Low-Overhead CSI Prediction via Gaussian Process Regression

Syed Luqman Shah, *Graduate Student Member, IEEE*, Nurul Huda Mahmood, *Member, IEEE*, and Italo Atzeni, *Senior Member, IEEE*

***Abstract*—Accurate channel state information (CSI) is critical for current and next-generation multi-antenna systems. Yet conventional pilot-based estimators incur prohibitive overhead as antenna counts grow. In this letter, we address this challenge by developing a novel framework based on Gaussian process regression (GPR) that predicts full CSI from only a few observed entries, thereby reducing pilot overhead. The correlation between data points in GPR is defined by the covariance function, known as kernel. In the proposed GPR-based CSI estimation framework, we incorporate three kernels, i.e., radial basis function, Matérn, and rational quadratic, to model smooth and multi-scale spatial correlations derived from the antenna array geometry. The proposed approach is evaluated across two channel models with three distinct pilot probing schemes. Results show that the proposed GPR with 50% pilot saving achieves the lowest prediction error, the highest empirical 95% credible-interval coverage, and the best preservation of spectral efficiency relative to the benchmarks.**

***Index Terms*—MIMO channel estimation, Gaussian process regression, reduced training overhead, covariance modeling.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) systems are expected to play a pivotal role in 6G wireless networks, but their promised gains rely on accurate channel state information (CSI) [1]. The pilot overhead required by classical estimators, such as least squares (LS) and minimum mean squared error (MMSE), scales unfavorably with the number of transmit antennas [2]. Compressed sensing (CS) techniques reduce this burden by exploiting channel sparsity, though they suffer from power leakage due to off-grid points [2], [3]. For instance, low-rank matrix completion mitigates overhead through methods like nuclear-norm minimization or the alternating direction method of multipliers [4], including extensions with subgradient descent and joint nuclear- and ℓ_1-norm formulations [2]. However, these approaches are sensitive to rank errors, noise, and violations of low-rank assumptions in rich-scattering environments.

In contrast, Gaussian process regression (GPR) is a nonparametric, Bayesian estimator that provides calibrated uncertainty [5]. The covariance function provides domain knowledge, so GPR is agnostic to sparsity and bandlimitedness [4]. GPR exploits joint transmit-receive correlations and captures the nonseparable structure observed in measured channels [6]. Its performance does not depend on the restricted isometry property or low coherence [3], and it remains effective with equispaced or hardware-constrained activations [7]. Unlike CS, GPR requires only a second-order description, making it resilient to off-grid leakage, diffuse multipath, and rank/sparsity variability [2]. It also returns per-entry posterior uncertainty, enabling principled pilot budgeting, reliability targeting, and link adaptation: these capabilities are absent in interpolation-based and standard CS solvers. Recent works have used GPR, e.g., to incorporate electromagnetic priors into statistical channel models [6], predict interference power for proactive 6G resource allocation [5], design spatio-temporal kernels grounded in electromagnetic theory to improve CSI prediction [8], and model distortion induced by hardware impairments to enhance channel estimation [7]. Taken together, GPR learns spatial correlations from limited observations and quantifies uncertainty, enabling informed pilot placement and reconstruction confidence.

This letter addresses spatial interpolation in MIMO systems with limited pilot measurements. We formulate the CSI prediction as a Gaussian process (GP) inference, enabling full channel reconstruction with quantified uncertainty. A GPR framework is developed to reduce pilot overhead by estimating the complete CSI from a small subset of pilot observations. The correlation between data points in GPR are defined by the covariance function, known as kernel, which captures the smooth and multi-scale spatial correlations inherent in wireless channels. We investigate three geometric kernels in our proposed GPR framework, i.e., radial basis function (RBF), Matérn, and rational quadratic (RQ). As an additional contribution, we propose a novel antenna probing design to reduce the pilot length in each coherence interval.

Our proposed approach is validated on the Kronecker [9] and Weichselberger [10] channel models, with three reduced-pilot antenna-probing schemes assessed for each. Although these channel models naturally align with the GP prior, we prove that the resulting GP posterior mean is equivalent to the best linear unbiased predictor (BLUP) [11], without assuming channel Gaussianity. Simulation results under both models demonstrate that our proposed framework can achieve lower prediction error, higher empirical 95% credible-interval coverage across all kernels, and the closest spectral efficiency (SE) corresponding to the true channel compared to conventional LS and MMSE estimators while enabling up to 50% reduction in pilot overhead without compromising the SE.

Notations: Bold lowercase and uppercase letters denote vectors and matrices, respectively. $\mathbf{I}_P$ denotes the $P \times P$

Received 21 October 2025; revised 20 November 2025; accepted 21 December 2025. Date of publication 26 December 2025; date of current version 2 January 2026. This work was supported by the Research Council of Finland under Grant 336449 Profi6, Grant 348396 HIGH-6G, Grant 359850 6G-ConCoRSe, and Grant 369116 6G Flagship. The associate editor coordinating the review of this article and approving it for publication was M. Chen. *(Corresponding author: Syed Luqman Shah.)*

The authors are with the Centre for Wireless Communications, University of Oulu, 90014 Oulu, Finland (e-mail: syed.luqman@oulu.fi; nurul-huda.mahmood@oulu.fi; italo.atzeni@oulu.fi).

Data is available on-line at https://github.com/Syed-Luqman-Shah-19/MIMOGPR1.

