Online Prediction of Spatial Fields for Radio-Frequency Communication

Dave Zachariah
Dept. Information Technology
Div. Systems and Control
Uppsala University, Sweden
E-mail: dave.zachariah@it.uu.se

Niklas Jaldén
Ericsson Research
Kista, Sweden
Email: niklas.jalden@ericsson.com

Petre Stoica
Dept. Information Technology
Div. Systems and Control
Uppsala University, Sweden
E-mail: ps@it.uu.se

***Abstract*—In this paper we predict spatial wireless channel characteristics using a stochastic model that takes into account both distance dependent pathloss and random spatial variation due to fading. This information is valuable for resource allocation, interference management, design in wireless communication systems. The spatial field model is trained using a convex covariance-based learning method which can be implemented online. The resulting joint learning and prediction method is suitable for large-scale or streaming data. The online method is first demonstrated on a synthetic dataset which models pathloss and medium-scale fading. We compare the method with a state-of-the-art scalable batch method. It is subsequently tested in a real dataset to capture small-scale variations.**

I. Introduction

The spatially varying characteristics of radio-frequency (RF) electromagnetic fields critically affect the performance of wireless communication systems. The received signal power from a set of transmitters may vary significantly over space and exploiting such information can greatly improve wireless resource allocation, interference management, system design, etc. [1]–[3].

Methods for predicting spatial processes using a given set of training data have been successfully developed in geostatistics, building on the study of stochastic processes, cf. [4]–[6]. These methods have been applied to predict maps of received signal powers [7], [8]. They often assume an isotropic covariance structure, which is fitted to observed data in some manner. More sophisticated methods of learning the covariance structure have been developed in the kernel and Gaussian process literature, using cross-validation or maximum likelihood frameworks [9]–[11]. Their computional requirements, however, scale poorly with the number of datapoints and make them unsuitable for large-scale or online data collection of RF fields. Moreover, the learning methods often require solving nonconvex problems which can be riddled with local minima issues.

By contrast, a recently developed framework based on covariance fitting addresses both of these important limitations, see [12] for a detailed derivation which is omitted here. We apply this framework to a spatial model of received signal strength, resulting in joint learning and prediction method for spatial RF fields that can be implemented by solving a convex problem online. It is tested on synthetic as well as real data and exhibits a promising performance.

II. Problem formulation

Let $\mathbf{x} \in \mathcal{X}$ denote spatial coordinates, where $\mathcal{X} \subseteq \mathbb{R}^d$ is a bounded region and $d = 2$ or 3. We consider T transmitters at known locations $\{\bar{\mathbf{x}}_k\}_{k=1}^T$ and let $y(\mathbf{x}) \in \mathbb{R}$ denote the received signal power at $\mathbf{x}$ in decibel scale. We model $y(\mathbf{x})$ as a stochastic process, affected by distance-dependent pathloss from the transmitters as well as fading and other random factors [1]–[3], [13], [14]. Given a stream of observations at different points,

$$(\mathbf{x}_1, y(\mathbf{x}_1)),\ (\mathbf{x}_2, y(\mathbf{x}_2)),\ \ldots,\ (\mathbf{x}_N, y(\mathbf{x}_N)),$$

our goal is to predict $y(\mathbf{x})$ at points $\mathbf{x} \neq \mathbf{x}_n$. As $\mathbf{x}$ is varied, the prediction $\hat{y}(\mathbf{x})$ provides a map of the predicted channel gain. An example of the general setup is illustrated in Fig. 1.

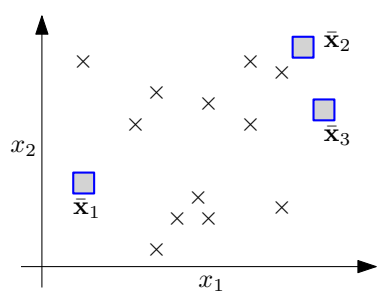

Fig. 1: Example setup. Three transmitters (boxes) located at known coordinates $\bar{\mathbf{x}}_1$, $\bar{\mathbf{x}}_2$ and $\bar{\mathbf{x}}_3$. The observed signal power $y(\mathbf{x})$ at various points $\mathbf{x}_n$, denoted by $\times$. The goal is to predict $y(\mathbf{x})$ at points $\mathbf{x} \neq \mathbf{x}_n$.

