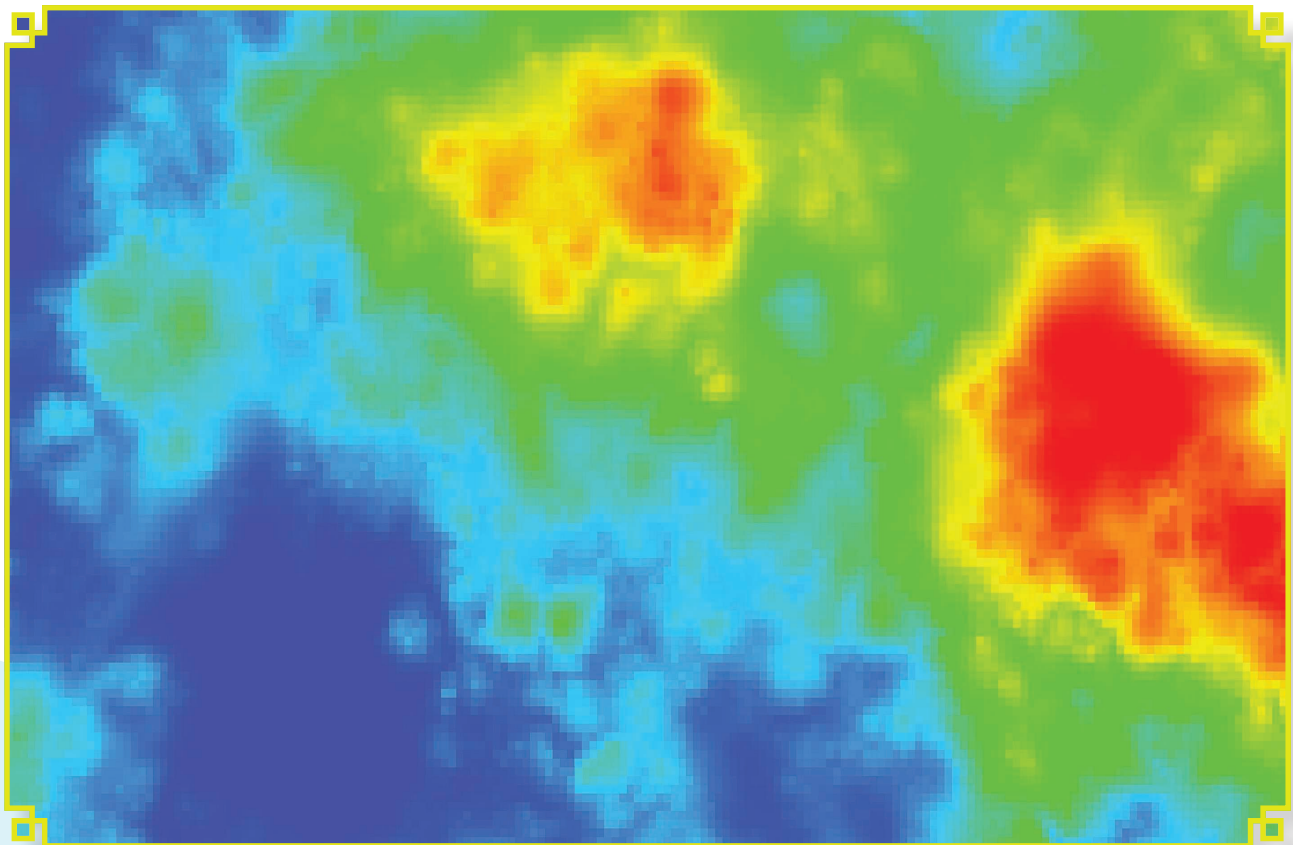Daniel Romero and Seung-Jun Kim

Radio Map Estimation

A data-driven approach to spectrum cartography

SHUTTERSTOCK.COM/GISKILLUP

adio maps characterize quantities of interest in radio communication environments, such as the received signal strength and channel attenuation, at every point of a geographical region. Radio map estimation (RME) typically entails interpolative inference based on spatially distributed measurements. In this tutorial article, after presenting some representative applications of radio maps, the most prominent RME methods are discussed. Starting from simple regression, the exposition gradually delves into more sophisticated algorithms, eventually touching upon state-of-the-art techniques. To gain insight into this versatile toolkit, illustrative toy examples will also be presented.

Introduction

Spectrum cartography comprises a collection of techniques used to construct and maintain radio maps, which provide useful information on the radio-frequency (RF) landscape, such as the received signal power, interference power, power spectral density (PSD), electromagnetic absorption, and channel gain across a geographic area; see, e.g., [1], [2], and [3]. A quick overview of the most representative types of radio maps is provided in Table 1.

Radio maps find a myriad of applications in wireless communications and networking, such as network planning, interference coordination and mitigation, power control, resource allocation, handoff management, multihop routing, dynamic spectrum access, and cognitive radio networking tasks; see [4]

and [5] and the references therein. Radio maps are also useful for localization [2] and tomography [6].
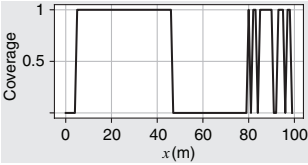
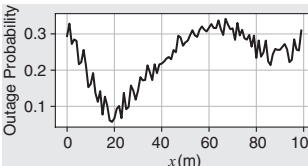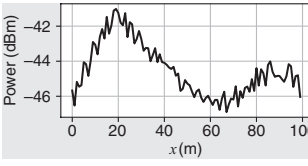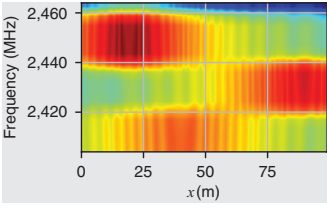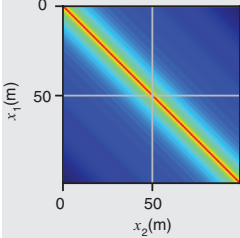Arguably, spectrum cartography can be traced back to the application of Maxwell's equations to characterize the propagation of radio waves across space. However, due to insufficient computational capacity, this approach has been traditionally confined to problems involving relatively simple geometries, such as determining the electromagnetic field radiated by a dipole. To analyze more complex environments, numerous empirical models have been developed, such as the well-known P recommendations from the International Telecommunication Union–Radiocommunication Sector. Unfortunately, this kind of model often fails to provide estimates that are accurate enough for a given application [7].

With the advent of modern computational resources, finite-element analysis and ray-tracing techniques paved the way for effectively approximating the solutions of Maxwell's equations in complex environments. However, besides their high computational complexity, their main limitation is that an accurate description of the propagation environment is required through 3D models of all objects and obstacles along with their electromagnetic properties.

To mitigate such limitations, RME was proposed, originally in the context of cognitive radios [1]. In RME, a collection of measurements acquired by spatially distributed sensors is used together with their locations to construct a map of the relevant RF descriptors, typically by applying some form of interpolation. As this approach does not require physical modeling of the propagation environment, it constitutes a data-driven alternative to the model-based techniques mentioned earlier. Since its conception, a sizable body of literature has emerged on the estimation of a variety of kinds of radio maps

Table 1. The most prominent types of radio maps.

Type of Map	Illustration in a 1D Scenario	Example Applications	Construction Complexity	Changes if . . .
Coverage map		Find coverage holes	The base station only needs to know if the mobile user can receive data.	• the environment changes. • the transmission activity changes. • the transmitter location or orientation changes.
Outage probability map		Improve the reliability of a cellular network		
Power map		Unveil regions of high interference; Determine appropriate locations for new base stations	The mobile user reports power measurements.	
PSD Map		Maximize frequency reuse	The mobile user reports power (density) measurements for each frequency (e.g., a periodogram).	
Channel gain map		Resource allocation for device-to-device communications	The user at position x_1 sends a pilot sequence. The user at x_2 sends an estimate of the received power to the base station after normalizing by the transmitted power.	• the environment changes.

Although radio maps find applications in many domains, this table exemplifies their applicability in cellular communications for specificity. The x-coordinate indexes a point on a road or railway.

for a wide range of application scenarios; see, e.g., [2], [4], [5], [8], and [9] as well as the references therein. Recently, the work in this area has intensified thanks to the boom of deep learning [10], [11], [12], [13].

This article provides an introduction to RME by guiding the reader on the foundations and applications of RME as well as on recent advances in this rapidly growing research area. To this end, the most common types of radio maps are first described. Afterward, RME methods for signal strength and propagation maps are expounded in a tutorial fashion. Practical considerations and future directions are also discussed.

Radio maps and their applications

The signal received at a certain location is determined by 1) the transmitted signal and 2) the communication channel between the transmitter and the receiver. Depending on whether the focus is on the combined effect of the two or, rather, on the effect of the propagation channel itself, two families of radio maps can be considered: signal strength maps and propagation maps.

For simplicity, unless stated otherwise, it will be assumed that the maps do not change significantly within the time interval under consideration. In practice, the length of the interval for which this assumption remains valid depends not only on the speed of variation but also on the specific application.

Signal strength maps

Signal strength maps focus on metrics of the received signal, which are determined by the aggregate effects of the channel upon the signals transmitted by all active sources. This is the case, for instance, if the goal is to map interference power levels. Constructing such maps does not require knowledge of the number, locations, and power of the transmitters, which is appealing in scenarios involving a large number of mobile transmitters, as in device-to-device communications or cellular uplink channels. Different kinds of signal strength maps are presented next by the increasing level of detail they capture.

Coverage maps

The coarsest characterization of the radio environment can be provided by a map that takes only binary values for coverage indication. Specifically, let $p(\boldsymbol{x})$ denote the signal power that a radio with an isotropic antenna (the case of nonisotropic antenna patterns is discussed later) receives at a spatial location $\boldsymbol{x} \in \mathcal{X}$ where $\mathcal{X}$ represents a geographical region of interest, typically a subset of $\mathbb{R}$, $\mathbb{R}^2$, or $\mathbb{R}^3$. A coverage map is a function $s : \mathcal{X} \to \{0, 1\}$ that takes the value $s(\boldsymbol{x}) = 1$ if $p(\boldsymbol{x}) \geq \gamma$ and zero otherwise, where γ is a given threshold. This threshold may correspond to the minimum signal power necessary to guarantee a prescribed communication rate. Coverage maps may also be constructed by replacing $p(\boldsymbol{x})$ in this definition with the signal-to-noise-power ratio or the signal-to-interference-plus-noise-power ratio.

Coverage maps are often used by cellular and TV broadcast network operators to find areas of weak coverage, which allows them to determine suitable sites for deploying new base stations and relay antennas. A more recent application is mission planning for autonomous mobile robots or vehicles that require network connectivity, where coverage maps may assist in, e.g., minimizing the time and distance traversed without connectivity.

Outage probability maps

A soft version of coverage maps can be constructed by adopting a probabilistic perspective, as the effects of the channel, such as fading and shadowing, are often modeled as random. An outage probability map $q(\boldsymbol{x})$ is a function $q : \mathcal{X} \to [0, 1]$ that provides the probability that $p(\boldsymbol{x}) < \gamma$. Since outage probability maps capture more detailed information than coverage maps, the former can be readily employed in the applications of the latter. However, the additional information provided by outage probability maps allows more sophisticated decision making, as in route planning [14].

Power maps

A substantially finer characterization of the signal strength is obtained by a power map, defined as a function $p : \mathcal{X} \to \mathbb{R}$, which returns the received power $p(\boldsymbol{x})$ at every spatial location $\boldsymbol{x} \in \mathcal{X}$ As the information contained in power maps is richer than that in coverage or outage probability maps, power maps can be used not only for tasks such as network planning and trajectory optimization but also for localizing transmitters [2]. Also, in fingerprint-based localization, a mobile device can measure the received powers of nearby access points and determine its own position by matching the measurements with the values of the map.

PSD maps

One is sometimes interested in the power distribution not only across space but also across the frequency domain. A PSD map is a function $p : \mathcal{X} \times \mathcal{F} \to \mathbb{R}$ that provides the PSD $p(\boldsymbol{x}, f)$ of the received signal at each location $\boldsymbol{x} \in \mathcal{X}$. Here, $f \in \mathcal{F}$ is the frequency variable and the set $\mathcal{F} \subset \mathbb{R}$ contains the frequencies of interest. If the latter is discretized as $\mathcal{F} = \{f_1, \ldots, f_{N_f}\}$, constructing a PSD map is tantamount to constructing a collection of power maps proportional to $p(\boldsymbol{x}, f_1), \ldots, p(\boldsymbol{x}, f_{N_f})$.

In addition to the applications mentioned for the previous kinds of signal strength maps, PSD maps enable additional use cases. For example, they can be used for speeding up handoff procedures in cellular networks by providing the quality of the relevant channels at a given location, obviating the need for time-consuming channel measurement or feedback processes. PSD maps can also be utilized for interference coordination

> **Signal strength maps focus on metrics of the received signal, which are determined by the aggregate effects of the channel upon the signals transmitted by all active sources.**

where concurrent transmissions are assigned to different frequency band channels based on the transceiver locations, promoting efficient spectrum reuse. In cognitive radio networks, PSD maps can unveil underutilized "white spaces" in the space/frequency/time domains, which can be exploited opportunistically by unlicensed users [15].

Propagation maps

Whereas signal strength maps capture the aggregate effect of the transmitted signals and the channels, propagation maps focus exclusively on the channel. Each parameter of interest gives rise to a different kind of propagation map. As described next, channel gain maps constitute the simplest kind. Suppose that p^{RX} denotes the power received at location x^{RX} due to a transmitter with power p^{TX} at location x^{TX}. A channel gain map is a function $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ of the transmitter and receiver locations that provides the channel gain $h(x^{\mathrm{TX}}, x^{\mathrm{RX}}) = p^{\mathrm{RX}}/p^{\mathrm{TX}}$. More sophisticated propagation maps arise by accounting for frequency selectivity. For example, the power gain that each subcarrier sees in an orthogonal frequency division multiplexing system can be mapped. For simplicity, this article focuses on channel gain maps, which provide the overall gain that affects a single narrow frequency band.

Clearly, given a channel gain map $h(x^{\mathrm{TX}}, x^{\mathrm{RX}})$ together with the locations $(x_1^{\mathrm{TX}}, \ldots, x_S^{\mathrm{TX}})$ and transmit powers $p_1^{\mathrm{TX}}, \ldots, p_S^{\mathrm{TX}}$ of S sources in a region, one can obtain the power map as $p(x) = \Sigma_s h(x_s^{\mathrm{TX}}, x) p_s^{\mathrm{TX}}$, provided that the signals transmitted by different sources are uncorrelated, as generally occurs in practice, except, e.g., in single-frequency networks, such as the ones utilized by digital television broadcasting. Thus, propagation maps can be readily used in the applications of signal strength maps provided that the locations and transmit powers of the sources are known. On the other hand, propagation maps offer more versatile information than signal strength maps: whereas a signal strength map may provide the total interference at each location, a propagation map reveals the contribution of each source. This enhanced flexibility is instrumental for tasks such as interference coordination or network planning.

Observe that changes in the locations and transmit powers of the sources give rise to changes in signal strength maps, whereas propagation maps remain unaffected. On the other hand, alterations in the scattering environment, such as the construction of new buildings or seasonal changes of foliage, affect both propagation and signal strength maps. Thus, the time scale of the variations of signal strength maps is never greater than that of propagation maps. Hence, propagation maps can be used to construct signal strength maps in highly dynamic setups, such as the uplinks of cellular networks, where mobile users rapidly change their positions and activity patterns.

Propagation maps can also help address the classic problem of predicting the potential interference inflicted to passive receivers, which arises in the context of cognitive radios [15]. For example, when reusing the TV spectrum, the challenge is to carry out unlicensed transmissions without introducing detrimental interference to TV receivers. With a propagation map, one can ensure that no receivers in a certain area will be negatively affected without the need to know their precise locations [16].

Yet another application is the problem of aerial base station placement, where a propagation map of the air-to-ground channel can be constructed to determine the best set of locations to deploy unmanned aerial vehicle (UAV)-mounted base stations to serve ground users [17].

> **Whereas signal strength maps capture the aggregate effect of the transmitted signals and the channels, propagation maps focus exclusively on the channel.**

Estimation of signal strength maps

In a typical RME formulation, the goal is to construct a radio map using a set of measurements acquired by spatially dispersed sensors together with their locations. For signal strength maps, consider N measurements, where the nth measurement m_n is acquired by a sensor at location x_n. In the case of power maps, m_n may be the average power measured in a certain band within a given time interval, which can be modeled as $m_n = p(x_n) + z_n$. Here, z_n denotes measurement noise, which is caused, e.g., by the finite length of the averaging time interval. For estimating PSD maps, m_n can contain power spectrum measurements, such as periodograms. The RME problem becomes constructing the desired signal strength map given the pairs $\{(x_n, m_n)\}_{n=1}^N$.