Digital Object Identifier 10.1109/LWC.2025.3648532

identity matrix and $\mathbf{e}_i$ is the ith canonical basis vector. $\mathrm{vec}(\cdot)$ denotes vectorization. $(\cdot)^{\mathsf{H}}$ denotes the Hermitian transpose. $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. $\mathrm{tr}(\cdot)$ denotes the trace and $\det(\cdot)$ represents the determinant. $\mathrm{diag}(\cdot)$ converts a diagonal matrix to a vector and a vector to a diagonal matrix. $\otimes$ denotes the Kronecker product. $\mathcal{CN}(\cdot,\cdot)$ denotes a circularly symmetric complex Gaussian distribution. $k_{\boldsymbol{\theta}}(\cdot,\cdot)$ is a kernel function and $\mathbf{K}$ is the corresponding kernel matrix. γ, ℓ, α, and ν denote the kernel's scaling, lengthscale, exponential-decay, and smoothness parameters, respectively (model-specific where applicable). $\mathbb{E}[\cdot]$ denotes expectation.

II. System Model

We consider a narrowband point-to-point MIMO link with N_{t} transmit and N_{r} receive antennas. During training, n_{t} of the N_{t} transmit antennas are active and transmit the pilot matrix $\mathbf{S} \in \mathbb{C}^{n_{\mathrm{t}} \times T}$, whose rows are orthonormal, i.e., $\mathbf{S}\mathbf{S}^{\mathsf{H}} = \mathbf{I}_{n_{\mathrm{t}}}$, which implies $T \geq n_{\mathrm{t}}$. Let $\mathbf{F} \in \{0,1\}^{N_{\mathrm{t}} \times n_{\mathrm{t}}}$ be the antenna selection matrix satisfying $\mathbf{F}^{\mathsf{H}}\mathbf{F} = \mathbf{I}_{n_{\mathrm{t}}}$. The transmitted signal is $\mathbf{X} = \mathbf{F}\mathbf{S} \in \mathbb{C}^{N_{\mathrm{t}} \times T}$ and the received signal is given by $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N} \in \mathbb{C}^{N_{\mathrm{r}} \times T}$, where $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ is the channel matrix (with (i,j)th entry $H_{i,j}$) and $\mathbf{N} \in \mathbb{C}^{N_{\mathrm{r}} \times T}$ is the noise matrix with independent $\mathcal{CN}(0, \sigma_{\mathrm{n}}^2)$ entries. After decorrelation via right-multiplication by $\mathbf{S}^{\mathsf{H}}$, we obtain $\tilde{\mathbf{Y}} = \mathbf{Y}\mathbf{S}^{\mathsf{H}} = \mathbf{H}\mathbf{F} + \mathbf{N}\mathbf{S}^{\mathsf{H}} \in \mathbb{C}^{N_{\mathrm{r}} \times n_{\mathrm{t}}}$, whose columns correspond to the observed channel vectors associated with the n_{t} transmit antennas.

We adopt both the Kronecker [9] and Weichselberger [10] channel models. Under the Kronecker model, we have $\mathbf{H} = \mathbf{R}_{\mathrm{r}}^{1/2}\mathbf{G}\mathbf{R}_{\mathrm{t}}^{1/2}$, where $\mathbf{R}_{\mathrm{t}} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{t}}}$ and $\mathbf{R}_{\mathrm{r}} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{r}}}$ are the transmit and receive covariance matrices, and $\mathbf{G} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ has independent $\mathcal{CN}(0,1)$ entries. Under the Weichselberger model, we have $\mathbf{H} = \mathbf{U}_{\mathrm{r}}\tilde{\mathbf{A}}\mathbf{U}_{\mathrm{t}}^{\mathsf{H}}$, where $\mathbf{U}_{\mathrm{t}}$ and $\mathbf{U}_{\mathrm{r}}$ are unitary matrices from the eigen decompositions $\mathbf{R}_{\mathrm{t}} = \mathbf{U}_{\mathrm{t}}\boldsymbol{\Lambda}_{\mathrm{t}}\mathbf{U}_{\mathrm{t}}^{\mathsf{H}}$ and $\mathbf{R}_{\mathrm{r}} = \mathbf{U}_{\mathrm{r}}\boldsymbol{\Lambda}_{\mathrm{r}}\mathbf{U}_{\mathrm{r}}^{\mathsf{H}}$, respectively. The matrix $\tilde{\mathbf{A}} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ has independent entries $\tilde{A}_{ij} \sim \mathcal{CN}(0, \Omega_{ij})$, where $\Omega_{ij} \in \mathbb{R}_+$ represents the power coupling coefficient between the ith receive and jth transmit eigenmodes. Collecting the coefficients $\{\Omega_{ij}\}$ into the matrix $\boldsymbol{\Omega} \in \mathbb{R}_+^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$, we have $\mathbf{R}_{\mathrm{H}} = (\mathbf{U}_{\mathrm{t}}^{*} \otimes \mathbf{U}_{\mathrm{r}})\mathrm{diag}(\mathrm{vec}(\boldsymbol{\Omega}))(\mathbf{U}_{\mathrm{t}}^{\mathsf{T}} \otimes \mathbf{U}_{\mathrm{r}}^{\mathsf{H}})$.

