URSI GASS 2021, Rome, Italy, 28 August - 4 September 2021

Multi-Output Variable-Fidelity Surrogate Modeling for Microwave Components Design

Rodrigo Silva Rezende*(1), Mirsad Hadžiefendić(1), and Rolf Schuhmann(1)
(1) Technical University of Berlin, Germany

Abstract

This paper proposes a Gaussian processes-based modeling technique for handling multi-output (frequency-dependent vector-valued) microwave systems, in which variable-fidelity data is available. This approach assumes that each frequency point has its own mean and covariance, which are independent of the ones of other frequencies, and a unique correlation matrix describes the full frequency range. Using these assumptions can significantly reduce the number of matrix calculations required by the maximum likelihood estimation of the data. The notable advantage of this proposed approach is the capability of modeling systems with multi-output responses without requiring any pre-conditioning of the data while keeping a high accuracy, which is especially useful in the microwave design. The proposed methodology is demonstrated for a parameterized PCB connector design and compared against usual surrogate modeling approaches.

1 Introduction

The design of microwave components based on full-wave electromagnetic (EM) simulations is a well-established approach due to trustful calculations that these methods can achieve. However, in many cases, such as optimization [1] or statistical analysis [2], the computational costs required tend to exceed the available capability when direct methods, e.g., Monte Carlo analysis, are used.
Surrogate modeling is a way to mitigate this problem since once this model is created with a certain amount of information of the original simulation, it can be evaluated many times at a low cost and reasonable accuracy when compared to the full-wave EM simulation [1], [4]. In recent years, some surrogate modeling approaches have been employing variable-fidelity (also called multi-fidelity) data, i.e., both high and low-fidelity, in order to speed up the modeling process, while maintaining the accuracy [3].
Although these modeling methods do speed up EM simulations in some cases, they are often not truly suitable for the design of microwave systems. Usually, microwave components outputs are vectors in a frequency range, e.g., the $|S_{11}|$ parameter of a structure. Nevertheless, the usual surrogate approaches consider these outputs as a set of scalars, which make them efficient only for systems with a low number of input variables and reduced size of modeling data [4], [5].
The most direct procedure to solve this problem is to incorporate the extra dimension, e.g., the frequency, in the design variables and adapt the vector-output into a scalar-output to construct a conventional surrogate. Another way is to build a model for each frequency point of the output and estimate each desired scalar with a different model [6]. The drawback of these methods is that they can be still too expensive to be evaluated depending on the size of the vector output. Recently, some progress has been made in the surrogate modeling based on Gaussian processes (GPs), in which the construction of a correlation matrix that relates each frequency point of the output allows a more correct prediction of systems with vector-valued outputs [7]. However, this method still requires a considerable number of matrix inversions, which does not make it any more computationally efficient. An alternative approach convolves the GPs with different kernels in order to create multiple outputs and adds a cross-correlation among them, reducing the computational costs in some cases [8].
In this paper, we propose an approach based on GPs to deal with vector-valued output microwave systems, for which both high and low-fidelity data are available. By considering that a unique correlation matrix describes each point of the vector-valued output and that each of these points has a unique mean and covariance, the number of matrix calculations required by the maximum likelihood estimation of the data can be extremely reduced, without significantly affecting the accuracy of the model. Our approach and its efficiency are demonstrated using a PCB connector operating at 12-18 GHz, whose training data are computed with CST MWS [9].

2 Vector-Output Variable-fidelity Modeling

In this section, we formulate the proposed vector-output surrogate technique by modifying the usual Co-Kriging model, a variable-fidelity Gaussian process-based approach [3]. Co-Kriging is a multi-level adaption of the Kriging method, which exploits both high-fidelity model (HFM) and low-fidelity model (LFM) to enhance the accuracy of the surrogate model and reduce its construction time [5]. A big drawback of this method, is that it can only model functions, whose outputs are scalars. However, often in the microwave design, we want to model a function, w.r.t. the HFM, $f_e := \boldsymbol{x} \mapsto \boldsymbol{y}_e : \mathbb{R}^k \to \mathbb{R}^l$, with $k,l \in \mathbb{N}$, where $\boldsymbol{x}$ is the input and $\boldsymbol{y}_e$ is the vector-output, whose observations of a cheaper function, i.e., responses of a LFM, $f_c : \boldsymbol{x} \mapsto \boldsymbol{y}_c : \mathbb{R}^k \to \mathbb{R}^l$, with input $\boldsymbol{x}_c$ and vector-output $\boldsymbol{y}_c$

