Compressed Vector-Valued Kernel Ridge Regression for Complex-Valued Parametric Modeling of High-Speed Interconnects

(Article begins on next page)

Compressed Vector-Valued Kernel Ridge Regression for Complex-Valued Parametric Modeling of High-Speed Interconnects

Nazanin Soleimani*, Igor S. Stievano, Riccardo Trinchero
Dept. Electronics and Telecommunications, Politecnico di Torino, Italy
* corresponding: {nazanin.soleimani@polito.it}

***Abstract*—This paper investigates a compressed formulation of Vector-Valued Kernel Ridge Regression (VV-KRR) for complex-valued modeling. The method is applied to the parametric modeling of the frequency response of a high-speed interconnect, as a function of four design parameters. Its effectiveness is assessed in terms of prediction accuracy and model compactness and compared with an uncompressed formulation.**

***Index Terms*—Signal and power integrity, high-speed interconnects, machine learning, kernel regression, surrogate model.**

I. Introduction

Machine Learning (ML) techniques are increasingly being adopted in the field of signal and power integrity for the parametric modeling of high-speed interconnects [1]–[4]. They provide accurate and computationally efficient surrogates for circuit and full-wave solvers [5], allowing the construction of fast-to-evaluate models capable of predicting the parametric behavior of the frequency response characterizing complex interconnect structures. Once trained, such data-driven surrogate models enable rapid evaluations that significantly accelerate design-space exploration, optimization, and variability analysis, which would otherwise be computationally prohibitive [6].

Among the available ML approaches, kernel methods are particularly attractive due to their convex training formulations, which make them data-efficient and simple to train [7], [8]. However, conventional kernel formulations are typically designed for scalar and real-valued data, limiting their applicability to multi-output and complex-valued problems. that frequently arise in electronic applications [1]–[3].

Complex-valued data are common in electronic applications for frequency-domain analysis (e.g., frequency responses, S responses) [1]–[4]. A common strategy to extend real-valued ML techniques to complex problems is the dual-channel formulation [1], [9], where the complex data are split into their real and imaginary components, resulting in two uncorrelated real-valued regressions. Although this approach enables the direct use of standard real-valued methods, it ignores the intrinsic correlation between the real and imaginary parts and, when applied to vector-valued problems, requires training a large number of independent single-output models.

To address these limitations, several complex-valued formulations have been proposed for various ML techniques, including Artificial Neural Networks (ANN) [10], Support Vector Machines (SVM) [11], kernel least-squares regression [9], Least-Squares SVM (LS-SVM) [1], [4], and Gaussian Process Regression (GPR) [2], [12]. However, for kernel-based methods such as SVM, LS-SVM, and GPR, the design of suitable complex-valued kernels remains challenging, as they are mathematically more involved than their real-valued counterparts and often require the tuning of multiple hyperparameters [1], [4].

As a result, complex-valued formulations are mainly used when detailed system information is available (e.g., when the response can be expressed in terms of poles and residues), enabling the design of physics-based custom kernels [2], [4]. For more general problems, the dual-channel approach remains a practical and widely adopted choice [13].

This work explores an alternative strategy based on the Vector-Valued Kernel Ridge Regression (VV-KRR) framework [3], [14], where the real and imaginary parts of complex-valued data are combined into a single vector-valued output learned by one model. This formulation captures correlations among output components, implicitly accounting for possible coupling between the real and imaginary parts.

Although multi-output formulations such as VV-KRR effectively exploit dependencies among outputs, their training cost scales rapidly with both the number of samples and the output dimensions [3]. To address this limitation, this work investigates a compressed VV-KRR implementation for the parametric modeling of complex-valued data derived from the frequency response of a high-speed interconnect. The proposed approach combines automatic, data-driven output-kernel learning with a compressed training scheme that reduces memory and computational requirements while maintaining high prediction accuracy. The effectiveness of the compressed modeling scheme is assessed in terms of prediction accuracy and number of regression coefficients for an increasing number of training samples, and compared against a plain uncompressed formulation.

II. Compressed Vector-Valued Kernel Ridge Regression & Complex-Valued Problems

Here, the VV-KRR framework is employed to learn a vector-valued map $\hat{\mathbf{f}} : \mathcal{X} \to \mathcal{Y}$, which serves as the surrogate model for the complex-valued data under study.

