IEICE Communications Express, Vol.13, No.8, 315–318

ComEX

LETTER

On hyperparameter determination for GPR-based channel prediction in IRS-assisted wireless communication systems

Norisato Suga [1, 2, a), b)], **Kazuto Yano** [1], **Yafei Hou** [1, 3], **and Toshikazu Sakano** [1]

Abstract Intelligent reflecting surface (IRS), which can reflect radio waves controlling the phase of incident radio waves, is being investigated for wireless communication in high-frequency bands. To control the reflection characteristic, it is necessary to separately estimate a large number of channel coefficients between transmitting and receiving antennas through each IRS element. This causes significant overhead for the channel estimation. We have proposed a channel prediction method to reduce the overhead using Gaussian process regression with spectral mixture kernel. In Gaussian process regression, the determination of the hyperparameters used to calculate the kernel matrix has a significant impact on prediction accuracy. In this study, we propose validation-based hyperparameter determination for GPR-based channel prediction and evaluate the performance difference between the gradient method and validation.

Keywords: IRS, RIS, channel prediction, Gaussian process regression
Classification: Wireless communication technologies

1. Introduction

The expansion of communication coverage is being pursued through the utilization of high-frequency band. Intelligent reflecting surfaces (IRS) a.k.a reconfigurable intelligent surface (RIS) is one of the promising approach to realize extremely high throughput in the 6G era [1]. In wireless communication systems employing IRS, the optimization of the reflection coefficients for all IRS elements requires the estimation of huge amount of propagation paths, comprising combinations of transmitting antennas, individual reflection elements, and receiving antennas [2, 3]. In general, channel acquisition relies on known training signals, and the numerous propagation paths formed by the IRS considerably increase the overhead of channel estimation.

In order to mitigate the significant overhead of channel estimation in IRS-assisted systems, one promising approach is channel prediction which forecasts the channel from the observed past channel values. Various methods exist for channel prediction, including classical approaches assuming linear models [4] and deep learning-based methods [5]. In this study, we investigate a channel prediction method based on Gaussian process regression (GPR), which offers relatively low computational complexity and high prediction accuracy [6]. In GPR-based prediction, the target time series are regarded as a Gaussian process, allowing for interpolation and extrapolation based on observations at multiple non-uniform time points. The selection of the kernel function in GPR significantly influences prediction accuracy, and in this study, we consider adopting the spectral mixture (SM) kernel, known for its suitability for extrapolation, for channel prediction.

In GPR, the prediction accuracy highly depends on hyperparameters embedded in the kernel, and it is general to optimize the hyperparameters through gradient descent. However, gradient-based optimization, which aims to maximize the likelihood function, does not guarantee a consistent improvement in prediction accuracy. Therefore, in this study, we propose determining the hyperparameters that yield the highest prediction accuracy through validation. Furthermore, GPR-based predictions for the time series of the channel corresponding to each reflective element entail substantial computational costs due to individual predictions for each element. To address this, we also propose the commonization of kernel matrices to reduce computational costs and evaluate its effectiveness through simulation.

In this letter, the following notations are employed. Matrices are represented in uppercase bold, vectors in lowercase bold, $\mathbb{C}$ denotes the set of complex numbers, $\boldsymbol{A}^{M\times N}$ denotes a complex matrix of size $M \times N$, $(\cdot)^{\mathsf{T}}$ denotes the transpose of a matrix, $(\cdot)^{-1}$ denotes the inverse matrix and $\angle(\cdot)$ denotes the phase of a complex number.

2. Preliminaries

2.1 System model

The channel from the transmit antenna, passing through the $(n_{\mathrm{h}}, n_{\mathrm{v}})$th element on the planar shaped-IRS, to the received antenna at the time t is denoted as $g_{n_{\mathrm{h}},n_{\mathrm{v}}}(t)$. The channel $g_{n_{\mathrm{h}},n_{\mathrm{v}}}(t)$ can be factorized into cascaded channel $h_{n_{\mathrm{h}},n_{\mathrm{v}}}(t)$ and reflective coefficient $\theta_{n_{\mathrm{h}},n_{\mathrm{v}}}$, i.e. $g_{n_{\mathrm{h}},n_{\mathrm{v}}}(t) = h_{n_{\mathrm{h}},n_{\mathrm{v}}}(t)\theta_{n_{\mathrm{h}},n_{\mathrm{v}}}$. At the receiver side, as paths through each reflective element are synthesized, the observed signal can be expressed as follows.