are available. Here k is the number of design input parameters and l the size of the vector output. The sampling data, i.e., the observed data, of f_e is described as $\boldsymbol{X}_e = \{\boldsymbol{x}_e^{(1)}, \cdots, \boldsymbol{x}_e^{(n_e)}\}^T$ and

$$\boldsymbol{Y}_e := \begin{pmatrix} \boldsymbol{y}_{e1} & \boldsymbol{y}_{e2} & \cdots & \boldsymbol{y}_{el} \\ y_{e1}^{(1)} & y_{e2}^{(1)} & \cdots & y_{el}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{e1}^{(n_e)} & y_{e2}^{(n_e)} & \cdots & y_{el}^{(n_e)} \end{pmatrix}, \tag{1}$$

Each row of the matrix $\boldsymbol{Y}_e$ represents a vector-valued output of the function f_e, e.g., the $|S_{11}|$ parameter of a microwave system with l frequency points in the frequency range ω_1 to ω_l Hz. Note that, the same data arrangement is valid for the cheap data $\boldsymbol{X}_c$ and $\boldsymbol{Y}_c$.

In order to deal with this case, we must modify the usual Co-Kriging approach, resulting in the proposed algebraic vectorial (AV) Co-Kriging. The first step of this adaption is to consider both functions' observations as realizations of Gaussian Processes $\boldsymbol{Y}_e^{\mathscr{G}\mathscr{P}}(\boldsymbol{X}_e) \sim \mathscr{N}(\mu_e, \Sigma_e)$, $\boldsymbol{Y}_c^{\mathscr{G}\mathscr{P}}(\boldsymbol{X}_c) \sim \mathscr{N}(\mu_c, \Sigma_c)$, i.e., vectors of random variables, where μ and Σ represent the mean and covariance matrix of the respectives GPs. The model's construction starts by defining a GP for the difference data as

$$\boldsymbol{Y}_d^{\mathscr{G}\mathscr{P}}(\boldsymbol{x}) := \boldsymbol{Y}_e^{\mathscr{G}\mathscr{P}}(\boldsymbol{x}) - \rho\, \boldsymbol{Y}_c^{\mathscr{G}\mathscr{P}}(\boldsymbol{x}), \tag{2}$$

where $\rho \in \mathbb{R}$ is the scaling parameter that dictates the difference between coarse and expensive data. From this point on, we describe the expensive data using the cheap and difference data, i.e., $\boldsymbol{Y}_d := \boldsymbol{Y}_e - \rho \boldsymbol{Y}_c$.

Then, we compute the correlation of the data and organize this information in a matrix, called correlation matrix $\boldsymbol{\Psi}$. Note that, there are different ways to define the structure of this correlation matrix given the data arrangement within matrices $\boldsymbol{Y}_d$ and $\boldsymbol{Y}_c$. We propose considering each columns of both matrices $\boldsymbol{y}_{di}$ and $\boldsymbol{y}_{ci}$, $i = 1, \cdots, l$ as random vectors (RV)s of Gaussian variables, and assume two conditions for these RVs:

1. Unique correlation matrices $\boldsymbol{\Psi}_d$ and $\boldsymbol{\Psi}_c$ describes all RVs separately (for each different and cheap data)
2. There exist single μ_d, σ_d^2 and μ_c and σ_c^2 for each RV and they are independent from the ones of other RVs

In proposition two, σ^2 means the variance of the respectively RV. In a practical example, we are saying that a geometry input parameter should affect the response of $|S_{11}|$ in two different frequencies ω_i and ω_j for $i, j = 1, \cdots, l$ in similar ways and that these responses $|S_{11}|_{\omega=\omega_i}$ and $|S_{11}|_{\omega=\omega_j}$ are not correlated. This may appear contradictory, however, these conditions do alleviate the matrix calculations, while affecting the accuracy insignificantly, as it will be shown in this section and the illustration design example in Section 3.