It is worth noting that each sensor may collect measurements at multiple locations provided that they are taken within a time window whose length is small relative to the scale of variations of the target map. Thus, the number of sensors may be much smaller than N. In fact, the RME formulation can be extended to accommodate the decision on where to acquire the measurements sequentially, as discussed in the "Spectrum Surveying" section. Furthermore, a sensor need not be a special-purpose device. For example, a user terminal in a cellular network may function as a sensing device. The rest of the section presents the main approaches for constructing signal strength maps.

Estimation of power maps

This section introduces the main power map estimation methods, a comparison of which can be found in Table 2.

Linear parametric RME

Let us start from the simple yet illustrative scenario where there is a single transmitter with known location x_1^{TX} in free space. As per Friis's transmission equation, the received power at location x is inversely proportional to the squared distance $\|x - x_1^{\mathrm{TX}}\|^2$. In other words, $p(x)$ can be written as $p(x) = \alpha_1 \psi_1(x)$ where $\psi_1(x) := 1/\|x - x_1^{\mathrm{TX}}\|^2$, and α_1 depends on the (unknown) transmit power. Therefore, to estimate $p(x)$ everywhere, it suffices to obtain α_1. Clearly, this

could be accomplished from a single noiseless measurement $m_1 = p(\boldsymbol{x}_1)$ at $\boldsymbol{x}_1$ by setting $\alpha_1 = m_1/\psi_1(\boldsymbol{x}_1)$.

Similarly, if S transmitters with known locations $\boldsymbol{x}_1^{\text{TX}}, \ldots, \boldsymbol{x}_S^{\text{TX}}$ are active in a certain region, one can let $\psi_s(\boldsymbol{x}) \coloneqq 1/\|\boldsymbol{x} - \boldsymbol{x}_s^{\text{TX}}\|^2$, to write $p(\boldsymbol{x})$ as

$$p(\boldsymbol{x}) = \alpha_1 \psi_1(\boldsymbol{x}) + \cdots + \alpha_S \psi_S(\boldsymbol{x}) \qquad (1)$$

so long as the transmitted waveforms are uncorrelated. Based on (1), one can typically estimate the S coefficients $\{\alpha_s\}$ from S noiseless measurements by solving the system of equations

$$m_1 = \alpha_1 \psi_1(\boldsymbol{x}_1) + \cdots + \alpha_S \psi_S(\boldsymbol{x}_1)$$
$$\vdots$$
$$m_S = \alpha_1 \psi_1(\boldsymbol{x}_S) + \cdots + \alpha_S \psi_S(\boldsymbol{x}_S). \qquad (2)$$

In practice, however, the measurements are noisy, and one may use more than S of them to estimate the coefficients. Upon defining $\boldsymbol{\alpha} \coloneqq [\alpha_1, \ldots, \alpha_S]^\top$, $\boldsymbol{m} \coloneqq [m_1, \ldots, m_N]^\top$, $(\boldsymbol{\Psi})_{n,s} \coloneqq \psi_s(\boldsymbol{x}_n)$, and $\boldsymbol{z} \coloneqq [z_1, \ldots, z_N]^\top$, (2) can be extended to the case with $N > S$ measurements as $\boldsymbol{m} = \boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{z}$. The least-squares (LS) estimate of $\boldsymbol{\alpha}$ is, therefore, $\hat{\boldsymbol{\alpha}} = \operatorname{argmin}_{\boldsymbol{\alpha}} \|\boldsymbol{m} - \boldsymbol{\Psi}\boldsymbol{\alpha}\|^2$. Because the number S of parameters to be estimated does not depend on the number N of measurements, this approach is termed *para-*

metric. Further parametric and *nonparametric* estimators are discussed in the rest of this section.

Figure 1 illustrates a setup where a map needs to be estimated on a line; i.e., the region of interest is given by $\mathcal{X} \subset \mathbb{R}$, which may correspond, e.g., to a road or a railway. The true map and the estimated map obtained by substituting $\hat{\boldsymbol{\alpha}}$ into the right-hand side (RHS) of (1) are compared. The estimated map is seen to be reasonably accurate and
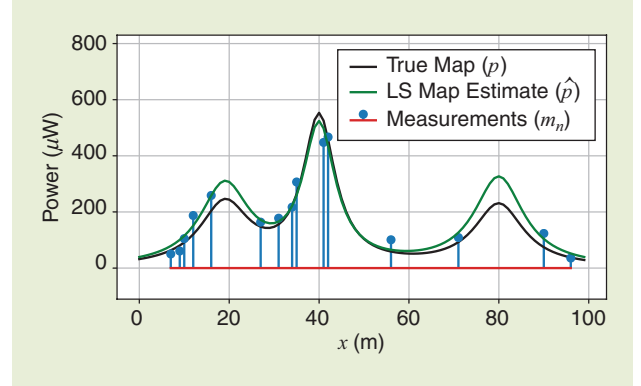

FIGURE 1. An example of map estimation in 1D using a parametric estimator that knows the transmitter locations. The estimate is reasonably accurate despite the low number of measurements.

Table 2. A comparison of the power map estimation methods discussed in this tutorial.

Method	Input (Besides Measurements)	Strengths	Limitations
Linear parametric RME	• Transmitter locations $\boldsymbol{x}_1^{\text{TX}}, \ldots, \boldsymbol{x}_S^{\text{TX}}$ • Path loss law: e.g., $\psi_s(\boldsymbol{x}) \coloneqq 1/\|\boldsymbol{x} - \boldsymbol{x}_s^{\text{TX}}\|^2$	• Simplicity • Closed form • Accuracy in LOS conditions • Can easily accommodate knowledge of transmit antenna patterns	• Inaccurate in non-LOS conditions • Requires transmitter locations
Kernel-based learning	• Reproducing kernel $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ • Loss $\mathcal{L}$ • Regularization parameter λ	• High flexibility • Does not require transmitter locations	• Sensitive to the choice of the kernel • Depending on $\mathcal{L}$, may require a numerical solver • λ must be tuned, e.g., via cross validation
Kriging	• Mean $\mu_p(\boldsymbol{x})$ and covariance $\operatorname{Cov}[p(\boldsymbol{x}), p(\boldsymbol{x}')]$ of p • Measurement noise variance σ_z^2	• LMMSE optimality • Closed form • Naturally suited to the customary log-normal shadowing model • Estimation error that can be quantified	• Accurate covariance structure may be hard to obtain • Requires user locations
Sparsity-based methods	• Discrete grid • Regularization parameter λ	• Efficient algorithms available for obtaining a solution • Recovered sparse solution readily interpretable	• Prior knowledge on propagation characteristics needed • Errors due to grid mismatch
Matrix completion	• Regular grid • Regularization parameter λ	• Agnostic to propagation characteristics • Spatial correlation structure exploited	• Critical low-rank condition • Sufficient number of measurements required for stable interpolation
Dictionary learning	• Dictionary size Q • Regularization parameters λ_s and λ_L	• Powerful union-of-subspace prior for spatial patterns • Can accommodate rapid temporal dynamics	• Nonconvex optimization • Hyperparameter tuning necessary
Deep learning	• Terrain maps • Vegetation maps • Building height maps • Network architecture • Training parameters • Others	• Can learn propagation patterns from a dataset • More accurate than other methods if sufficient data are available [18]	• Large amount of data required • Training that is computationally intensive

LMMSE: linear minimum mean-square error; LOS: line of sight.

can be shown to converge to the true map for $N \to \infty$ under mild conditions.

So far, it was assumed that propagation takes place in free space. If this is not the case, then the basis functions $\psi_s(\boldsymbol{x}) = 1/\|\boldsymbol{x} - \boldsymbol{x}_s^{\mathrm{TX}}\|^2$ may not yield a satisfactory fit. Although one can, in principle, adopt other families of basis functions, such as those determined by the well-known Okumura–Hata model, the flexibility of such an approach is rather limited. In addition, the location of the sources is required, which may not be a realistic assumption in some applications. These observations suggest generalizing (1) to

$$p(\boldsymbol{x}) = \alpha_1 \tilde{\psi}_1(\boldsymbol{x}) + \cdots + \alpha_B \tilde{\psi}_B(\boldsymbol{x}) \tag{3}$$

where $\tilde{\psi}_b(\boldsymbol{x})$ can take an arbitrary form and need not even be linked to any particular transmitter. For example, in the case where a map needs to be constructed on a line, $\{\tilde{\psi}_b(\boldsymbol{x})\}_b$ could form a polynomial basis by setting $\tilde{\psi}_b(\boldsymbol{x}) = \tilde{\psi}_b(x) = x^{b-1}$. The coefficients $\{\alpha_b\}$ can again be found by LS estimation. However, despite the appealing simplicity of this approach, the quality of the estimates is often poor. As illustrated
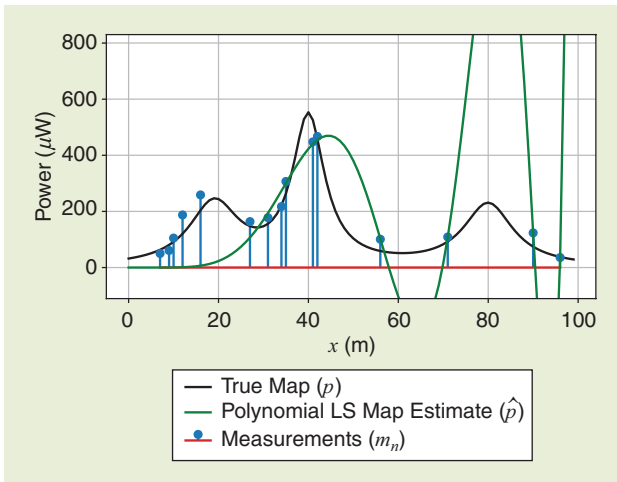

FIGURE 2. An example of map estimation by fitting a polynomial of degree 13 via LS. The estimate is clearly unsatisfactory despite the fact that the estimate accurately fits most of the measurements.
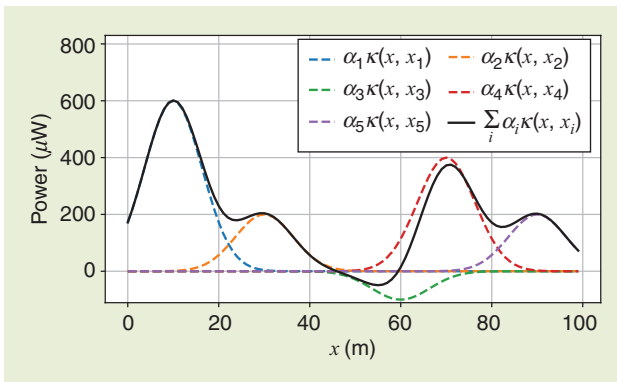

FIGURE 3. An example of a function in an RKHS obtained with the expansion in (4) with only five terms.

by Figure 2 for the same setup as in Figure 1, this kind of regression method may be sensitive to the choice of the basis functions.

Kernel-based learning

The main challenge faced by the parametric methods described in the previous section lies in the difficulty in selecting suitable basis functions. This difficulty is further exacerbated in higher dimensions, such as when $\mathcal{X} = \mathbb{R}^2$ or $\mathbb{R}^3$. Kernel-based learning can sidestep this issue while enjoying simplicity, universality, and good performance [19].

Upon postulating a family of functions $\mathcal{G}$, the goal is to select, based on the data $\{(\boldsymbol{x}_n, m_n)\}_{n=1}^N$, a function $\hat{p}$ in $\mathcal{G}$ that satisfies $\hat{p}(\boldsymbol{x}) \approx p(\boldsymbol{x}) \ \forall \boldsymbol{x}$. In kernel-based learning, $\mathcal{G}$ is a special class of functions termed *reproducing-kernel Hilbert space* (*RKHS*), given by

$$\mathcal{G} := \left\{ g : g(\boldsymbol{x}) = \sum_{i=1}^{\infty} \alpha_i \kappa(\boldsymbol{x}, \boldsymbol{x}_i'), \ \boldsymbol{x}_i' \in \mathcal{X}, \ \alpha_i \in \mathbb{R} \ \forall i \right\}. \tag{4}$$

Here, $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a reproducing kernel [19, Ch. 2], which is a function that is 1) symmetric—i.e., $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa(\boldsymbol{x}', \boldsymbol{x}) \ \forall \boldsymbol{x}', \boldsymbol{x}$—and 2) positive definite, meaning that the matrix $\bar{\boldsymbol{K}}$ with entries $(\bar{\boldsymbol{K}})_{i,j} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is positive definite for any set of points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$. A common choice is the so-called Gaussian radial basis function $\kappa(\boldsymbol{x}, \boldsymbol{x}') := \exp\left(-\|\boldsymbol{x} - \boldsymbol{x}'\|^2/2\sigma^2\right)$, where $\sigma > 0$ is a prescribed parameter. Seen as a function of $\boldsymbol{x}$, $\kappa(\boldsymbol{x}, \boldsymbol{x}_i')$ is in this case a bell-shaped surface centered at $\boldsymbol{x}_i'$. Thus, it can be observed from (4) that a function in $\mathcal{G}$ is a superposition of (a possibly infinite number of) Gaussian bells with different centers and amplitudes, as illustrated in Figure 3.

A typical approach for finding a suitable estimate $\hat{p}$ in $\mathcal{G}$ is to solve

$$\hat{p} = \underset{g \in \mathcal{G}}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}(m_n, g(\boldsymbol{x}_n)) + \lambda \|g\|_{\mathcal{G}}^2 \tag{5}$$

where $\lambda > 0$ is a predetermined regularization parameter, and $\mathcal{L}$ is a loss function quantifying the deviation between the observations $\{m_n\}_{n=1}^N$ and the predictions $\{g(\boldsymbol{x}_n)\}_{n=1}^N$ produced by a candidate g. If the square loss $\mathcal{L}(m_n, g(\boldsymbol{x}_n)) = (m_n - g(\boldsymbol{x}_n))^2$ is adopted, (5) becomes kernel ridge regression (KRR) [19, Ch. 4] (Figures 4 and 5).

The RKHS norm of $g(\boldsymbol{x}) = \sum_{i=1}^{\infty} \alpha_i \kappa(\boldsymbol{x}, \boldsymbol{x}_i')$ is given by

$$\|g\|_{\mathcal{G}} := \sqrt{\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \alpha_i \alpha_j \kappa(\boldsymbol{x}_i', \boldsymbol{x}_j')}. \tag{6}$$

The term $\|g\|_{\mathcal{G}}^2$ in (5) can be replaced by other increasing functions of $\|g\|_{\mathcal{G}}$, but this explanation considers just the special case of $\|g\|_{\mathcal{G}}^2$ for simplicity.

To understand the role of the regularization term $\lambda \|g\|_{\mathcal{G}}^2$ in (5), first note that $\mathcal{L}$ is typically designed so that its minimum is attained when $g(\boldsymbol{x}_n) = m_n$. Thus, in the absence of the regularization term, owing to the infinite degrees of freedom

of g [compare with (4)], the solution $\hat{p}$ to (5) would achieve a perfect fit for all measurements. However, such a $\hat{p}$ would typically be highly irregular since it would fit even the noise component of the measurements and, thus, likely differ significantly from p at the locations where no measurements were taken. The regularization term helps avoid such overfitting by promoting smoothness in $\hat{p}$. The reason is that, since κ is positive definite, $\|g\|_{\mathcal{G}}^2$ penalizes large values of $\{\alpha_i\}$, which tend to occur in overfitted solutions. Parameter λ is adjusted to achieve the "sweet spot" between data fitting and regularization.