We use the following approximate model of the expected value of $y(\mathbf{x})$:

$$\begin{aligned}\mu(\mathbf{x}) &= \mathrm{E}[y(\mathbf{x})] \\ &= \eta_0 - \sum_{k=1}^{T} \eta_k 10 \log_{10}(\|\mathbf{x} - \bar{\mathbf{x}}_k\|),\end{aligned} \tag{1}$$

where η_0 captures transmit powers, propagation gains, the noise floor, etc. This constant as well as the pathloss coefficients η_k from each transmitter are unknown.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

The random variation of $y(\mathbf{x})$ over space captures for instance spatially correlated shadow fading that arises from obstacles in the propagation medium. Here we assume only that the covariance function of $y(\mathbf{x})$ is locally well described as stationary and rotation-invariant. Using Mercer's theorem, the following model

$$\begin{aligned} r(\mathbf{x},\mathbf{x}') &= \mathrm{Cov}[y(\mathbf{x}), y(\mathbf{x}')] \\ &= \sum_{k=1}^{q} \theta_k \phi_k(\mathbf{x})\phi_k(\mathbf{x}') + \theta_0 \delta(\mathbf{x},\mathbf{x}'), \end{aligned} \tag{2}$$

can approximate any covariance function up to an arbitrary accuracy determined by q, where $\theta_k > 0$. For stationary covariance functions, Bochner's theorem provides one choice for $\boldsymbol{\phi}(\mathbf{x})$, namely the Fourier basis for which the coefficients θ_k have to be determined [11]. Another choice for $\boldsymbol{\phi}(\mathbf{x})$, appropriate for bounded regions $\mathcal{X} = [-L_1, L_1] \times \cdots \times [-L_d, L_d]$, is the Laplace operator basis [15]. For this basis, the elements of $\boldsymbol{\phi}(\mathbf{x})$ are given by

$$\phi_{k_1,\ldots,k_d}(\mathbf{x}) = \prod_{i=1}^{d} \frac{1}{\sqrt{L_i}} \sin\left(\frac{\pi k_i (x_i + L_i)}{2L_i}\right). \tag{3}$$

where $k_i = 1, \ldots, m$ are the indices for dimension i. Then $\boldsymbol{\phi}(\mathbf{x}) = [\phi_{1,\ldots,1}(\mathbf{x}) \ \cdots \ \phi_{m,\ldots,m}(\mathbf{x})]^\top$ has dimension $q = m^d$. We will use (3) in the examples, but the derivations below are general and do not depend on the choice of $\boldsymbol{\phi}(\mathbf{x})$.

III. Learning and prediction

We write the observed signal powers up to sample N in vector form:

$$\mathbf{y}_N = \begin{bmatrix} y(\mathbf{x}_1) & \cdots & y(\mathbf{x}_N) \end{bmatrix}^\top$$

and make use of the following statistical moments,

$$\mathrm{E}[\mathbf{y}_N], \quad \mathrm{Cov}[\mathbf{y}_N] \quad \text{and} \quad \mathrm{Cov}[\mathbf{y}_N, y(\mathbf{x})], \tag{4}$$

to derive a learning and prediction method.

For the mean of $\mathbf{y}_N$, we can write

$$\mathrm{E}[\mathbf{y}_N] = \underbrace{\begin{bmatrix} \mathbf{u}^\top(\mathbf{x}_1) \\ \vdots \\ \mathbf{u}^\top(\mathbf{x}_N) \end{bmatrix}}_{\triangleq \mathbf{U}_N} \underbrace{\begin{bmatrix} \eta_0 \\ \eta_1 \\ \vdots \\ \eta_T \end{bmatrix}}_{\boldsymbol{\eta}}, \tag{5}$$

where we define the vector

$$\mathbf{u}(\mathbf{x}) = \begin{bmatrix} 1 \\ -10\log_{10}(\|\mathbf{x} - \bar{\mathbf{x}}_1\|) \\ \vdots \\ -10\log_{10}(\|\mathbf{x} - \bar{\mathbf{x}}_T\|) \end{bmatrix}.$$

Note that $\boldsymbol{\eta}$ is a vector of unknown coefficients.

