POLITECNICO DI TORINO
Repository ISTITUZIONALE

Efficient Implementation of the Vector-Valued Kernel Ridge Regression for the Uncertainty Quantification of the Scattering Parameters of a 2-GHz Low-Noise Amplifier

*Original*
Efficient Implementation of the Vector-Valued Kernel Ridge Regression for the Uncertainty Quantification of the Scattering Parameters of a 2-GHz Low-Noise Amplifier / Soleimani, N., Manfredi, P., Trinchero, R.. - ELETTRONICO. - (2023), pp. 143-146. (IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO'2023) Winnipeg, MB (CAN) 28-30 June 2023) [10.1109/NEMO56117.2023.10202518].

*Availability:*
This version is available at: 11583/2980404 since: 2023-07-26T08:36:40Z

*Publisher:*
IEEE

*Published*
DOI:10.1109/NEMO56117.2023.10202518

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*
IEEE postprint/Author's Accepted Manuscript

©2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

19 November 2026

# Efficient Implementation of the Vector-Valued Kernel Ridge Regression for the Uncertainty Quantification of the Scattering Parameters of a 2-GHz Low-Noise Amplifier

Nastaran Soleimani, Paolo Manfredi and Riccardo Trinchero
Politecnico di Torino, Torino, Italy; e-mail: {nastaran.soleimani,paolo.manfredi,riccardo.trinchero}@polito.it

***Abstract*—This paper focuses on the application of an efficient implementation of the vector-valued kernel Ridge regression (KRR) to the uncertainty quantification (UQ) of the scattering parameters of a low-noise amplifier (LNA). Specifically, the performance of the proposed technique have been investigated for the statistical assessment of the mean value, variance and probability density function (PDF) of the $S_{11}$ and $S_{21}$ parameters of a 2-GHz LNA induced by $25$ stochastic input parameters and compared with the corresponding reference results computed via a plain Monte Carlo (MC) simulation.**

***Index Terms*—Machine learning, kernel machine, vector-valued KRR, stochastic analysis, amplifiers.**

## I. Introduction

Uncertainty quantification (UQ) represents an essential task in the early-stage design of microwave components and devices, since it allows to account for the impact of process variation in modern mass-production electronics. In the above scenario, Monte Carlo (MC) simulation can be seen as the most straightforward way to deal with the inherently statistical nature of the problem at hand. The MC sampling allows analyzing the impact of random input variables on a given set of outputs of interest by considering the results of a *large* number of deterministic parametric simulations (usually in the order of thousands), in which the value of the stochastic parameters is drawn according to their probability density function (PDF). Such brute force approach turns out to be extremely robust and easy to implement within the simulation flow used by most of the commercial circuital and full-wave solvers, but it is also characterized by a low efficiency and by a non-negligible computational cost [1].

In the last decades, surrogate modeling techniques have been proposed for the UQ of the responses of electronic devices and circuits as more efficient alternatives to direct MC sampling [2]–[9]. A surrogate model is "a model of a model" acting as a closed-form and fast-to-evaluate emulator of the input-output map characterizing the true underlying model, also known as computational model. A surrogate model is built by solving a regression problem based on the information provided by a *small* set of training samples calculated via the computationally expensive computational model. After the training, the surrogate model can be used as an efficient alternative to the plain computational model in computationally expensive task such as the UQ and optimization. It is important to stress that the performance of the resulting surrogate model, in terms of accuracy with respect of the number of training samples, unavoidably depend on the model structure and the regression technique used to build it [1].

In the above framework, kernel-based machine learning regressions have shown several advantages with respect to standard regression techniques relying on basis expansion (e.g., ordinary least-squares, ridge and LASSO regressions, polynomial chaos [7]), since they allow building non-parametric surrogate models in which the number of unknowns is independent from the number of input parameters considered by the model [1], [5]. Moreover, they rely on a linear model structure in which the model unknowns can be estimated from the solution of a convex optimization problem, with several benefits in term of training time and accuracy, when a small set of training samples is available [3], [6]. On the other hand, different from artificial neural network structures, state-of-the-art kernel machine regressions are limited to scalar-output problem, thus making their direct application to vector-valued regression problems, usually encountered in microwave and electronic applications, rather cumbersome [5]. A possible way to overcome the above limitation is to combine single-output kernel machine regressions with data compression techniques, like the principal component analysis (PCA) [8]. The underlying idea is to apply a compression technique on the available data in order to reduce the number of output components to be modeled and thus the number of single-output regression to be trained. Such approach has shown promising results in UQ problems [8], [9].

