputs:

$$Y = [M_{21} M_{32} M_{41} M_{43}]^T \quad (18)$$

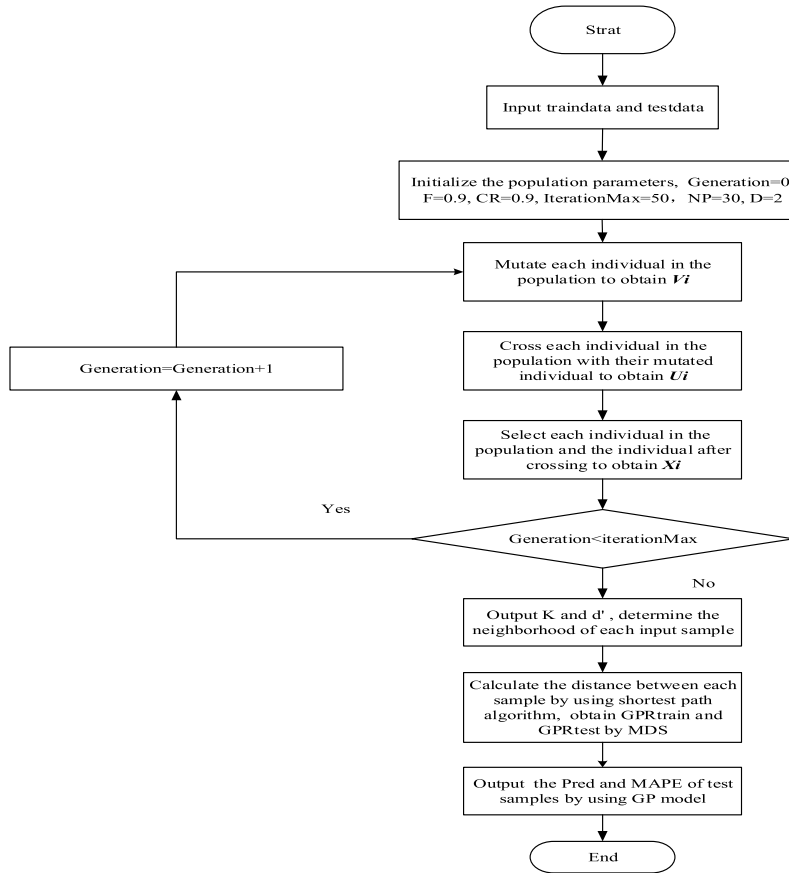


FIGURE 1. The modeling process of DE-MGP.

In this case, the dimension reduction parameter of the established MGP is set as (d', K), where d' is the target dimension of manifold dimension reduction, and K is the number of neighborhood individuals of dimension reduction samples. We will use the DE optimization algorithm to model the DE-MGP in the case of 300 samples and 600 samples respectively, and compare with the original model. Our evaluation of experimental results is based on an ideal coupling coefficient matrix. In this way, we can clearly find that under the same sample size, the training model proposed in this study has obvious advantages in high-dimensional parameter extraction. In the following, we will introduce the S-parameter extraction experiments of the ideal coupling coefficient matrix of the proposed DE-MGP model and the GP model under the conditions of 300 and 600 samples. The calculation of S-parameter is based on the method studied by predecessors and related electromagnetic simulation softwares [34].

Ideal coupling coefficient matrix is given by

$$M_{ideal} = \begin{bmatrix} 0 & 0.8480 & 0 & -0.2759 \\ 0.8480 & 0 & 0.7660 & 0 \\ 0 & 0.7660 & 0 & 0.8391 \\ -0.2759 & 0 & 0.8391 & 0 \end{bmatrix} \quad (19)$$

coupling coefficient matrix predicted by the DE-MGP in 300 training samples is

$$M_{pre_300} = \begin{bmatrix} 0 & 0.8416 & 0 & -0.2663 \\ 0.8416 & 0 & 0.7927 & 0 \\ 0 & 0.7927 & 0 & 0.8668 \\ -0.2663 & 0 & 0.8668 & 0 \end{bmatrix} \quad (20)$$

coupling coefficient matrix predicted by the DE-MGP in 600 training samples is

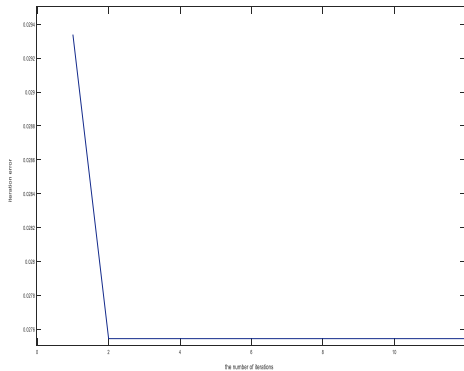
$$M_{pre_600} = \begin{bmatrix} 0 & 0.8458 & 0 & -0.2780 \\ 0.8458 & 0 & 0.7713 & 0 \\ 0 & 0.7713 & 0 & 0.8530 \\ -0.2780 & 0 & 0.8530 & 0 \end{bmatrix} \quad (21)$$