The dataset is defined as $\mathcal{D} = \{(\mathbf{x}_l, \mathbf{y}_l)\}_{l=1}^{L}$, where each input $\mathbf{x}_l \in \mathcal{X} \subseteq \mathbb{R}^p$ is real-valued, while every complex output $\mathbf{y}^C \in \mathcal{Y} \in \mathbb{C}^D$ is reformulated in real coordinates through

$$\mathbf{y}_l = \begin{bmatrix} \mathrm{Re}\{\mathbf{y}_l^C\} \\ \mathrm{Im}\{\mathbf{y}_l^C\} \end{bmatrix} \in \mathbb{R}^{2D}.$$

This conversion provides a straightforward real-valued representation of the complex outputs and allows the learning problem to be handled within standard vector-valued settings [3], [15]. The augmented dimensionality ($2D$ instead of D) does not alter the structure of the problem but simply makes both real and imaginary components accessible to the model.

Under this formulation, learning amounts to estimating $2D$ real-valued functions $\hat{f}^{(d)} : \mathcal{X} \to \mathbb{R}$, $d = 1, \ldots, 2D$, by minimizing the empirical risk

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

where λ is the regularization parameter used to balance data fidelity and model smoothness.

According to [15], the solution of (1) can be written in the compact vector-valued form:

$$\hat{\mathbf{f}}(\mathbf{x}) = \sum_{l=1}^{L} k_{\mathbf{x}}(\mathbf{x}, \mathbf{x}_l)\, \mathbf{B}\, \mathbf{c}_l, \qquad (2)$$

where $\mathbf{c}_l \in \mathbb{R}^{2D}$ are trainable coefficient vectors. The scalar kernel $k_{\mathbf{x}}$ acts on the input space and may be chosen among classical options such as Gaussian or polynomial kernels depending on the application, whereas $\mathbf{B} \in \mathbb{R}^{2D \times 2D}$ is a positive semidefinite matrix describing correlations among output components, including interactions between real and imaginary parts. In practice, $\mathbf{B}$ can either be specified a priori based on domain knowledge or estimated from the empirical covariance structure of the outputs [16].

Using the separable structure in (2), training reduces to solving the discrete Sylvester equation

$$\mathbf{K}_{\mathbf{x}} \mathbf{C} \mathbf{B} + \lambda\, \mathbf{C} = \mathbf{Y}, \qquad (3)$$

where $\mathbf{K}_{\mathbf{x}} \in \mathbb{R}^{L \times L}$ is the input Gram matrix (with entries $[\mathbf{K}_{\mathbf{x}}]_{ij} = k_{\mathbf{x}}(\mathbf{x}_i, \mathbf{x}_j)$), $\mathbf{C} \in \mathbb{R}^{L \times 2D}$ contains the coefficient vectors $\mathbf{c}_l$ and $\mathbf{Y} \in \mathbb{R}^{L \times 2D}$ stacks the real-valued training outputs. In practice, this linear matrix equation clearly highlights how the interaction between input and output kernels influences the model coefficients.

To reduce the computational cost associated with solving (3), low-rank approximations of the input and output kernel matrices are introduced:

$$\mathbf{K}_{\mathbf{x}} \approx \mathbf{U}_r \boldsymbol{\Lambda}_r \mathbf{U}_r^{\mathsf{T}}, \qquad \mathbf{B} \approx \mathbf{T}_n \mathbf{M}_n \mathbf{T}_n^{\mathsf{T}}. \qquad (4)$$

The matrices $\mathbf{U}_r$ and $\mathbf{T}_n$ contain the dominant eigenvectors of $\mathbf{K}_{\mathbf{x}}$ and $\mathbf{B}$, respectively, while $\boldsymbol{\Lambda}_r$ and $\mathbf{M}_n$ store the associated eigenvalues. Truncating to the leading components retains the most relevant spectral structure of the kernels and significantly reduces the dimensionality of the problem. In practice, the truncation ranks r and n can be selected so as to preserve a prescribed percentage (e.g., 99%) of the cumulative spectral energy of $\mathbf{K}_{\mathbf{x}}$ and $\mathbf{B}$, respectively, thereby keeping the approximation error negligible.