As an alternative to the above techniques, this paper investigates the performance in the UQ scenario of a generalized vector-valued formulation of the kernel Ridge regression (KRR) able to directly account for the multi-output nature of the regression problem, without requiring any data manipulation or compression [4], [5], [10], [11], [12]. Specifically, this paper deals with the development an efficient implementation of the vector-valued KRR based on a diagonalization process applied on its constitutive equations [11]. The effectiveness and the performance of the proposed technique are investigated for the UQ of the scattering parameters of a low-noise amplifier (LNA) affected by $25$ stochastic parameters and compared with the ones obtained by a MC simulation.

## II. Vector-Valued Kernel Ridge Regression

Let us consider the problem of building a generic vector-valued surrogate model $\hat{\boldsymbol{f}} : \mathcal{X} \rightarrow \mathcal{Y}$, starting from the information available on the training set $\mathcal{D} = \{(\boldsymbol{x}_l, \boldsymbol{y}_l\}_{l=1}^{L}$, where $\boldsymbol{x}_l \in \mathcal{X} \subseteq \mathbb{R}^p$ represents the generic $l$-th configuration of the input parameters and $\boldsymbol{y}_l \in \mathcal{Y} \subseteq \mathbb{R}^D$ represents the corresponding vector-valued output. The above problem turns out to be equivalent to learn $D$ scalar functions $\hat{f}^{(d)} : \mathcal{X} \rightarrow \mathbb{R}$ with $d = 1, \ldots, D$, such that $\hat{\boldsymbol{f}} = [\hat{f}^{(1)}, \ldots, \hat{f}^{(D)}]^T$, minimizing the following empirical risk functional:

$$\hat{\boldsymbol{f}} = \underset{\tilde{\boldsymbol{f}} \in \mathcal{H}}{\arg\min} \sum_{d=1}^{D} \sum_{l=1}^{L} (y_l^{(d)} - \tilde{f}^{(d)}(\boldsymbol{x}_l))^2 + \lambda \|\tilde{\boldsymbol{f}}\|_{\mathcal{H}}^2 \quad (1)$$

where $\lambda$ is the regularizer hyperparameter providing a trade-off between the model flatness and accuracy on the training set, whilst $y_l^{(d)}$ and $\tilde{f}^{(d)}(\boldsymbol{x}_l)$ represent the $d$-th component of the $l$-th training output and the corresponding model prediction, respectively.

According to the representer theorem for vector-valued regression problems presented in [12], any optimal solution $\hat{\boldsymbol{f}}$ of (1) for a generic $d'$-th output component, takes the form:

$$\hat{f}^{(d')}(\boldsymbol{x}) = \sum_{d=1}^{D} \sum_{l=1}^{L} k\left((\boldsymbol{x}, d'), (\boldsymbol{x}_l, d)\right) c_{d,l}, \quad (2)$$

where $c_{d,l}$ are unknown coefficients to be estimated during the model training and $k((\boldsymbol{x}, d'), (\boldsymbol{x}_l, d)) : \mathbb{R}^{p \times p} \times \mathbb{R}^{\{1,\ldots,D\} \times \{1,\ldots,D\}} \rightarrow \mathbb{R}$ represents a *new* kernel function acting on both the input parameters $\boldsymbol{x} \in \mathbb{R}^p$ and the output dimensions $\{1, \ldots, D\}$. The above model can be rewritten in its matrix formulation, which writes:

$$\hat{\boldsymbol{f}}(\boldsymbol{x}) = \sum_{l=1}^{L} \mathbf{K}(\boldsymbol{x}, \boldsymbol{x}_l) \boldsymbol{c}_l, \quad (3)$$

where $\mathbf{K}(\cdot, \cdot) : \mathbb{R}^{p \times p} \rightarrow \mathbb{R}^{D \times D}$ is the multi-output kernel matrix and $\boldsymbol{c}_l = [c_{1,l}, \ldots, c_{D,l}]^T \in \mathbb{R}^D$ are column vectors collecting the regression unknowns.