coupling coefficient matrix predicted by original GP in 300 training samples is

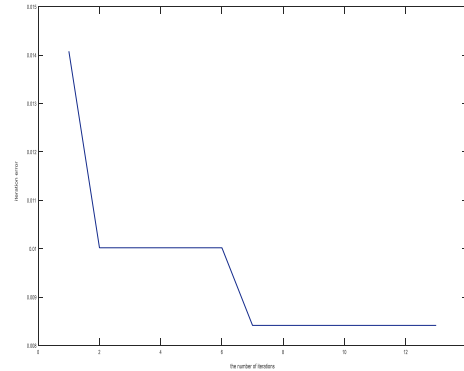
$$M_{pre_300}^* = \begin{bmatrix} 0 & 0.8114 & 0 & -0.2306 \\ 0.8114 & 0 & 0.7909 & 0 \\ 0 & 0.7909 & 0 & 0.8581 \\ -0.2306 & 0 & 0.8581 & 0 \end{bmatrix} \quad (22)$$

TABLE 1. Experimental results of the fourth order coupled filter.

	Training model	Goal dimension	Goal dimensional space	MAPE
300 samples	The DE-MGP	4	4	2.75%
	Original GP			6.56%
600 samples	The DE-MGP	7	7	0.84%
	Original GP			5.24%



a. optimization results of 300 training samples



b. optimization results of 600 training samples

FIGURE 2. Convergence curves of differential evolution algorithm on fourth-order coupling filter.

And coupling coefficient matrix predicted by original GP in 600 training samples is

$$M_{pre_600}^* = \begin{bmatrix} 0 & 0.8214 & 0 & -0.2879 \\ 0.8214 & 0 & 0.8542 & 0 \\ 0 & 0.8542 & 0 & 0.8223 \\ -0.2879 & 0 & 0.8223 & 0 \end{bmatrix} \quad (23)$$

Equations (20) and (21) are the coupling coefficient prediction matrix of the DE-MGP under 300 training samples and 600 training samples respectively. The results obtained by the original GP are shown in equation (22) and (23).

As shown in Table 1, in the experiment of 300 groups of samples, the optimization result of the DE algorithm for dimensionality reduction parameters (d', k) is (4, 4), in which the dimension of the low-dimensional space is 4 and the number of samples in each neighborhood is 4. The iterative convergence curve of the optimization algorithm is shown in Figure 2a, where the abscissa is the number of iterations and the ordinate is the iteration error of each step. The error of the first iteration is 2.93%, and the error of the final convergence result is 2.75%. That is the MAPE of the proposed model with the condition of optimal dimensionality reduction parameters is 2.75%, which is 6.14% lower than that of the first step iteration. The error rate of the original GP model is 6.56%, which is more than twice the error rate of the proposed model. In the experiment of 600 groups of samples, the dimension of the low-dimensional space output by the DE algorithm is 7, and the number of individuals in each neighborhood is 7. In this case, optimization algorithm

convergence curve is shown in figure 2b, where the abscissa is the number of iterations and the ordinate is the iteration error of each step. The first step iterative error is 1.41%, and the final convergence results of about 0.84%. Compared with the first step iterative result error rate, it decreases 40.43%. The prediction error of the original GP model is 5.24%, and the error rate of the proposed algorithm is more than five times better. Through these two experiments, we find that under the same sample sets, the proposed DE-MGP can effectively improve the regression effect of the inverse model of the filter.

Figure 3 and Figure 4, respectively, show the S-parameter fitting curves corresponding to the predicted coupling coefficient matrix obtained by the DE-MGP model under the conditions of 300 groups of samples and 600 groups of samples. Their abscissa is the frequency, ranging from 3900MHz to 4100MHz, and the ordinate is the dB value of S-parameter. The blue curve in the figure represents the S-parameter of the ideal coupling coefficient matrix, and the red circle represents the S-parameter of the predicted coupling coefficient matrix. It can be seen from the figure that the proposed DE-MGP in this study has a good fitting result in these two S-parameter extraction experiments.

