Augmenting Channel Estimation via Loss Field: Site-trained Bayesian Modeling and Comparative Analysis

Jie Wang[a], Meles G. Weldegebriel[a], Neal Patwari[b]

[a]*McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO, USA*
[b]*Kahlert School of Computing, University of Utah, Salt Lake City, UT, USA*

Abstract

Future wireless networks that share spectrum dynamically among groups of mobile users will require fast and accurate channel estimation in order to guarantee varying signal-to-interference-plus-noise ratio (SINR) requirements for co-channel links. There is a need for channel models with low computational complexity and high accuracy that adapt to the particular area of deployment while preserving explainability. In this work, we propose the *Channel Estimation via Loss Field (CELF)* model, which augments existing channel models using channel loss measurements from a deployed network and a Bayesian linear regression method to estimate a site-specific loss field for the area. The loss field is explainable as a site map of additional radio 'shadowing', compared to the channel base model, but it requires no site-specific terrain or building information. For an arbitrary pair of transmitter and receiver positions, CELF sums the loss field near the link line to estimate its shadowing loss. We use extensive indoor and outdoor measurements to show that CELF lowers the modeling error variance of the log-distance path loss base model by up to 68% for prediction, and outperforms 3 popular Machine Learning (ML) methods in variance reduction and training efficiency. To validate CELF's robustness, it is applied to a different channel base model, the terrain-integrated rough earth model (TIREM), and numerical results show that CELF can reduce the test variance by up to 63%. We further discuss two spatial multipath models for a weight matrix in CELF and observe similar accuracy improvement. To summarize, CELF offers a new type of explainable learning model for accurate and fast site-specific radio channel loss estimation.

Keywords: dynamic spectrum management, channel modeling, shadowing prediction, mobile networks

1. Introduction

Spectrum allocation is becoming increasingly dynamic and shareable in order to meet the growing demand [1, 2]. Examples include the citizens broadband radio service (CBRS) band [3] and the radio dynamic zone [4, 5]. A major part of the challenge to achieve reliable dynamic spectrum allocation is to accurately and efficiently predict signal and interference powers between

Email addresses: jie.w@wustl.edu (Jie Wang), g.weldegebriel@wustl.edu (Meles G. Weldegebriel), neal.patwari@utah.edu (Neal Patwari)

[0]This research is supported in part by the NSF grants #2232464 and #1827940, and the PAWR Project Office.

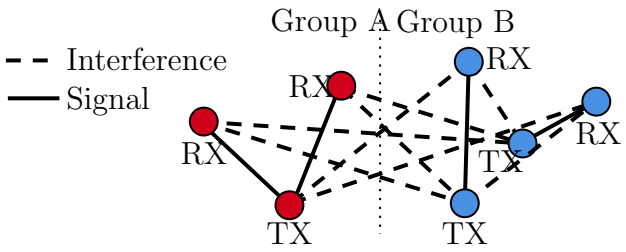


Figure 1: Assigning channels and transmit powers to ensure the required SINRs among links between T transmitters (TX) and R receivers (RX) demands RT co-channel channel loss estimation and recomputation as users move.

all pairs of proximate mobile transmitters and receivers, as shown in Fig. 1, to ensure that signal-to-interference-plus-noise ratios (SINRs) are sufficient for all groups.

Current channel models are not well-matched to the needs of dynamic spectrum management in mobile networks. Many path loss prediction models require the computation of multiple propagation loss mechanisms such as reflection and diffraction in the particular geometry of the network deployment area. For example, the terrain-integrated rough earth model (TIREM) [6] computes diffraction losses based on the terrain features and building heights extracted for each transmitter and receiver pair. Ray tracing models [7] additionally require high-resolution environmental databases and are highly computationally complex. Such site-specific models have high accuracy compared to general-purpose models that curve-fit to empirical data, such as Okumura-Hata [8] and log-distance path loss [9]. However, if real-time dynamic spectrum management requires high-resolution site clutter data and significant computational and memory resources, it will limit who can perform this management [10].