For the covariance matrix of $\mathbf{y}_N$, we can write

$$\mathrm{Cov}[\mathbf{y}_N] = \{r(\mathbf{x}_n, \mathbf{x}_m)\}_{n,m} = \underbrace{\boldsymbol{\Phi}_N \boldsymbol{\Theta} \boldsymbol{\Phi}_N^\top + \theta_0 \mathbf{I}_N}_{\triangleq \mathbf{R}_N}, \tag{6}$$

where

$$\boldsymbol{\Phi}_N = \begin{bmatrix} \boldsymbol{\phi}^\top(\mathbf{x}_1) \\ \vdots \\ \boldsymbol{\phi}^\top(\mathbf{x}_N) \end{bmatrix}, \quad \boldsymbol{\Theta} = \mathrm{diag}(\boldsymbol{\theta}) \quad \text{and} \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_q \end{bmatrix}.$$

Finally, the covariance between $\mathbf{y}_N$ and the unknown $y(\mathbf{x})$ at test point $\mathbf{x} \neq \mathbf{x}_n$ can be written as

$$\mathrm{Cov}[\mathbf{y}_N, y(\mathbf{x})] = \{r(\mathbf{x}_n, \mathbf{x})\}_n = \underbrace{\boldsymbol{\Phi}_N \boldsymbol{\Theta} \boldsymbol{\phi}(\mathbf{x})}_{\triangleq \mathbf{r}_N}. \tag{7}$$

For notational simplicity, we drop the subindex N in (5), (6) and (7), and write $\mathbf{U}$, $\mathbf{R}$, $\boldsymbol{\Phi}$ and $\mathbf{r}$.

A. Prediction with given parameters

As we do not make any distributional assumptions about $y(\mathbf{x})$, but rely only on the first and second-order statistical moments in (4), we constrain the predictor to be a linear function of the data,

$$\hat{y}(\mathbf{x}) = \mathbf{p}^\top \mathbf{y}. \tag{8}$$

Assuming the covariance parameters $\boldsymbol{\theta}$ are given, we seek the unbiased linear predictor $\mathbf{p}$ which minimizes the mean square error (MSE) of the prediction. That is, the solution to the following problem:

$$\begin{aligned} \min_{\mathbf{p}} \quad & \mathrm{E}[(y(\mathbf{x}) - \hat{y}(\mathbf{x}))^2] \\ \text{subject to} \quad & \mathrm{E}[\hat{y}(\mathbf{x})] = \mu(\mathbf{x}), \end{aligned} \tag{9}$$

where the constraint ensures unbiasedness given model (1).

The optimization problem (9) can be solved using Lagrange multipliers, and the solution is

$$\mathbf{p} = \mathbf{R}^{-1}\mathbf{U}(\mathbf{U}^\top \mathbf{R}^{-1}\mathbf{U})^- \mathbf{u}(\mathbf{x}) + \mathbf{R}^{-1}\boldsymbol{\Pi}^\perp \mathbf{r}, \tag{10}$$

where $(\cdot)^-$ denotes the generalized inverse and $\boldsymbol{\Pi}^\perp$ is the orthogonal projector onto $\mathcal{R}(\mathbf{U})^\perp$ with respect to the inner product weighted by $\mathbf{R}^{-1}$. See [12] and [6, ch. 3.4.2] for derivations and expressions of the solution to (9).

The optimal predictor coefficients (10) are both a function of the test point $\mathbf{x}$ and the unknown covariance parameters $\boldsymbol{\theta}$, via (6) and (7).