$$y(t) = \left(\sum_{n_{\mathrm{h}},n_{\mathrm{v}}} g_{n_{\mathrm{h}},n_{\mathrm{v}}}(t)\right) x(t) + n(t)$$

[1] ATR Wave Engineering Laboratories, 2–2–2 Hikaridai, Seika, Souraku, Kyoto, 619–0288, Japan
[2] Faculty of Engineering, Shibaura Institute of Technology, 3–7–5 Toyosu, Koto, Tokyo 135–8548, Japan
[3] Faculty of Environmental, Life, Natural Science and Technology, Okayama University, 3–1–1 Tsushima-Naka, Kita-ku, Okayama City, Okayama, 700–8530, Japan
[a)] norisato.suga@atr.jp
[b)] nsuga@shibaura-it.ac.jp

DOI: 10.23919/comex.2024XBL0058
Received March 19, 2024
Accepted April 30, 2024
Publicized June 11, 2024
Copyedited August 1, 2024

$$= \left(\sum_{n_\mathrm{h}, n_\mathrm{v}} h_{n_\mathrm{h}, n_\mathrm{v}}(t) \theta_{n_\mathrm{h}, n_\mathrm{v}} \right) x(t) + n(t)$$
$$= \boldsymbol{h}(t)^\mathsf{T} \boldsymbol{\theta} x(t) + n(t) \quad (1)$$

where $x(t)$ represents the transmitted symbol, and $n(t)$ denotes complex white noise with zero mean and variance σ^2, $\boldsymbol{h}(t) = [h_{1,1}(t), \ldots, h_{1,N_\mathrm{v}}(t), h_{2,1}(t), \ldots, h_{N_\mathrm{h},N_\mathrm{v}}(t)]^\mathsf{T} \in \mathbb{C}^{N_\mathrm{h}N_\mathrm{v} \times 1}$ and $\boldsymbol{\theta} = [\theta_{1,1}, \ldots, \theta_{1,N_\mathrm{v}}, \theta_{2,1}, \ldots, \theta_{N_\mathrm{h},N_\mathrm{v}}]^\mathsf{T} \in \mathbb{C}^{N_\mathrm{h}N_\mathrm{v} \times 1}$.

During channel estimation, known training signal $x(t)$ is transmitted. By dividing the observed signal by $x(t)$, we obtain

$$z(t) = \frac{y(t)}{x(t)} = \boldsymbol{h}(t)^\mathsf{T} \boldsymbol{\theta} + \bar{n}(t) \quad (2)$$

where $\bar{n}(t) = \frac{n(t)}{x(t)}$. Since the symbol period T_s is generally very short, we assume that the channel remains unchanged during the transmission of $N_\mathrm{h}N_\mathrm{v}$ symbols. If the reflection coefficient $\boldsymbol{\theta}$ changes symbol by symbol (with the interval of T_s), the vector $\boldsymbol{z}(t) = [z(t), \ldots, z(t + N_\mathrm{h}N_\mathrm{v} - 1)]^\mathsf{T}$ can be expressed as follows

$$\boldsymbol{z}(t) = \underbrace{[\boldsymbol{\theta}(1), \cdots, \boldsymbol{\theta}(N_\mathrm{h}N_\mathrm{v})]^\mathsf{T}}_{\boldsymbol{\Omega}} \boldsymbol{h}(t) + \bar{\boldsymbol{n}}(t)$$
$$= \boldsymbol{\Omega} \boldsymbol{h}(t) + \bar{\boldsymbol{n}}(t) \quad (3)$$

where $\bar{\boldsymbol{n}}(t) = [\bar{n}(t), \ldots, \bar{n}(t + N_\mathrm{h}N_\mathrm{v} - 1)]^\mathsf{T}$. If we transmit sufficient number of training symbols and switch the reflective coefficient appropriately, $\boldsymbol{\Omega}$ is invertible. Then, the channel estimation can be realized for IRS-assisted systems [2].