The coefficient matrix is accordingly projected onto these reduced subspaces:

$$\mathbf{C} \approx \mathbf{U}_r\, \widetilde{\mathbf{C}}_{r \times n}\, \mathbf{T}_n^{\mathsf{T}}. \qquad (5)$$

This representation allows the model to operate within the most informative regions of both the input and output spaces, while dramatically reducing the number of unknowns.

Substituting (4) and (5) into (3) equation leads to

$$\begin{aligned} &\left(\mathbf{U}_r \boldsymbol{\Lambda}_r \mathbf{U}_r^{\mathsf{T}}\right)\left(\mathbf{U}_r \widetilde{\mathbf{C}}_{r \times n} \mathbf{T}_n^{\mathsf{T}}\right)\left(\mathbf{T}_n \mathbf{M}_n \mathbf{T}_n^{\mathsf{T}}\right) \\ &\quad + \lambda\, \mathbf{U}_r \widetilde{\mathbf{C}}_{r \times n} \mathbf{T}_n^{\mathsf{T}} = \mathbf{Y}. \end{aligned} \qquad (6)$$

Multiplying from the left by $\mathbf{U}_r^{\mathsf{T}}$ and from the right by $\mathbf{T}_n$ and using orthogonality properties yields the reduced Sylvester system

$$\boldsymbol{\Lambda}_r\, \widetilde{\mathbf{C}}_{r \times n}\, \mathbf{M}_n + \lambda\, \widetilde{\mathbf{C}}_{r \times n} = \widetilde{\mathbf{Y}}, \qquad (7)$$

where $\widetilde{\mathbf{Y}} = \mathbf{U}_r^{\mathsf{T}} \mathbf{Y} \mathbf{T}_n$ is the projected target matrix.

Since both $\boldsymbol{\Lambda}_r$ and $\mathbf{M}_n$ are diagonal, the entries of $\widetilde{\mathbf{C}}_{r \times n}$ can be computed in closed form as [16]:

$$\widetilde{c}_{ij} = \frac{\widetilde{y}_{ij}}{\lambda_i m_j + \lambda}, \qquad (8)$$

for $i = 1, \ldots, r$ and $j = 1, \ldots, n$.

This yields all entries of $\widetilde{\mathbf{C}}_{r \times n}$ without requiring matrix inversion and provides a numerically stable solution.

Finally, the prediction for a new input $\boldsymbol{x}^*$ is computed using

$$\hat{\mathbf{f}}(\boldsymbol{x}^*)^{\mathsf{T}} = \mathbf{k}_{\mathbf{x}}(\boldsymbol{x}^*)\, \mathbf{U}_r\, \widetilde{\mathbf{C}}_{r \times n}\, \mathbf{T}_n^{\mathsf{T}} \mathbf{B}, \qquad (9)$$

where $\mathbf{k}_{\mathbf{x}}(\boldsymbol{x}^*) = \begin{bmatrix} k_{\mathbf{x}}(\boldsymbol{x}^*, \mathbf{x}_1) & \cdots & k_{\mathbf{x}}(\boldsymbol{x}^*, \mathbf{x}_L) \end{bmatrix} \in \mathbb{R}^{1 \times L}$ is the row vector of kernel evaluations between $\boldsymbol{x}^*$ and the training inputs. This formulation allows efficient model evaluation while exploiting the reduced structure of both the input and output kernels.

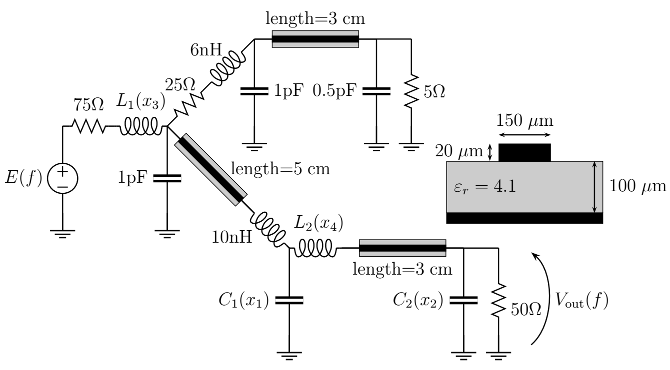

Fig. 1. Schematic of the considered parametric high-speed interconnect link.