It can be found from Table 1, Figure 3 and Figure 4 that increasing the sample size can improve the fitting accuracy of the algorithm. In the case of the same sample size, the precision of the proposed DE-MGP model is obviously better than that of the original GP model, and the MAPE of the MGP after parameter selection is at least one times lower than that

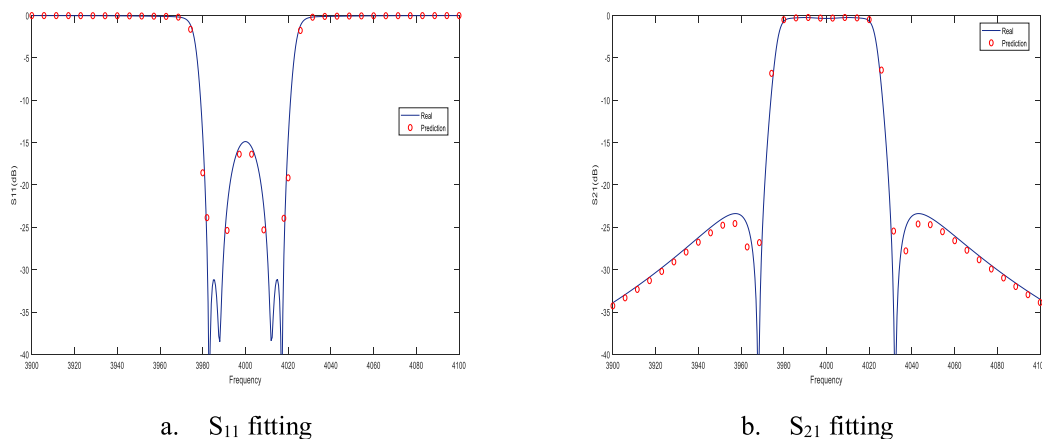


FIGURE 3. S-parameter fitting results corresponding to the predicted results of the proposed DE-MGP for 300 groups of samples (fourth-order coupling filter).

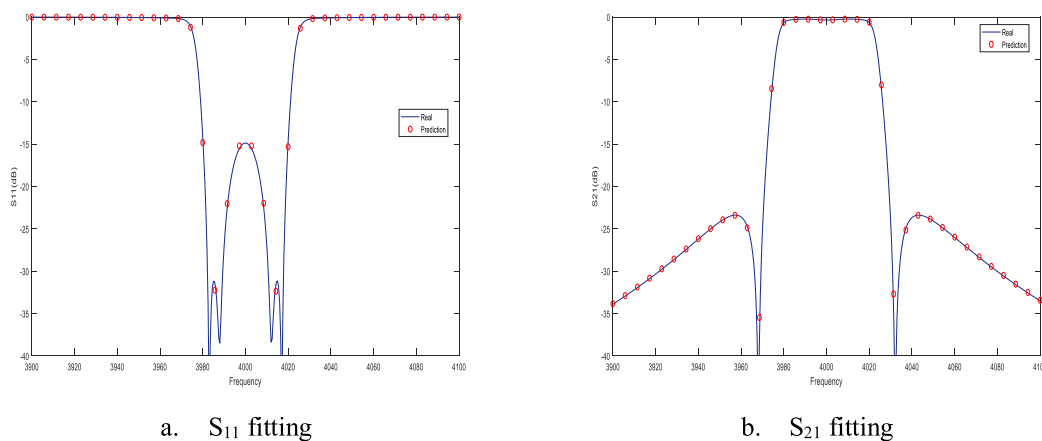


FIGURE 4. S-parameter fitting results corresponding to the predicted results of the proposed DE-MGP for 600 groups of samples (fourth-order coupling filter).

of the original GP. By comparing the 300 groups of samples of the DE-MGP and 600 groups of samples of the original GP modeling results (2.75% for DE-MGP in 300 samples, 5.24% for GP in 600 samples), we can clearly find that the dependence of the DE-MGP on the number of samples is obviously reduced. In other words, through our proposed DE-MGP model, we can make the model become more accurate with fewer training samples, which is very great significance for us to establish surrogate models in the field of electromagnetics.

B. THE SIX-ORDER COUPLING FILTER

In order to make our conclusion more convincing, we will carry out further parameter extraction experiments in the

same environment through a six-order coupling filter. The central frequency of the filter is 4GHz, the bandwidth is 36MHz, and the initial coupling coefficient matrix [33] is shown in equation (24), as shown at the bottom of the page.

Using the same experimental method as the above fourth-order coupling filter, we randomly generate a set of coupling coefficients from the seven different non-zero terms of the initial matrix within the tolerance of ±0.3, and then calculate their corresponding S-parameters respectively. We take the S-parameter as the training input and the seven different non-zero coupling coefficients as the sample labels. In the same way as above, we sample the curve at 35 frequency points, so that we get the system input (25) and output (26)

$$M = \begin{bmatrix} 0 & 0.88336 & 0 & 0 & 0 & 0.04863 \\ 0.88336 & 0 & 0.56859 & 0 & -0.26420 & 0 \\ 0 & 0.56859 & 0 & 0.78178 & 0 & 0 \\ 0 & 0 & 0.78178 & 0 & 0.56859 & 0 \\ 0 & -0.26420 & 0 & 0.56859 & 0 & 0.88336 \\ 0.04863 & 0 & 0 & 0 & 0.88336 & 0 \end{bmatrix} \tag{24}$$

we needed:

$$X = [dB(S_{11}(f_1))dB(S_{11}(f_2)) \dots dB(S_{11}(f_{35}))] \quad (25)$$

$$Y = [M_{21}M_{32}M_{43}M_{52} M_{54}M_{61}M_{65}]^T \quad (26)$$

In this case, we also used an ideal coupling coefficient matrix to test our model, and extracted S-parameters of the sixth order coupling filter in the conditions of 300 and 600 groups of samples respectively.