B. Learning the parameters via covariance fitting

We now turn to the problem of learning $\boldsymbol{\theta}$ from the data, to be used in the linear predictor (10). As in the previous section, we rely only on the statistical moments in (4).

Let the sample covariance matrix be denoted as

$$\mathbf{S} = (\mathbf{y} - \mathbf{U}\boldsymbol{\eta})(\mathbf{y} - \mathbf{U}\boldsymbol{\eta})^\top \tag{11}$$

and let $\rho = \|\mathbf{y} - \mathbf{U}\boldsymbol{\eta}\|$ be a normalizing factor. Then we seek to learn $\boldsymbol{\theta}$ by fitting the covariance matrix $\mathbf{R}$ in (6) to $\mathbf{S}/\rho$, subject to the normalization constraint $\mathrm{tr}\{\mathbf{R}\} = \rho$. The covariance-fitting problem can be formulated as

$$\begin{aligned} \min_{\boldsymbol{\theta}} \quad & \left\| \frac{1}{\rho}\mathbf{S} - \mathbf{R} \right\|_{\mathbf{R}^{-1}}^2 \\ \text{subject to} \quad & \mathrm{tr}\{\mathbf{R}\} = \rho, \end{aligned} \tag{12}$$

where $\|\cdot\|^2_{\mathbf{R}^{-1}}$ is a weighted norm. Problem (12) fits $\mathbf{R}$ to the normalized sample covariance, while taking into account the correlation of the residuals, cf. [16]–[18].

Remark: The covariance-fitting criterion (12) is convex in $\boldsymbol{\theta}$. By contrast, if we use a Gaussian likelihood function for $\mathbf{y}$ to learn $\boldsymbol{\theta}$, the problem is nonconvex. Furthermore, as we will see next, the former learning approach can be readily solved online, unlike the latter.

IV. Online solution

By inspection, it can be verified that $\mathbf{p}$ in (10) is invariant to any scaling of $\boldsymbol{\theta}$. In other words, $\boldsymbol{\theta}$ and $c\boldsymbol{\theta}$, yield the same $\mathbf{p}$ for all $c > 0$.

Now the covariance-fitting criterion in (12) can be expanded as follows

$$\begin{aligned}\left\|\frac{1}{\rho}\mathbf{S}-\mathbf{R}\right\|^2_{\mathbf{R}^{-1}} &= \operatorname{tr}\left\{\left(\frac{1}{\rho}\mathbf{S}-\mathbf{R}\right)\mathbf{R}^{-1}\left(\frac{1}{\rho}\mathbf{S}-\mathbf{R}\right)\right\}\\ &= \frac{1}{\rho^2}\operatorname{tr}\left\{\mathbf{S}\mathbf{R}^{-1}\mathbf{S}\right\}+\operatorname{tr}\{\mathbf{R}\}-\frac{1}{\rho}2\operatorname{tr}\{\mathbf{S}\}\\ &= \operatorname{tr}\left\{\mathbf{R}^{-1}\mathbf{S}\right\}+\operatorname{tr}\{\mathbf{R}\}-2\rho,\end{aligned}$$

where the last equality follows from inserting (11). Thus we can write an equivalent cost function to that in (12) as:

$$\operatorname{tr}\left\{\mathbf{R}^{-1}\mathbf{S}\right\}+\operatorname{tr}\{\mathbf{R}\}. \tag{13}$$

In [12] we prove that omitting the constraint in (12) only rescales the optimal solution.

Let $\boldsymbol{\theta}_\star$ and $\hat{\boldsymbol{\theta}}$ denote the minimizers of (12) and (13), respectively. Since they are equal up to scale, using either in (10) yields the same linear predictor coefficients, which we denote $\mathbf{p}_\star$. Using this result, we also prove in [12] that the linear predictor (8) using covariance-fitted parameters can be written as

$$\hat{y}(\mathbf{x}) = \mathbf{p}_\star^\top\mathbf{y} = \boldsymbol{\psi}^\top(\mathbf{x})\hat{\mathbf{z}} \tag{14}$$

where $\boldsymbol{\psi}^\top(\mathbf{x}) \triangleq [\mathbf{u}^\top(\mathbf{x})\ \boldsymbol{\phi}^\top(\mathbf{x})]^\top$ is given with fixed dimension. The coefficient vector $\hat{\mathbf{z}}$ is obtained as the solution to a convex problem