Without loss of generality, the matrix kernel function $\mathbf{K}(\boldsymbol{x}, \boldsymbol{x}')$ can be built as a separable kernel function defined as the product between two scalar kernels acting either on the input space or output dimensions, such that:

$$\begin{aligned} [\mathbf{K}(\boldsymbol{x}, \boldsymbol{x}')]_{[d,d']} &= k((\boldsymbol{x}, d), (\boldsymbol{x}_l, d')) \\ &= k_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{x}') k_o(d, d'), \end{aligned} \quad (4)$$

where $k_{\boldsymbol{x}}$ and $k_o$ are scalar kernels acting on the input space (i.e., $k_{\boldsymbol{x}} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$) and output dimensions (i.e., $k_o : \{1, \ldots, D\} \times \{1, \ldots, D\} \rightarrow \mathbb{R}$). In our implementation, a radial basis function (RBF) kernel is used for the scalar kernel $k_{\boldsymbol{x}}$ and $k_o$, respectively.

By using (3) and (4), the empirical functional minimization in (1) can be recast in terms of the following discrete-time Sylvester equation:

$$\mathbf{K}_{\boldsymbol{x}} \mathbf{C} \mathbf{B} + \lambda \mathbf{C} = \mathbf{Y}, \quad (5)$$

where $\mathbf{K}_{\boldsymbol{x}}$ is a $L \times L$ Gram matrix computed from the input samples $\{\boldsymbol{x}_l\}_{l=1}^{L}$ (i.e., $[\mathbf{K}_{\boldsymbol{x}}]_{ij} = k_{\boldsymbol{x}}(\boldsymbol{x}_i, \boldsymbol{x}_j)$), $\mathbf{B}$ is a $D \times D$ Gram matrix computed on the output dimensions $\{1, \ldots, D\}$ (i.e., $[\mathbf{B}]_{ij} = k_o(d_i, d_j)$), $\mathbf{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_L]^T \in \mathbb{R}^{L \times D}$ is a matrix collecting the model unknowns and $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L]^T$ is a $L \times D$ matrix associated to the training output.

By using the properties of the Kronecker product, the solution of the discrete-time Sylvester in (5) can be obtained as the solution of a linear system [13]. Unfortunately, this approach leads to a *huge* linear system with $(LD) \times (LD)$ equations, whose inversion would require a computational cost proportional to $\mathcal{O}(L^3 D^3)$.

Such a prohibitive training cost can be heavily reduced by applying a diagonalization procedure on the kernel matrices $\mathbf{K}_{\boldsymbol{x}}$ and $\mathbf{B}$ [11], i.e.,

$$\mathbf{K}_{\boldsymbol{x}} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T \text{ and } \mathbf{B} = \mathbf{T} \mathbf{M} \mathbf{T}^T, \quad (6)$$

where $\mathbf{U} \in \mathbb{R}^{L \times L}$ and $\mathbf{T} \in \mathbb{R}^{D \times D}$ are matrices collecting the eigenvectors of the matrices $\mathbf{K}_{\boldsymbol{x}}$ and $\mathbf{B}$, respectively, whereas $\boldsymbol{\Lambda} \in \mathbb{R}^{L \times L}$ and $\mathbf{M} \in \mathbb{R}^{D \times D}$ are diagonal matrices collecting the corresponding eigenvalues.

Substituting (6) into the original Sylvester equation in (5), leads to the following formulation:

$$\boldsymbol{\Lambda} \tilde{\mathbf{C}} \mathbf{M} + \lambda \tilde{\mathbf{C}} = \tilde{\mathbf{Y}} \quad (7)$$

where $\tilde{\mathbf{C}} = \mathbf{U}^T \mathbf{C} \mathbf{T}$ and $\tilde{\mathbf{Y}} = \mathbf{U}^T \mathbf{Y} \mathbf{T}$ are new transformed matrices collecting a transformed version of regression unknowns and source terms.

Due to the diagonal structure of (7), a generic $ij$-entry of the unknown matrix $[\tilde{\mathbf{C}}]_{ij} = c_{ij}$ can be suitably computed via a scalar equation using the diagonal eigenvalue matrices $\boldsymbol{\Lambda}$ and $\mathbf{M}$, as:

$$\tilde{c}_{ij} = \frac{\tilde{y}_{ij}}{[\boldsymbol{\Lambda}]_{ii} [\mathbf{M}]_{jj} + \lambda}. \quad (8)$$

Once all the entries of the matrix $\tilde{\mathbf{C}}_{ij}$ has been computed via (8), the original unknown matrix $\mathbf{C}$ is reconstructed as:

$$\mathbf{C} = \mathbf{U} \tilde{\mathbf{C}} \mathbf{T}^T. \quad (9)$$

For the case of large multi-output kernel matrices, the above diagonalization procedure for solving the discrete-time Sylvester equation turns out to more efficient than the equivalent vectorized solution based on the Kronecker formulation presented in [4], [5]. Indeed, since the diagonalization is applied on the matrices $\mathbf{K}_{\boldsymbol{x}}$ and $\mathbf{B}$ separately, the overall computational cost required for the model training reduces from $\mathcal{O}(L^3 D^3)$ to $\mathcal{O}(L^3 + D^3 + L^2 D + L D^2)$, with a beneficial effect on the training time when the product $L \times D$ is large.