Ideal coupling coefficient matrix is given in (27), as shown at the bottom of the page, coupling coefficient matrix predicted by the DE-MGP in 300 training samples is (28), as shown at the bottom of the page, coupling coefficient matrix predicted by the DE-MGP in 600 training samples is (29), as shown at the bottom of the page, coupling coefficient matrix predicted by original GP in 300 training samples is (30), as shown at the bottom of the page, And coupling coefficient matrix predicted by original GP in 600 training samples is (31), as shown at the bottom of the page.

Equations (28) and (29) are respectively the prediction junctions of the proposed DE-MGP model on the ideal matrix in the conditions of 300 groups of training samples and 600 groups of training samples of the sixth order coupling

filter. The predicted values of the ideal sixth order coupling coefficient matrix by the original GP model in the conditions of 300 groups of samples and 600 groups of training samples are shown in equations (30) and (31), respectively.

As shown in Table 2, in the experiment of 300 groups of samples, the optimization results of the DE algorithm are as follows: the optimal dimension number of low-dimensional space is 11, and the optimal number of neighborhood individuals is 17. In the case of 300 groups of samples, the convergence curve of the DE algorithm is shown in Figure 5a, where the abscissa is the number of iterations and the ordinate is the iteration error of each step. The first iteration error of the DE algorithm is 2.67%, and the convergence result is 2.22%, which means the MAPE obtained by the proposed model in the experiment of 300 training samples is 2.22%. The error rate is reduced by 16.85% of the first iteration. The original GP model is 5.04%, and the error rate is nearly double that of the proposed model. In the experiment of 600 groups of samples, the optimal dimension number of low dimension space is 7. The number of optimal dimensions of neighborhood individuals is 30. In this experiment, the convergence curve of the DE algorithm is shown in Figure 5b, where the

$$M = \begin{bmatrix} 0 & 0.88181 & 0 & 0 & 0 & 0.05035 \\ 0.88181 & 0 & 0.70565 & 0 & -0.31181 & 0 \\ 0 & 0.70565 & 0 & 0.95344 & 0 & 0 \\ 0 & 0 & 0.95344 & 0 & 0.63067 & 0 \\ 0 & -0.31181 & 0 & 0.63067 & 0 & 0.94453 \\ 0.05035 & 0 & 0 & 0 & 0.94453 & 0 \end{bmatrix} \quad (27)$$

$$M_{pre_300} = \begin{bmatrix} 0 & 0.89487 & 0 & 0 & 0 & 0.05178 \\ 0.89487 & 0 & 0.71152 & 0 & -0.30604 & 0 \\ 0 & 0.71152 & 0 & 0.92775 & 0 & 0 \\ 0 & 0 & 0.92775 & 0 & 0.59917 & 0 \\ 0 & -0.30604 & 0 & 0.59917 & 0 & 0.93687 \\ 0.05178 & 0 & 0 & 0 & 0.93687 & 0 \end{bmatrix} \quad (28)$$

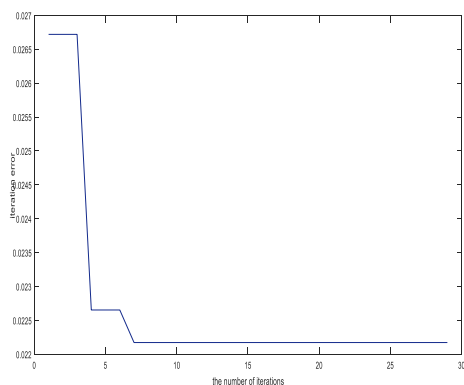
$$M_{pre_600} = \begin{bmatrix} 0 & 0.89369 & 0 & 0 & 0 & 0.05081 \\ 0.89369 & 0 & 0.67018 & 0 & -0.30785 & 0 \\ 0 & 0.67018 & 0 & 0.94200 & 0 & 0 \\ 0 & 0 & 0.94200 & 0 & 0.63531 & 0 \\ 0 & -0.30785 & 0 & 0.63531 & 0 & 0.94665 \\ 0.05081 & 0 & 0 & 0 & 0.94665 & 0 \end{bmatrix} \quad (29)$$

$$M_{pre_300}^* = \begin{bmatrix} 0 & 0.89201 & 0 & 0 & 0 & 0.05404 \\ 0.89201 & 0 & 0.71287 & 0 & -0.28703 & 0 \\ 0 & 0.71287 & 0 & 0.92141 & 0 & 0 \\ 0 & 0 & 0.92141 & 0 & 0.57466 & 0 \\ 0 & -0.28703 & 0 & 0.57466 & 0 & 0.89175 \\ 0.05404 & 0 & 0 & 0 & 0.89175 & 0 \end{bmatrix} \quad (30)$$