> **It is worth mentioning that RME based on kernel methods is best suited for scenarios where no prior knowledge of the propagation environment is available.**

To solve (5), one could initially think of substituting the expansion (4) into (5) and optimizing over the infinitely many coefficients $\{\alpha_i\}$ and centroids $\{x_i'\}$. However, this approach is obviously intractable. Instead, the so-called representer theorem can be invoked [19, Th. 4.2], which states that the solution to (5) must be of the form

$$\hat{p}(\boldsymbol{x}) = \sum_{n=1}^{N} \alpha_n \kappa(\boldsymbol{x}, \boldsymbol{x}_n) \tag{7}$$

for some $\{\alpha_n\}_{n=1}^{N}$. Observe that the centroids in (7) are precisely the measurement locations. This effectively reduces an optimization problem with infinitely many variables to a problem with just the N variables $\alpha_1, \ldots, \alpha_N$. For example, if one adopts the square loss, substituting (7) into (5) yields

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{N} \|\boldsymbol{m} - \boldsymbol{K}\boldsymbol{\alpha}\|^2 + \lambda \boldsymbol{\alpha}^\top \boldsymbol{K}\boldsymbol{\alpha} \tag{8}$$

where (with some abuse of notation) $\boldsymbol{\alpha} := [\alpha_1, \ldots, \alpha_N]^\top$, and $\boldsymbol{K}$ is an N-by-N matrix with $(\boldsymbol{K})_{i,j} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$. It should be noted that now the number of parameters to be determined depends on the number of measurements N, which is why these kinds of methods are called *nonparametric*. Problem (8) admits the closed-form solution

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{K} + \lambda N \mathbf{I}_N)^{-1} \boldsymbol{m} \tag{9}$$

from which $\hat{p}$ can be obtained via (7). Figures 4 and 5 show the KRR-based map estimates in the same setup as in Figures 1 and 2. It can be seen that, as the number of measurements increases, the estimated map becomes closer to the true map.

It is worth mentioning that RME based on kernel methods is best suited for scenarios where no prior knowledge of the propagation environment is available. When prior information, such as the transmitter locations or the path loss exponent, is indeed available, it is also possible to combine the flexibility of nonparametric kernel methods with the ability of parametric methods to capture prior information by means of appropriate basis functions. To this end, one can postulate that p can be represented as the sum of a function in the form of (3) and a function in an RKHS [20]. Such an approach also generalizes the so-called thin-plate spline regression, which has well-documented merits in RME [4], [9].

Another limitation of kernel-based methods is the need for choosing the kernel (including its parameters), which may affect the estimation performance significantly. This difficulty may be alleviated through multikernel learning, where a dictionary of kernels can be specified and a suitably designed algorithm uses the measurements to construct a kernel by combining the kernels in the dictionary; see the references in [18].

Kriging

RME can also be formulated in a statistical framework, where $p(\boldsymbol{x})$ is treated as a random process. A popular approach is kriging, which is a linear spatial interpolator based on the linear minimum mean-square error (LMMSE) criterion [1], [21], [22]. In simple kriging, the mean and the covariance of $p(\boldsymbol{x})$ are assumed to be known. That is, $\mu_p(\boldsymbol{x}) := \mathbb{E}[p(\boldsymbol{x})]$ and $\mathrm{Cov}[p(\boldsymbol{x}), p(\boldsymbol{x}')]$ are given for all $\boldsymbol{x}$ and $\boldsymbol{x}'$. How to obtain these functions is discussed later.

Under the measurement model $m_n = p(\boldsymbol{x}_n) + z_n$, $n = 1, 2, \ldots, N$, assume that z_n is zero mean with variance σ_z^2 and uncorrelated with $z_{n'}$ for all $n' \neq n$ and with $p(\boldsymbol{x})$ for all $\boldsymbol{x}$. Thus, the mean and covariance of the measurements are, respectively,
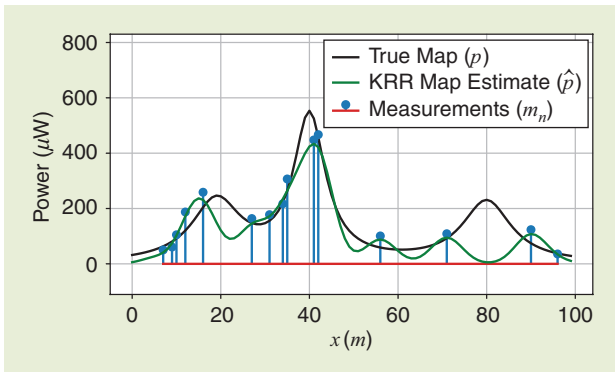

FIGURE 4. An example of a KRR estimate. As expected, the quality of the fit is higher in regions with higher measurement density.
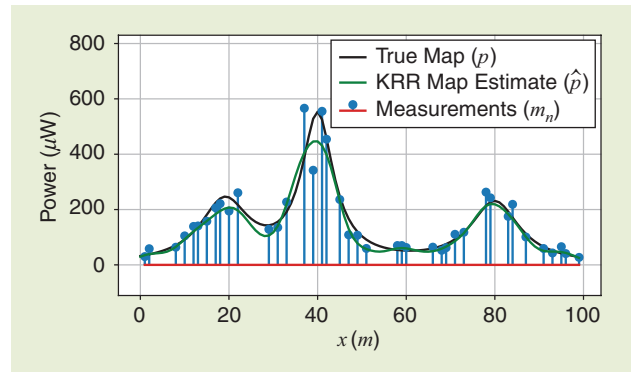

FIGURE 5. An example of a KRR estimate with more measurements than in Figure 4. The fit is considerably better.

$\mathbb{E}[m_n] = \mu_p(\boldsymbol{x}_n)$ and $\mathrm{Cov}[m_n, m_{n'}] = \mathrm{Cov}[p(\boldsymbol{x}_n), p(\boldsymbol{x}_{n'})] + \sigma_z^2 \delta_{n,n'}$, where $\delta_{n,n'}$ equals one if $n = n'$ and zero otherwise. It can also be verified that $\mathrm{Cov}[p(\boldsymbol{x}), m_n] = \mathrm{Cov}[p(\boldsymbol{x}), p(\boldsymbol{x}_n)]$. Then, it can be shown that the LMMSE estimator of $p(\boldsymbol{x})$ based on the measurements $\boldsymbol{m} := [m_1, \ldots, m_N]^\top$ is given by

$$\hat{p}(\boldsymbol{x}) = \mu_p(\boldsymbol{x}) + \mathrm{Cov}[p(\boldsymbol{x}), \boldsymbol{m}] \mathrm{Cov}^{-1}[\boldsymbol{m}, \boldsymbol{m}](\boldsymbol{m} - \mathbb{E}[\boldsymbol{m}]) \quad (10)$$

where $\mathrm{Cov}[\boldsymbol{m}, \boldsymbol{m}]$ is the $N \times N$ matrix whose (n, n')th entry is $\mathrm{Cov}[m_n, m_{n'}]$, and $\mathrm{Cov}[p(\boldsymbol{x}), \boldsymbol{m}]$ is the $1 \times N$ vector with the nth entry equal to $\mathrm{Cov}[p(\boldsymbol{x}), m_n]$.

It is worth comparing (10) with (7) and (9). It can be easily seen that, except for the mean terms in (10), the estimators provided by (10) and (7) coincide if one sets $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}[p(\boldsymbol{x}), p(\boldsymbol{x}')]$, and λ is adjusted properly. This is a manifestation of the well-known fact that a reproducing kernel can be thought of as a generalization of covariance. As a result, some of the practical issues and corresponding mitigation strategies for kernel-based learning apply to kriging as well.

To obtain the mean $\mu_p(\boldsymbol{x})$ and the covariance $\mathrm{Cov}[p(\boldsymbol{x}), p(\boldsymbol{x}')]$ of the map $p(\boldsymbol{x})$ to be estimated, one can rely on historic measurement data. Given the covariance function, universal kriging also provides a framework to estimate $\mu_p(\boldsymbol{x})$ as a part of the kriging estimator.

> **In many practical RME problems, estimation performance can be significantly improved by incorporating prior information.**

Next, a simple example with a single transmitter at location $\boldsymbol{x}^{\mathrm{TX}}$ transmitting with power p^{TX} will be used to illustrate how the mean and covariance can be derived from common propagation models; a more sophisticated example involving the idea of universal kriging and incorporating temporal variations as well will be presented in the "Nontomographic Approaches" section. To this end, note that the received power in the logarithmic scale can be written as $p_{\mathrm{dB}}(\boldsymbol{x}) = p_{\mathrm{dB}}^{\mathrm{TX}} + h_{\mathrm{dB}}(\boldsymbol{x}^{\mathrm{TX}}, \boldsymbol{x})$, where $p_{\mathrm{dB}}^{\mathrm{TX}}$ and $h_{\mathrm{dB}}(\boldsymbol{x}^{\mathrm{TX}}, \boldsymbol{x})$, are expressed in decibels. A common decomposition for the latter is $h_{\mathrm{dB}}(\boldsymbol{x}^{\mathrm{TX}}, \boldsymbol{x}) = h^{\mathrm{PL}}(\boldsymbol{x}) - a^{\mathrm{SF}}(\boldsymbol{x}) - a^{\mathrm{FF}}(\boldsymbol{x})$, where $h^{\mathrm{PL}}(\boldsymbol{x})$ is the path loss, $a^{\mathrm{SF}}(\boldsymbol{x})$ is the attenuation due to shadow fading, and $a^{\mathrm{FF}}(\boldsymbol{x})$ is the attenuation due to fast fading. The dependence on $\boldsymbol{x}^{\mathrm{TX}}$ and the subscript "dB" on the RHS have been omitted for brevity. Recall that shadow fading is produced by obstructions in the line of sight between the transmitter and the receiver, whereas fast fading is due to the constructive and destructive interference between the different multipath components arriving at the receiver.

With this decomposition, it is common to model $h^{\mathrm{PL}}(\boldsymbol{x})$ as a deterministic function of $\boldsymbol{x}$. Furthermore, $a^{\mathrm{SF}}(\boldsymbol{x})$ and $a^{\mathrm{FF}}(\boldsymbol{x}')$ can be assumed to be uncorrelated for all $\boldsymbol{x}$ and $\boldsymbol{x}'$ and to have means μ^{SF} and μ^{FF} respectively. The spatial structure of $a^{\mathrm{SF}}(\boldsymbol{x})$ is often captured by a simple correlation model, such as the Gudmundson model [23], which prescribes that $\mathrm{Cov}[a^{\mathrm{SF}}(\boldsymbol{x}), a^{\mathrm{SF}}(\boldsymbol{x}')] = \sigma_{\mathrm{SF}}^2 2^{-\|\boldsymbol{x}-\boldsymbol{x}'\|/d^{\mathrm{SF}}}$. Here, σ_{SF}^2 is a constant, and d^{SF} is the distance at which the correlation decays by 50%. On the other hand, due to the rapid spatial variability of $a^{\mathrm{FF}}(\boldsymbol{x})$, it is reasonable to set $\mathrm{Cov}[a^{\mathrm{FF}}(\boldsymbol{x}), a^{\mathrm{FF}}(\boldsymbol{x}')] = \sigma_{\mathrm{FF}}^2 \delta_{\boldsymbol{x}, \boldsymbol{x}'}$.

Then, we have $\mu_{p,\mathrm{dB}}(\boldsymbol{x}) = p_{\mathrm{dB}}^{\mathrm{TX}} + h^{\mathrm{PL}}(\boldsymbol{x}) - \mu^{\mathrm{SF}} - \mu^{\mathrm{FF}}$ and $\mathrm{Cov}[p_{\mathrm{dB}}(\boldsymbol{x}), p_{\mathrm{dB}}(\boldsymbol{x}')] = \sigma_{\mathrm{SF}}^2 2^{-\|\boldsymbol{x}-\boldsymbol{x}'\|/d^{\mathrm{SF}}} + \sigma_{\mathrm{FF}}^2 \delta_{\boldsymbol{x}, \boldsymbol{x}'}$.

Leveraging sparsity

In many practical RME problems, estimation performance can be significantly improved by incorporating prior information. The sparsity prior has played a critical role in compressive sensing, in which framework RME problems can often be formulated. Moreover, depending on the choice of the basis functions, the sparsity prior can be physically interpreted in terms of the spatial, temporal, and spectral scarceness of the RF energy distribution [2], [9].

Consider once more the linear parametric RME model (1) but, rather than assuming that the number S and locations $\{\boldsymbol{x}_s^{\mathrm{TX}}\}$ of the transmitters are known, simply discretize the map area using N_g grid points $\{\boldsymbol{x}_{n_g}^{\mathrm{grid}}\}_{n_g} \subset \mathcal{X}$ representing the possible locations of the transmitters. Then, upon defining $\tilde{\boldsymbol{\alpha}} := [\tilde{\alpha}_1, \ldots, \tilde{\alpha}_{N_g}]^\top$ and $\tilde{\boldsymbol{\Psi}} \in \mathbb{R}^{N \times N_g}$ with $(\tilde{\boldsymbol{\Psi}})_{n,n_g} = \psi_{n_g}(\boldsymbol{x}_n) := 1/\|\boldsymbol{x}_n - \boldsymbol{x}_{n_g}^{\mathrm{grid}}\|^2$ for $n = 1, \ldots, N$ and $n_g = 1, \ldots, N_g$, one has the model $\boldsymbol{m} = \tilde{\boldsymbol{\Psi}} \tilde{\boldsymbol{\alpha}} + \boldsymbol{z}$. In practical scenarios, it is expected that only a small subset of the grid points are actually occupied by transmitters; that is, $S \ll N_g$. Thus, one can impose the sparsity prior on $\tilde{\boldsymbol{\alpha}}$. For example, a least absolute shrinkage and selection operator problem can be formulated as $\hat{\tilde{\boldsymbol{\alpha}}} := \operatorname{argmin}_{\tilde{\boldsymbol{\alpha}}} \|\boldsymbol{m} - \tilde{\boldsymbol{\Psi}} \tilde{\boldsymbol{\alpha}}\|_2^2 + \lambda \|\tilde{\boldsymbol{\alpha}}\|_1$, where $\lambda > 0$, and the term $\|\tilde{\boldsymbol{\alpha}}\|_1 := \sum_{n_g=1}^{N_g} |\tilde{\alpha}_{n_g}|$ is known to promote sparsity in $\tilde{\boldsymbol{\alpha}}$. The non-zero entries of the obtained $\hat{\tilde{\boldsymbol{\alpha}}}$ reveal the (grid-based) locations $\{\boldsymbol{x}_s^{\mathrm{TX}}\}$ and the number S of the transmitters. Then, one can reconstruct the desired power map $p(\boldsymbol{x})$ using (1).

As in the linear parametric RME approach, the adopted basis functions $\psi_{n_g}(\boldsymbol{x}) = 1/\|\boldsymbol{x} - \boldsymbol{x}_{n_g}^{\mathrm{grid}}\|^2$ may not accurately capture the actual propagation characteristics. Possible remedies for this issue include sparse total LS [8], kernel-based learning [9], and sparse Bayesian learning techniques [24]. In particular, the basis mismatch issue due to the grid-based discretization of space can be mitigated in the atomic norm minimization framework.

Matrix completion

Another useful framework for RME is low-rank matrix completion. For instance, consider building a power map over a rectangular area $\mathcal{X} \subset \mathbb{R}^2$. By discretizing $\mathcal{X}$ using a regular grid $\{\boldsymbol{x}_{(i,j)}^{\mathrm{grid}} : i = 1, \ldots, I, j = 1, \ldots, J\}$, one can obtain a power map matrix $\boldsymbol{P} \in \mathbb{R}^{I \times J}$ where $(\boldsymbol{P})_{i,j} := p(\boldsymbol{x}_{(i,j)}^{\mathrm{grid}})$. Of course, only a small subset of the entries will be actually observed by the sensors. However, when the grid is dense enough compared to the spatial variability of the map, adjacent entries of $\boldsymbol{P}$ will be similar, which will, in turn, manifest itself as an approximate rank deficiency of $\boldsymbol{P}$; that is, $\mathrm{rank}(\boldsymbol{P}) \ll \min\{I, J\}$. Matrix completion, thus, tries to estimate the unobserved entries of $\boldsymbol{P}$ under a low-rank prior. Since directly promoting low rank gives rise to nonconvex problems, tractable formulations are typically pursued by penalizing the nuclear norm of the

estimate, which is the sum of its singular values. Denote the set of indices of the observed entries as $O \subset \{1, \ldots, I\} \times \{1, \ldots, J\}$ and the nuclear norm of P as $\|P\|_*$. Also, let M be the matrix whose (i, j)th element equals the sensor measurement at $x_{(i,j)}^{\text{grid}}$ if $(i, j) \in O$ and zero otherwise. A matrix completion problem for the power map can be posed as

$$\underset{P}{\text{minimize}} \frac{1}{2} \sum_{(i,j) \in O} [(P)_{(i,j)} - (M)_{(i,j)}]^2 + \lambda \|P\|_*. \quad (11)$$

With a sufficient number of observed entries, which depends on the rank and the incoherence of P, the desired map can be reconstructed reliably.

When $\mathcal{X}$ grows large, the rank of P may increase, as the power distribution may become more diverse. In this case, local matrix completion on submatrices of P may be a viable approach [25]. The matrix completion idea can also be extended to tensors when the maps in a 3D space are desired [26] or when the time and frequency domains are considered together with space [27].