## III. Application Example: 2-GHz LNA

The performance of the proposed technique has been evaluated on the UQ of the scattering parameters of the 2-GHz LNA shown in Fig. 1 by considering 25 Gaussian stochastic variables affecting the parasitic resistances, capacitances and inductances of the BJT, its forward current gain,

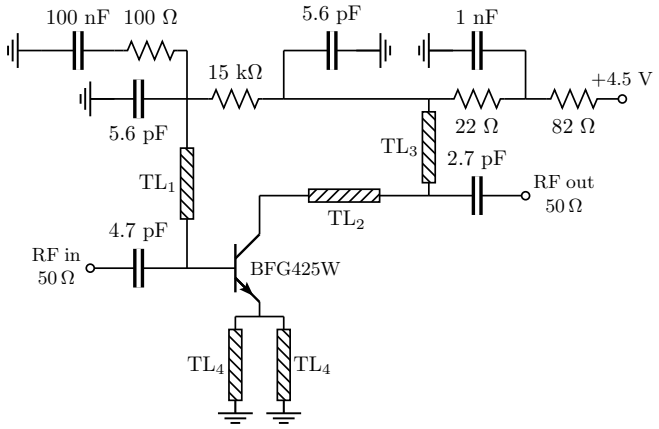



Fig. 1. 2-GHz BJT LNA.

TABLE I
TRAINING TIME AND RELATIVE L2-ERROR COMPUTED FROM THE FREQUENCY-DOMAIN SAMPLES OF THE PREDICTIONS OF THE SURROGATE MODELS TRAINED WITH AN INCREASING NUMBER OF TRAINING SAMPLES VIA THE PROPOSED IMPLEMENTATION OF THE VECTOR-VALUED KRR FOR THE $S_{11}$ AND $S_{21}$ PARAMETERS BY CONSIDERING 1000 TEST SAMPLES.

| Parameter | $L = 30$ | | $L = 50$ | | $L = 100$ | |
|---|---|---|---|---|---|---|
| | $t_{\text{train}}$ | L2-error | $t_{\text{train}}$ | L2-error | $t_{\text{train}}$ | L2-error |
| $S_{11}$ | 74s | 6.15% | 115s | 3.58% | 205s | 2.70% |
| $S_{21}$ | 78s | 1.96% | 118s | 1.54% | 216s | 1.16% |

all the lumped components in the amplifier schematic, and the widths of the microstrip lines, each with a 10% relative standard deviation [7], [9]. The considered test case has been implemented as a parametric small-signal ac analysis in HSPICE. Such implementation allows computing the two-port scattering parameters $S_{11}$ and $S_{21}$ of the LNA at 201 frequency points for any configuration of the 25 random parameters. The HSPICE simulations have been used to generate three training sets with an increasing number of samples (i.e., $L = 30$, 50, and 100) based on a latin hypercube sampling scheme and to run a 1000-sample MC simulation that is used hereafter as a reference for the proposed statistical analysis. Two independent surrogate models, one for $S_{11}$ and one for $S_{21}$, have been trained via the proposed implementation of the vector-valued KRR presented in Section II.

Table I reports the performance of the obtained surrogate models in terms of training time and relative L2-norm error computed on a test set collecting the results of a 1000-sample MC simulation for the parameters $S_{11}$ and $S_{21}$ by considering an increasing number of the training samples (i.e., $L = 30$, 50, and 100). The results show a constant reduction of the model error (i.e., the relative L2-error) with respect to the number of training samples (i.e., $L$), thus highlighting the capability of the proposed vector-valued KRR of learning the actual information provided by the training set. Concerning the computational cost, the training time required to build the proposed vector-valued surrogate models is less than 7 min for all the considered modeling scenarios. After the