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}}\ \underbrace{\|\mathbf{y}-\boldsymbol{\Psi}\mathbf{z}\|_2+\|\mathbf{w}\odot\mathbf{z}\|_1}_{\triangleq V(\mathbf{z})}, \tag{15}$$

where

$$\begin{aligned}\boldsymbol{\Psi} &= [\mathbf{U}\ \boldsymbol{\Phi}]\\ \mathbf{w} &= \frac{1}{\sqrt{N}}[0,\ldots,0,\|[\boldsymbol{\Phi}]_1\|_2\ \cdots\ \|[\boldsymbol{\Phi}]_q\|_2]^\top.\end{aligned}$$

Thus the learning problem is reduced to solving (15).

The convex cost function $V(z_1,\ldots,z_{T+q+1})$ can be minimized cyclically. That is, one element z_i at a time until convergence to the global minimum. More importantly, the cyclic minimization method can be performed using only recursively computed quantities. The derivation is beyond the scope of this short paper and we refer to [12] for details. In sum, as a new data sample $(\mathbf{x}, y(\mathbf{x}_n))$ is obtained, $\hat{\mathbf{z}}$—and therefore the predictor (14)—can be updated in an online manner. The memory requirement for the method is constant and the computational complexity grows linearly in the number of data points N. Since it is based on (12), we refer to it as 'sparse iterative covariance-based estimation' (SPICE) [18].

Remark: In the interest of reproducible research, we have made a MATLAB implementation available at http://www.it.uu.se/katalog/davza513.

V. Experiments

We evaluate the SPICE predictor for spatial RF fields using both synthetic and real data. The former dataset models a scenario with pathloss effects as well as medium-scale fading characteristics for $d = 2$. The latter dataset focuses on small-scale variations for $d = 3$.

A. Synthetic data

We consider a scenario with three transmitters located at

$$\bar{\mathbf{x}}_1 = \begin{bmatrix}-10\\-10\end{bmatrix},\quad \bar{\mathbf{x}}_2 = \begin{bmatrix}15\\5\end{bmatrix}\quad\text{and}\quad \bar{\mathbf{x}}_3 = \begin{bmatrix}15\\15\end{bmatrix}.$$

Their respective pathloss coefficients are $\eta_1 = 2$, $\eta_2 = 1.5$ and $\eta_3 = 3$. The constant coefficient η_0 is set to 70 dBm. The received power is modeled as a Gaussian process, i.e. $y(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), r(\mathbf{x},\mathbf{x}'))$, where $\mu(\mathbf{x})$ is given in (1). The covariance function captures medium-scale fading here and we use an isotropic model that belongs to the Matérn class [11]:

$$r(\mathbf{x},\mathbf{x}') = \sigma^2\left(1+\frac{\sqrt{3}\|\mathbf{x}-\mathbf{x}'\|}{\ell}\right)\exp\left(-\frac{\sqrt{3}\|\mathbf{x}-\mathbf{x}'\|}{\ell}\right),$$

where the length scale is $\ell = 10$ m and $\sigma = 5$ dBm. We consider predicting the process over a region $\mathcal{X}$ spanning 40×40 m. An example realization of $y(\mathbf{x})$ is given in Figure 2.

For the SPICE predictor we use order $m = 16$ in the approximative model (2) and (3). An example prediction, using a set of $N = 250$ randomly sampled observations, is also given in Figure 2. As can be seen, the predictor is capable of capturing the important spatial characteristics of the wireless channel.