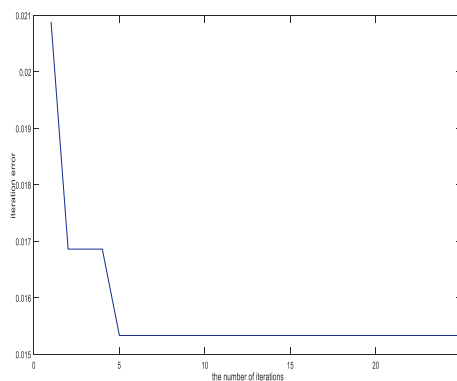
$$M_{pre_600}^* = \begin{bmatrix} 0 & 0.89050 & 0 & 0 & 0 & 0.05583 \\ 0.89050 & 0 & 0.68227 & 0 & -0.30097 & 0 \\ 0 & 0.68227 & 0 & 0.91972 & 0 & 0 \\ 0 & 0 & 0.91972 & 0 & 0.57147 & 0 \\ 0 & -0.30097 & 0 & 0.57147 & 0 & 0.95474 \\ 0.05583 & 0 & 0 & 0 & 0.95474 & 0 \end{bmatrix} \quad (31)$$

TABLE 2. Experimental results of the sixth order coupling filter.

	Training model	Goal dimension	Goal dimensional space	MAPE
300 samples	DE-MGP	11	17	2.22%
	Original GP			5.04%
600 samples	DE-MGP	7	30	1.53%
	Original GP			4.66%

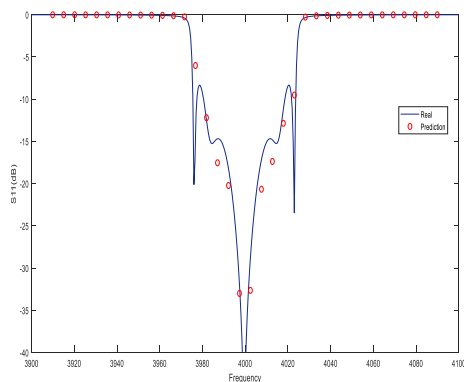


a. Optimization results of 300 training samples

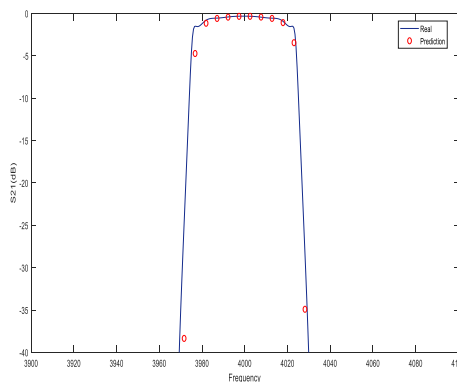


b. Optimization results of 600 training samples

FIGURE 5. Convergence curve of differential evolution algorithm on sixth order coupling filter.



a. S₁₁ fitting



b. S₂₁ fitting

FIGURE 6. S-parameter fitting results corresponding to prediction results of the proposed DE-MGP in 300 groups of samples (six-order coupling filter).

abscissa is the number of iterations and the ordinate is the iteration error of each step. The error rate of the DE algorithm in the first calculation is 2.09%, and the convergence result is 1.53%, which means the error rate is reduced by 26.76% of the first calculation result. Compared with the original GP model, the error rate is only about 1/3 of the original model. This shows that in the experiment of the sixth order coupled filter, the proposed model also works very well. In the case of the same training samples, the proposed DE-MGP has better parameter extraction effect than that of the original GP. If we carefully observe Table 2, we can also find that, like

the fourth-order coupling filter, the precision of the DE-MGP model in the condition of 300 groups of training samples is better than that of the GP model in the condition of 600 groups of samples, which also indicates that the model we proposed has a low dependence on the number of samples when dealing with high-dimensional problems. This has huge implications when we're dealing problems without enough samples.