Based on the obtained channel estimate $\hat{\boldsymbol{h}}(t)$, we determine the optimal reflection coefficients for each reflective element. In this study, we assume that each element can only control the phase, and we determine the reflection phase according to the strongest-CIR maximization (SCM) criterion [2]. In SCM, for each element of the estimated cascade channel $\hat{\boldsymbol{h}}(t)$, the inverse phase is adopted as the weight for the reflection coefficient as

$$\angle \hat{\theta}_{n_\mathrm{h}, n_\mathrm{v}} = -\angle \hat{h}_{n_\mathrm{h}, n_\mathrm{v}}(t). \quad (4)$$

2.2 GPR-based channel prediction

In this study, we adopt two phases frame transmission: (Training phase) the prediction model is trained from the channel values obtained by channel estimation; (prediction phase) the future channels are predicted based on the trained model without transmitting training symbols (Fig. 1). In the training and prediction phases, N_t and N_p frames are transmitted, respectively. The frame in the training phase contains $N_\mathrm{h}N_\mathrm{v}$ training symbols at the beginning, and channel estimation is performed. When N_t channel estimates are obtained during the training phase, the prediction model is trained to predict the channel values during the prediction phase. From the predicted channel values, the appropriate reflection coefficient at each transmission timing are determined in the prediction phase according to (4).

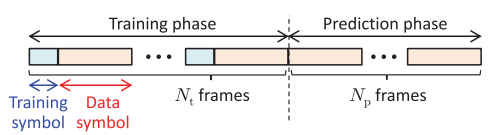


Fig. 1 Frame format in training and prediction phases.

Training and prediction are performed for each element of the cascaded channel $\boldsymbol{h}(t)$ and for each real and imaginary part. The real (or imaginary) part of an element $h_{n_\mathrm{h}, n_\mathrm{v}}(t)$ is denoted by $h(t)$, omitting the subscripts. The set of time indexes when channel estimation is performed in the training phase is denoted by $\hat{\mathcal{T}} = \{t_1, \cdots, t_{N_\mathrm{t}}\}$, and the set of predicted times is also denoted by $\tilde{\mathcal{T}} = \{\tilde{t}_1, \cdots, \tilde{t}_L\}$, where L is the number of time points to be predicted. In GPR, the vector $\boldsymbol{f} = [\hat{h}(t_1), \cdots, \hat{h}(t_{N_\mathrm{t}}), \tilde{h}(\tilde{t}_1), \cdots, \tilde{h}(\tilde{t}_L)]^\mathsf{T}$ is modeled as a Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \mathcal{K})$, where $\mathcal{K}$ is a covariance matrix called kernel matrix. The vector consisted of the estimated channel is $\hat{\boldsymbol{g}} = [\hat{h}(t_1), \cdots, \hat{h}(t_{N_\mathrm{t}})]^\mathsf{T}$, and the vector gathering the future channel values is also defined as $\tilde{\boldsymbol{g}} = [\tilde{h}(\tilde{t}_1), \cdots, \tilde{h}(\tilde{t}_L)]^\mathsf{T}$. If the vector $\boldsymbol{f}$ follows a Gaussian process, the predicted channel is given as [8]

$$\tilde{\boldsymbol{g}} = \boldsymbol{K}_*^\mathsf{T} \boldsymbol{K}^{-1} \hat{\boldsymbol{g}} \quad (5)$$

where each element of the matrices $\boldsymbol{K} \in \mathbb{R}^{N_\mathrm{t} \times N_\mathrm{t}}$ and $\boldsymbol{K}_* \in \mathbb{R}^{N_\mathrm{t} \times L}$ are calculated as: $\kappa(t_i, t_j)$ for $i = 1, \cdots, N_\mathrm{t}$ and $j = 1, \cdots, N_\mathrm{t}$; $\kappa(t_i, \tilde{t}_j)$ for $i = 1, \cdots, N_\mathrm{t}$ and $j = 1, \cdots, L$. The function $\kappa(\cdot, \cdot)$ is called kernel function.

3. Validation-based hyperparameter determination

The kernel function $\kappa(\cdot, \cdot)$ determines the shape of the covariance matrix $\mathcal{K}$ of the Gaussian process $\mathcal{N}(\boldsymbol{0}, \mathcal{K})$. By selecting an appropriate kernel function, we can construct a model that well represents the entire data vector $\boldsymbol{f}$. Spectral mixture (SM) kernel is known to be suitable for extrapolation task [9], and we adopt this kernel function for the GPR-based channel prediction adding the term $\epsilon \delta(i, j)$ representing the channel estimation error.

$$\kappa(\tau) = \sum_{q=1}^{Q} w_q e^{-2\pi^2 \tau^2 v_q} \cdot \cos\left(2\pi\tau\mu_q\right) + \epsilon\delta(i, j) \quad (6)$$

where $\tau = t_i - t_j$, and w_q, v_q, μ_q are hyperparameters that control the shape of the kernel function, $\delta(i, j)$ is the Kronecker delta function, and ϵ is also hyperparameter representing the strength of the channel estimation error.