In order to evaluate the performance, we consider a uniform grid over $\mathcal{X}$ consisting of $N_{\text{tot}} = 10^4$ points. We use a random subset of N points to learn the process and predict $\mathbf{y}(\mathbf{x})$ at the remaining $P = N_{\text{tot}} - N$ points. The performance is evaluated using the normalized mean square error

$$\text{NMSE} = \frac{\mathrm{E}[\|\mathbf{y}_P-\hat{\mathbf{y}}_P\|^2]}{\mathrm{E}[\|\mathbf{y}_P\|^2]},$$

where $\mathbf{y}_P$ and $\hat{\mathbf{y}}_P$ are the vectors of the data and predicted values, respectively. The NMSE is computed using 10^3 Monte Carlo runs. For a comparison, we consider a (batch) oracle predictor, which is given the true coefficients $\boldsymbol{\eta}$ as well as the true Matern covariance model above. The oracle predictor produces the MSE-optimal prediction of the random medium-scale fading. The comparison is also contrasted with a linear predictor which learns only the deterministic model (1) using a recursive least-squares (RLS) fitting of $\boldsymbol{\eta}$.

For the applications considered in this paper, N can become very large, e.g. $N > 10^3$ when collecting data with mobile

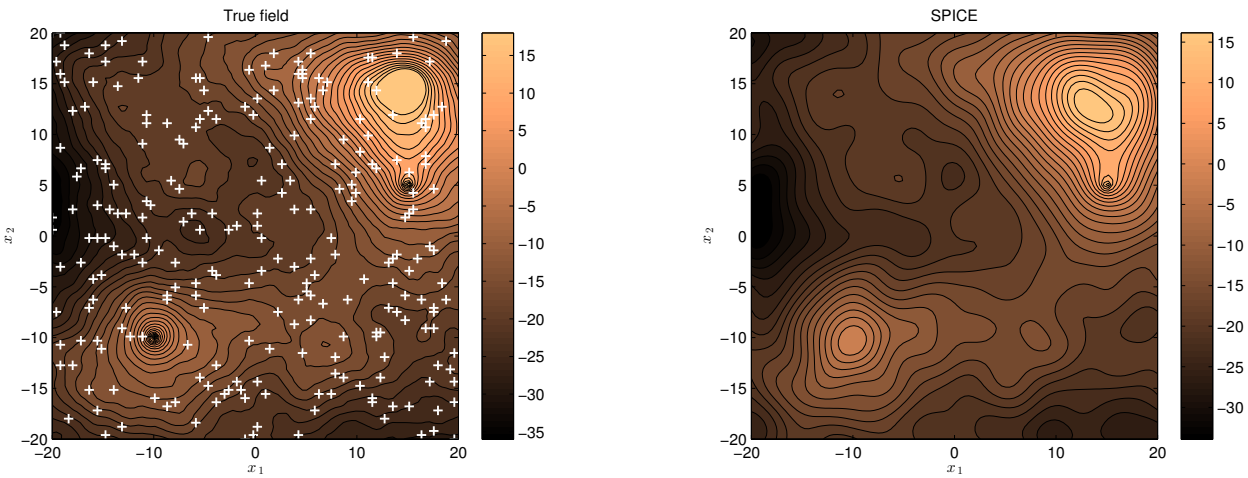

Fig. 2: Left: Example realization of $y(\mathbf{x})$ [dBm] and spatial coordinates $\mathbf{x}$ in [m]. Training points are marked as white crosses. Right: Prediction $\hat{y}(\mathbf{x})$ from SPICE predictor after learning from the training points. The contour plots are generated using a very fine spatial grid.

receivers, which renders the standard Gaussian process regression (GPR) techniques computationally intractable [11]. Here we use a direct implementation of the sparse spectrum Gaussian process regression (GPR-SS) method which is a state-of-the-art scalable batch method whose memory requirements and computational complexity scales linearly in N. We ran GPR-SS using the same number of parameters to learn as in SPICE.

The performances of the above methods are summarized in Table I. It is clear that the fading component is non-neglible here, since RLS does not improve its predictions as N increases. The batch method GPR-SS improves its spectral representation of the covariance structure with increasing N, which leads to reduced prediction errors. The SPICE predictor uses both the deterministic pathloss model (1) as well as the approximate covariance model (2) in an online manner and provides a more accurate prediction.