After assuming these conditions, we calculate the correlation of the cheap and difference data separately and build with this information the correlation matrices $\boldsymbol{\Psi}_c \in \mathbb{R}^{n_c \times n_c}$ and $\boldsymbol{\Psi}_d \in \mathbb{R}^{n_e \times n_e}$. The correlation between any two elements of the rows $i, h = 1, \cdots, n_c$ of the cheap output matrix, for example, is calculated with the function

$$\boldsymbol{\Psi}_{c(i,h)} = cor[\boldsymbol{Y}_c(\boldsymbol{x}^{(i)}), \boldsymbol{Y}_c(\boldsymbol{x}^{(h)})] := exp\Big(-\sum_{j=1}^{k} \theta_{cj} \big| x_j^{(i)} - x_j^{(h)} \big|^{p_c} \Big). \tag{3}$$

Note that two parameters, namely $\boldsymbol{\theta}_c \in \mathbb{R}^{k_c}$ and $p_c \in \mathbb{R}$ control the shape of this function. Naturally, there are many possible values, which one can use for these parameters. However, we should calculate the values that most probably would describe the original sampling data the best. For this task we take use of the Maximum Likelihood Estimation (MLE). The usual MLE, however, considers the output of a function as a scalar. Therefore, we propose an average MLE such that the best parameters for an output matrix as $\boldsymbol{Y}_c$ can be found by maximizing the modified concentrated log-likelihood of the cheap data

$$\underset{\boldsymbol{\theta}_c, p_c \in \mathbb{R}^{k_c+1}}{\arg\max} \Big(\ln(\mathfrak{L}(\boldsymbol{Y}_c)) \Big) \approx \sum_{i=1}^{l} \Big(-\frac{n_c}{2} ln(\hat{\sigma}_{ci}^2) - \frac{1}{2} ln(\det(\boldsymbol{\Psi}_c)) \Big) \frac{1}{l}, \tag{4}$$

where the vector $\hat{\boldsymbol{\sigma}}_c^2 = (\hat{\sigma}_{c1}^2, \cdots, \hat{\sigma}_{cl}^2)$ is calculated by extracting the diagonal elements of:

$$\mathrm{diag}(\hat{\boldsymbol{\sigma}}_c^2) = \frac{(\boldsymbol{Y}_c - \mathbf{1}\hat{\boldsymbol{\mu}}_c)^T \boldsymbol{\Psi}_c^{-1} (\boldsymbol{Y}_c - \mathbf{1}\hat{\boldsymbol{\mu}}_c)}{n_c} \cdot \boldsymbol{I}, \tag{5}$$

where $\boldsymbol{I}$ is the identity matrix, $\mathbf{1}$ a vector of ones and with

$$\hat{\boldsymbol{\mu}}_c = \frac{\mathbf{1}^T \boldsymbol{\Psi}_c^{-1} \boldsymbol{Y}_c}{\mathbf{1}^T \boldsymbol{\Psi}_c^{-1} \mathbf{1}}. \tag{6}$$

For the difference data, we compute the following

$$\underset{\boldsymbol{\theta}_d, p_d, \rho \in^{k_c+1+1}}{\arg\max} \Big(\ln(\mathfrak{L}(\boldsymbol{Y}_d)) \Big) \approx \sum_{i=1}^{l} \Big(-\frac{n_e}{2} ln(\hat{\sigma}_{di}^2) - \frac{1}{2} ln(\det(\boldsymbol{\Psi}_d)) \Big) \frac{1}{l}, \tag{7}$$