Dictionary learning

When it is desired to capture the temporal variations of the power map, e.g., to exploit unused spectral resources over both time and space, it is useful to learn a library of power maps from which the suitable one can be chosen to explain the power distribution at a given time. Dictionary learning is an unsupervised learning method that seeks a possibly overcomplete basis, termed a *dictionary*, such that the data vectors can be expressed as linear combinations of a small number of vectors in the dictionary.

Denote the power measurements of the N sensors at time t as $m(t) \coloneqq [m_1(t), \ldots, m_N(t)]^\top$ for $t = 1, \ldots, T$. Dictionary learning postulates that $m(t)$ can be represented using a dictionary $D \in \mathbb{R}^{N \times Q}$ as $m(t) \approx Ds(t)$, where $s(t) \in \mathbb{R}^Q$ is a sparse vector of coefficients for the measurements at time t. The columns of D are called the *atoms*. Collecting the data samples into a matrix $M \coloneqq [m(1), \ldots, m(T)] \in \mathbb{R}_+^{N \times T}$, one can appreciate that finding such a dictionary can be viewed as a matrix factorization task since $M \approx DS$, where $S \coloneqq [s(1), \ldots, s(T)]$ is a sparse matrix. There are various optimization formulations to learn D from M [28].

In the present context of power map estimation, consider the case where the sensors do not report their measurements every time due to, e.g., energy-saving sleep modes or congested signaling channels. Thus, the network controller must apply an appropriate interpolation technique to estimate the missing observations. A helpful piece of side information is the topology of the network of sensors, which is typically maintained for various network control tasks, such as routing. To leverage this topology information, let $A \in \{1, 0\}^{N \times N}$ be the adjacency matrix of the network topology; i.e., the (n, n')th entry $a_{n,n'}$ of A is equal to one if nodes n and n' can communicate directly with each other and zero otherwise. The Laplacian matrix L is defined as $L \coloneqq \text{diag}\{A\mathbf{1}\} - A$, where $\mathbf{1}$ is the all-one vector. As seen next, this matrix can be used to promote spatial

smoothness in the sense that the power estimates at adjacent sensors are similar.

For training, at each time t, a subset $\mathcal{N}^{\text{obs}}(t) \subset \mathcal{N} \coloneqq \{1, \ldots, N\}$ of sensors acquires power measurements, which are stacked in vector $m^{\text{obs}}(t) \in \mathbb{R}_+^{|\mathcal{N}^{\text{obs}}(t)|}$. Also, let $O(t)$ denote the matrix that contains the nth row of the $N \times N$ identity matrix if and only if $n \in \mathcal{N}^{\text{obs}}(t)$. Then, upon defining

$$f(s, D; m^{\text{obs}}(t), O(t)) \coloneqq \frac{1}{2} \|m^{\text{obs}}(t) - O(t)Ds\|_2^2$$
$$+ \lambda_s \|s\|_1 + \frac{\lambda_L}{2} s^\top D^\top LDs \quad (12)$$

the dictionary can be learned via

$$\hat{D} \coloneqq \underset{D \in \mathcal{D}\{s(t)\}}{\text{argmin}} \sum_{t=1}^{T} f(s(t), D; m^{\text{obs}}(t), O(t)) \quad (13)$$

where $\mathcal{D} \coloneqq \{[d_1, \ldots, d_Q] \in \mathbb{R}^{N \times Q} : \|d_q\|_2^2 \leq 1, q = 1, \ldots, Q\}$. The first term in (12) promotes the fitness of the reconstruction to the training datum in an LS sense; the second term, with an adjustable weight $\lambda_s > 0$, is an ℓ_1-norm-based regularizer encouraging sparsity in s; and the third term, with weight $\lambda_L \geq 0$, captures the prior information that the power levels at the neighboring sensor nodes should be similar since it holds that $v^\top L v = (1/2) \sum_{n=1}^{N} \sum_{n'=1}^{N} a_{nn'}(v_n - v_{n'})^2$ for any $v \coloneqq [v_1, \ldots, v_N]^\top \in \mathbb{R}^N$. Problem (13) can be solved efficiently via a block coordinate descent algorithm [29].

In the operational phase, once the dictionary $\hat{D}$ is obtained from (13), given a (new) set of measurements $\bar{m}^{\text{obs}}$ and the corresponding observation matrix $\bar{O}$ (corresponding to the observation set $\bar{\mathcal{N}}^{\text{obs}}$), one first finds the sparse coefficients by solving $\bar{s} \coloneqq \text{argmin}_s f(s, \hat{D}; \bar{m}^{\text{obs}}, \bar{O})$. Then, the missing power levels for sensors $n \in \bar{\mathcal{N}}^{\text{miss}} \coloneqq \mathcal{N} \backslash \bar{\mathcal{N}}^{\text{obs}}$ can be obtained by first reconstructing the whole $\hat{\bar{m}} = \hat{D}\bar{s}$ and extracting the entries $\{\hat{\bar{m}}_n\}_{n \in \bar{\mathcal{N}}^{\text{miss}}}$. A practical challenge is to implement the algorithm for online and distributed operation to handle large-scale real-time computation [29], [30]. Additionally, tuning the hyperparameters, such as the dictionary size, and the regularization parameters may require cross validation based on historic measurements.

Deep learning

A deep neural network (DNN) is a function g_w that can be expressed as the composition of more elementary functions called *layers*, which are parameterized by a vector w. Training a DNN involves finding w so that g_w fits the given dataset. DNNs feature a large learning capacity and can be efficiently trained via stochastic optimization methods. Spatial structures in the data can be readily exploited utilizing convolutional layers, in which case the DNN is called a *convolutional neural network* (*CNN*). Next, multiple approaches for using DNNs for signal strength map estimation are described.

Pointwise DNN estimators

The simplest approach is to use a DNN to construct a function where the input is the sensor location, and the output is the

signal strength at that location. This approach was pursued in [10], where the input was encoded using a spherical coordinate system located at the (single) transmitter. Since the dimensionality of the input is small, the network architecture can be kept simple, and the resulting estimator is not affected by the so-called curse of dimensionality [19, Sec. 4.3].

However, such an approach cannot easily capture the spatial structure of the map using CNNs. Furthermore, the DNN needs to be retrained for each specific RF environment. Therefore, it cannot benefit from measurements previously collected in other scenarios, such as different cities.

Local DNN estimators

To alleviate the aforementioned limitations of pointwise DNN estimators, the network input can be replaced with a collection of matrices that capture information about the local environment of the sensor. These matrices, typically stacked as slabs of a tensor, can be thought of as local maps defined over a rectangular grid centered at the sensor. A transmitter (alternatively, a sensor) distance map, for example, is a matrix whose (i, j)th entry equals the distance from the (i, j)th grid point to the transmitter (sensor) [11], [12]. It is also possible to include a local terrain map that indicates the altitude of the terrain at each grid point. Further kinds of local maps include building indicator maps [18], building height maps [31], [32], or foliage maps [31]. One can also use aerial or satellite images of the surroundings of the sensor as local maps [33]. Figure 6 provides an illustration of this kind of setup.

This input format lends itself to CNN architectures that leverage spatial information in the vicinity of the sensor to predict the received power. To learn across different environments where the transmitters possibly employ different transmit power, one can set the output of the network to be the gain between each transmitter and the sensor and work out the received power afterward. This effectively sets this approach halfway between signal strength and propagation map estimation.

The practical limitation of this approach is that it requires knowledge of the locations (and the transmit powers if one wishes to estimate the gains) of all transmitters, and the measurements must be obtained separately for each transmitter. Furthermore, it only exploits the information in the vicinity of the sensor, but, in practice, obstacles or scatterers far away from the sensor may also affect the received power significantly. In addition, networks designed in this way provide the received power (or channel gain) only at a single location per evaluation (also known as forward pass). To construct the entire map, the estimator needs to be evaluated repeatedly for each point on a grid, resulting in significant computational complexity.

Global DNN estimators

To accommodate global, rather than local, environment information, one can create a regular grid across the region where the map needs to be constructed and formulate the RME problem as a matrix or tensor completion task [13], [18], [31], [32], [34]. To this end, each measurement is associated with the nearest grid point and a matrix is constructed with an entry per grid point. If a single measurement is assigned to a grid point, the corresponding entry contains the measurement. If multiple measurements are assigned to a grid point, the corresponding entry may contain their average. Those points with no associated measurements can simply be filled with physically unlikely values [32], [34], [35], or a separate binary mask matrix can be included in the input [18], [36].

Other maps with side information, such as the ones used in local DNN estimators, can also be appended to the input tensor to the network. However, note that, now, these maps must be global in the sense that they capture the entire region of interest.
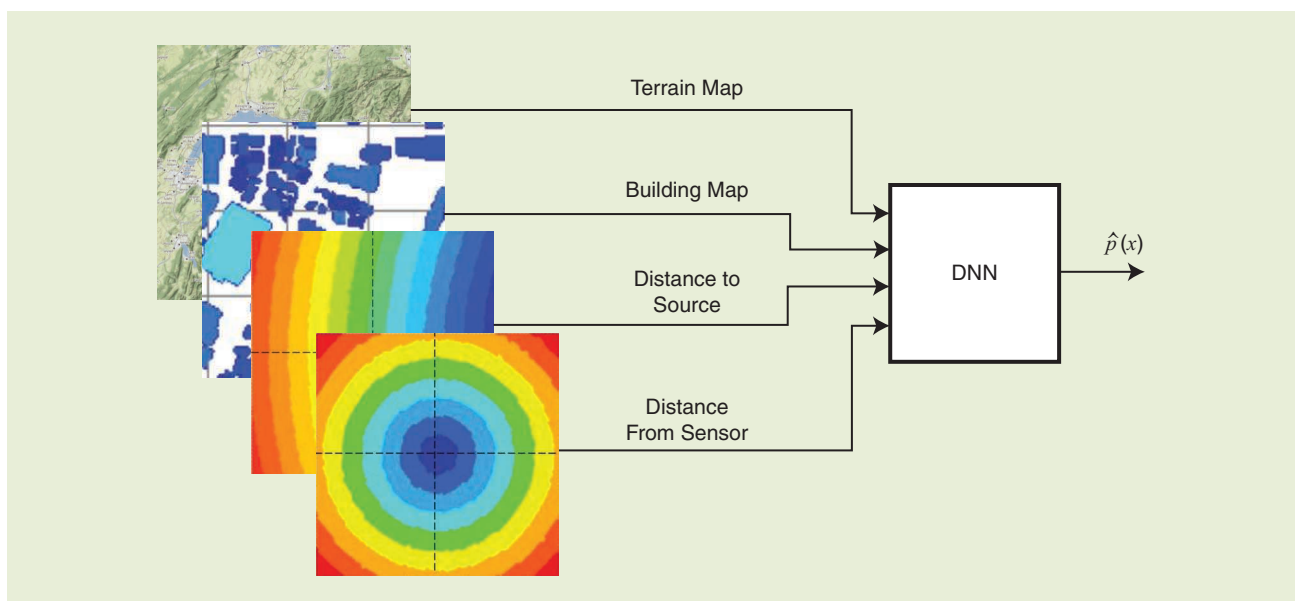

FIGURE 6. A local DNN estimator, which provides $\hat{p}(x)$ at a single x.

The global input can naturally be processed by CNN architectures. The most common ones are autoencoders [18], [35] and UNets [13], [32]. The motivation for the former is described in "Manifold Structure of Power Maps." A global DNN estimator is illustrated in Figure 7.

Unlike local DNN estimators, a single forward pass of the DNN produces the entire map. Furthermore, using map measurements collected in multiple environments, the architecture can readily learn across different RF scenarios. On the other hand, collecting a sufficiently large dataset to train such a net-work may be challenging. To alleviate this difficulty, one may resort to data augmentation or incorporate synthetic data from ray-tracing simulators [18]. Another limitation deals with the spatial resolution of the constructed maps. A high-resolution map requires a dense grid, significantly increasing the computational complexity.

Estimation of PSD maps

PSD maps describe how the power is distributed not only across space but also across the frequency domain. To estimate

---

Manifold Structure of Power Maps

Autoencoder networks are attuned to situations where the data lies on a low-dimensional manifold embedded in a high-dimensional space. To see that this is the case of radio maps, consider the values of a power map in 2D produced by two sources radiating with a fixed height and power in free space. A dataset can be generated where each map is obtained by placing the sources at random locations on the horizontal plane. Each map is, therefore, uniquely identified by the four scalars corresponding to the locations of the sources. If the maps are defined on a 32×32 grid, they comprise $32^2 = 1,024$ points, which means that these maps lie on a manifold of dimension four embedded in a space of dimension 1,024.

This observation is corroborated in [18] by training an autoencoder on the aforementioned dataset. An autoencoder is the concatenation of an encoder and a decoder. In this case, the encoder takes a 32×32 map and produces a code vector λ of length four. The decoder takes this vector at its input and aims at reconstructing the original 32×32 map. For a properly trained encoder and decoder, the output of the decoder closely resembles the input of the encoder, which means that the code effectively condenses the information of the map in just four numbers.

Each value of the code identifies a point in the manifold. The top panel of Figure S1 shows the output of the decoder when its input equals the average of the codes associated with each map in the dataset. The rest of the panels show the output of the decoder applied to the result of perturbing the entries of this average code, indicated by index set $\mathcal{S}$, by an amount equal to the standard deviation of that entry across the dataset. This procedure yields different points in the manifold. All panels approximately correspond to maps of the kind composing the dataset, which supports the manifold hypothesis.

If propagation does not take place in free space, or if the power or height of the sources is variable, a longer code needs to be utilized to capture all information in the maps. Experiments with other datasets reveal that, in the presence of propagation phenomena such as shadowing and fading, radio maps lie close to a manifold of low dimension [18].
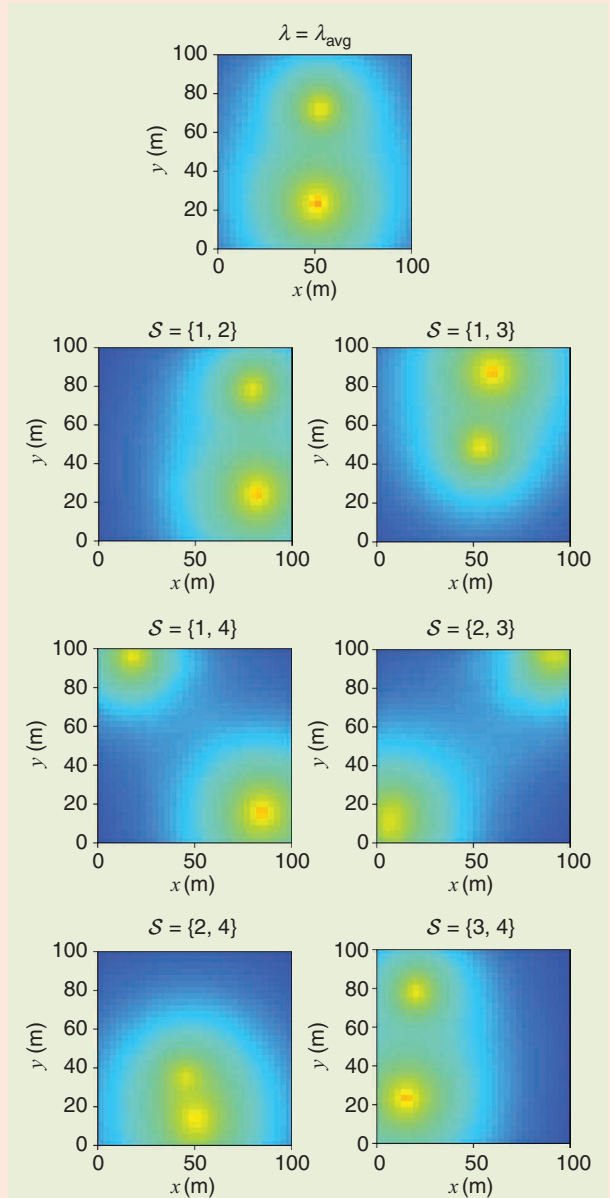

FIGURE S1. The decoder outputs for the average code and its perturbed versions obtained with an autoencoder with a code length of four [18].