When performing predictions with GPR, the determination of hyperparameters (in the case of the SM kernel, w_q, v_q, μ_q, ϵ) significantly affects the prediction accuracy. Here, we define the vector $\boldsymbol{\psi}$ which includes all hyperparameters in each element. In general, the hyperparameters of the Gaussian process are determined by maximizing the probability density function (marginal likelihood) $p(\hat{\boldsymbol{g}}|\hat{\mathcal{T}}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{\psi}})$ conditioned on the training data $\hat{\mathcal{T}}$ with respect to $\boldsymbol{\psi}$. However, since optimization problem on marginal likelihood does not have a closed-form solution, it is common to determine the hyperparameters as local solutions through iterative algorithms based on gradient methods. Figure 2 illustrates the likelihood value on $p(\hat{\boldsymbol{g}}|\hat{\mathcal{T}})$ with respect to the hyperparameters μ_q and ϵ_q when predicting the channel (for detailed simulation conditions, please refer to the next section). As seen in Fig. 2, $p(\hat{\boldsymbol{g}}|\hat{\mathcal{T}})$ is a multimodal function with respect to μ. When determining the hyperparameters using

gradient-based algorithms, there is a possibility of falling into local optima depending on the initial values. The function $p(\hat{\boldsymbol{g}}|\hat{\mathcal{T}})$ also has vast flat regions, forcing a large number of iterations for gradient-based methods. Furthermore, as gradient methods are primarily aimed at maximizing the marginal likelihood, there is no guarantee that the prediction accuracy will necessarily improve.

Therefore, in this letter, we consider determining the hyperparameters through validation by dividing the training data. Among the training samples, the most recent N_q samples are used as validation samples, and the hyperparameters are selected with the remaining $N_\mathrm{t} - N_\mathrm{q}$ samples. The hyperparameter candidates are prepared in advance, and for each candidate, training (calculation of kernel matrix) is performed with $N_\mathrm{t} - N_\mathrm{q}$ samples, and predictions are made for the validation sample points using the obtained kernel matrix. Subsequently, the prediction accuracy is calculated based on the obtained prediction results and validation samples. After calculating the prediction accuracy for all candidates, the hyperparameters corresponding to the best prediction accuracy are adopted. The kernel matrix is then calculated again using all N_t training samples, and predictions are made for the target time points $\tilde{\mathcal{T}}$.

Next, we discuss the reduction in computational cost achieved through the commonization of the kernel matrix. The conventional method [6] assumes the GPR-based prediction is performed for the time series of cascade channels corresponding to each element. However, GPR involves inverse matrices calculation, leading to a significant computational cost. To reduce computational complexity, we commonize the kernel matrix in the prediction of cascade channels for all elements. With this commonization, the calculation of the kernel matrix for all cascade channels, including both the real and imaginary parts, can be reduced from $2N_\mathrm{h}N_\mathrm{v}$ times (including validation) to a single calculation. Since individual adjustments of hyperparameters for each cascade channel are no longer possible, the prediction accuracy may decrease. To mitigate the degradation of prediction accuracy due to the commonization, validation is performed across the validation samples corresponding to the multiple reflective elements. In more detail, for each hyperparameter candidate, the kernel matrix is calculated from $\hat{\mathcal{T}}$ and $\tilde{\mathcal{T}}$, and predictions are made for cascade channels corresponding to all reflective elements. The prediction errors are calculated from the obtained predictions and validation samples for each cascade channel, and the total prediction error is computed. By adopting the candidate with the smallest total prediction error among all candidates, it is possible to avoid selecting hyperparameters that overfit to a specific reflective element.

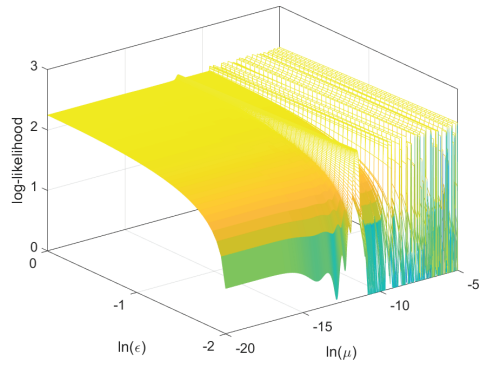


Fig. 2 Marginal likelihood on ϵ and μ.