TABLE I: Normalized MSE [dB] versus N

N	Oracle	RLS	GPR-SS	SPICE
250	−30.61	−13.17	−18.90	−21.43
500	−35.00	−13.41	−20.92	−26.22
1000	−39.86	−13.11	−24.18	−30.90

B. Real data

We now turn to an experiment with small-scale variations using a single transmitter in a non-line-of-sight scenario. The setup is illustrated in Figure 3, both transmit and receive antennas are omni-directional in the horizontal plane. The transmitter is located approximately 20 m apart from the receiver, which is placed on a robot that moves within a volume $\mathcal{X}$ of $1475 \times 475 \times 475$ mm. The position accuracy of the robot is 0.1 mm. In these small-scale displacements, the pathloss effect is nearly constant across $\mathcal{X}$ and therefore

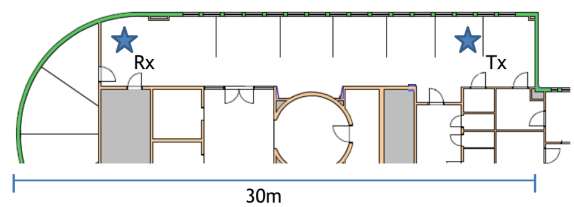

Fig. 3: Non-line-of-sight experiment setup in office environment. The thin lines in between the receiver (RX) and transmitter (TX) signify obstructing office dividers.

the distance dependence is neglected here. The measurements were conducted in the 2 GHz band and the bandwidth was 50 MHz. For a given frequency, we have $N_{\text{tot}} = 24\ 000$ data points sampled in a uniform grid in $\mathcal{X}$. We use a random subset of $N = 10\ 000$ points to train the SPICE predictor with $m = 16$ as before.

In Figure 4, we illustrate the predicted spatial field along two planes in $\mathbb{R}^3$ located at two different heights. The predictions show a distinct periodic pattern along the x_2-axis, which faces the transmitter. For an evaluation of the prediction errors on this dataset, we use the remaining $P = N_{\text{tot}} - N = 14\ 000$ points and compute

$$\text{RMSE} = \sqrt{\frac{1}{P}\|\mathbf{y}_P - \hat{\mathbf{y}}_P\|^2} \approx 2.06 \text{ dBm}.$$

This figure can be compared with the dynamic range of $y(\mathbf{x})$ which is $[-14.87, 8.86]$. Figure 5 shows the empirical distribution of the prediction errors that is centered near 0 but slightly skewed. These results indicate that the SPICE predictor method using the approximate covariance model (2) can also capture small-scale fading.

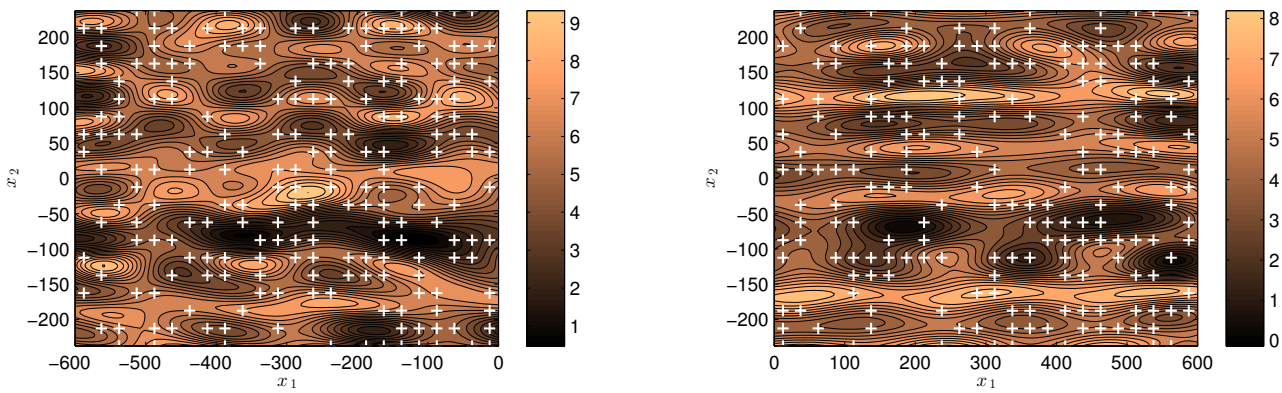

Fig. 4: Output from SPICE predictor $\hat{y}(\mathbf{x})$ [dBm]. Training points are marked as white crosses with spatial coordinates given in [mm]. Left: $x_3 = -87.5$ mm. Right: $x_3 = 237.5$ mm.