---

a PSD map $p(\boldsymbol{x}, f)$, most schemes assume that the sensors measure the power that they receive at a set of frequencies $f_1, \ldots, f_{N_f}$. The nth measurement is, therefore, a vector $\boldsymbol{m}_n = [\tilde{p}(\boldsymbol{x}_n, f_1), \ldots, \tilde{p}(\boldsymbol{x}_n, f_{N_f})]^\top$, where $\tilde{p}(\boldsymbol{x}, f)$ denotes the measured PSD at location $\boldsymbol{x}$ and frequency f, possibly obtained by using a periodogram or Welch's method. Relying on these measurements, the goal is to obtain a PSD map estimate $\hat{p}$ such that $\hat{p}(\boldsymbol{x}, f)$ is as close to the true $p(\boldsymbol{x}, f)$ as possible. To this end, several alternatives are explored next.

Separate estimation per frequency

The simplest approach is to consider each frequency separately and essentially decompose the problem of estimating a PSD map at N_f frequencies as N_f problems of estimating a single power map [35]. More specifically, the nth power map is estimated from PSD measurements of $p(\boldsymbol{x}_1, f_n), \ldots, p(\boldsymbol{x}_N, f_n)$ using the techniques described earlier. The main limitation of this approach is that it disregards any structure in the frequency domain, making it more sensitive to measurement noise than other schemes explored later. On the upside, these approaches are simple and do not require prior knowledge on the channel or transmit PSD characteristics. Moreover, a twofold benefit arises in terms of the sizes of the training set and the number of parameters in schemes such as deep learning estimators. First, provided that the propagation environment affects all frequencies in a similar fashion, considering each frequency separately will increase the number of training examples by a factor of N_f. On the other hand, if the neural network takes per-frequency measurements as the input rather than processing all frequencies jointly, the number of parameters to be learned can be significantly reduced [18, Sec. III-C1].

> **PSD maps describe how the power is distributed not only across space but also across the frequency domain.**

Estimation in narrowband channels

When the width of the band of interest is small or moderate, it makes sense to assume that the channel is not frequency selective [27], [36]. This means that the true PSD map can be written as $p(\boldsymbol{x}, f) = \Sigma_s h_s(\boldsymbol{x}) p_s^{\mathrm{TX}}(f)$, where $h_s(\boldsymbol{x}) = h(\boldsymbol{x}_s^{\mathrm{TX}}, \boldsymbol{x})$ is the channel gain from the sth transmitter to location $\boldsymbol{x}$ and $p_s^{\mathrm{TX}}(f)$ is the transmit PSD of the sth source. This implies that the measurements essentially provide N_f noisy linear combinations of the S functions $h_1(\boldsymbol{x}), \ldots, h_S(\boldsymbol{x})$. Therefore, when $N_f \gg S$, one can effectively exploit the structure in the frequency domain, improving robustness to measurement noise. One of the main benefits of this approach is that no knowledge of the transmit PSD is required, as it can often be estimated using tools, such as nonnegative matrix factorization, without requiring any prior knowledge [36].

Estimation in wideband channels

For a wideband channel, one cannot realistically assume that the channel response is flat. To exploit the frequency domain structure, one can utilize prior knowledge on the transmitter waveforms. Specifically, the PSDs of the transmitted waveforms are typically constrained by communication standards and spectrum regulations, which specify the bandwidth, carrier frequencies, transmission masks, roll-off factors, number of subcarriers, and so forth [37]. Therefore, the transmit PSD of a source can be approximated by a basis expansion model (BEM) as $p_s^{\mathrm{TX}}(f) = \Sigma_c \beta_{s,c} \phi_c(f)$, where ϕ_c denotes the cth basis function, and $\beta_{s,c}$ is a nonnegative quantity. This decomposition is illustrated in Figure 8.

If the signals transmitted by different sources are uncorrelated, the received PSD at a location $\boldsymbol{x}$ can be expressed as $p(\boldsymbol{x}, f) = \Sigma_s h(\boldsymbol{x}_s^{\mathrm{TX}}, \boldsymbol{x}, f) p_s^{\mathrm{TX}}(f)$, where $h(\boldsymbol{x}_s^{\mathrm{TX}}, \boldsymbol{x}, f)$ is
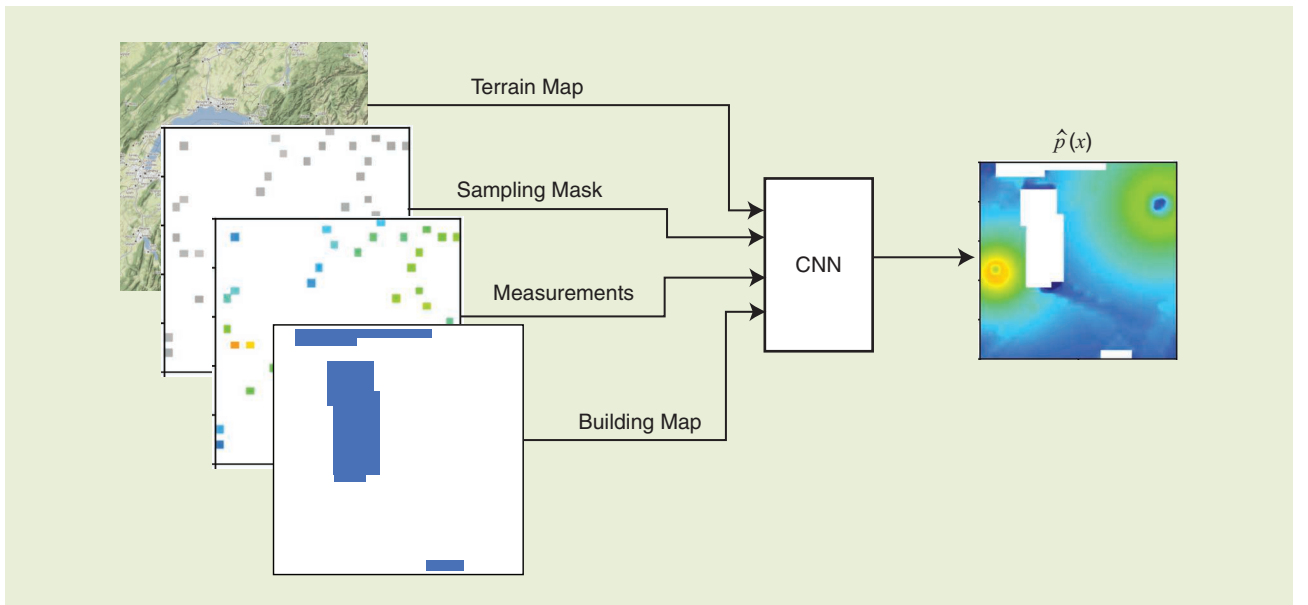

FIGURE 7. A global DNN estimator, which provides $\hat{p}(\boldsymbol{x})$ for all values of on a grid.

the channel gain at frequency f. Then, using the BEM, one arrives at $p(\boldsymbol{x}, f) = \Sigma_c \Sigma_s \beta_{s,c} h(\boldsymbol{x}_s^{\text{TX}}, \boldsymbol{x}, f) \phi_c(f)$. If the bandwidths of the basis functions are small relative to the entire band, it is reasonable to assume that h is approximately frequency flat in the band of each basis function. This yields $h(\boldsymbol{x}_s^{\text{TX}}, \boldsymbol{x}, f) \phi_c(f) \approx h(\boldsymbol{x}_s^{\text{TX}}, \boldsymbol{x}, \tilde{f}_c) \phi_c(f)$, where $\tilde{f}_c$ is the central frequency of the cth basis function. With this approximation, one can write $p(\boldsymbol{x}, f) = \Sigma_c p_c(\boldsymbol{x}) \phi_c(f)$, where $p_c(\boldsymbol{x}) = \Sigma_s h(\boldsymbol{x}_s^{\text{TX}}, \boldsymbol{x}, \tilde{f}_c) \beta_{s,c}$ constitutes the power captured by the cth basis function at location $\boldsymbol{x}$.

Observe that introducing the BEM has reduced the problem of estimating N_f power maps to the problem of estimating the $C \ll N_f$ power maps $p_1, \ldots, p_C$. Clearly, the smaller C, the smaller the sensitivity to measurement noise. The approaches in the preceding sections can be seen as the extreme cases of choosing $C = N_f$ and $C = 1$, respectively. To estimate a PSD map, the aforementioned technique can be used in combination with virtually any of the approaches for power map estimation discussed earlier [2], [4], [8], [9]. A recent example is [18], where a BEM is used in the last layer of a DNN for RME.

The main limitation of estimators that rely on a BEM is a manifestation of the well-known bias–variance tradeoff. In particular, if the number of basis functions is small, the approximation $h(\boldsymbol{x}_s^{\text{TX}}, \boldsymbol{x}, f) \phi_c(f) \approx h(\boldsymbol{x}_s^{\text{TX}}, \boldsymbol{x}, \tilde{f}_c) \phi_c(f)$ may not hold, which will generally result in estimation bias. On the other hand, if the number of basis functions is large, the representation capacity of the BEM is large, which results in a small bias, but a larger variance must be expected as the result of the increase in the number of scalar maps to be learned for a fixed number of samples.

Estimation of propagation maps

Propagation maps quantify channel effects, such as channel gains, for links between arbitrary pairs of locations where no sensors may have been deployed. The nth measurement is collected by a pair of sensors, one at location $\boldsymbol{x}_n$ and the other at $\boldsymbol{x}_n'$. The channel gain of the link between them can possibly be measured by employing pilot signals. The resulting measurement can be expressed as $m_n = h(\boldsymbol{x}_n, \boldsymbol{x}_n') + z_n$, where h is the true map, and z_n represents the measurement noise. The RME problem is to obtain an estimate $\hat{h}$ of h given $\{(\boldsymbol{x}_n, \boldsymbol{x}_n', m_n)\}_{n=1}^N$. A good RME algorithm should have good generalization properties, meaning that $\hat{h}(\boldsymbol{x}, \boldsymbol{x}') \approx h(\boldsymbol{x}, \boldsymbol{x}')$ for all location pairs $(\boldsymbol{x}, \boldsymbol{x}')$, even those for which no measurements have been collected.

Like signal strength RME, propagation RME is a function estimation problem. Therefore, the techniques described in section "Estimation of Signal Strength Maps" can again be employed in principle. The key difference is that now the function to be estimated depends on two locations rather than one. If $\boldsymbol{x}$ denotes a location in 3D space, it is clear that h is a

Propagation maps quantify channel effects, such as channel gains, for links between arbitrary pairs of locations where no sensors may have been deployed.

function of a 6D input, namely, the entries of $\boldsymbol{x}$ and $\boldsymbol{x}'$. This means that the number of measurements necessary to attain a given accuracy may be considerably greater than for estimating a signal strength map—a manifestation of the curse of dimensionality. Thus, as explored next, a number of algorithms have been tailor made for propagation RME to alleviate such a difficulty.

Nontomographic approaches

In the nontomographic approaches, channel gains are directly modeled based on basic wireless propagation models without introducing any underlying auxiliary map. To maintain tractability, however, the RME problem is often simplified by fixing one end of a link. For example, one may consider estimating the maps $\{h_n(\boldsymbol{x}) := h(\boldsymbol{x}, \boldsymbol{x}_n)\}_n$ for fixed positions $\{\boldsymbol{x}_n\}_{n=1}^N$ where the sensors are located. The individual functions $\{h_n(\boldsymbol{x})\}_n$ can be estimated using methods employed for signal strength maps. Since (static) signal strength map estimation techniques have been explained in the preceding sections, here, we extend the RME problem to include the time domain to capture the temporal variation of channel gains. Needless to say, static channel gain maps can also be constructed in a nontomographic fashion.

Consider the channel gain $h_n(\boldsymbol{x}, t)$ between locations $\boldsymbol{x}$ and $\boldsymbol{x}_n$ at time t [5]. Suppose that the effect of small-scale fading has been averaged out, allowing $h_n(\boldsymbol{x}, t)$ to be expressed in decibels as $h_{n,\text{dB}}(\boldsymbol{x}, t) = h_n^{\text{PL}}(\boldsymbol{x}) - a_n^{\text{SF}}(\boldsymbol{x}, t)$, where $h_n^{\text{PL}}(\boldsymbol{x})$ is the known path loss from $\boldsymbol{x}_n$ to $\boldsymbol{x}$, and $a_n^{\text{SF}}(\boldsymbol{x}, t)$ is the shadow fading between $\boldsymbol{x}$ and $\boldsymbol{x}_n$ at time t. Note that $h_n^{\text{PL}}(\boldsymbol{x})$ can be assumed known whenever both $\boldsymbol{x}_n$ and $\boldsymbol{x}$ as well as the antenna gains are known. Thus, the problem becomes tracking the time-varying shadow fading map $a_n^{\text{SF}}(\boldsymbol{x}, t)$.
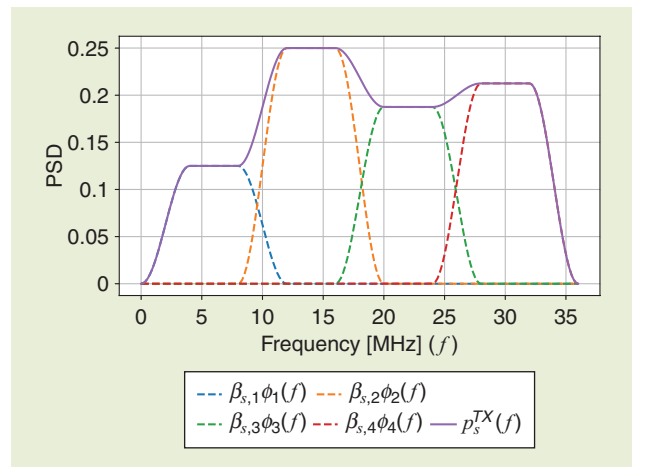

FIGURE 8. A basis expansion model can be used to decompose a PSD as a linear combination of functions in a basis. In this case, the basis functions are raised cosine functions, each one corresponding to a transmission in a different band. This makes it possible to exploit prior information about bandwidths, central frequencies, and transmission pulse shapes.

To do this, shadow fading measurements are needed, which can be obtained by subtracting the transmit power and the path loss from the received power measurement. By letting $\mathcal{N}_{-n} := \{1, \ldots, n-1, n+1, \ldots, N\}$, the noisy measurements $\{\check{a}_n^{\mathrm{SF}}(\boldsymbol{x}_j, t)\}_j$ of shadow fading obtained at time t by the sensor at $\boldsymbol{x}_n$ using the pilot signals sent from the radios at $\{\boldsymbol{x}_j\}_{j \in \mathcal{N}_{-n}}$ can be expressed as

$$\check{a}_n^{\mathrm{SF}}(\boldsymbol{x}_j, t) = a_n^{\mathrm{SF}}(\boldsymbol{x}_j, t) + z_n(\boldsymbol{x}_j, t), \quad j \in \mathcal{N}_{-n} \qquad (14)$$

where $z_n(\boldsymbol{x}_j, t)$ is the zero-mean Gaussian measurement noise. Upon defining $\check{a}_n^{\mathrm{SF}}(t) := [\check{a}_n^{\mathrm{SF}}(\boldsymbol{x}_1, t), \ldots, \check{a}_n^{\mathrm{SF}}(\boldsymbol{x}_{n-1}, t), \check{a}_n^{\mathrm{SF}}(\boldsymbol{x}_{n+1}, t), \ldots, \check{a}_n^{\mathrm{SF}}(\boldsymbol{x}_N, t)]^\top$, the problem is to estimate $h_{n,\mathrm{dB}}(\boldsymbol{x}, t)$ for arbitrary $\boldsymbol{x}$ based on the measurements $\check{\mathcal{A}}_n^{\mathrm{SF}}(t) := \{\check{a}_n^{\mathrm{SF}}(\tau)\}_{\tau=1}^t$ up to time t.

This problem can be tackled in the framework of kriged Kalman filtering, also known as *space-time Kalman filtering* [38]. Employing the log-normal shadowing model, it is assumed that $a_n^{\mathrm{SF}}(\boldsymbol{x}, t)$ is a Gaussian process with spatiotemporal dynamics [5], [38]:

$$a_n^{\mathrm{SF}}(\boldsymbol{x}, t) = \mu_n^{\mathrm{SF}}(\boldsymbol{x}, t) + \nu_n(\boldsymbol{x}, t) \qquad (15)$$

$$\mu_n^{\mathrm{SF}}(\boldsymbol{x}, t) = \int w_n(\boldsymbol{x}, \boldsymbol{u}) \mu_n^{\mathrm{SF}}(\boldsymbol{u}, t-1) d\boldsymbol{u} + \eta_n(\boldsymbol{x}, t) \qquad (16)$$

where $\mu_n^{\mathrm{SF}}(\boldsymbol{x}, t)$ is the spatiotemporally correlated component, $w_n(\boldsymbol{x}, \boldsymbol{u})$ captures the interaction of this component at location $\boldsymbol{x}$ at time t and at location $\boldsymbol{u}$ at time $(t-1)$, and $\nu_n(\boldsymbol{x}, t)$ and $\eta_n(\boldsymbol{x}, t)$ are spatially correlated but temporally white zero-mean Gaussian processes. Process $\nu_n(\boldsymbol{x}, t)$ is uncorrelated with $z_n(\boldsymbol{u}, \tau)$, and $\eta_n(\boldsymbol{x}, t)$ is uncorrelated with $\nu_n(\boldsymbol{u}, \tau)$ and $z_n(\boldsymbol{u}, \tau)$ for all $\boldsymbol{u}$ and τ. Moreover, $\mathbb{E}\{\nu_n(\boldsymbol{x}, t) \mu_n^{\mathrm{SF}}(\boldsymbol{u}, t)\} = \mathbb{E}\{\eta_n(\boldsymbol{x}, t) \mu_n^{\mathrm{SF}}(\boldsymbol{u}, t-1)\} = 0$ for all $\boldsymbol{x}, \boldsymbol{u}$, and t.