4. Evaluation

4.1 Simulation settings

In this study, we assume the use of IRS in the high-frequency band and evaluate the channel prediction accuracy in the 60 GHz band according to the channel model presented in [7]. Figure 3 illustrates the geometry among the transmitter, IRS, and receiver assumed in this simulation. A metallic object exists between the transmitter and receiver, and then there is no direct path. We also assume that only the receiver is moving, with movement speeds of 0, 3, or 5 km/h. The transmit and receive antennas are oriented towards the IRS direction, ensuring a constant maximum antenna gain. The spacing between each reflective element is set to 0.5 or 1.5 wavelengths, and we verify the difference in prediction accuracy due to the element spacing. The symbol interval T_s is set to 50 ns, and each wireless frame contains 12,000 BPSK symbols. The number of reflective elements is $N_\mathrm{h} = N_\mathrm{v} = 10$, resulting in a total of 100 elements on the IRS. Therefore, during the training phase, 100 training symbols are added at the beginning of each frame for channel estimation. The number of frames in training and prediction phase both are set to 40.

For GPR-based prediction, we set $Q = 1$, $w = 1$, and $v = e^{-15.3}$. Since the hyperparameters μ and ϵ significantly affect prediction accuracy, we determine these value by gradient-based optimization (conventional) and validation (proposed). For the gradient-based optimization, we use Adam [10] as an optimizer. The initial value of μ_0 is randomly determined as e^{α} where α is uniformly random variable from the range $[-15, -10]$, and ϵ_0 is uniformly randomly selected from the ranges $[0.1, 1]$, respectively, for the case denoted as "Random Init.". In the case denoted as "Fix Init.", the initial values are fixed intentionally close to the optimal values as $\mu_0 = e^{-13}$ and $\epsilon_0 = 0.1$. For the validation-based hyperparameter determination, we consider candidate values for $\ln(\mu_0)$ in the range of $-15.00, -14.99, \ldots, -10.00$, and for $\log_{10}(\epsilon_0)$ in the range of $-1, -0.9, \ldots, 0$. The evaluation metrics are the normalized mean square error (NMSE) calculated as $\mathrm{NMSE}(t) = \frac{|\boldsymbol{h}(t)-\tilde{\boldsymbol{h}}(t)|^2}{|\boldsymbol{h}(t)|^2}$ and bit error rate (BER) at each time point during the prediction phase. We also evaluate signal to noise ratio after IRS control defined as $\mathrm{SNR}(t) = \frac{|\boldsymbol{h}(t)^\mathrm{T}\tilde{\boldsymbol{\theta}}|^2}{\sigma^2}$, where $\tilde{\boldsymbol{\theta}}$ is reflection coefficient obtained from channel prediction. We conduct 20 trials and show the average values of the aforementioned performance metrics.

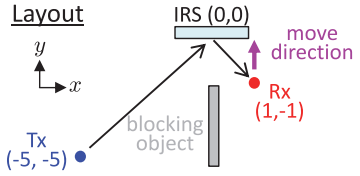
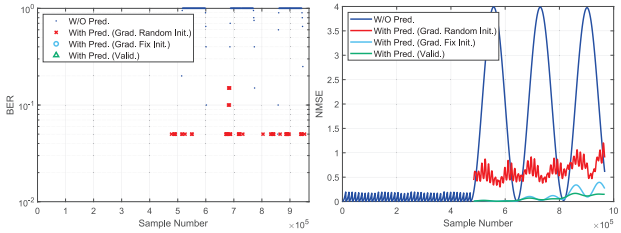
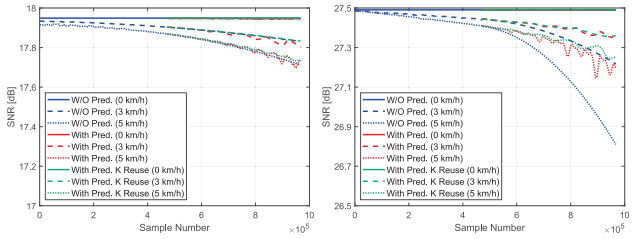


Fig. 3 Tx, IRS, and Rx positions (x-y coordinates [m]).