To relax the constraint on the pilot length from $T \geq N_{\mathrm{t}}$ to $T \geq n_{\mathrm{t}}$ within a coherence block, we introduce an antenna-probing design in which the transmitter activates the antennas in the set $\mathcal{A} \subseteq \{1, \ldots, N_{\mathrm{t}}\}$, with $n_{\mathrm{t}} = |\mathcal{A}|$, while all the N_{r} receive antennas listen. The corresponding columns of $\mathbf{H}$ are observed and serve as training data for GPR, which infers the remaining entries of $\mathbf{H}$ as described next in Section III. The proposed GPR-based predictor applies to any probing scheme; here we evaluate it under three illustrative geometries:

- *Case I:* only the first transmit antenna is active, i.e., $n_{\mathrm{t}} = 1$, so we observe the first channel column $\mathbf{H}(:,1)$ using all the receive antennas (see Fig. 1(a)). This serves as a stress test (failure mode).
- *Case II:* we activate $n_{\mathrm{t}} = \lceil N_{\mathrm{t}}/2 \rceil$ equispaced transmit antennas, forming the set $\mathcal{A}(n_{\mathrm{t}}) = \{1 + \lfloor (m-1)\frac{N_{\mathrm{t}}-1}{n_{\mathrm{t}}-1} \rfloor : m = 1, \ldots, n_{\mathrm{t}}\}$, and observe the channel columns $\mathbf{H}(:, \mathcal{A}(n_{\mathrm{t}}))$ (see Fig. 1(b)). This design offers a favorable pilot-efficiency trade-off.
- *Case III:* a diagonal-anchored scheme that activates $n_{\mathrm{t}} = \min(N_{\mathrm{t}}, N_{\mathrm{r}})$ transmit antennas as in full training, but retains only the diagonal channel entries $\{H_{i,i}\}$ for training. This emulates one-to-one transmit-receive pairing while discarding all the off-diagonal samples (see Fig. 1(c)). This is not a pilot-saving design; rather, it isolates sampling-geometry effects under the same feasibility condition $T \geq n_{\mathrm{t}}$.

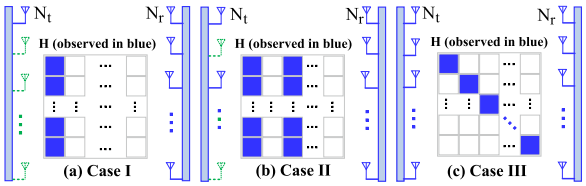


Fig. 1. Illustrative geometries used for the evaluation of the proposed framework, with n_{t} active transmit antennas shown in blue; inactive (green) antennas do not transmit pilots. (a) Case I: $n_{\mathrm{t}} = 1$ active transmit antenna; (b) Case II: $n_{\mathrm{t}} = \lceil N_{\mathrm{t}}/2 \rceil$ active transmit antennas; (c) Case III: $n_{\mathrm{t}} = \min(N_{\mathrm{t}}, N_{\mathrm{r}})$ active transmit antennas with one-to-one transmit-receive antenna pairing.

Algorithm 1 GPR-Based MIMO Channel Prediction

Require: Training points $\mathbf{Z} \in \mathbb{N}^{P \times 2}$, observations $\mathbf{h} \in \mathbb{C}^{P}$, test points $\mathbf{Z}_* \in \mathbb{N}^{M \times 2}$, $k_{\boldsymbol{\theta}}(\cdot,\cdot)$, initial $\boldsymbol{\theta}$

Ensure: Predicted complex entries $\mathbf{h}_* \in \mathbb{C}^{M}$ and posterior covariances $\boldsymbol{\Sigma}_*^{\mathrm{Re/Im}} \in \mathbb{R}^{M \times M}$

1: Split the observations, i.e., $\mathbf{h}^{\mathrm{Re}} = \mathrm{Re}(\mathbf{h})$, $\mathbf{h}^{\mathrm{Im}} = \mathrm{Im}(\mathbf{h})$
2: Maximize $\log p(\mathbf{h}^{\mathrm{Re}}|\mathbf{Z}, \boldsymbol{\theta}) + \log p(\mathbf{h}^{\mathrm{Im}}|\mathbf{Z}, \boldsymbol{\theta})$ to get $\boldsymbol{\theta}^{\star}$
3: Build $\mathbf{K}$, $\mathbf{K}_*$, and $\mathbf{K}_{**}$ from $k_{\boldsymbol{\theta}}(\cdot,\cdot)$ and $\boldsymbol{\theta}^{\star}$
4: Compute the posteriors $\boldsymbol{\mu}_*^{\mathrm{Re/Im}}$ and $\boldsymbol{\Sigma}_*^{\mathrm{Re/Im}}$
5: Form complex prediction: $\mathbf{H}_* = \boldsymbol{\mu}_*^{\mathrm{Re}} + \mathrm{j}\,\boldsymbol{\mu}_*^{\mathrm{Im}}$
6: **return** $\mathbf{H}_*$ and $\boldsymbol{\Sigma}_*^{\mathrm{Re/Im}}$

III. GPR for Channel Prediction

Define the antenna grid $\mathcal{G} = \{(i,j) : i \in \{1, \ldots, N_{\mathrm{r}}\}, j \in \{1, \ldots, N_{\mathrm{t}}\}\} \subset \mathbb{N}^2$. We model the channel entries as a GP over $\mathcal{G}$ and use GPR to predict the missing entries from the observed ones via kernel-based spatial correlations. This section outlines the training set and prediction, kernel selection, hyperparameter optimization, and complex channel reconstruction. The procedure is summarized in Algorithm 1.

Training set and prediction: Let $\mathcal{X} \subset \mathcal{G}$ be the set of observed indices (corresponding to the training points), with $|\mathcal{X}| = P$. Moreover, let $\mathcal{X}_* = \mathcal{G} \setminus \mathcal{X}$ be the set of indices to be predicted (corresponding to the test points), with $|\mathcal{X}_*| = M$. Define the matrix of training points $\mathbf{Z} = [\mathbf{z}_1^{\mathsf{T}}, \ldots, \mathbf{z}_P^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{N}^{P \times 2}$, with $\mathbf{z}_n = (r_n, t_n) \in \mathcal{X}$, and the corresponding output points given by $\mathbf{h} = [h_1, \ldots, h_P]^{\mathsf{T}} \in \mathbb{C}^{P}$, with $h_n = H_{r_n, t_n}$. Likewise, define the matrix of test points $\mathbf{Z}_* = [\mathbf{z}_{*,1}^{\mathsf{T}}, \ldots, \mathbf{z}_{*,M}^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{N}^{M \times 2}$, with $\mathbf{z}_{*,m} = (r_m, t_m) \in \mathcal{X}_*$, and the corresponding output points given by $\mathbf{h}_* = [h_{*,1}, \ldots, h_{*,M}]^{\mathsf{T}} \in \mathbb{C}^{M}$, with $h_{*,m} = H_{r_m, t_m}$. Our goal is to predict $\mathbf{h}_*$ from $\mathbf{h}$ via GPR.