where we calculate $\hat{\boldsymbol{\sigma}}_d^2$ and $\boldsymbol{\mu}_d$ with (5) and (6) using the difference data and correlation matrix. These two MLEs result in the estimated values of the five parameters $\boldsymbol{\theta}_c, p_c, \boldsymbol{\theta}_d, p_d$, and ρ, with which we can compute the extended correlation matrices C_i, $i = 1, \cdots, l$, correlating both high and low-fidelity data for each RV with

$$\boldsymbol{C}_i = \begin{pmatrix} \sigma_{ci}^2 \boldsymbol{\Psi}_c(\boldsymbol{X}_c, \boldsymbol{X}_c) & \rho\, \sigma_{ci}^2 \boldsymbol{\Psi}_c(\boldsymbol{X}_c, \boldsymbol{X}_e) \\ \rho\, \sigma_{ci}^2 \boldsymbol{\Psi}_c(\boldsymbol{X}_e, \boldsymbol{X}_c) & \rho^2 \sigma_{ci}^2 \boldsymbol{\Psi}_c(\boldsymbol{X}_e, \boldsymbol{X}_e) + \sigma_{di}^2 \boldsymbol{\Psi}_d(\boldsymbol{X}_e, \boldsymbol{X}_e) \end{pmatrix}. \tag{8}$$

The notation $\boldsymbol{\Psi}_c(\boldsymbol{X}_c, \boldsymbol{X}_e)$, for example, means the matrix of correlations constructed with the parameters of the cheap data, correlating expensive and cheap data.

Mind that, even though we construct l different C matrices, the most expensive part of it, i.e., computing the inner correlation matrices, is done only once. This means, performing only two MLEs to find the unique parameters $\boldsymbol{\theta}_c$, p_c, $\boldsymbol{\theta}_d$, p_d and ρ that characterize $\boldsymbol{\Psi}_c$ and $\boldsymbol{\Psi}_d$, keeps the number of matrix calculations low. This is the main property that makes the AV Co-Kriging a fast and suitable tool for modeling vector-valued output systems.

Finally, we maximize the likelihood of the RVs with respect

Table 1. Modeling error and time for four different cases

	Number of point samples vector samples (expensive/cheap)	Method	Training + Testing Time [s]	Error (NRMSE)
Case 1	100/300	Kriging	102.30	31.36%
		Co-Kriging	258.04	31.50%
	5/15	AV Co-Kriging	200.07	11.26%
Case 2	200/600	Kriging	111.49	19.78%
		Co-Kriging	376.82	19.75%
	10/30	AV Co-Kriging	125.85	8.76%
Case 3	400/1000	Kriging	134.31	18.21%
		Co-Kriging	592.65	20.14%
	20/50	AV Co-Kriging	206.55	6.92%
Case 4	800/2100	Kriging	182.54	13.34%
		Co-Kriging	1791.41	12.01%
	40/105	AV Co-Kriging	164.08	6.62%

to the predicted output, given the whole data and the estimated parameters. This is done in an element-wise way, since we have a $\boldsymbol{C}_i$ for each RV, resulting in the AV Co-Kriging predictor

$$\hat{y}_{ei}(\boldsymbol{x}) = \hat{\mu}_i + \boldsymbol{c}_i^T \boldsymbol{C}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{1}\hat{\mu}_i), \tag{9}$$

where each $\boldsymbol{c}_i$ is the correlation vector between the predicted point $\hat{y}_{ei}(\boldsymbol{x})$ and the expensive and cheap data for each RV

$$\boldsymbol{c}_i = \begin{pmatrix} \rho\hat{\sigma}_{ci}^2 \boldsymbol{\psi}_c(\boldsymbol{X}_c, \boldsymbol{x}) \\ \hat{\rho}_i^2\hat{\sigma}_{ci}^2 \boldsymbol{\psi}_c(\boldsymbol{X}_e, \boldsymbol{x}) + \hat{\sigma}_{di}^2 \boldsymbol{\psi}_d(\boldsymbol{X}_e, \boldsymbol{x}) \end{pmatrix}. \tag{10}$$

After concatenation, (9) results in the vector-valued output $\hat{\boldsymbol{y}}_e := \{\hat{y}_{e1}, \hat{y}_{e2}, \cdots, \hat{y}_{el}\}$. The elements of the vector $\hat{\boldsymbol{\mu}} = \{\hat{\mu}_1, \cdots, \hat{\mu}_l\}$ are calculated with (6) using both data instead of only the cheap output matrix.