Figure 6 shows the parameter extraction results of the DE-MGP for the sixth order coupling filter in 300 samples, and Figure 7 shows the parameter extraction results of the DE-MGP in 600 groups of samples. Figure 6a and Figure 6b

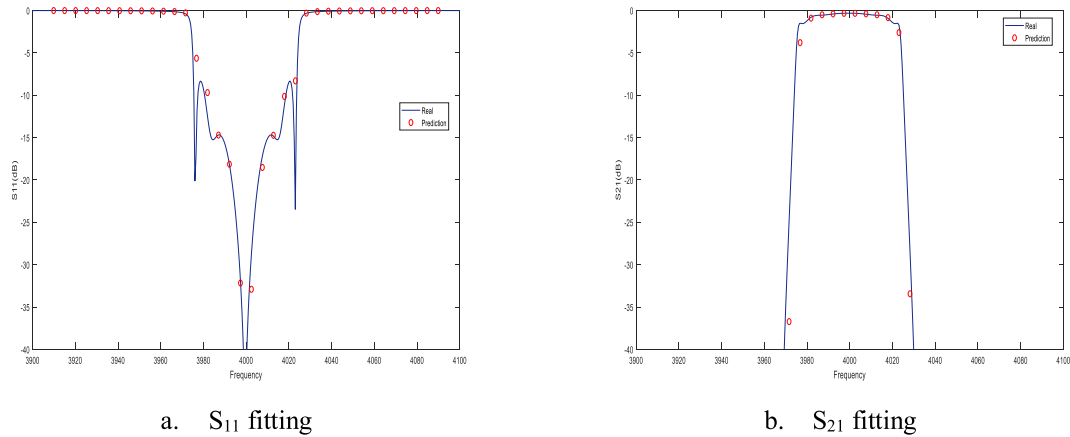


FIGURE 7. S-parameter fitting results corresponding to prediction results of the proposed DE-MGP in 600 groups of samples (six-order coupling filter).

TABLE 3. Comparison of filter parameter extraction results of different models.

Filters	Number of samples	SNN	DNN	SVM	KELM	DE-MGP
Four order	300	9.38%	6.25%	5.32%	6.39%	2.75%
	600	3.21%	2.02%	4.80%	4.82%	0.84%
Six order	300	12.68%	9.30%	9.37%	8.82%	2.22%
	600	8.15%	4.25%	6.52%	5.95%	1.53%

are the fitting results of S_{11} and S_{21} under 300 groups of samples, respectively. Figure 7a and Figure 7b are the fitting results of S_{11} and S_{21} under 600 groups of samples, respectively. The abscissa is the frequency range from 3900MHz to 4100MHz, and the ordinate is the dB value of S-parameter. The blue curve is the S-parameter curve of the ideal coupling matrix, and the red circle is the S-parameter predicted by the proposed DE-MGP.

It can be found from Table 2, Figure 6 and Figure 7 that, generally, increasing the sample size can improve the accuracy of the training model. However, we can reasonably reduce the training dimension and set the dimensionality reduction parameters to improve the accuracy of our training model under the same sample size. According to Table 2, we also find that for the sixth order filter, the DE-MGP training model we proposed can also effectively save the training samples needed in our experiment.

C. COMPARISON WITH OTHER MODELS

In this part, we implement several classical surrogate models respectively and apply them to the filter design we studied. In the case of the same data set, their experimental results are shown in Table 3. The models include shallow neural network (SNN), deep neural network (DNN), support vector machine (SVM), and kernel extreme learning machine (KELM). Among them, the modeling process of SNN and DNN refers to the method in [35], the modeling of SVM

refers to the method in [14], and the modeling of KELM refers to the method in [16]. From Table 3, we can find that compared with other surrogate models, the proposed DE-MGP has better extraction accuracy in both fourth-order filter and sixth-order filter under the same experimental data. In all the methods mentioned above, DNN can improve the accuracy of the model obviously when there are more training data, but compared with the proposed DE-MGP, it still has a large error.

V. CONCLUSION

In this paper, we propose a new machine learning model named differential evolution based manifold Gaussian process (DE-MGP). Compared with the original GP model, the proposed model has a higher accuracy in dealing with high dimensional problems. After setting reasonable low dimensional parameters, the DE-MGP can obtain more concise low dimensional spatial mapping through a small information loss when processing data, which greatly reduces the spatial complexity of the data in training the model, and also greatly reduces the amount of calculation. By comparing the DE-MGP with the original GP in different sample sizes, we can also find that the DE-MGP with less training samples can achieve the modeling effect of the original GP model. This indicates that we can save our training samples largely when we use the proposed model to solve the problems with high dimension and small sample size. And it also greatly

reduces the cost of establishing the inverse surrogate model of microwave filter. To sum up, the model proposed in this study can well solve the problem of high sample acquisition cost and large sample space dimension in the microwave field, which is of great significance to the optimization design of our microwave components.

In the following research, we will continue our study in two directions. Firstly, we will further explore how to avoid the local optimal risk of manifold parameters in the DE training, and we will improve the generalization ability of our model through more reliable optimization methods. Secondly, we will further study the multivalued problem of inverse surrogate model of microwave components, and try to use a new surrogate model structure to model the problem.