Because the channel entries are complex, we model the real (Re) and imaginary (Im) parts as two different real-valued GPs that share the same covariance $k_{\boldsymbol{\theta}}(\cdot,\cdot)$. For each Re/Im part, the joint real multivariate distribution is

$$\begin{bmatrix} \mathbf{h}^{\mathrm{Re/Im}} \\ \mathbf{h}_*^{\mathrm{Re/Im}} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_{\mathrm{n}}^2\mathbf{I}_P & \mathbf{K}_*^{\mathsf{T}} \\ \mathbf{K}_* & \mathbf{K}_{**} \end{bmatrix}\right), \tag{1}$$

with $\mathbf{K} = k_{\boldsymbol{\theta}}(\mathbf{Z},\mathbf{Z}) \in \mathbb{R}^{P\times P}$, $\mathbf{K}_* = k_{\boldsymbol{\theta}}(\mathbf{Z},\mathbf{Z}_*) \in \mathbb{R}^{P\times M}$, and $\mathbf{K}_{**} = k_{\boldsymbol{\theta}}(\mathbf{Z}_*,\mathbf{Z}_*) \in \mathbb{R}^{M\times M}$. The posterior is given by

$$p\left(\mathbf{h}_*^{\mathrm{Re/Im}} \mid \mathbf{h}^{\mathrm{Re/Im}}\right) = \mathcal{N}\big(\boldsymbol{\mu}_*^{\mathrm{Re/Im}}, \boldsymbol{\Sigma}_*^{\mathrm{Re/Im}}\big), \tag{2}$$

with $\boldsymbol{\mu}_*^{\mathrm{Re/Im}} = \mathbf{K}_*^{\mathsf{T}}(\mathbf{K}+\sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1}\mathbf{h}^{\mathrm{Re/Im}}$ and $\boldsymbol{\Sigma}_*^{\mathrm{Re/Im}} = \mathbf{K}_{**} - \mathbf{K}_*^{\mathsf{T}}(\mathbf{K}+\sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1}\mathbf{K}_*$.

Kernel selection for GPR: The kernel function $k_{\boldsymbol{\theta}}(\cdot,\cdot)$ encodes prior beliefs about the spatial structure of $\mathbf{H}$. We evaluate our proposed framework considering three different kernels, namely, RBF, Matérn, and RQ. These are chosen as they do not require an explicit channel model and span a wide range of smoothness and scale behaviors. The RBF kernel is $k_{\mathrm{RBF}}(\mathbf{z},\mathbf{z}') = \gamma_{\mathrm{RBF}}\exp(-\frac{\|\mathbf{z}-\mathbf{z}'\|^2}{2\ell_{\mathrm{RBF}}^2})$, modeling smooth spatial variations via the lengthscale ℓ_{RBF} and scaling γ_{RBF}. The Matérn kernel is $k_{\mathrm{Mat}}(\mathbf{z},\mathbf{z}') = \gamma_{\mathrm{Mat}}\frac{2^{1-\nu}}{\Gamma(\nu)}(\frac{\sqrt{2\nu}\|\mathbf{z}-\mathbf{z}'\|}{\ell_{\mathrm{Mat}}})^{\nu}\mathrm{K}_{\nu}(\frac{\sqrt{2\nu}\|\mathbf{z}-\mathbf{z}'\|}{\ell_{\mathrm{Mat}}})$, controlling smoothness via ν. Lastly, the RQ kernel is $k_{\mathrm{RQ}}(\mathbf{z},\mathbf{z}') = \gamma_{\mathrm{RQ}}(1+\frac{\|\mathbf{z}-\mathbf{z}'\|^2}{2\alpha\ell_{\mathrm{RQ}}^2})^{-\alpha}$, generalizing RBF by multiple scales through α.

In the Kronecker [9] and Weichselberger [10] models, $\mathrm{vec}(\mathbf{H})$ is distributed as $\mathcal{CN}(\cdot,\cdot)$, which aligns naturally with a GP prior. To motivate the proposed GPR framework with RBF, Matérn, or RQ kernels, Proposition 1 shows that the GP posterior mean equals the BLUP [11]. Thus, optimality holds under second-order structure without requiring channel Gaussianity.

Proposition 1 (GP Posterior Mean Equals BLUP): Let $\mathbf{u} = \mathrm{vec}(\mathbf{H}) \in \mathbb{C}^{N_{\mathrm{r}}N_{\mathrm{t}}}$ be a zero-mean, proper complex random vector with covariance $\mathbf{R}_{\mathrm{u}} \succeq \mathbf{0}$ induced by any positive semidefinite kernel (e.g., RBF, Matérn, or RQ). Let $\mathbf{B} \in \mathbb{C}^{P\times N_{\mathrm{r}}N_{\mathrm{t}}}$ select P entries and consider the observation $\mathbf{H} = \mathbf{B}\mathbf{u}+\boldsymbol{\varepsilon}$, with $\boldsymbol{\varepsilon} \sim \mathcal{CN}(\mathbf{0},\sigma_{\mathrm{n}}^2\mathbf{I}_P)$ and $\boldsymbol{\varepsilon} \perp\!\!\!\perp \mathbf{u}$. Then, the BLUP of $\mathbf{u}$ from $\mathbf{H}$ under the second-order model $(\mathbf{R}_{\mathrm{u}},\sigma_{\mathrm{n}}^2)$ is

$$\widehat{\mathbf{u}}_{\mathrm{BLUP}} = \mathbf{R}_{\mathrm{u}}\,\mathbf{B}^{\mathsf{H}}\big(\mathbf{B}\,\mathbf{R}_{\mathrm{u}}\,\mathbf{B}^{\mathsf{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P\big)^{-1}\mathbf{H}. \tag{3}$$