Since the state–space model in (15) and (16) is infinite dimensional, adopt a BEM for tractability, as in universal kriging. For a set of K orthonormal basis functions $\{\psi_k(\boldsymbol{x})\}_k$, μ_n^{SF} and w_n are, respectively, approximated as $\mu_n^{\mathrm{SF}}(\boldsymbol{x}, t) \approx \sum_{k=1}^K \alpha_{n,k}(t) \psi_k(\boldsymbol{x})$ and $w_n(\boldsymbol{x}, t) \approx \sum_{k=1}^K \beta_{n,k}(t) \psi_k(\boldsymbol{x})$ with expansion coefficients $\{\alpha_{n,k}(t)\}$ and $\{\beta_{n,k}(t)\}$. Substituting these expansions into (14)–(16) and evaluating the resulting equations at $\{\boldsymbol{x}_j\}_{j \in \mathcal{N}_{-n}}$ yields the finite dimensional state–space model:
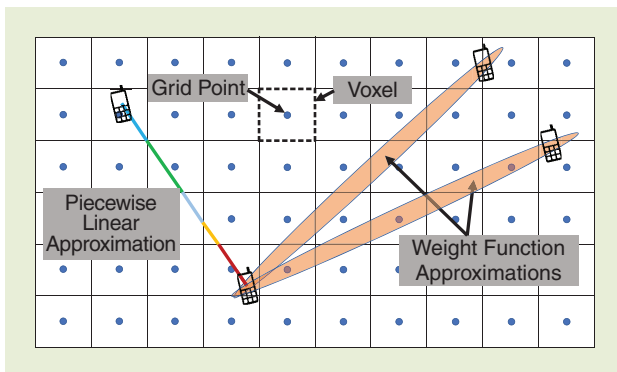

FIGURE 9. The possible approximations of the tomographic integral.

$$\check{a}_n^{\mathrm{SF}}(t) = \boldsymbol{\Psi}_n \boldsymbol{\alpha}_n(t) + \boldsymbol{\nu}_n(t) + \boldsymbol{z}_n(t) \qquad (17)$$

$$\boldsymbol{\alpha}_n(t) = \boldsymbol{\Psi}_n^\dagger \boldsymbol{B}_n \boldsymbol{\alpha}_n(t-1) + \boldsymbol{\Psi}_n^\dagger \boldsymbol{\eta}_n(t). \qquad (18)$$

Here, $\boldsymbol{\psi}(\boldsymbol{x}) := [\psi_1(\boldsymbol{x}), \ldots, \psi_K(\boldsymbol{x})]^\top$, $\boldsymbol{\alpha}_n(t) := [\alpha_{n,1}(t), \ldots, \alpha_{n,K}(t)]^\top$, and $\boldsymbol{\beta}_n(\boldsymbol{x})$ is defined likewise. Vectors $\boldsymbol{\nu}_n(t)$, $\boldsymbol{z}_n(t)$, and $\boldsymbol{\eta}_n(t)$ are constructed in a similar fashion from $\{\nu_n(\boldsymbol{x}_j, t)\}_{j \in \mathcal{N}_{-n}}$, $\{z_n(\boldsymbol{x}_j, t)\}_{j \in \mathcal{N}_{-n}}$, and $\{\eta_n(\boldsymbol{x}_j, t)\}_{j \in \mathcal{N}_{-n}}$, respectively. $\boldsymbol{B}_n$ and $\boldsymbol{\Psi}_n$ are matrices constructed by, respectively, arranging $\boldsymbol{\beta}_n(\boldsymbol{x}_j)^\top$ and $\boldsymbol{\psi}(\boldsymbol{x}_j)^\top$ as rows for $j \in \mathcal{N}_{-n}$.

Based on (17) and (18), the LMMSE (or simply MMSE in this case due to the joint Gaussianity assumption) estimate $\hat{\boldsymbol{\alpha}}_n(t \mid t)$ of $\boldsymbol{\alpha}_n(t)$ given $\check{\mathcal{A}}_n^{\mathrm{SF}}(t)$ can be obtained via ordinary Kalman filtering, from which the temporally dynamic component $\mu_n^{\mathrm{SF}}(\boldsymbol{x}, t)$ can be estimated as $\mathbb{E}\{\mu_n^{\mathrm{SF}}(\boldsymbol{x}, t) \mid \check{\mathcal{A}}_n^{\mathrm{SF}}(t)\} = \boldsymbol{\psi}(\boldsymbol{x})^\top \hat{\boldsymbol{\alpha}}_n(t \mid t)$. To capture $\nu_n(\boldsymbol{x}, t)$ as well, a kriging estimator is employed (compare with the "Kriging" section). Overall, the LMMSE estimate $\hat{a}_n^{\mathrm{SF}}(\boldsymbol{x}, t) := \mathbb{E}\{a_n^{\mathrm{SF}}(\boldsymbol{x}, t) \mid \check{\mathcal{A}}_n^{\mathrm{SF}}(t)\}$ can be obtained exploiting the covariance structure [5]. Once $\hat{a}_n^{\mathrm{SF}}(\boldsymbol{x}, t)$ is obtained, the channel gain map estimate $\hat{h}_{n,\mathrm{dB}}(\boldsymbol{x}, t)$ can be constructed as $\hat{h}_{n,\mathrm{dB}}(\boldsymbol{x}, t) = h_n^{\mathrm{PL}}(\boldsymbol{x}) - \hat{a}_n^{\mathrm{SF}}(\boldsymbol{x}, t)$.

Tomographic approaches

The radio tomographic model can be used to estimate shadow fading maps. It postulates that the attenuation due to shadowing can be expressed in terms of an underlying auxiliary map termed the *spatial loss field* (*SLF*) [16], [39]. The SLF characterizes how much radio waves attenuate when passing through each location and, hence, is specific to each propagation environment [40]. Specifically, the radio tomographic model prescribes that the shadowing attenuation between locations $\boldsymbol{x}$ and $\boldsymbol{x}'$ is given by the line integral [40]

$$a^{\mathrm{SF}}(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{\sqrt{\|\boldsymbol{x} - \boldsymbol{x}'\|}} \int_{\boldsymbol{x}}^{\boldsymbol{x}'} F(\bar{\boldsymbol{x}}) d\bar{\boldsymbol{x}} \qquad (19)$$

where $F : \mathcal{X} \to \mathbb{R}_+$ is the SLF. This naturally captures the notion that nearby radio links generally experience similar shadowing due to the presence of common obstacles. Since this integral provides the shadowing attenuation between two arbitrary locations, one does not need to fix one end of a link, as in the nontomographic approach explained earlier. Remarkably, time-varying maps can be readily accommodated in the tomographic approach [16]. Furthermore, as the SLF can reveal the locations of obstacles, the SLF itself can be useful for various applications, such as device-free passive localization [6], surveillance monitoring for intrusion detection [41], and through-the-wall imaging for emergency or military operations [6].

To approximate the line integral in (19), F can be discretized on a regular grid of 2D or 3D spatial locations. One common approach is to approximate this integral as a weighted sum of the SLF values on the grid points that lie inside an ellipse or an ellipsoid with foci at $\boldsymbol{x}$ and $\boldsymbol{x}'$, as shown in Figure 9. The intuition is that the attenuation between the two endpoints should

be heavily affected by the obstacles around the line of sight or, more specifically, within the so-called Fresnel zone, which is an ellipse whose geometry is dictated by the wavelength. Several functions have been proposed in the literature [41] to generate such weights, mainly based on heuristics. Alternatively, the weights can be learned from the data through blind schemes [6].

While easy to implement, this approximation yields shadowing maps with discontinuities, as small changes in x and x' may lead to a change in the set of grid points that lie in the ellipse. Even more, if the ellipse misses all of the grid points, as shown by the left ellipse in Figure 9, the approximation becomes zero. Thus, to attain good accuracy, the grid must be dense enough.

This motivates an alternative approach, where the SLF is approximated as a piecewise constant function, taking a constant value within each grid cell (or voxel) [17]. The integral can then be computed as the weighted sum of the SLF values in the cells that the line of sight traverses. The weight simply corresponds to the distance traversed in each cell. This is illustrated by the colored line in Figure 9. This approximation involves less computational burden than the one based on the ellipse, is continuous in x and x', and does not vanish unless the SLF vanishes. Thus, the need for a dense grid is relaxed—which is particularly attractive in 3D [17].

In either approach, the shadowing attenuation is a linear function of the SLF values at the grid points. Thus, the SLF can be estimated via (nonnegative) LS.

However, this requires that the number of measurements be significantly larger than the number of grid points. One can mitigate this through appropriate regularizers [43] or by specifying a prior distribution in a Bayesian framework [44]. Another limitation of tomographic approaches is that only the attenuation due to absorption (shadowing) is accounted for. Other propagation effects, such as reflection, refraction, and diffraction, are completely ignored.

> **The main limitation of estimators that rely on a BEM is a manifestation of the well-known bias–variance tradeoff.**

Spectrum surveying

To collect the measurements needed to build radio maps, traditionally, technicians in a vehicle with measurement equipment would drive around the site. With the advances in mobile robotics of the last decade, it is now possible to employ an autonomous UAV with an onboard sensor to collect the desired measurements. This is clearly more efficient in terms of time and personnel cost.

An important task is to plan the path traversed by the autonomous UAV for acquiring measurements. A common approach is to define a grid and take measurements at each grid point. However, visiting each grid point can be very time-consuming and puts a strain on the limited battery capacity, especially when the grid is dense. A more efficient approach is to collect measurements at a small set of highly informative locations and apply the interpolation techniques discussed in previous sections to construct the entire map. To this end, in addition to the map estimate, RME algorithms need to provide an uncertainty map that indicates how informative a measurement would be at each location given the measurements collected so far [22]. Based on the uncertainty map, a route planning algorithm can produce a trajectory through areas of high uncertainty. This approach achieves a much higher estimation quality in a given surveying time (or requires a much shorter time for a given quality) compared to the naive grid-based approach.

Figure 10 illustrates an example of a surveying operation using a ray-tracing dataset in a region of downtown Rosslyn, VA. The three panels show the UAV trajectory seen from above. White boxes correspond to space occupied by buildings, where no measurements can be taken. Red and white crosses denote measurement locations. Figure 10(a) shows the ground truth power map in a setup with two transmitters. Figure 10(b) and (c) respectively show the estimated power map and the uncertainty map when only the measurements marked by red crosses have been collected. At that point in time, the UAV
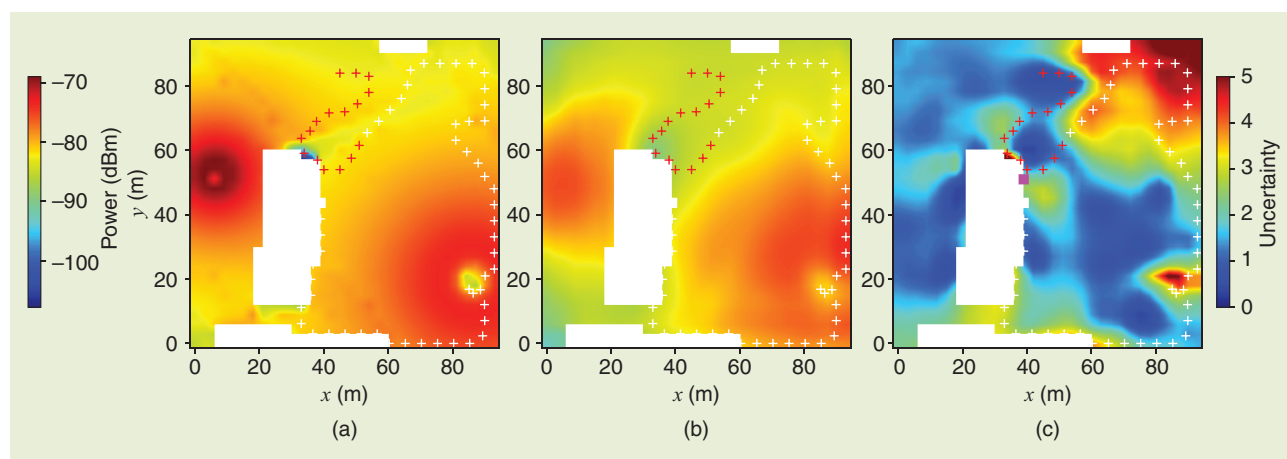

FIGURE 10. An example of a surveying operation with an autonomous UAV in an urban environment seen from above: the (a) true power, (b) estimated power, and (c) uncertainty metric. White boxes denote buildings. Red and white crosses denote measurement locations.

plans a trajectory through areas of high uncertainty, represented by white crosses. The estimator, in this case, is a global DNN estimator capable of learning the nature of propagation phenomena from a dataset (see the "Deep Learning" section).

Practical considerations

In this section, some challenges that arise when implementing RME techniques in practical setups are discussed. These include coping with localization errors, nonisotropic antenna patterns, decentralized implementation, and reducing the bandwidth required to collect measurements.

Localization errors

The RME schemes described earlier typically require accurate knowledge of the measurement locations. In practice, the sensor locations are themselves estimated based on localization systems, such as GPS, in the following way. A number of transmitters with known locations, such as satellites or cellular base stations, regularly transmit signals termed *localization pilots*. Each sensor then extracts certain features from the pilots to estimate their locations. For example, the received signal strength or the propagation delay, which contain information on the distance to the transmitters, are used to produce location estimates based on geometric principles. Thus, the quality of the estimates can be significantly degraded due to multipath propagation, as in indoor and dense urban scenarios, where localization errors may reach tens of meters.

This phenomenon is illustrated in Figure 11, where the x-coordinates of the location estimates are compared in a scenario without multipath [Figure 11(a)] and with multipath [Figure 11(b)]. The localization algorithm is based on the time difference of arrival between the pilot signals arriving from each pair of transmitters; see [45] for details. The poor quality of the location estimates in Figure 11(b) hinders the application of conventional RME techniques. This is because the maps are indexed by the locations (e.g., the input for the power map p is x) and, thus, the localization error in x propagates to the output $p(x)$.

The key realization is, therefore, that x is not suitable as the "index" of the map. To mitigate this issue, one can resort to the so-called location-free cartography framework [45].

To motivate this framework, it is worth stepping back and recalling that the location estimates are produced by a localization algorithm based on the pilot features. It is sensible, therefore, to bypass this step and directly use the pilot features to index the map since these features evolve more smoothly across space than the location estimates.

Once such a map has been estimated, there are two approaches to evaluate it at a given location. If a terminal is present at that location, it can directly employ the features of the pilot signals. If no sensor is present, one can interpolate the features, e.g., based on the low-rank prior.