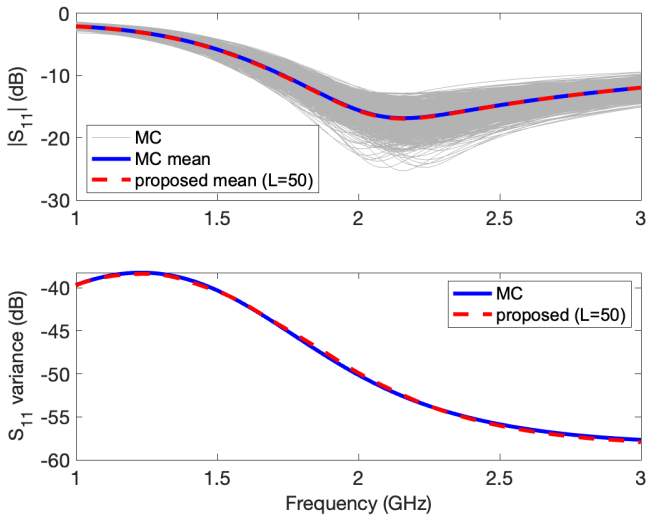



Fig. 2. $S_{11}$ parameter of the LNA in Fig. 1. Top panel: the gray lines show the magnitude of the $S_{11}$ computed from a 1000-sample MC simulation. The blue solid and the red dashed lines are the magnitude of the average $S_{11}$ obtained from the MC samples and the proposed model, respectively. Bottom panel: the blue solid and the red dashed lines are the magnitude of the $S_{11}$ variance obtained from the MC samples and the proposed models, respectively.

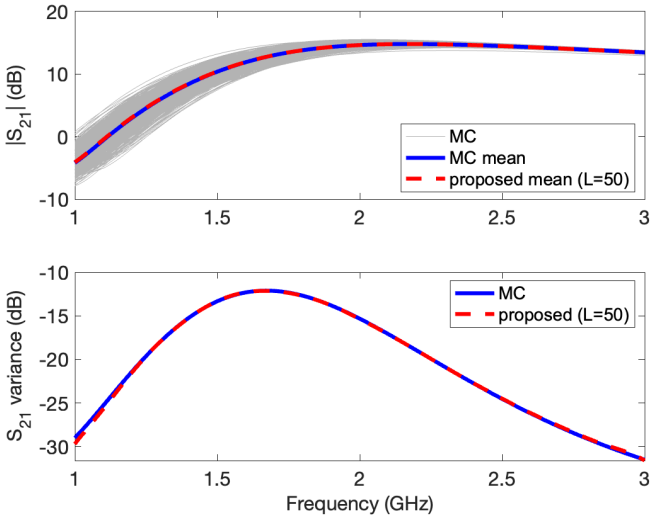



Fig. 3. $S_{21}$ parameter of the LNA in Fig. 1. Top panel: the gray lines show the magnitude of the $S_{21}$ computed from a 1000-sample MC simulation. The blue solid and the red dashed lines are the magnitude of the average $S_{21}$ obtained from the MC samples and the proposed models, respectively. Bottom panel: the blue solid and the red dashed lines are the magnitude of the $S_{21}$ variance obtained from the MC samples and the proposed models, respectively.

training, the evaluation of the obtained model on the 1000 test samples required less than 20 s, while the corresponding MC simulation requires 1131 s.

Figures 2 and 3 compare the mean (top panels) and the variance (bottom panels) of the absolute value of the $S_{11}$ and $S_{21}$ parameters predicted by the proposed surrogate models trained with $L = 50$ training samples against the corresponding results computed via the MC samples. The results

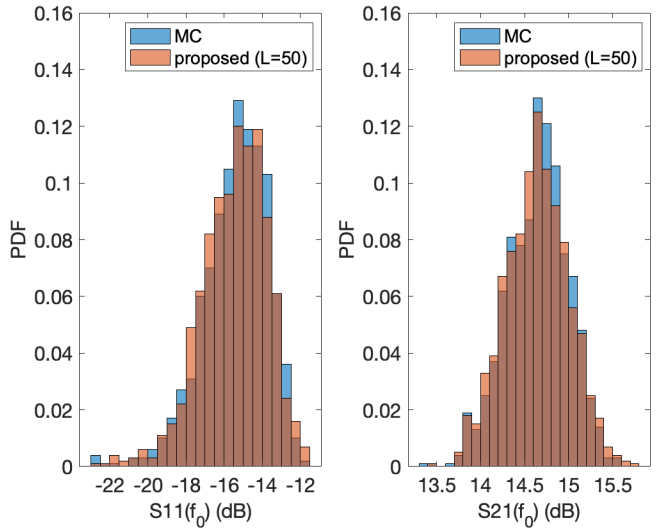



Fig. 4. Comparison of among the PDFs of the $S_{11}$ and $S_{21}$ parameters at the frequency $f_0 = 2\,\text{GHz}$ computed from the predictions of the surrogate models built via the proposed vector-valued KRR evaluated on 1000 test samples and the corresponding ones computed from a 1000-sample MC simulation.