Moreover, if $\mathbf{u} \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_{\mathrm{u}})$ and $\boldsymbol{\varepsilon} \sim \mathcal{CN}(\mathbf{0},\sigma_{\mathrm{n}}^2\mathbf{I}_P)$, then (3) coincides with the GP posterior mean estimator. Importantly, this equality is entirely determined by the second-order model $(\mathbf{R}_{\mathrm{u}},\sigma_{\mathrm{n}}^2)$ and does not require $\mathbf{u}$ to be Gaussian.

Proof: Consider the linear predictors $\widehat{\mathbf{u}} = \mathbf{L}\mathbf{H}$, with $\mathbf{L} \in \mathbb{C}^{(N_{\mathrm{r}}N_{\mathrm{t}})\times P}$. Since we have $\mathbb{E}[\mathbf{u}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{H}] = \mathbf{0}$, every linear predictor is unbiased. The mean squared error is $J(\mathbf{L}) = \mathbb{E}\|\mathbf{u}-\mathbf{L}\mathbf{H}\|_2^2 = \mathrm{tr}(\mathbf{R}_{\mathrm{u}}) - 2\,\Re\{\mathrm{tr}(\mathbf{L}\,\mathbf{R}_{hu})\} + \mathrm{tr}(\mathbf{L}\,\mathbf{R}_{hh}\,\mathbf{L}^{\mathsf{H}})$, with $\mathbf{R}_{hu} = \mathbb{E}[\mathbf{H}\mathbf{u}^{\mathsf{H}}] = \mathbf{B}\mathbf{R}_{\mathrm{u}}$ and $\mathbf{R}_{hh} = \mathbb{E}[\mathbf{H}\mathbf{H}^{\mathsf{H}}] = \mathbf{B}\mathbf{R}_{\mathrm{u}}\mathbf{B}^{\mathsf{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P$. Using complex matrix calculus, $\partial J/\partial \mathbf{L}^* = -\mathbf{R}_{hu}^{\mathsf{H}} + \mathbf{L}\,\mathbf{R}_{hh}$, so the unique minimizer (since $\mathbf{R}_{hh} \succ \mathbf{0}$ when $\sigma_{\mathrm{n}}^2 > 0$) is $\mathbf{L}^\star = \mathbf{R}_{hu}^{\mathsf{H}}\mathbf{R}_{hh}^{-1} = \mathbf{R}_{\mathrm{u}}\,\mathbf{B}^{\mathsf{H}}(\mathbf{B}\mathbf{R}_{\mathrm{u}}\mathbf{B}^{\mathsf{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1}$, which yields (3). Under a Gaussian prior and Gaussian noise, the joint law of $(\mathbf{u},\mathbf{H})$ is zero-mean with covariances $\mathbf{R}_{\mathrm{u}}$, $\mathbf{B}\mathbf{R}_{\mathrm{u}}$, and $\mathbf{B}\mathbf{R}_{\mathrm{u}}\mathbf{B}^{\mathsf{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P$. Finally, conditioning gives $\mathbb{E}[\mathbf{u} \mid \mathbf{H}] = \mathbf{R}_{\mathrm{u}}\,\mathbf{B}^{\mathsf{H}}(\mathbf{B}\mathbf{R}_{\mathrm{u}}\mathbf{B}^{\mathsf{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1}\mathbf{H}$, which yields the BLUP. ■

Hyperparameter optimization: Let $\boldsymbol{\theta} = [\gamma,\ell,\alpha,\nu]$ denote the kernel hyperparameters. For each Re/Im part, the log marginal likelihood is $\log p(\mathbf{h}^{\mathrm{Re/Im}} \mid \mathbf{Z},\boldsymbol{\theta}) = -\frac{1}{2}(\mathbf{h}^{\mathrm{Re/Im}})^{\mathsf{T}}(\mathbf{K}+\sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1}\mathbf{h}^{\mathrm{Re/Im}} - \frac{1}{2}\log\det(\mathbf{K}+\sigma_{\mathrm{n}}^2\mathbf{I}_P) - \frac{P}{2}\log(2\pi)$. Its derivative with respect to each hyperparameter $\theta_i \in \{\gamma,\ell,\alpha,\nu\}$ is $\frac{\partial}{\partial\theta_i}\log p = \frac{1}{2}\mathrm{tr}[(\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathsf{T}} - (\mathbf{K}+\sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1})\frac{\partial\mathbf{K}}{\partial\theta_i}]$, with $\boldsymbol{\alpha} = (\mathbf{K}+\sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1}\mathbf{h}^{\mathrm{Re/Im}}$. The optimal $\boldsymbol{\theta}^\star$ is obtained by maximizing $\log p(\mathbf{h}^{\mathrm{Re}} \mid \mathbf{Z},\boldsymbol{\theta}) + \log p(\mathbf{h}^{\mathrm{Im}} \mid \mathbf{Z},\boldsymbol{\theta})$ using gradient-based methods [6].