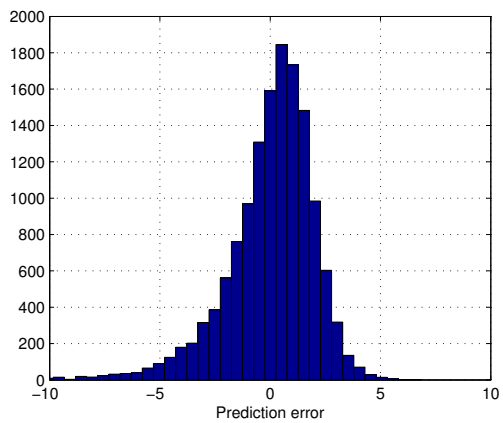

Fig. 5: Histogram of prediction errors in [dBm]. The RMSE is 2.06 [dBm]. The dynamic range of $y(\mathbf{x})$ is $[-14.87, 8.86]$.

VI. Conclusion

We have developed an online learning and prediction method for spatially varying RF fields that models distance dependent pathloss as well as random variations using an approximately isotropic covariance function. The learning method is based on a convex covariance-fitting approach. Its online capability enables prediction and analysis of the RF fields in large-scale datasets as well as streaming data scenarios.

References

[1] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Trans. Wireless Communications*, vol. 8, no. 8, pp. 4024–4036, 2009.

[2] M. Malmirchegini and Y. Mostofi, "On the spatial predictability of communication channels," *IEEE Trans. Wireless Communications*, vol. 11, no. 3, pp. 964–978, 2012.

[3] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Processing Magazine*, vol. 31, pp. 102–112, Nov 2014.

[4] N. Cressie, "The origins of kriging," *Mathematical Geology*, vol. 22, no. 3, pp. 239–252, 1990.

[5] M. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics, Springer New York, 1999.

[6] J. Chilès and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics, Wiley, 2012.

[7] M. Wellens, J. Riihijärvi, and P. Mähönen, "Spatial statistics and models of spectrum use," *Computer Communications*, vol. 32, no. 18, pp. 1998–2011, 2009.

[8] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *IEEE Int. Symp. Dynamic Spectrum Access Networks (DYSPAN)*, pp. 422–433, IEEE, 2012.

[9] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive computation and machine learning, MIT Press, 2002.

[10] C. Bishop, *Pattern Recognition and Machine Learning*. Information Science and Statistics, Springer, 2006.

[11] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Adaptative computation and machine learning series, MIT Press, 2006.

[12] D. Zachariah, P. Stoica, and T. B. Schön, "Online learning and prediction using linear covariance structures," 2016. Submitted.

[13] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics letters*, vol. 27, no. 23, pp. 2145–2146, 1991.

[14] N. Jaldén, P. Zetterberg, B. Ottersten, A. Hong, and R. Thomä, "Correlation properties of large scale fading based on indoor measurements," in *Wireless Communications and Networking Conference, 2007. WCNC 2007. IEEE*, pp. 1894–1899, IEEE, 2007.

[15] A. Solin and S. Särkkä, "Hilbert space methods for reduced-rank Gaussian process regression," 2014. arXiv preprint arXiv:1401.5508.

[16] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.

[17] T. W. Anderson, "Linear latent variable models and covariance structures," *Journal of Econometrics*, vol. 41, no. 1, pp. 91–119, 1989.

[18] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Processing*, vol. 59, no. 2, pp. 629–638, 2011.