Emerging machine learning (ML) channel models can be both accurate and fast during testing but require very large datasets and computational resources during model training [11, 12]. Further, ML models suffer from the black-box problem, in which no human-understandable explanation or reasoning for their predictions is possible [13]. This prevents system engineers from diagnosing problems when a model performs poorly. Updating an ML channel model over time does not allow engineers to explain how (or if) the model has been impacted by changes in the environment, e.g., a new building. Current and future regulations may require model explanations for legal purposes [14] — if a system is harmed by path loss prediction model errors, a human-understandable explanation must be provided.

In this paper, we develop and validate a new type of channel learning model, *Channel Estimation via Loss Field (CELF)*, which is simultaneously explainable, is less computationally complex to train, and is more accurate than current ML channel models trained with the same data. CELF formulates modeling errors after an arbitrary channel base model as a linear function of a shadowing loss field. This loss field is connected to the underlying wave propagation physics in that it accounts for the physical mechanism of shadowing due to obstacles in the spatial domain, and is viewable as a simple image map. A long history of radio propagation research has applied this linear additive loss modeling to estimate particular objects [15] or image motion [16]. However, we are unaware of any work using it as a foundation for a site-agnostic learning-based channel model. As shown in Fig. 2, the loss field is learned from training measurements via Bayesian linear regression, but training is lower in computation requirements compared to a general-purpose ML model. Using training measurements allows the model to fit the particular site of deployment. Sensors deployed as part of a radio dynamic zone, or the dynamic spectrum access protocol, can be used

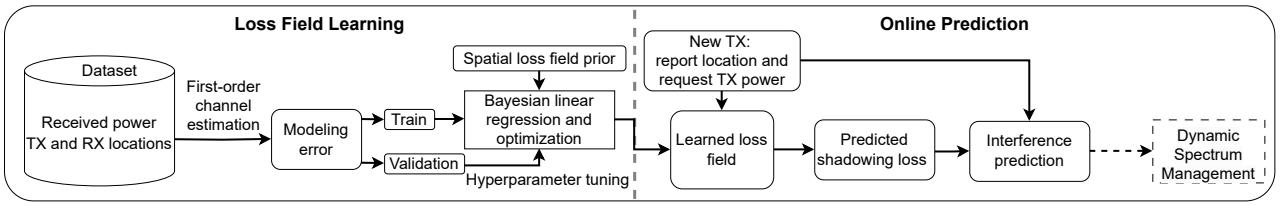


Figure 2: Overview of the proposed CELF model and the online shadowing prediction process. Interference predictions can be used for dynamic spectrum management, but its implementation is not within this paper's scope.

to collect these measurements. Training data quantities can be low in comparison with other ML methods. We also discuss Bayesian regression's stability and optimization for more robust and efficient learning of the loss field. To predict shadowing loss for a new link, CELF computes a weighted sum of a small number of pixels of the learned loss field image. The implementation of CELF is at [17].

We use one indoor and three outdoor datasets to experimentally quantify how accurately and efficiently CELF performs. We compare CELF with three general-purpose ML channel models: support vector regression (SVR), random forests, and multi-layer perceptron (MLP)-ANN, in terms of (1) variance reduction compared to a channel base model, (2) training efficiency and (3) prediction efficiency. CELF reduces the variance of modeling errors on the test dataset by up to 68% outdoors and 40% indoors. In comparison to other methods, CELF achieves larger test variance reductions. The MLP-ANN model is the most accurate model out of the three ML methods, but it requires three times more time than CELF for model training. For loss prediction, CELF is faster than SVR but slower than MLP-ANN as the test dataset size and the loss field size impact CELF's prediction efficiency.

We comprehensively explore CELF's accuracy improvement for two different channel base models and two different weight matrix models. First, CELF should, regardless of the base model, be able to learn a loss field from the modeling errors and improve the shadowing predictions. In this work, we include the popular *log-distance path loss* [9] and TIREM [6] to verify the robustness of CELF. Results show that CELF reduces the *log-distance path loss* model variance by up to 68% and reduces TIREM variance by up to 63% across all test datasets. Second, a weight model represents the geometric shape near and between the transmitter and receiver for which the spatial loss field has an impact on the link loss. Different weight matrix models formulate different multipath environments, e.g., ellipse for reflected multipath [18, 16] and Cassini oval for scattered multipath [19, 20]. We hence incorporate the ellipse and Cassini oval models in CELF for comparison. Numerical results show that both models are valid weight matrix models and result in similar accuracy improvement on the test datasets.