Complex channel reconstruction: The Re and Im components of the channel are modeled by two GPs, $h_{\mathrm{Re}}(r,t)$ and $h_{\mathrm{Im}}(r,t)$, sharing the same kernel and hyperparameters. For a test location $\mathbf{z}_* = (r,t)$, the complex prediction is $[\widehat{H}_{\mathrm{GPR}}]_{r,t} = \hat{h}_{\mathrm{Re}}(\mathbf{z}_*) + \mathrm{j}\hat{h}_{\mathrm{Im}}(\mathbf{z}_*)$, where the posterior means $\hat{h}_{\mathrm{Re}}(\mathbf{z}_*)$ and $\hat{h}_{\mathrm{Im}}(\mathbf{z}_*)$ are obtained from (2). This preserves the shared second-order structure while avoiding cross-correlation assumptions between the Re and Im components.

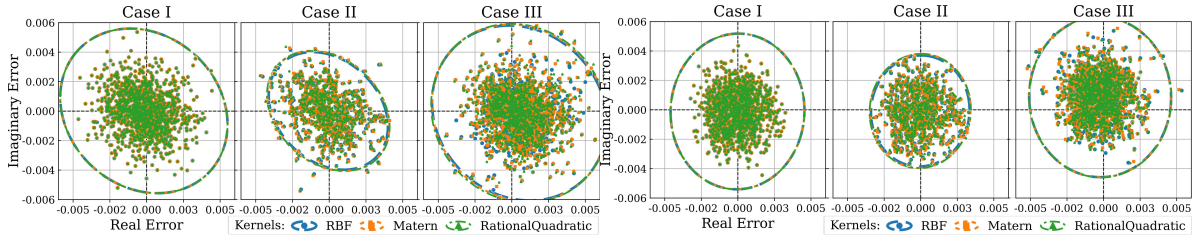
IV. Simulation Results and Discussion

We simulate the proposed schemes on a 36×36 link, under both Kronecker [9] and Weichselberger [10] channel models. Unless stated otherwise, the simulation parameters summarized in Table I are adopted. We consider the three pilot probing schemes of Fig. 1. Specifically: Case I observes a single column of $\mathbf{H}$ (i.e., N_{r} entries) and predicts the remaining $N_{\mathrm{t}}-1$ columns via GPR; Case II observes half of the columns (i.e., approximately $N_{\mathrm{r}}N_{\mathrm{t}}/2$ entries) and predicts the unobserved half; Case III observes only the diagonal entries $(\min(N_{\mathrm{r}},N_{\mathrm{t}}))$ and predicts all the off-diagonal entries from the correlation structure. Note that Case I and Case II lead to a $\frac{(N_{\mathrm{t}}-1)}{N_{\mathrm{t}}}$ and 50% pilot-reduction, respectively.

To assess the prediction accuracy, uncertainty calibration, and communication impact, we employ three metrics. First, we compute the complex entrywise prediction error $\epsilon_{ij} = H_{ij} - \widehat{H}_{ij}$, visualized via scatter plots of $\mathrm{Re}(\epsilon_{ij})$ versus $\mathrm{Im}(\epsilon_{ij})$ with 95% error ellipses. Second, to quantify how well each estimator preserves the channel structure relevant for data transmission, we evaluate the SE of a multi-stream MIMO link under a linear receiver designed from the estimated channel. For a given channel estimate $\widehat{\mathbf{H}} \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{t}}}$, we adopt the linear MMSE detector $\mathbf{W}(\widehat{\mathbf{H}}) = (\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} + \frac{N_{\mathrm{t}}}{\rho}\mathbf{I}_{N_{\mathrm{r}}})^{-1}\widehat{\mathbf{H}} = [\mathbf{w}_1,\ldots,\mathbf{w}_{N_{\mathrm{t}}}] \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{t}}}$. Let $\mathbf{h}_k$ and $\mathbf{w}_k$ denote the kth columns of the true channel $\mathbf{H}$ and the detector $\mathbf{W}$, respectively. The post-equalization signal-to-interference-plus-noise ratio (SINR) of stream k is then

$$\mathrm{SINR}_k\left(\widehat{\mathbf{H}}\right) = \frac{\left|\mathbf{w}_k^{\mathsf{H}}\mathbf{h}_k\right|^2}{\sum_{j\neq k}\left|\mathbf{w}_k^{\mathsf{H}}\mathbf{h}_j\right|^2 + \frac{N_{\mathrm{t}}}{\rho}\|\mathbf{w}_k\|^2}, \tag{4}$$

where the true channel $\mathbf{H}$ is always used inside the SINR expression, while the estimate $\widehat{\mathbf{H}}$ affects performance only through the detector $\mathbf{W}(\widehat{\mathbf{H}})$. The corresponding SE is $\mathrm{SE}(\widehat{\mathbf{H}}) = \sum_{k=1}^{N_{\mathrm{t}}}\log_2(1+\mathrm{SINR}_k(\widehat{\mathbf{H}}))$, which we evaluate for the true channel $\mathbf{H}$, the GPR-predicted channel $\widehat{\mathbf{H}}_{\mathrm{GPR}}$, and the LS and MMSE estimates $\widehat{\mathbf{H}}_{\mathrm{LS}}$ and $\widehat{\mathbf{H}}_{\mathrm{MMSE}}$, respectively. Only for the LS and MMSE baselines, we adopt full-array orthogonal training with length $T \geq N_{\mathrm{t}} = n_{\mathrm{t}}$ and pilot matrix $\mathbf{S} \in \mathbb{C}^{N_{\mathrm{t}}\times T}$. The LS estimate equals $\widehat{\mathbf{H}}_{\mathrm{LS}} = \mathbf{Y}\mathbf{S}^{\mathsf{H}}(\mathbf{S}\mathbf{S}^{\mathsf{H}})^{-1}$. Let $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$. The MMSE estimator equals $\widehat{\mathbf{u}}_{\mathrm{MMSE}} = \mathbf{R}_{\mathrm{H}}\mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{R}_{\mathrm{H}}\mathbf{A}^{\mathsf{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_{TN_{\mathrm{r}}})^{-1}\mathbf{y}$, with $\mathbf{A} = \mathbf{S}^{\mathsf{T}} \otimes \mathbf{I}_{N_{\mathrm{r}}}$ and $\widehat{\mathbf{H}}_{\mathrm{MMSE}} = \mathrm{unvec}(\widehat{\mathbf{u}}_{\mathrm{MMSE}})$. Finally, we evaluate the uncertainty calibration via the empirical coverage of marginal 95% credible intervals $\widehat{H}_{ij} \pm 1.96\sigma_{ij}$, i.e., the fraction of the true