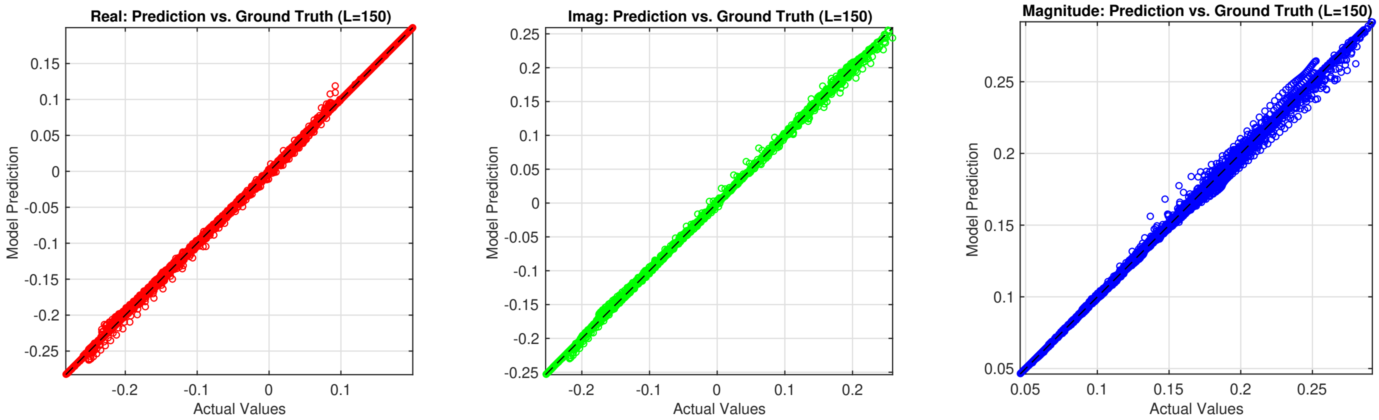

Fig. 2. Scatter plots comparing the reference data and model predictions for the real (left panel), imaginary (center panel), and magnitude (right panel) components of the frequency response. The surrogate model was trained with $L = 150$ samples. The magnitude scatter plot is obtained by comparing the reference magnitude with the prediction computed from the real and imaginary parts estimated by the model.

III. Application Example

The proposed compressed formulation of the VV-KRR has been employed to construct a surrogate model capable of reproducing the parametric behavior of the frequency response of the high-speed link shown in Fig. 1, in terms of both its real and imaginary components. The considered structure represents a typical example of a signal distribution network on a printed circuit board (PCB).

The capacitances $C_1(x_1) = (1 \pm 0.5x_1)\,$pF and $C_2(x_2) = (0.5 \pm 0.25x_2)\,$pF, together with the inductances $L_1(x_3) = (10 \pm 5x_3)\,$nH and $L_2(x_4) = (10 \pm 5x_4)\,$nH, are treated as bounded variables exhibiting a $\pm 50\%$ variation around their nominal values. These quantities are defined in terms of four uniformly distributed random variables $\boldsymbol{x} = [x_1, x_2, x_3, x_4]^\mathsf{T} \in [-1, 1]^4$.

The goal is to investigate the effect of these parameters on the real and imaginary parts of the frequency response, defined as the ratio between the phasors of the output and input voltages:

$$y(f; \boldsymbol{x}) = \frac{\hat{V}_{out}(f; \boldsymbol{x})}{\hat{E}(f)}. \tag{10}$$

The parametric structure was analyzed through Monte Carlo (MC) simulations carried out in MATLAB. A total of 1000 random configurations were generated, providing, for each parameter realization $\boldsymbol{x}$, the corresponding frequency response within the $1\,$MHz–$2\,$GHz bandwidth, sampled at $D = 500$ logarithmically spaced frequency points.

The resulting responses were standardized and divided into training and test sets after stacking their real and imaginary parts, with an additional zero-padding of 10 samples. A subset of L realizations was then used to train the VV-KRR model, with a number of output dimension of $2 \times 500 + 10 = 1010$, which is slightly larger than $2D$, corresponding to twice the number of frequency samples plus the zero padding.