show an excellent agreement between the mean value and the variance predicted by the proposed surrogate model and the ones computed from the MC samples, being the two sets of curves almost perfectly overlapped. Furthermore, Fig. 4 shows the probability density function (PDF) of the $S_{11}$ and $S_{21}$ parameters computed at $f_0 = 2\,\text{GHz}$. Also in this case, the histograms predicted by the proposed models are in perfect agreement with the corresponding ones calculated from the MC simulation, thus confirming the excellent performance of the proposed approach in the UQ scenario.

## IV. Conclusions

This paper investigated the effectiveness and the accuracy of a surrogate model built via an efficient implementation of the vector-valued KRR for the UQ of the scattering parameter of a LNA. Specifically, the performance of the proposed technique have been investigated for the statistical assessment of the mean value, variance, and PDF of the $S_{11}$ and $S_{21}$ parameters of a 2-GHz LNA induced by 25 stochastic parameters and compared with the corresponding reference results computed via a MC simulation.

## References

[1] R. Trinchero and F. Canavero, "Machine Learning Regression Techniques for the Modeling of Complex Systems: An Overview," *IEEE Electromagnetic Compatibility Magazine*, vol. 10, no. 4, pp. 71-79, 4th Quarter 2021.

[2] R. Trinchero and F. G. Canavero, "Modeling of eye diagram height in high-speed links via support vector machine," in Proc. of 2018 IEEE 22nd Workshop on Signal and Power Integrity (SPI), Brest, 2018, pp. 1–4.

[3] N. Soleimani and R. Trinchero, "Compressed complex-valued least squares support vector machine regression for modeling of the frequency-domain responses of electromagnetic structures," *Electronics*, vol. 11, no. 4, 2022.

[4] N. Soleimani, R. Trinchero and F. Canavero, "Vector-Valued Kernel Ridge Regression for the Modeling of High-Speed Links," in Proc. IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO 2022), Limoges (France), July 6-8.

[5] N. Soleimani, R. Trinchero and F. Canavero, "Bridging the Gap Between Artificial Neural Networks and Kernel Regressions for Vector-Valued Problems in Microwave Applications," *IEEE Transactions on Microwave Theory and Techniques*, doi: 10.1109/TMTT.2022.3232895 (early access).

[6] S. Kushwaha, et. al,"Comparative Analysis of Prior Knowledge-Based Machine Learning Metamodels for Modeling Hybrid Copper–Graphene On Chip Interconnects," *IEEE Transactions on Electromagnetic Compatibility*, vol. 64, no. 6, pp. 2249-2260, Dec. 2022.

[7] P. Manfredi, D. Vande Ginste, D. De Zutter and F. G. Canavero, "Generalized Decoupled Polynomial Chaos for Nonlinear Circuits With Many Random Parameters," *IEEE Microwave and Wireless Components Letters*, vol. 25, no. 8, pp. 505-507, Aug. 2015.

[8] P. Manfredi and R. Trinchero, "A data compression strategy for the efficient uncertainty quantification of time-domain circuit responses," *IEEE Access*, vol. 8, pp. 92019–92027, 2020.

[9] P. Manfredi, "Probabilistic Uncertainty Quantification of Microwave Circuits Using Gaussian Processes," *IEEE Transactions on Microwave Theory and Techniques*, doi: 10.1109/TMTT.2022.3228953 (early access).

[10] A. Mauricio, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," *Foundations and Trends in Machine Learning* 4(3) pp. 195-266, 2012.

[11] L. Baldassarre, et al., "Multi-output learning via spectral filtering", *Machine Learning*, vol. 87, pp. 259–301, 2012.

[12] C.A. Micchelli and M. Pontil, "On learning vector–valued functions," *Neural Computation*, vol. 17, pag. 177-–204, 2005.

[13] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge: Cambridge University Press, 1991.