(a) Kronecker model. (b) Weichselberger model.

Fig. 2. Scatter plots of the prediction error illustrating the Re and Im error components for different training scenarios.

TABLE I
SIMULATION PARAMETERS AND HYPERPARAMETERS

System Parameter	Symbol	Value
Channels		
Receive × transmit antennas	$N_r \times N_t$	36×36
Element spacing	d/λ	0.5
Carrier frequency	f_c	28 GHz
Angle spread	σ_ϕ	$\pi/6$
Training (LS/MMSE baselines)		
Pilot length (orthogonal)	T	$T \geq N_t$, $\mathbf{X} \in \mathbb{C}^{N_t \times T}$
SNR (unit power)	ρ	$1/\sigma_n^2$
GPR configuration		
Kernel scale (init; bounds)	γ	1.0; $[10^{-2}, 10^2]$
RBF lengthscale (init; bounds)	ℓ_{RBF}	0.5; $[10^{-2}, 10]$
Matérn lengthscale	ℓ_{Mat}	0.5; $[10^{-2}, 10]$
Matérn smoothness	ν	1.5 (fixed)
RQ lengthscale	ℓ_{RQ}	0.5; $[10^{-2}, 5]$
RQ shape	α	0.5; $[10^{-1}, 5]$

H_{ij} that falls within these intervals derived from the GPR posterior variance.

A. Prediction Error With 95% Confidence Tolerance

Fig. 2 shows, for each kernel, the joint scatter plots of the Re and Im components of the per-entry prediction error ϵ_{ij} under the three pilot-probing schemes (Cases I-III) for both the Kronecker (Fig. 2(a)) and Weichselberger (Fig. 2(b)) channel models. Each scatter cloud is overlaid with a 95% confidence ellipse computed from the empirical covariance of the error pairs, representing the region where 95% of the bivariate errors are expected to lie. The ellipse area reflects the total error variance, with its orientation and eccentricity indicating the correlation and anisotropy between the Re and Im parts.

In Case I, the prediction errors are largest across all kernels, the scatter is widely spread, and the confidence ellipses are expansive. In Case II, activating 50% of the transmit antennas produces a clear contraction of the error clouds toward the origin. The corresponding ellipses shrink in area and become more circular, indicating improved prediction accuracy due to the larger number of observed entries. The Kronecker model retains more elongated ellipses (especially in Cases I and III), signifying persistent anisotropy, and stronger directional correlation. In contrast, the Weichselberger model exhibits nearly circular ellipses, reflecting its richer spatial coupling that balances the error variance more evenly between components.

The prediction errors in Case III are higher than in Case II but lower than in Case I. Although Cases I and III use the same number of observations, diagonal sampling provides

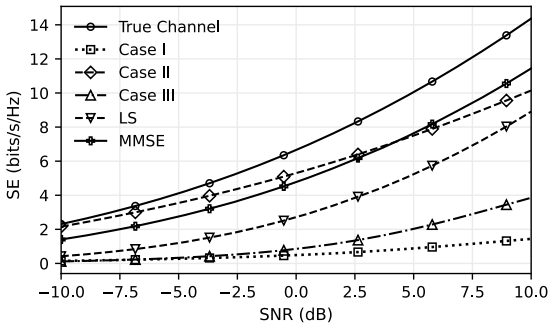


Fig. 3. SE versus SNR for Cases I–III with the Matérn kernel, against LS/MMSE and the true channel.

two-dimensional coverage and better anchors the channel energy/principal subspace that GPR exploits, albeit at the cost of higher pilot overhead, whereas a single column lacks cross-row diversity. Case III is considered to highlight the important fact that the performance of GPR not only depends on the number of observations, but also on how those observations are sampled. Case II performs best because its equispaced half-column design offers both more pilots and richer cross-column information for GPR. The Kronecker model continues to exhibit anisotropic errors, while the Weichselberger model maintains balanced, isotropic error distributions. These results support three main conclusions: (i) increasing the number of directly observed channel entries consistently improves the prediction accuracy; (ii) the Weichselberger model's richer spatial correlation yields more balanced error statistics than the separable Kronecker model; and (iii) all the considered kernels deliver comparable and well-calibrated predictions across the different pilot-probing schemes. Subsequent analyses are presented exclusively using the Weichselberger channel due to its greater physical realism [10].