For perspective, path loss models that have no fine-grained site-specific information *do not predict small-scale fading effects*, i.e., those caused by sub-wavelength (cm-level) changes in the position of the transmitter or receiver. Small-scale fading is severe, e.g., more than 20 dB below the mean power 1% of the time in a Rayleigh fading channel [9]. Path loss models do not know the device and environmental obstruction positions to the required level of accuracy. Instead, channel path loss models, like our proposed CELF, predict large-scale fading (caused by increasing distance) and medium-scale or shadow fading (caused by obstructions) [21]. Since training and testing measurements include small-scale fading which our model cannot predict, we cannot reduce

the path loss variance to zero. Instead, we judge models by how much they can reduce fading variance compared to a standard statistical channel model. We find that CELF shows larger variance reductions across all of our experiments than any other model, ML or otherwise.

The initial work of CELF appeared at IEEE DySPAN 2024 [22]. The major additional contributions in comparison to [22] are as follows: (1) A hybrid empirical/physical model, the terrain-integrated rough earth model (TIREM), is included as a different channel base model. The analysis verifies that CELF can robustly improve the accuracy of various channel models. (2) Two spatial multipath models for a weight matrix in CELF are presented to discuss different geometric shapes for which the spatial loss field contributes to the link loss. Numerical results validate both spatial models and they result in similar accuracy improvements. (3) Two more real-world datasets, one collected in a shared frequency band and the other from a large urban environment, are experimentally analyzed to evaluate CELF's accuracy and efficiency performance under the new scenarios. (4) We provide a detailed discussion on the explainability of CELF modeling and the learned loss field. Numerical tests verify that the increase in the loss field near known obstructions is statistically more significant in both indoor and outdoor environments, and hence the loss field can be explained by the locations of the obstructions in the area.

2. Related Work

Path loss prediction has an extensive disciplinary history over several decades. Models used today rely on [22]:

1. the physical mechanisms of radio propagation, e.g., reflection and diffraction;
2. information about the site, e.g., terrain and building geometry data;
3. curve-fitting of empirical measurements recorded in a different area;
4. fitting or learning using empirical data collected in the area of deployment.

2.1. Physics-based models

Physics-based models aim to accurately characterize radio wave propagation effects such as reflection and diffraction. The most fundamental is the *free-space path loss model* [9], but it models only unobstructed channels, and is thus limited to satellite or deep space communication and unobstructed microwave relay links. The *two-ray ground reflection model* accounts for both the line-of-sight (LOS) and the ground-reflected paths [23], and is typically used in flat clutter-free areas like plains [24]. When more multipath must be modeled, *ray tracing* is a highly accurate but highly computationally complex model for path loss [7]. Ray tracing requires site-specific building databases, i.e., building layout, heights, and dielectric properties, as well as detailed terrain and ground use data, so that each wave path can be traced using geometrical optics [25]. Its computational complexity and need for high-resolution site-specific data make it impractical for large-scale, real-time applications.

2.2. General empirical models

General empirical models are based on an analysis of measurements taken from an environment similar in use to the area of interest, e.g., urban or suburban. The *Okumura-Hata model* is based on measurements from Tokyo in the 1960s as formulated by Hata [8]. It uses curve-fitting to model the effect of signal frequency, antenna heights, path length, and environment type on the channel loss. The *COST-231 Hata model* extends the Okumura-Hata model to data from some European cities [26]. The benefits of statistical models are the simple closed-form formula and no need for data from the site of interest. However, they are restricted to certain frequency and distance ranges, and most critically, they are most accurate in the environments from which the measurements are sourced [24].

2.3. Hybrid empirical/physical models

The *Longley-Rice model*, also known as the irregular terrain model (ITM), combines empirical modeling and physical principles for ground reflection, knife-edge and far-field diffraction, and troposcatter predictions [27]. This model considers environmental factors including surface refractivity, ground conductivity, atmospheric parameters, and terrain irregularities for path loss prediction [7]. It is in use today in systems like Spectrum Access System (SAS) [28]. The TIREM model [6] considers a profile of the terrain features and building heights [29]. The last hybrid model is the International Telecommunication Union's (ITU)-R P.1812 model, which uses detailed terrain profiles to target path-specific predictions. It has been widely used for terrestrial wireless systems [30].