REFERENCES

- [1] W. P. Mason and R. A. Sykes, "The use of coaxial and balanced transmission lines in filters and wide-band transformers for high radio frequencies," *Bell Syst. Tech. J.*, vol. 16, no. 3, pp. 275–302, Jul. 1937.
- [2] S. B. Cohn, "Direct-coupled-resonator filters," *Proc. IRE*, vol. 45, no. 2, pp. 187–196, Feb. 1957.
- [3] J. D. Rhodes, "The theory of generalized interdigital networks," *IEEE Trans. Circuit Theory*, vol. CT-16, no. 3, pp. 280–288, Aug. 1969.
- [4] A. E. Atia and A. E. Williams, "Narrow-bandpass waveguide filters," *IEEE Trans. Microw. Theory Techn.*, vol. MTT-20, no. 4, pp. 258–265, Apr. 1972, doi 10.1109/tmtt.1972.1127732.
- [5] J. D. Rhodes and R. J. Cameron, "General extracted pole synthesis technique with applications to low-loss TE₀₁₁ mode filters," *IEEE Trans. Microw. Theory Techn.*, vol. MTT-28, no. 9, pp. 1018–1028, Sep. 1980.
- [6] R. J. Cameron, "General prototype network-synthesis methods for microwave filters," *ESA J.*, vol. 6, no. 2, pp. 193–206, 1982.
- [7] R. J. Cameron, "General coupling matrix synthesis methods for Chebyshev filtering functions," *IEEE Trans. Microw. Theory Techn.*, vol. 47, no. 4, pp. 433–442, Apr. 1999.
- [8] S. Amari, "Synthesis of cross-coupled resonator filters using an analytical gradient-based optimization technique," *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 9, pp. 1559–1564, Sep. 2000.
- [9] T. Yu-Bo, D. Yue, X. Zhi-Bin, S. Sha, and P. Tao, "Frequency characteristics of electromagnetic bandgap structure with bow-tie cells and its optimal design based on particle swarm optimization," *IEEE Trans. Electr. Electron. Eng.*, vol. 8, no. 1, pp. 63–68, Jan. 2013.
- [10] H. Otani, S. Yamaguchi, Y. Maeda, and M. Omiya, "High-speed, automatic and optimal design of microwave filter using genetic algorithms," in *Proc. Int. Symp. Antennas Propag. Conf.*, Dec. 2014, pp. 163–164.
- [11] J. C. Melgarejo, J. Ossorio, S. Cogollos, M. Guglielmi, V. E. Boria, and J. W. Bandler, "On space mapping techniques for microwave filter tuning," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 4860–4870, Dec. 2019.
- [12] B. Liu, H. Yang, and M. J. Lancaster, "Global optimization of microwave filters based on a surrogate model-assisted evolutionary algorithm," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 6, pp. 1976–1985, Jun. 2017.
- [13] F. Feng, C. Zhang, J. Ma, and Q.-J. Zhang, "Parametric modeling of EM behavior of microwave components using combined neural networks and Pole-Residue-Based transfer functions," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 1, pp. 60–77, Jan. 2016.
- [14] G. Angiulli, M. Cacciola, and M. Versaci, "Microwave devices and antennas modelling by support vector regression machines," *IEEE Trans. Magn.*, vol. 43, no. 4, pp. 1589–1592, Apr. 2007.
- [15] J. L. Chávez-Hurtado and J. E. Rayas-Sanchez, "Polynomial-based surrogate modeling of RF and microwave circuits in frequency domain exploiting the multinomial theorem," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 12, pp. 4371–4381, Dec. 2016.
- [16] L.-Y. Xiao, W. Shao, X. Ding, and B.-Z. Wang, "Dynamic adjustment kernel extreme learning machine for microwave component design," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 10, pp. 4452–4461, Oct. 2018.
- [17] R. B. Gramacy and H. K. H. Lee, "Bayesian treed Gaussian process models with an application to computer modeling," *J. Amer. Stat. Assoc.*, vol. 103, no. 483, pp. 1119–1130, Sep. 2008.
- [18] J. P. de Villiers and J. P. Jacobs, "Gaussian process modeling of CPW-fed slot antennas," *Prog. Electromagn. Res.*, vol. 98, no. 4, pp. 233–249, 2009.
- [19] J. P. Jacobs and J. P. de Villiers, "Gaussian-process-regression-based design of ultrawide-band and dual-band CPW-FED slot antennas," *J. Electromagn. Waves Appl.*, vol. 24, no. 13, pp. 1763–1772, Jan. 2010.
- [20] J. P. Jacobs and S. Koziel, "Gaussian process modeling of multi-variable microwave filters using non-standard covariance functions," in *Proc. Int. Rev. Prog. Appl. Comput. Electromagn.*, 2012, pp. 1082–1086.
- [21] J. P. Jacobs and S. Koziel, "Two-stage framework for efficient Gaussian process modeling of antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 62, no. 2, pp. 706–713, Feb. 2014.