(a) BER (b) NMSE

Fig. 4 BER and NMSE performance at moving speeds of 3 km/h with reflective element spacing 1.5λ.



(a) 0.5λ (b) 1.5λ

Fig. 5 SNR performance.

4.2 Results

Figure 4 illustrates the BER and NMSE characteristics at movement speed 3 km/h. The blue line shows the case without channel prediction in which the channel estimate at the last frame in the training phase is used instead of the prediction. The performance of GPR with validation shown here involves the commonization of the kernel matrices $\boldsymbol{K}$ and $\boldsymbol{K}_*$. The channel varies as the receiver moves, and bit errors begin to occur in prediction phase. When optimization is performed using gradient descent with a randomly chosen initial value, a significant degradation in prediction accuracy and the bit errors can be observed. On the other hand, both gradient descent with appropriate initial values and validation do not result in bit errors. Since the initial values must be carefully selected for gradient-based hyperparameter optimization, validation provides a more robust approach for determining hyperparameters, making it a more reliable method.

Next, we evaluate the impact of commonization of kernel matrices $\boldsymbol{K}$ and $\boldsymbol{K}_*$. We compare the two cases where validation is performed for each cascade channel corresponding to each reflecting element and the case where validation is performed once due to the commonization of kernel matrices (denoted as "$\boldsymbol{K}$ Reuse"). Figures 5(a) and 5(b) show the SNR performance when the spacing between reflecting elements is set to 0.5λ and 1.5λ, respectively. In case of without prediction, the SNR deteriorates over time due to the outdated reflection coefficient which is determined according to the last channel estimation in the training phase. On the other hand, by predicting the channel and determining appropriate reflection coefficients at each time point, the degradation of SNR can be suppressed. Furthermore, when comparing the cases with and without commonization of kernel matrices, no significant degradation is observed. Especially, focusing on the SNR performance for a movement speed of 5 km/h in Fig. 5(b), a slight improvement is observed when commonization is applied even reducing the computational complexity. This is attributed to the increased number of validation samples by performing validation for all cascade channels of all reflecting elements during commonization, leading to more robust determination of hyperparameters.

5. Conclusion

In this letter, we investigated GPR-based channel prediction in the IRS-assisted wireless communication system. The hyperparameter determination is an issue in GPR-based channel prediction, and we addressed it through validation and commonization of the kernel matrix. The simulation results confirmed that the validation-based approach is more robust on the hyperparameter determination. Additionally, commonization of the kernel matrix not only reduces computational costs, but also results an improvement in prediction accuracy due to an increase of validation samples.

Acknowledgments

These research results were obtained from the commissioned research (No. JPJ012368C03401) by National Institute of Information and Communications Technology (NICT), Japan.

References

[1] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, 2020. DOI: 10.1109/mcom.001.1900107

[2] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, 2020. DOI: 10.1109/lwc.2019.2961357

[3] B. Zheng, C. You, and R. Zhang, "Fast channel estimation for IRS-Assisted OFDM," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 580–584, 2021. DOI: 10.1109/lwc.2020.3038434

[4] W. Jiang and H.Di. Schotten, "A comparison of wireless channel predictors: Artificial intelligence versus Kalman filter," IEEE Int. Conf. Commun., 2019. DOI: 10.1109/icc.2019.8761308

[5] W. Jiang and H.D. Schotten, "Deep learning for fading channel prediction," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 320–332, April, 2020. DOI: 10.1109/ojcoms.2020.2982513

[6] N. Suga, K. Yano, Y. Hou, and T. Sakano, "Study on channel prediction in IRS-assisted wireless communication systems," *IEICE Commun. Express*, vol. 12, no. 7, pp. 374–378, 2023. DOI: 10.1587/comex.2023xbl0035

[7] W. Tang, M.Z. Chen, X. Chen, J.Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T.J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, 2021. DOI: 10.1109/twc.2020.3024887

[8] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006. DOI: 10.7551/mitpress/3206.001.0001

[9] A.G. Wilson and R.P. Adams, "Gaussian process kernels for pattern discovery and extrapolation," 30th Int. Conf. Mach. Learn. ICML 2013, vol. 28, no. PART 3, pp. 2104–2112, 2013.

[10] D.P. Kingma and J.L. Ba, "Adam: a method for stochastic optimization," Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015.