The automatic relevance determination (ARD) Matérn 5/2 kernel was employed as the input kernel $k_{\boldsymbol{x}}$, while the empirical correlation matrix computed over the training samples

TABLE I
Comparison between the plain and proposed compressed implementations of VV-KRR in terms of relative L_2 error and number of regression coefficients for an increasing number of training samples (L).

	Metric	Proposed	Full
$L = 50$	ε_{L2}	0.70%	1.13%
	Coeff.	882	2350
$L = 100$	ε_{L2}	0.32%	0.39%
	Coeff.	1740	5600
$L = 150$	ε_{L2}	0.19%	0.17%
	Coeff.	2751	8700

was used as the output kernel, i.e., $\mathbf{B} = \mathrm{corr}(\mathbf{Y})$ [16]. Three surrogate models were trained using an increasing number of training samples ($L = 50$, 100, and 150), randomly selected from the dataset, and their performance was evaluated on a common test set of 500 unseen samples.

Table I summarizes the relative L_2 error ε_{L2} of the frequency-response magnitude, computed from the overall real and imaginary parts, and the number of regression coefficients used by the models trained with both the proposed compressed VV-KRR and the full (uncompressed) VV-KRR formulation. The latter was obtained by simply discarding the zero singular values of the Gramian kernel matrices $\mathbf{K}_{\boldsymbol{x}}$ and $\mathbf{B}$.

The results clearly highlight the advantages of the proposed compressed formulation, both in terms of model compactness (i.e., the number of coefficients to be computed and stored during training) and, in some cases, even in terms of prediction accuracy. In particular, the proposed compression approach, based on diagonalization and eigenvalue truncation, removes negligible eigenvalues, thereby improving the conditioning of the training problem in (3) and consequently enhancing the numerical stability of the solution and the tuning of the model hyperparameters compared to the *full* formulation. Regarding training cost, the training time with $L = 150$ (worst-case scenario) for the compressed complex-valued model is approximately $30\,$s, which is comparable to that of the full model

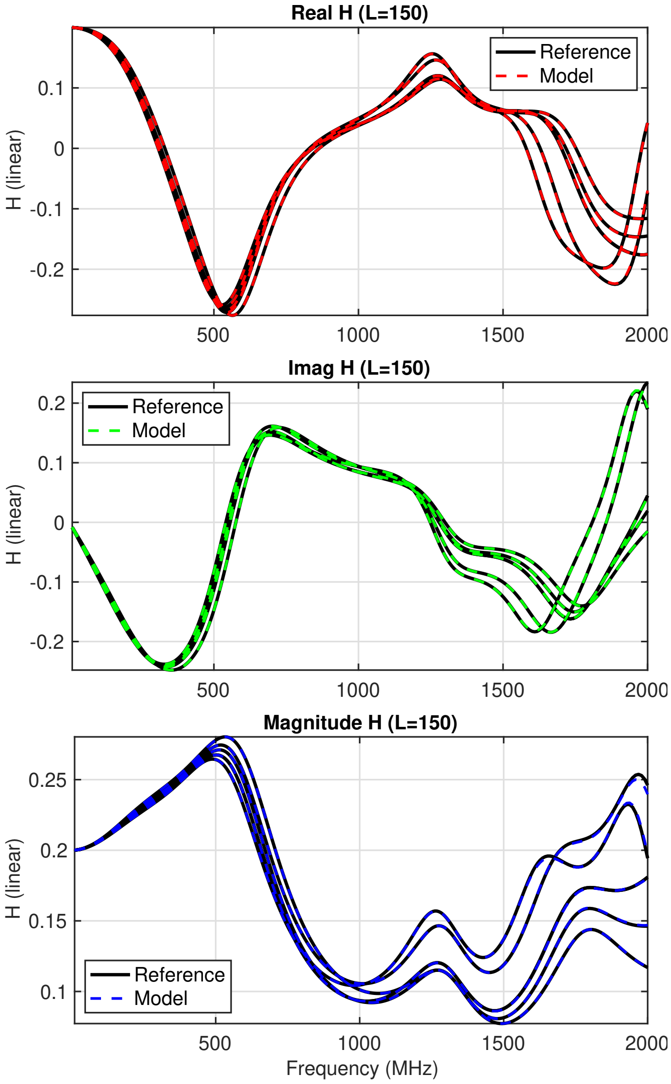

Fig. 3. Parametric plots comparing the reference data and model predictions for the real (top panel), imaginary (middle panel), and magnitude (bottom panel) components of the frequency response. The surrogate model was trained with $L = 150$ samples, and the plots correspond to five different realizations of the input parameters in the test set.

trained. In both cases, the prediction time is around 10 ms.