B. Spectral Efficiency

Fig. 3 shows the SE of the GPR-predicted channels for Cases I–III with the Matérn kernel, compared against the true Weichselberger channel and the full-pilot LS and MMSE baselines. Case II consistently achieves the highest SE among the pilot-saving schemes across the whole SNR range, and clearly outperforms LS. It also exceeds MMSE at low-to-moderate SNR and remains close to it at higher SNR. In relative terms, Case II preserves about 94% of the perfect-CSI SE at SNR $= -10$ dB and about 71% at SNR $= 10$ dB, while using only 50% of the pilots. This follows from probing half of the transmit columns, which yields dense and well-distributed

TABLE II
PILOT SAVING AND RELATIVE SE AT 0 DB RELATIVE TO THE TRUE CHANNEL FOR GPR-RQ, LS, AND MMSE

Metric	Case I	Case II	Case III	LS	MMSE
Pilot length T	1	18	36	36	36
Pilot saving [%]	97.22	50.00	0.00	0.00	0.00
Relative SE [%]	7.23	80.24	4.05	39.52	71.19

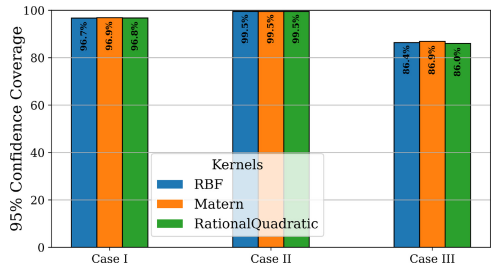


Fig. 4. Empirical 95% credible-interval coverage rates.

observations over the antenna grid; the GPR posterior then acts as a data-adaptive spatial regularizer that learns the correlation structure, suppresses noise, and preserves the dominant modes that drive SE under the adopted linear receiver.

The MMSE baseline provides the strongest conventional performance because it is built from the true second-order channel statistics derived from the joint eigenstructure of the Weichselberger model, but it requires perfect knowledge of these statistics and full-array pilots. In contrast, Case II attains comparable SE without assuming a parametric channel model and with only half of the pilot overhead. On the other hand, LS is highly noise-sensitive and does not exploit spatial correlation, which results in substantially lower SE. At 0 dB, for example, LS attains only 39.52% of the perfect-CSI SE, compared with 80.24% for Case II and 71.19% for MMSE, as summarized in Table II for the RQ kernel.

Cases I and III use the same number of observations but different sampling geometries. Case III consistently achieves higher SE than Case I, highlighting the role of spatial coverage: diagonal sampling provides joint transmit–receive supervision, whereas column sampling excites only a single transmit dimension. However, in both cases the available supervision remains insufficient to identify the full joint channel structure required for spatial multiplexing, leading to poor stream separability and residual multi-stream interference.

For all the three cases, the RBF, Matérn, and RQ kernels produce very similar SE curves against the SNR, indicating robustness to the kernel choice under the considered sampling geometries. After hyperparameter optimization, the learned lengthscales become large relative to the array aperture, so the induced covariance structures are nearly indistinguishable and impose similar spatial priors. Consequently, the SE trends observed with the Matérn kernel in Fig. 3 are consistent with those reported for the RQ kernel in Table II.

C. Uncertainty Calibration

To evaluate uncertainty quantification, Fig. 4 reports the empirical coverage of the marginal 95% credible intervals, i.e., the fraction of true entries that fall within $\widehat{H}_{ij} \pm 1.96\sigma_{ij}$ from the GPR posterior. A well-calibrated model attains coverage close to the nominal 95% (where higher indicates conservative uncertainty and lower indicates overconfidence). The RBF, Matérn, and RQ kernels generally exhibit nearly identical coverage. The learned hyperparameters drive ℓ to large values relative to the array aperture, making these isotropic kernels induce almost constant correlation over the observed antenna grid. As a result, their posterior means and variances, and thus their marginal credible intervals, are effectively the same. This behavior is reinforced by the sampling geometries in Cases I/III (e.g., predominantly collinear/diagonal observations), under which minor differences in kernel shape become negligible.

V. CONCLUSION

This letter proposed a GPR-based framework for MIMO channel estimation that reconstructs full CSI from a small subset of pilot observations. By employing RBF, Matérn, and RQ kernels, the framework models smooth and multi-scale spatial correlations without requiring prior channel statistics. Simulations on Kronecker and Weichselberger channels show that, in Case II, the method reduces pilots by 50% while preserving over 80.24% of the true-channel SE, outperforming LS and MMSE baselines. These results indicate that GPR enables accurate, uncertainty-aware channel reconstruction with substantial training-overhead savings for next-generation MIMO systems. Future work will design kernel structures that more precisely encode channel statistics to further improve estimation performance.

REFERENCES

[1] N. Rajatheva et al., "White paper on broadband connectivity in 6G," 6G Res. Vis., 2020. [Online]. Available: https://urn.fi/URN:ISBN:9789526226798
[2] Z. Chen et al., "Efficient spatial channel estimation in extremely large antenna array communication systems: A subspace approximated matrix completion approach," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 1216–1230, 2025.
[3] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.
[4] E. Vlachos et al., "Massive MIMO channel estimation for millimeter wave systems via matrix completion," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1675–1679, Nov. 2018.
[5] S. L. Shah et al., "Interference prediction using Gaussian process regression and management framework for critical services in local 6G networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2025, pp. 1–6.
[6] J. Zhu et al., "Electromagnetic information theory-based statistical channel model for improved channel estimation," *IEEE Trans. Inf. Theory*, vol. 71, no. 3, pp. 1777–1793, Mar. 2025.
[7] A. Arjas et al., "Nonlinear sparse Bayesian learning methods with application to massive MIMO channel estimation with hardware impairments," 2025, *arXiv:2506.03775*.
[8] J. Li et al., "Spatio-temporal electromagnetic kernel learning for channel prediction," 2024, *arXiv:2412.17414*.
[9] J. P. Kermoal et al., "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.
[10] W. Weichselberger et al., "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.
[11] S. J. Haslett et al., "The fundamental BLUE equation in linear models revisited," *Stat. Appl.*, vol. 22, no. 3, pp. 95–118, 2024.