Note that, since we are using the average MLE for each data shown in (4) and (7), the (9) is not equal to applying the usual scalar Co-Kriging l times in a simple element-wise way. Here, we make use of single $\boldsymbol{\Psi}_c$ and $\boldsymbol{\Psi}_d$ for the whole vector-valued output, which can alleviate the matrix calculations tremendously, especially for a high number of sample points l per vector output.

3 Application Example

The proposed multi-output variable-fidelity approach is demonstrated for the design of a PCB connector shown in Fig. 1. This system connects two PCBs with FR4 substrate $\varepsilon_r = 4.4$, $h = 0.32mm$, $\delta = 0.02$. The chosen design variables are $\boldsymbol{x} = [s_1, h_1, r_1, r_2]$. Our goal is to build a surrogate model of the connector's $|S_{11}|$ parameter in the frequency band from 12 to 18 GHz. In this case, the $|S_{11}|$ is a vector of 20 points in the frequency dimension. The lower and upper bounds for the design parameters are $\boldsymbol{x}_{low} = [2, 1, 0.2, 1.1]$ and $\boldsymbol{x}_{up} = [6, 2.5, 0.5, 2.5]$, measured in *mm*. We perform two sets of finite element (FE) simulations, in order to compute the training data of the surrogate model. The first is the HFM with a high discretization factor ($\approx 80\,000$ tetrahedra) and the second an LFM with a coarser discretization ($\approx 30\,000$ tetrahedra) of the geometry shown in Fig. 1. The benchmark of the proposed technique is made against the usual Kriging and Co-Kriging methodologies [3]. For these usual approaches, we consider the frequency as a design parameter. Therefore, these techniques are modeling a function with five input parameters, namely $\boldsymbol{x}^* = [s_1, h_1, r_1, r_2, \omega]$, while AV Co-Kriging estimates a vector-valued function, which has four inputs, i.e., the original $\boldsymbol{x} = [s_1, h_1, r_1, r_2]$. This means that for the AV Co-Kriging the training information is made of vectors that come directly from the FE solver, and for Kriging and Co-Kriging the training samples are points, which are extracted from the vectors of the FE solver. The comparison between the methods is done for four different cases, in which we train each model with an increasing number of expensive and cheap samples, as shown in Table 1. Naturally, the Kriging model will use only the expensive data, while the other techniques use both data. Note that, for each case, we write the total number of training points and vectors. For instance, 15 cheap vector samples for the AV Co-Kriging mean 300 cheap points samples for both Kriging and Co-

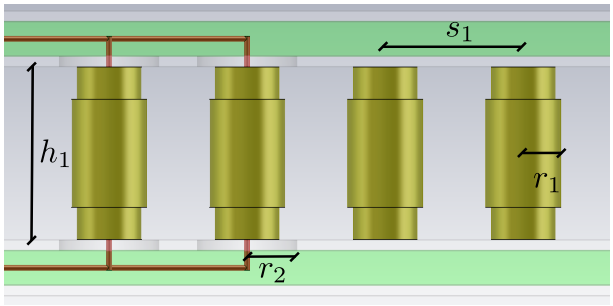

Figure 1. Geometry of the PCB connector

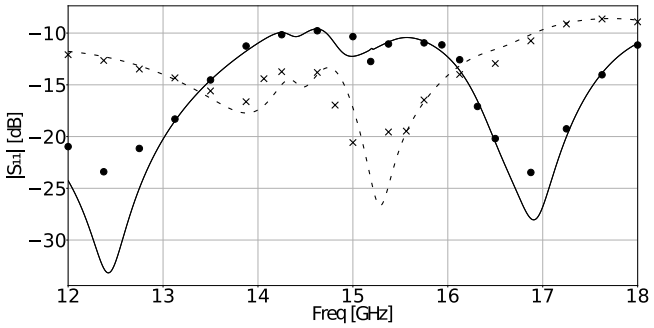

Figure 2. The responses of two geometries of the PCB connector: FE references (-) and (- -), and AV Co-Kriging (•) and (×) respectively.