Due to the larger input dimension of the map function, location-free RME requires a larger number of measurements than location-based approaches in the absence of localization errors. Another difficulty is that the availability of the features depends on the availability of the pilot signals. However, one can reconstruct the missing features [45] or define features that can be extracted from regular communication signals, e.g., the ones broadcast by cellular base stations, rather than from dedicated localization pilots [46].
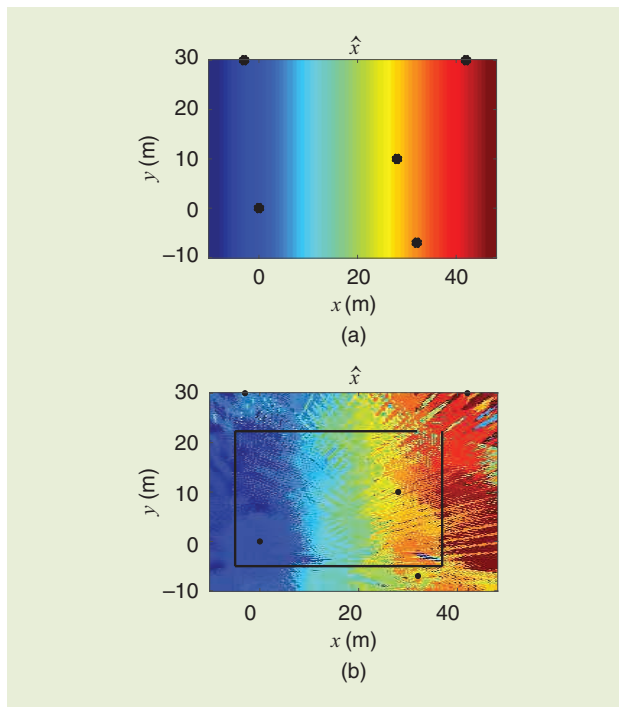

FIGURE 11. (a) Free space. (b) The scenario with four walls. The color of each point indicates the x-coordinate of the location estimate obtained by a sensor at that location. The black circles indicate the positions of the transmitters. In (a), the estimate accurately matches the true coordinate when there is no multipath; thus, this image serves as a color bar. On the other hand, in (b), the estimation error is large in the presence of multipath. (Source: [45]; used with permission.)

Antenna patterns

So far, we assumed that a signal strength map $p(x)$ provides the power received by a sensor with an isotropic antenna at location x. If the antenna pattern is not isotropic, the measured power will depend on the sensor orientation. For this reason, it may be convenient to estimate the angular spectrum map $p(x, \theta)$, which provides the angular power density received by a sensor at location x from direction θ. Here, θ parameterizes the direction through, e.g., the azimuth and elevation angles.

If $\Gamma(\theta - \theta')$ denotes the antenna gain along direction θ for a sensor with orientation θ', it follows that the power received by such a sensor when placed at x will be $\int \Gamma(\theta - \theta')p(x, \theta)d\theta$. If the sensor orientations associated with all measurements are known, then each measurement is a noisy linear observation of $p(x, \theta)$ and, therefore, the latter can be estimated. The techniques described earlier for PSD map estimation in the frequency domain can be adapted to this end, possibly upon discretizing the aforementioned integral. Specifically, $p(x, \theta)$ can be estimated for a discrete set of angle bins separately or

by parameterizing $p(x, \theta)$ by means of a BEM with standard or tailored basis functions along the lines of [4, Sec. III-A], although the choice of suitable basis functions seems to warrant further research.

The challenges emerging in this approach are twofold. First, due to the curse of dimensionality and the fact that function $p(x, \theta)$ takes the additional input θ, a significantly larger number of measurements may be required to estimate $p(x, \theta)$ relative to $p(x)$. Second, sensors need to be able to measure their orientation, e.g., through accelerometers and magnetometers, which affects the cost and introduces additional error sources.

A pragmatic alternative is to treat the sensor orientations as random variables with uniform distribution over orientations θ. This implies that the isotropic power map $p(x)$ equals the expectation of $p(x, \theta)$ and, thus, one can still estimate $p(x)$ using the procedures described in previous sections upon disregarding orientation. The uncertainty introduced by the directionality of the antennas translates into additional measurement noise, which, therefore, increases the number of measurements required to estimate $p(x)$ with a target accuracy. This is the price to be paid for circumventing the aforementioned limitations.

Decentralized implementation

Unlike conventional spectrum sensing techniques, which often assume a common spectrum occupancy over the entire area of interest [15], spectrum cartography accounts for spatial variability. Thus, it is necessary that the measurements be obtained at various locations $\{x_n\}_{n=1}^N$ within the region, which then must be processed jointly. While this can be achieved, in theory, by collecting the measurements at a fusion center (FC) for centralized processing, the feedback overhead and the associated delay can be significant in practice. Moreover, the FC must operate with higher resource and security require-ments. An alternative is to employ distributed in-network processing, where all sensors collaboratively estimate the map via local interactions; i.e., the nth sensor, $n \in \mathcal{N} := \{1, \dots, N\}$, exchanges information only with its set of single-hop neighbors $\mathcal{N}_n \subset \mathcal{N}$ [2], [5], [9], [16], [47]. The key idea is that RME tasks often boil down to a regression problem of the form

$$\underset{\theta}{\text{minimize}} \frac{1}{2} \| y - X\theta \|_2^2 + \xi(\theta) \tag{20}$$

where $y \in \mathbb{R}^M$ and $X \in \mathbb{R}^{M \times \Theta}$ represent the targets and the regressors, respectively; $\theta \in \mathbb{R}^\Theta$ contains the regression coefficients; and $\xi(\cdot)$ is a convex regularizer that captures prior information; see, e.g., (8). It is often the case that the data X and y consist of the collection of the data $\{X_n\}$ and $\{y_n\}$ from the individual sensors. That is, $y = [y_1^\top, \dots, y_N^\top]^\top$, where $y_n \in \mathbb{R}^{M_n}$ for $n \in \mathcal{N}$ and $\Sigma_{n=1}^N M_n = M$. Likewise, $X = [X_1^\top, \dots, X_N^\top]^\top$ with $X_n \in \mathbb{R}^{M_n \times \Theta}$ for $n \in \mathcal{N}$.

To solve (20) in a decentralized manner, consider first an undirected graph $\mathbb{G} := (\mathcal{N}, \mathcal{E})$ with vertex set $\mathcal{N}$ and edge set $\mathcal{E}$, where vertices represent sensors, and the edge (n, n') is in $\mathcal{E}$ whenever sensors n and n' can communicate in a single hop—i.e., $n' \in \mathcal{N}_n$. If $\mathbb{G}$ is connected—i.e., there is a (possibly multihop) path between every pair of sensors—it can be easily shown that (20) is equivalent to

$$\underset{\{\theta_n, \gamma_n, \gamma_{(n,n')}\}}{\text{minimize}} \sum_{n=1}^N \left[\frac{1}{2} \| y_n - X_n \gamma_n \|_2^2 + \frac{1}{N} \xi(\theta_n) \right] \tag{21a}$$

$$\text{subject to } \gamma_n = \theta_n, n \in \mathcal{N} \tag{21b}$$

$$\theta_n = \gamma_{(n,n')} = \theta_{n'}, n' \in \mathcal{N}_n, n \in \mathcal{N}, \tag{21c}$$

where $\{\gamma_n\}$ and $\{\gamma_{(n,n')}\}$ are auxiliary variables. Per (21b), γ_n is just a copy of θ_n. In addition, $\{\gamma_{(n,n')}\}$ facilitate the derivation of simple update rules and are eventually eliminated. A decentralized algorithm can be derived by applying the
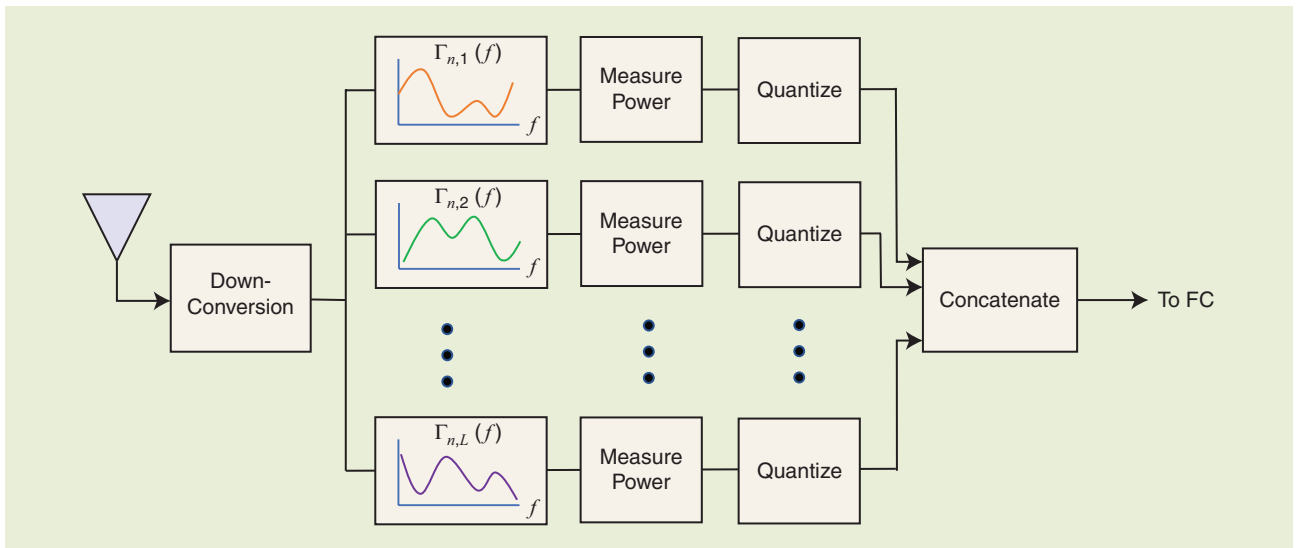

FIGURE 12. To reduce the rate necessary to report measurements, sensors may use a bank of random filters. The energy of each filter is measured, quantized, and sent to an FC that performs RME.

alternating-direction method of multipliers to (21a)–(21c). Following steps similar to those in [9, Appendix D], one can obtain the decentralized update rules for iteration k as

$$\boldsymbol{u}_n^{[k]} = \boldsymbol{u}_n^{[k-1]} + \rho \sum_{n' \in \mathcal{N}_n} (\boldsymbol{\theta}_n^{[k]} - \boldsymbol{\theta}_{n'}^{[k]}) \tag{22a}$$

$$\boldsymbol{\lambda}_n^{[k]} = \boldsymbol{\lambda}_n^{[k-1]} + \rho (\boldsymbol{\theta}_n^{[k]} - \boldsymbol{\gamma}_n^{[k]}) \tag{22b}$$

$$\boldsymbol{\theta}_n^{[k+1]} = \underset{\boldsymbol{\theta}}{\arg\min} \frac{1}{Nc_n} \xi(\boldsymbol{\theta}) + \frac{1}{2} \| \boldsymbol{\theta} - \boldsymbol{a}_n \|_2^2 \tag{22c}$$

$$\boldsymbol{\gamma}_n^{[k+1]} = (\rho \boldsymbol{I}_\Theta + \boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1} (\boldsymbol{X}_n^\top \boldsymbol{y}_n + \rho \boldsymbol{\theta}_n^{[k+1]} + \boldsymbol{\lambda}_n^{[k]}) \tag{22d}$$

where $\rho > 0$ is the step size, $c_n := \rho(1 + 2|\mathcal{N}_n|)$, and

$$\boldsymbol{a}_n := \frac{1}{c_n} \Big(\rho \sum_{n' \in \mathcal{N}_n} (\boldsymbol{\theta}_n^{[k]} + \boldsymbol{\theta}_{n'}^{[k]}) + \rho \boldsymbol{\gamma}_n^{[k]} - \boldsymbol{u}_n^{[k]} - \boldsymbol{\lambda}_n^{[k]} \Big) \tag{22e}$$

for $n \in \mathcal{N}$. As can be seen in (22a) and (22e), the updates involve only local communication with the neighbors. The proximal problem in (22c) admits a closed-form solution for various common choices of $\xi(\cdot)$. It can be proven that the iterate $\boldsymbol{\theta}_n^{[k]}$ for any $n \in \mathcal{N}$ converges to the solution of (20) as $k \to \infty$ [9].

Rate constraints
To maintain up-to-date maps, every certain time interval, the sensors need to collect new measurements, which are then sent to a FC or shared with other nodes. For maps that change rapidly over time or require high-dimensional measurements, such as PSD maps, the bandwidth required to report the measurements may be significant. To mitigate such an issue, compression and quantization can be employed [4].

The idea is twofold. First, instead of directly computing PSD estimates at the sensors, each sensor measures the powers at the outputs of a filter bank acting on the received signal. Then, only quantized versions of those measurements are reported; see Figure 12.

To simplify the exposition, assume, for now, that each sensor employs a single filter. Recall that $p(\boldsymbol{x}, f)$ denotes the PSD at location $\boldsymbol{x}$. If the received signal at location $\boldsymbol{x}_n$ is processed by a filter with frequency response $\Gamma_n(f)$, the output power is given by $\bar{p}_n := \int p(\boldsymbol{x}_n, f) |\Gamma_n(f)|^2 df$. Due to the measurement noise, the measured value $\tilde{p}_n$ will be generally different from the true $\bar{p}_n$. Subsequently, $\tilde{p}_n$ is quantized to $\check{m}_n$, which is then sent to the FC or other sensors. This clearly requires a much smaller bandwidth than sending, e.g., the entire periodogram.

To see how the map can be estimated from these linearly compressed and quantized measurements, recall the decomposition $p(\boldsymbol{x}, f) = \Sigma_c p_c(\boldsymbol{x}) \phi_c(f)$ from section "Estimation in Wideband Channels").

Since the basis functions ϕ_c are known, this decomposition reduces the problem of estimating p to that of estimating C functions $p_1, \dots, p_C$. It also follows that $\bar{p}_n$ can be written as $\bar{p}_n = \Sigma_c p_c(\boldsymbol{x}_n) \int \phi_c(f) |\Gamma_n(f)|^2 df = [p_1(\boldsymbol{x}_n), \dots, p_C(\boldsymbol{x}_n)] \boldsymbol{\phi}_n$, where the cth entry of vector $\boldsymbol{\phi}_n$ is $\int \phi_c(f) |\Gamma_n(f)|^2 df$. In other words, $\bar{p}_n$ is a linear combination of the values that the functions $p_1, \dots, p_C$ take at $\boldsymbol{x}_n$.

If the true powers $\bar{p}_1, \dots, \bar{p}_N$ were known exactly, one could seek RKHS functions $\hat{p}_1, \dots, \hat{p}_C$ such that $[\hat{p}_1(\boldsymbol{x}_n), \dots, \hat{p}_C(\boldsymbol{x}_n)] \boldsymbol{\phi}_n = \bar{p}_n \; \forall n$ using kernel-based learning.

Now, consider the case where, instead of $\bar{p}_1, \dots, \bar{p}_N$, one has the quantized measurements $\check{m}_1, \dots, \check{m}_N$, but it holds that $\tilde{p}_n = \bar{p}_n$ for all n; i.e., there is no measurement noise. Each $\check{m}_n$, therefore, indicates which quantization interval contains $\bar{p}_n$. Upon denoting the endpoints of the interval that contains $\check{m}_n$ as $a(\check{m}_n)$ and $b(\check{m}_n)$, it makes sense to now seek $\hat{p}_1, \dots, \hat{p}_C$ that satisfy $[\hat{p}_1(\boldsymbol{x}_n), \dots, \hat{p}_C(\boldsymbol{x}_n)] \boldsymbol{\phi}_n \in [a(\check{m}_n), b(\check{m}_n)]$ for all n.

Finally, in the case where there is measurement noise, $\tilde{p}_n$ is generally different from $\bar{p}_n$. If the noise is small relative to the width of the quantization interval, the result of quantizing either value will be often the same, but not always. This means that one cannot impose that $[g_1(\boldsymbol{x}_n), \dots, g_C(\boldsymbol{x}_n)] \boldsymbol{\phi}_n$ necessarily falls in the quantization interval $[a(\check{m}_n), b(\check{m}_n)]$. Instead, the condition must be encouraged in a soft manner by penalizing deviations from the interval. Interestingly, by penalizing deviations in a linear fashion, it can be shown that the resulting estimates can be obtained through support vector regression [4].

The previous considerations can be extended to the case where the filter bank at each sensor contains $L > 1$ filters, as depicted in Figure 12. Observe that now two subscripts are necessary to index each branch.

The power at the lth branch of the sensor at $\boldsymbol{x}_n$ is given by $\bar{p}_{n,l} = [p_1(\boldsymbol{x}_n), \dots, p_C(\boldsymbol{x}_n)] \boldsymbol{\phi}_{n,l}$. Since all of the vectors $\boldsymbol{\phi}_{n,1}, \dots, \boldsymbol{\phi}_{n,L}$ multiply the same $[p_1(\boldsymbol{x}_n), \dots, p_C(\boldsymbol{x}_n)]$, the values $\bar{p}_{n,l}$ are not fully informative about $p_1, \dots, p_C$ unless the vectors $\boldsymbol{\phi}_{n,1}, \dots, \boldsymbol{\phi}_{n,L}$ are linearly independent. This imposes a design constraint on the filters. For example, filters with pseudorandom impulse responses may be utilized and are expected to yield linearly independent vectors $\boldsymbol{\phi}_{n,1}, \dots, \boldsymbol{\phi}_{n,L}$ so long as $L \leq C$.