As a further validation, Fig. 2 shows the scatter plots comparing the real, imaginary, and magnitude components of the frequency responses predicted by the model for $L = 150$ with the corresponding reference data. The magnitude values are obtained in post-processing from the predicted real and imaginary parts. Moreover, Fig. 3 presents parametric comparisons of the real, imaginary, and magnitude components of the frequency responses predicted by the proposed model trained with $L = 150$ against the reference results for five different configurations in the test set. In this case as well, the magnitude values are obtained in post-processing from the predicted real and imaginary parts. The results confirm the accuracy of the proposed approach and are consistent with the errors reported in Table I.

IV. Conclusions

This paper investigated the performance of a preliminary compressed formulation of the Vector-Valued Kernel Ridge Regression (VV-KRR) in complex-valued regression problems. The proposed modeling approach was applied to the parametric modeling of the frequency response of high-speed interconnects characterized by four design parameters. The results demonstrate that the proposed VV-KRR models can achieve accurate predictions while maintaining compact and computationally efficient models.

References

[1] 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.

[2] Y. Lindemans, T. Ullrick, I. Couckuyt, D. Deschrijver, and T. Dhaene, "Bayesian optimization of microwave filters: A physics-informed approach using the szegő kernel," in *Proc. IEEE Workshop on Signal and Power Integrity (SPI)*, Gaeta, Italy, 2025, pp. 1–4.

[3] 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*, vol. 71, no. 6, pp. 2319–2332, Jun. 2023.

[4] F. Treviso, R. Trinchero, and F. G. Canavero, "Multiple delay identification in long interconnects via ls-SVM regression," *IEEE Access*, vol. 9, pp. 39 028–39 042, 2021.

[5] 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, 2021, 4th Quarter.

[6] K.-T. Fang, R. Li, and A. Sudjianto, *Design and Modeling for Computer Experiments*. London, U.K.: Taylor & Francis, 2016.

[7] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York, NY, USA: Springer, 1999.

[8] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.

[9] R. Boloix-Tortosa, J. J. Murillo-Fuentes, and S. A. Tsaftaris, "The generalized complex kernel least-mean-square algorithm," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5213–5222, Oct. 2019.

[10] A. Hirose, *Complex-Valued Neural Networks*. Hoboken, NJ, USA: Wiley–IEEE Press, 2013.

[11] P. Bouboulis, S. Theodoridis, C. Mavroforakis, and L. Evaggelatou-Dalla, "Complex support vector machines for regression and quaternary classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 6, pp. 1260–1274, Jun. 2015.

[12] R. Boloix-Tortosa, J. J. Murillo-Fuentes, F. J. Payán-Somet, and F. Pérez-Cruz, "Complex gaussian processes for regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5499–5511, 2018.

[13] S. Scardapane, S. Van Vaerenbergh, A. Hussain, and A. Uncini, "Complex-valued neural networks with nonparametric activation functions," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 2, pp. 140–150, 2018.

[14] V. Sindhwani, H. Q. Minh, and A. C. Lozano, "Scalable matrix-valued kernel learning for high-dimensional nonlinear multivariate regression and granger causality," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2013, pp. 1–10.

[15] C. A. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural Computation*, vol. 17, no. 1, pp. 177–204, 2005.

[16] N. Soleimani and R. Trinchero, "Efficient implementation of the vector-valued kernel ridge regression for the parametric modeling of the frequency-response of a high-speed link," in *Proceedings of the URSI General Assembly and Scientific Symposium (URSI GASS)*, Sapporo, Japan, 2023, pp. 1–4.