Kriging since each vector output has 20 points. The normalized root mean square (NRMSE) of each approach is calculated for every situation using extra 10 vector samples, i.e., 200 points, which were not used for the training of the model. The time required by each model to perform training, i.e., the MLE calculations using a genetic algorithm, which is limited to be 200 seconds, and testing has been also calculated. These two comparison parameters are shown in Table 1. The results show an accuracy advantage of the proposed AV Co-Kriging in all cases over Kriging and Co-Kriging, since the fast parameter estimation inside the AV Co-Kriging overcome the estimation of the other methods. Note that, the usual Co-Kriging has the highest NRMSE, even when compared to the Kriging. This is due to the insufficient training time, which is especially harmful to this method. Considering bigger sets of training data, as in the last case, the proposed AV technique is the fastest one, since high-dimensional matrices become too arduous for the usual methods. Fig. 2 shows two testing vectors of the FE reference, used to calculate the NRMSE, together with their correspondent AV Co-Kriging predictions of the last case of Table 1, as means of visualization of the model's accuracy.

4 Conclusion

A methodology to build fast and accurate surrogate models of vector-valued outputs, for which both high- and low-fidelity data are available, is proposed. In order to demonstrate the functionality of the proposed approach, the frequency-dependent $|S_{11}|$ parameter, i.e., a vector output, of a PCB connector is modeled. The estimation proved to be more accurate and in some cases faster than conventional modeling techniques when applied to this microwave component design.

References

[1] F. Feng and C. Zhang and W. Na and J. Zhang and W. Zhang and Q. Zhang, "Adaptive Feature Zero Assisted Surrogate-Based EM Optimization for Microwave Filter Design" *IEEE Microwave and Wireless Components Letters*, **29**, 1, 2019, pp. 2-4, doi:10.1109/LMWC.2018.2884643.

[2] M. Ahadi and M. Kabir and S. Roy and R. Khazaka, "Fast multidimensional statistical analysis of microwave networks via stroud cubature approach" *2015 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO)*, 2015, pp. 1-3, doi:10.1109/NEMO.2015.7415047.

[3] Slawomir Koziel and Stanislav Ogurtsov and Ivo Couckuyt and Tom Dhaene, "Variable-Fidelity Electromagnetic Simulations and Co-Kriging for Accurate Modeling of Antennas" *IEEE Transactions on Antennas and Propagation*, **61**, 3, 2013, pp. 1301-1308.

[4] Slawomir Koziel and Anna Pietrenko-Dabrowska and John W. Bandler, "Computationally Efficient Performance-Driven Surrogate Modeling of Microwave Components Using Principal Component Analysis" *IEEE/MTT-S International Microwave Symposium*, 2020, pp. 68-71, doi:10.1109/IMS30576.2020.9223805.

[5] Alexander I.J. Forrester and Andy J. Keane, "Recent advances in surrogate-based optimization" *Progess in Aerospace Sciences*, **45**,2009, pp. 50-79.

[6] Jin Yuan and Kesheng Wang and Tao Yu and Minglun Fang, "Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression" *International Journal of Machine Tools and Manufacture*, **48**, 1, 2008, pp. 47 - 60, doi:https://doi.org/10.1016/j.ijmachtools.2007.07.011.

[7] W. Ruan and A. B. Milstein and W. Blackwell and E. L. Miller, "Multiple output Gaussian process regression algorithm for multi-frequency scattered data interpolation" *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 3992-3995, doi:10.1109/IGARSS.2017.8127875.

[8] M. Alvarez and N. Lawrence, "Computationally efficient convolved multiple output Gaussian processes" *J. Mach. Learn. Res.*, **12**, 2011, pp. 1459–1500.

[9] Dassault Systems "CST Studio Suite®, version 2019", https://www.3ds.com/, 2019, 3DS.