Future directions
Although RME has been the subject of a sizable research body, a large number of open issues still remain. First of all, the potential of radio maps to endow applications with radio situational awareness is yet to be fully exploited. A large part of the progress in this regard has taken place in the context of device-free localization (see, e.g., [43]) and UAV communications (see, e.g., [17] and references therein), but a number of tasks arising in cellular networks, such as resource allocation, are yet to be explored. Radio maps can also be used as priors for enhanced channel estimation in mobile communications.

Improving inference biases in data-driven radio map estimators is also necessary. This can be achieved by collecting extensive datasets in multiple bands since most works so far rely on synthetic data generated with ray-tracing software. Such datasets would also open the door to devising improved uncertainty metrics for spectrum surveying. Remarkably, these can be used for improving spectrum surveying techniques [22]. Furthermore, hybrid model-based and data-driven approaches have the potential to combine the best of both worlds [33]. The

rationale is that radio propagation models may significantly reduce the amount of data required to train data-driven estimators, whereas learning from data can significantly improve the accuracy of model-based approaches.

Methods for coping with various sources of error must also be devised. For example, time variations may be better predicted by exploiting side information on the mobility of terminals. In this context, trajectories of ground vehicles on the road or UAVs in aerial corridors may be instrumental to reduce the effective dimensionality of propagation maps. One can also model how groups of persons or vehicles move to better predict signal strength maps as a whole. Other sources of error to counter include the use of antennas with two polarizations and nonisotropic gain patterns.

Recent developments adopt machine learning algorithms to predict the channel state information (CSI) of desired multiantenna channels based on pilot CSI. This approach can capture the characteristics of small-scale fading, going beyond the channel gain maps. In [46], the pilot CSIs are obtained from a set of links that are different from the target link. The optimal transmit beam pattern of the desired link is predicted based on the acquired CSIs. When the source and the target links are not colocated, the traditional assumption is that the CSIs are statistically independent. In reality, there can be significant dependency between the CSIs and the geometry of the propagation environment, transceiver locations, the line-of-sight path, and other multipaths within the coherence time of the channels. Given sufficiently rich pilot CSI measurements that capture the relevant geometry, an appropriate nonlinear mapping (e.g., via a DNN) can exploit this dependency. As a related idea, channel charting obtains, in an unsupervised fashion, low-dimensional embeddings of the high-dimensional CSIs that approximately provide the spatial locations of the measurements [48].

Finally, further types of radio maps may also be explored. For example, maps may be developed for massive multiple-input, multiple-output and millimeter-wave networks to benefit from reduced search time for beam selection. As another example, exploring delay Doppler maps could be instrumental in the context of resource allocation for the emerging orthogonal time frequency and space modulation.

Related work

The interested reader can delve deeper into RME through the surveys [42] and [49]. In [49], the focus is on occupancy maps, which are radio maps that provide the fraction of time that a certain frequency channel is used. On the other hand, the authors of [42] focus on power map estimation and review other methods that are not discussed here due to space limitations. Relative to these works, the present tutorial is more introductory in nature and considers more classes of maps as well as more recent methods.

Conclusions

Radio maps characterize important metrics of the RF spectrum landscape across a geographical area. Two families of radio maps were considered based on whether the received signal strength or the propagation channel effects are of interest, and a large number of representative applications were discussed. Tutorial expositions of various data-driven methods for RME have been presented, ranging from parametric, nonparametric, and probabilistic approaches to recent powerful deep learning techniques, incorporating useful priors, such as sparsity, low rank, and union-of-subspace structures. Practical issues related to spectrum surveying, noisy location estimates, decentralized implementation, and limited-rate measurements were also discussed. With the advent of ultradense and ultradynamic deployment scenarios often envisioned in future wireless networking, the role of data-driven spectrum cartography enabled via sophisticated RME techniques will likely become even more relevant.

Acknowledgment

Authors

Daniel Romero (daniel.romero@uia.no) received his M.Sc. and Ph.D. degrees in signal theory and communications from the University of Vigo, Spain, in 2011 and 2015, respectively. From 2015 to 2016, he was a postdoctoral researcher with the Digital Technology Center and Department of Electrical and Computer Engineering, University of Minnesota, USA. Then, he joined the Department of Information and Communication Technology, University of Agder, Grimstad 4879, Norway, as an associate professor. His research interests include machine learning, artificial intelligence, optimization, signal processing, and aerial communications.

Seung-Jun Kim (sjkim@umbc.edu) received his B.S. and M.S. degrees from Seoul National University in 1996 and 1998, respectively, and his Ph.D. degree from the University of California, Santa Barbara in 2005, all in electrical engineering. During 2005–2008, he was with NEC Labs America, Princeton, NJ. During 2008–2014, he was with the University of Minnesota, where his final title was research associate professor. He joined the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250 USA in 2014. He is a senior area editor for *IEEE Signal Processing Letters*. His research interests include signal processing, machine learning, and optimization with applications to wireless communication, power systems, and medical imaging.

References

[1] A. Alaya-Feki, S. B. Jemaa, B. Sayrac, P. Houze, and E. Moulines, "Informed spectrum usage in cognitive radio networks: Interference cartography," in *Proc. IEEE Int. Symp. Personal, Indoor Mobile Radio Commun.*, Cannes, France, Sep. 2008, pp. 1–5, doi: 10.1109/PIMRC.2008.4699911.

[2] J.-A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010, doi: 10.1109/TSP.2009.2038417.

[3] H. B. Yilmaz, T. Tugcu, F. Alagöz, and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 162–169, Dec. 2013, doi: 10.1109/MCOM.2013.6685772.

[4] D. Romero, S.-J. Kim, G. B. Giannakis, and R. López-Valcarce, "Learning power spectrum maps from quantized power measurements," *IEEE Trans. Signal*

Process., vol. 65, no. 10, pp. 2547–2560, May 2017, doi: 10.1109/TSP.2017.2666775.

[5] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using Kriged Kalman filtering," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 24–36, Feb. 2011, doi: 10.1109/JSTSP.2010.2053016.

[6] D. Romero, D. Lee, and G. B. Giannakis, "Blind radio tomography," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2055–2069, Jan. 2018, doi: 10.1109/TSP.2018.2799169.

[7] C. Phillips, D. Sicker, and D. Grunwald, "Bounding the practical error of path loss models," *Int. J. Antennas Propag.*, vol. 2012, pp. 71–82, Jun. 2012, doi: 10.1155/2012/754158.

[8] E. Dall'Anese, J.-A. Bazerque, and G. B. Giannakis, "Group sparse lasso for cognitive network sensing robust to model uncertainties and outliers," *Phy. Commun.*, vol. 5, no. 2, pp. 161–172, Jun. 2012, doi: 10.1016/j.phycom.2011.07.005.

[9] J.-A. Bazerque, G. Mateos, and G. B. Giannakis, "Group-lasso on splines for spectrum cartography," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4648–4663, Oct. 2011, doi: 10.1109/TSP.2011.2160858.

[10] C. Parera, Q. Liao, I. Malanchini, C. Tatino, A. E. C. Redondi, and M. Cesana, "Transfer learning for tilt-dependent radio map prediction," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 2, pp. 829–843, Jan. 2020, doi: 10.1109/TCCN.2020.2964761.

[11] T. Imai, K. Kitao, and M. Inomata, "Radio propagation prediction model using convolutional neural networks by deep learning," in *Proc. IEEE Eur. Conf. Antennas Propag.*, Krakow, Poland, Apr. 2019, pp. 1–5.

[12] M. Iwasaki, T. Nishio, M. Morikura, and K. Yamamoto, "Transfer learning-based received power prediction with ray-tracing simulation and small amount of measurement data," 2020, *arXiv:2005.00833.*

[13] R. Levie, Ç. Yapar, G. Kutyniok, and G. Caire, "RadioUNet: Fast radio map estimation with convolutional neural networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 4001–4015, 2021, doi: 10.1109/TWC.2021.3054977.

[14] D. Romero and G. Leus, "Non-cooperative aerial base station placement via stochastic optimization," in *Proc. IEEE Mobile Ad-Hoc Sensor Netw.*, Shenzhen, China, Dec. 2019, pp. 131–136.

[15] E. Axell, G. Leus, and E. G. Larsson, "Overview of spectrum sensing for cognitive radio," in *Proc. 2nd Int. Workshop Cogn. Inf. Process.*, 2010, pp. 322–327, doi: 10.1109/CIP.2010.5604136.

[16] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205–1211, Feb. 2011, doi: 10.1109/TVT.2011.2113195.

[17] D. Romero, P. Q. Viet, and G. Leus, "Aerial base station placement leveraging radio tomographic maps," May 2022, *arXiv:2109.07372.*

[18] Y. Teganya and D. Romero, "Deep completion autoencoders for radio map estimation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1710–1724, Aug. 2021, doi: 10.1109/TWC.2021.3106154.

[19] B. Schölkopf and A. J. Smola, "Learning with Kernels: Support vector machines," in *Regularization, Optimization, and Beyond*, Cambridge, MA, USA: MIT Press, 2001.

[20] D. Romero, S.-J. Kim, and G. B. Giannakis, "Stochastic semiparametric regression for spectrum cartography," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, Cancun, Mexico, Dec. 2015, pp. 513–516, doi: 10.1109/CAMSAP.2015.7383849.

[21] A. Agarwal and R. Gangopadhyay, "Predictive spectrum occupancy probability-based spatio-temporal dynamic channel allocation map for future cognitive wireless networks," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 8, p. e3442, Jun. 2018, doi: 10.1002/ett.3442.

[22] R. Shrestha, D. Romero, and S. P. Chepuri, "Spectrum surveying: Active radio map estimation with autonomous UAVs," *IEEE Trans. Wireless Commun.,* early access, Aug. 2022, doi: 10.1109/TWC.2022.3197087.

[23] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Lett.*, vol. 27, no. 23, pp. 2145–2146, Nov. 1991, doi: 10.1049/el:19911328.

[24] D.-H. Huang, S.-H. Wu, W.-R. Wu, and P.-H. Wang, "Cooperative radio source positioning and power map reconstruction: A sparse Bayesian learning approach," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2318–2332, Jun. 2015, doi: 10.1109/TVT.2014.2345738.

[25] B. Khalfi, B. Hamdaoui, and M. Guizani, "AirMAP: Scalable spectrum occupancy recovery using local low-rank matrixapproximation," in *Proc. IEEE GLOBECOM*, Abu Dhabi, UAE, Dec. 2018, pp. 206–212, doi: 10.1109/GLOCOM.2018.8647667.

[26] D. Schäufele, R. L. G. Cavalcante, and S. Mtanczak, "Tensor completion for radio map reconstruction using low rank and smoothness," in *Proc. IEEE SPAWC*, Cannes, France, Jul. 2019, pp. 1–5, doi: 10.1109/SPAWC.2019.8815495.

[27] G. Zhang, X. Fu, J. Wang, and M. Hong, "Coupled block-term tensor decomposition based blind spectrum cartography," in *Proc. Asilomar Conf. Signal, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 1644–1648, doi: 10.1109/IEEECONF44664.2019.9048667.

[28] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 37, no. 23, pp. 3311–3325, Dec. 2006, doi: 10.1109/TSP.2006.881199.

[29] S.-J. Kim and G. B. Giannakis, "Cognitive radio spectrum prediction using dictionary learning," in *Proc. IEEE Global Commun. Conf.*, Atlanta, GA, USA, Dec. 2013, pp. 3206–3211, doi: 10.1109/GLOCOM.2013.6831565.

[30] S.-J. Kim and G. B. Giannakis, "Dynamic learning for cognitive radio sensing," in *Proc. 5th IEEE Int. Workshop Comp. Adv. Multi-Sensor Adapt. Process.*, St. Martin, French Caribbean, Dec. 2013, pp. 388–391, doi: 10.1109/CAMSAP.2013.6714089.

[31] V. V. Ratnam et al., "FadeNet: Deep learning-based mm-wave large-scale channel fading prediction and its applications," *IEEE Access*, vol. 9, pp. 3278–3290, 2021, doi: 10.1109/ACCESS.2020.3048583.

[32] E. Krijestorac, S. Hanna, and D. Cabric, "Spatial signal strength prediction using 3D maps and deep learning," in *Proc. IEEE Int Conf. Commun.*, 2021, pp. 1–6, doi: 10.1109/ICC42927.2021.9500970.

[33] J. Thrane, D. Zibar, and H. L. Christiansen, "Model-aided deep learning method for path loss prediction in mobile communication systems at 2.6 GHz," *IEEE Access*, vol. 8, pp. 7925–7936, Jan. 2020, doi: 10.1109/ACCESS.2020.2964103.

[34] Q. Niu, Y. Nie, S. He, N. Liu, and X. Luo, "RecNet: A convolutional network for efficient radiomap reconstruction," in *Proc. IEEE Int. Conf. Commun.*, Kansas City, MO, USA, May 2018, pp. 1–7, doi: 10.1109/ICC.2018.8422971.

[35] X. Han, L. Xue, F. Shao, and Y. Xu, "A power spectrum maps estimation algorithm based on generative adversarial networks for underlay cognitive radio networks," *Sensors*, vol. 20, no. 1, p. 311, Jan. 2020, doi: 10.3390/s20010311.

[36] S. Shrestha, X. Fu, and M. Hong, "Deep generative model learning for blind spectrum cartography with NMF-based radio map disaggregation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 4920–4924, doi: 10.1109/ICASSP39728.2021.9413382.

[37] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6232–6246, Dec. 2013, doi: 10.1109/TSP.2013.2283473.

[38] K. V. Mardia, C. Goodall, E. J. Redfern, and F. J. Alonso, "The Kriged Kalman filter," *Test*, vol. 7, no. 2, pp. 217–285, Dec. 1998, doi: 10.1007/BF02565111.

[39] D. Lee, S.-J. Kim, and G. B. Giannakis, "Channel gain cartography for cognitive radios leveraging low rank and sparsity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5953–5966, Jun. 2017, doi: 10.1109/TWC.2017.2717822.

[40] N. Patwari and P. Agrawal, "NeSh: A joint shadowing model for links in a multi-hop network," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, NV, USA, Mar. 2008, pp. 2873–2876, doi: 10.1109/ICASSP.2008.4518249.

[41] B. R. Hamilton, X. Ma, R. J. Baxley, and S. M. Matechik, "Propagation modeling for radio frequency tomography in wireless networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 1, pp. 55–65, Feb. 2014, doi: 10.1109/JSTSP.2013.2287471.

[42] M. Pesko, T. Javornik, A. Kosir, M. Stular, and M. Mohorcic, "Radio environment maps: The survey of construction methods," *KSII Trans. Internet Inf. Syst.*, vol. 8, no. 11, pp. 3789–3809, Dec. 2014, doi: 10.3837/tiis.2014.11.008.

[43] J. Wilson, N. Patwari, and O. G. Vasquez, "Regularization methods for radio tomographic imaging," in *Proc. Virginia Tech Symp. Wireless Personal Commun.*, Blacksburg, VA, USA, Jun. 2009, pp. 1–9.

[44] D. Lee, D. Berberidis, and G. B. Giannakis, "Adaptive Bayesian channel gain cartography," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Calgary, Canada, Apr. 2018, pp. 3555–3558, doi: 10.1109/ICASSP.2018.8461412.

[45] Y. Teganya, D. Romero, L. M. Lopez-Ramos, and B. Beferull-Lozano, "Location-free spectrum cartography," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 4013–4026, Aug. 2019, doi: 10.1109/TSP.2019.2923151.

[46] Z. Jiang, S. Chen, A. F. Molisch, R. Vannithamby, S. Zhou, and Z. Niu, "Exploiting wireless channel state information structures beyond linear correlations: A deep learning approach," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 28–34, Mar. 2019, doi: 10.1109/MCOM.2019.1800581.

[47] S.-J. Kim, N. Jain, G. B. Giannakis, and P. Forero, "Joint link learning and cognitive radio sensing," in *Proc. Asilomar Conf. Signal, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2011, pp. 1415–1419, doi: 10.1109/ACSSC.2011.6190250.

[48] J. Deng, O. Tirkkonen, J. Zhang, X. Jiao, and C. Studer, "Network-side localization via semi-supervised multi-point channel charting," in *Proc. Int. Wireless Commun. Mobile Comput. Conf.*, Harbin, China, Jul. 2021, pp. 1654–1660, doi: 10.1109/IWCMC51323.2021.9498723.

[49] M. Höyhtyä et al., "Spectrum occupancy measurements: A survey and use of interference maps," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2386–2414, Apr. 2016, doi: 10.1109/COMST.2016.2559525.

SP