2.4. Statistical models

Statistical models characterize the statistical distribution of the channel losses, rather than only the average value. The most common model is the *log-normal shadowing model*, which models shadowing loss as normally distributed in dB [24]. Other models explain the statistical correlation between the shadowing loss on two proximate links [31, 32, 33], which become correlated by passing through the same or similar obstructions. CELF models this correlation implicitly via its loss field. Other distributions for shadowing include the *Gamma* [34] and *inverse Gamma* [35] distributions. We note that the most well-known distributions, Rayleigh and Rician, are models for small-scale fading loss, and are thus not further discussed in this paper.

2.5. ML channel models

Machine Learning (ML) channel models are designed using general-purpose ML architectures and extensive datasets [11, 36, 37]. We categorize these models as: (1) SVR, K-Nearest-Neighbors (KNN), and ensemble learning methods such as random forests [37]; (2) Artificial Neural Networks (ANN) models including MLP-ANN models [38, 39] and radial basis function-ANN models (RBF-ANN) [40], and (3) more complex DNN models [12, 41]. For example, the RadioUNet model in [12] utilizes large datasets and environmental geometry as input to Unet, a special Convoluted Neural Network (CNN) architecture for path loss modeling.

ML channel methods can provide higher prediction accuracy than the aforementioned models at the cost of extensive datasets or detailed environmental information. Additionally, the high complexity of model training and updating will result in significant latency. The lack of interpretability of ML methods is a particular challenge, as RF engineers can find it difficult to diagnose a problem

when the model performs poorly. Further, regulation increasingly requires businesses to be able to explain why an algorithm's prediction was made [14].

CELF is also a learning-based model which uses site measurements to train. It requires no knowledge about the environment and can be trained with fewer measurements than a general-purpose ML model. Further, CELF explains its estimates via the shadowing field image, which should show statistically more significant losses near the attenuating obstructions in the area.

3. Channel Estimation via Loss Field

In this section, we present the CELF model in three parts. First, we describe the idea of a *base model*, and describe what is used in this paper. Next, we describe how CELF augments the base model for better path loss estimation using a spatial loss field. Finally, we explain how to estimate the loss field from training measurements.

3.1. First-order channel estimation

CELF predicts the additional path loss compared to a *channel base model*. The channel base model could be any model described in the related work (Section 2). CELF's role is to augment the estimates from the base model by additionally accounting for the natural spatial correlations in the path loss that are not modeled by the base model. To verify the effectiveness of CELF, we explore two different models in this paper as the base model: *log-distance path loss* and *TIREM*.

Log-distance path loss model. It states that the received power estimate in dBm $\tilde{P}(d_l)$ along a link $l = (i, j)$ between node i and node j reduces in a logarithmic manner with increasing distance [9]:

$$\tilde{P}(d_l) = P_T - \Pi_0 - 10n_p \log \frac{d_l}{\Delta_0}, \tag{1}$$

where P_T is the transmitted power in dBm, d_l is the link distance, Π_0 is a constant specifying the dB loss at a reference distance Δ_0, and the path loss exponent n_p indicates the level of environmental clutter.

Given the same distance d_l, the received power measurements vary around the estimate $\tilde{P}(d_l)$ due to shadow fading and small-scale fading [24]. As a result, the modeling errors between the measurement $P(d_l)$ and the estimate $\tilde{P}(d_l)$ can be written as:

$$Z_{l,\text{log-dist.}} = \tilde{P}(d_l) - P(d_l), \tag{2}$$

where $Z_{l,\text{log-dist.}}$ is the modeling error which consists of independent shadowing loss and small-scale fading loss [16].