- [22] J. P. Jacobs and S. Koziel, "Single-model versus ensemble-model strategies for efficient Gaussian process surrogate modeling of antenna input characteristics," in *Proc. Int. Conf. Electromagn. Adv. Appl. (ICEAA)*, Sep. 2013, pp. 510–513.
- [23] J. P. Jacobs, "Efficient resonant frequency modeling for dual-band microstrip antennas by Gaussian process regression," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, no. 3, pp. 337–341, 2015.
- [24] J. P. Jacobs and S. Koziel, "Reduced-cost microwave filter modeling using a two-stage Gaussian process regression approach," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 25, no. 5, pp. 453–462, Jun. 2015.
- [25] V. Cardona and A. Mauricio, "Multi-output Gaussian processes for enhancing resolution of diffusion tensor fields," *IEEE Electr. Eng.*, vol. 4, no. 16, pp. 1111–1114, Aug. 2016.
- [26] Y. Chen, Y.-B. Tian, Z. Qiang, and L. Xu, "Optimisation of reflection coefficient of microstrip antennas based on KBNN exploiting GPR model," *IET Microw., Antennas Propag.*, vol. 12, no. 4, pp. 602–606, Mar. 2018.
- [27] X.-H. Fan, Y.-B. Tian, and Y. Zhao, "Rapid microwave antenna design optimization with Gaussian process exploiting fitness-estimation-based particle swarm optimization," *Int. J. Antennas Propag.*, vol. 2019, pp. 1–9, 2019, Art. no. 2176518.
- [28] Y.-X. Xu and Y.-B. Tian, "Optimal design of conical horn antenna based on GP model with coarse mesh," *Iranian J. Sci. Technol., Trans. Electr. Eng.*, vol. 43, no. 4, pp. 717–724, Dec. 2019, doi: 10.1007/s40998-019-00209-3.
- [29] J. Gao, Y. Tian, X. Zheng, and X. Chen, "Resonant frequency modeling of microwave antennas using Gaussian process based on semisupervised learning," *Complexity*, vol. 2020, May 2020, Art. no. 3485469.
- [30] G. Pan, Y. Wu, M. Yu, L. Fu, and H. Li, "Inverse modeling for filters using a regularized deep neural network approach," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 5, pp. 457–460, Apr. 2020.
- [31] Y.-Z. Chen, S.-F. Shen, T. Chen, and R. Yang, "Path optimization study for vehicles evacuation based on Dijkstra algorithm," *Procedia Eng.*, vol. 71, pp. 159–165, Jan. 2014.
- [32] Z. Wang, Q. Sun, H. Guo, and Y. Zhao, "Power management for plug-in hybrid electric vehicle with automated mechanical transmission using multiple dimensional scaling method," *Recent Patents Eng.*, vol. 14, no. 1, pp. 133–141, Jun. 2020.
- [33] T. Shen, H.-T. Hsu, K. A. Zaki, A. E. Atia, and T. G. Dolan, "Full-wave design of canonical waveguide filters by optimization," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 2, pp. 504–511, Feb. 2003.
- [34] H.-T. Hsu, Z. Zhang, K. A. Zaki, and A. E. Atia, "Parameter extraction for symmetric coupled-resonator filters," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 12, pp. 2971–2978, Dec. 2002.
- [35] J. Jin, C. Zhang, F. Feng, W. Na, J. Ma, and Q.-J. Zhang, "Deep neural network technique for high-dimensional microwave modeling and applications to parameter extraction of microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4140–4155, Oct. 2019.



XUEZHI CHEN was born in Nantong, China, in 1995. He received the B.Tech. degree from the Nanjing Institute of Technology. He is currently pursuing the master's degree with the Jiangsu University of Science and Technology. His research interest includes the rapid optimization design of electromagnetic microwave devices.



YUBO TIAN was born in Changtu, China, in 1971. He received the Ph.D. degree from Nanjing University, in 2004. He is currently a Full Professor with the School of Electronics and Information, Jiangsu University of Science and Technology. His research interests include the applications of computational intelligence to the electromagnetics field.



JING GAO was born in Huai'an, China, in 1995. She is currently pursuing the master's degree with the Jiangsu University of Science and Technology. Her research interests include signal processing theory and technology.

...



TIANLIANG ZHANG was born in Ma'anshan, China. He is currently pursuing the master's degree with the Jiangsu University of Science and Technology. His research interests include signal processing theory and technology.