TIREM. The second model, TIREM, is an extension of the widely-applied Longley-Rice model. It utilizes site-specific terrain features and electromagnetic theory to estimate path loss for frequencies between 1 MHz - 1 THz and distances up to 30 km [6]. One disadvantage, however, is that it over-predicts shadowing loss as obstacles are assumed to be infinitely long knife edges [29]. We denote the difference between the received power measurements and the TIREM estimates as:

$$Z_{l,\text{TIREM}} = \tilde{P}_{\text{TIREM}}(d_l) - P(d_l), \tag{3}$$

where $\tilde{P}_{\text{TIREM}}(d_l)$ is the estimated power given by TIREM, and $Z_{l,\text{TIREM.}}$ is the over-prediction error. To simplify the notation used throughout the rest of this paper, Z_l is used to represent the total modeling error by any arbitrary base model.

3.2. Network shadowing model for shadowing correlation

The total modeling error Z_l is commonly modeled as independent and identically distributed (*i.i.d.*) across links [42, 43, 44]. However, that simplification disagrees with the empirical observation that shadowing losses along two links are correlated due to obstructions, e.g., outdoor buildings and terrain variations, and indoor walls and furniture [45, 46, 32].

In order to simultaneously model the correlations in shadow fading that exist across multiple link pairs in a network, we use the network shadowing model [47]. Let $\mathcal{L}$ be a set of link pairs in a wireless network, and $L = |\mathcal{L}|$ where $|\cdot|$ counts the number of elements in the set. We assume that each link is different in either transmitter or receiver location from the other links in the set $\mathcal{L}$. The network shadowing model describes the joint link modeling error as:

$$z = Wp + \eta, \tag{4}$$

where $z = [Z_1, Z_2, \ldots, Z_L]^T \in \mathcal{R}^{L\times 1}$ is the total modeling error after the channel base model, $W \in \mathcal{R}^{L\times M}$ is a weight matrix, $p \in \mathcal{R}^{M\times 1}$ is a discretized loss field in dB, and $\eta \in \mathcal{R}^{L\times 1}$ is the linear model error. Their details are given below.

Spatial loss field p. The spatial loss field of [47, 46] characterizes the environment of interest as a Gaussian random field in dB that is isotropic wide-sense stationary. It has zero mean and an exponentially decaying spatial covariance function:

$$C_p(m,n) \triangleq \frac{\sigma_S^2}{\delta} \exp\left(-\frac{d_{m,n}}{\delta}\right), \tag{5}$$

where $d_{m,n}$ is the Euclidean distance between the centers of pixels m and n, σ_S^2 is the variance of the shadowing loss, and δ is a space constant.The modeling error Z_l on link l is then a weighted sum of the loss field p over the pixels that cross near the link l.

Weight matrix model for W. A weight matrix model formulates a spatial area near the link which has a non-zero contribution to the variation in modeling errors. In this paper, we consider a popular ellipse model [16] and a Cassini oval model [19] for the weight matrix W. The rationale is that these two models can theoretically describe reflection- and scattering-dominant modeling errors, respectively [18].

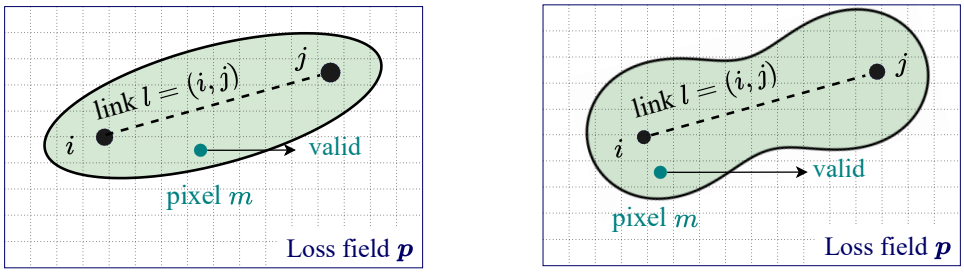
The ellipse model, as shown in Fig. 3a, considers the two ends i and j of link l as the foci and utilizes a tunable parameter λ_{ellipse} to determine the ellipse width. A pixel is viewed as valid if it falls within the ellipse, and the corresponding weight in W will have a non-zero contribution to the shadowing loss of link l. Past studies [47, 46, 16] construct the weight as:

$$w_{lm,\text{ellipse}} = \frac{1}{\sqrt{d_l}} \begin{cases} 1, & \text{if } d_{i,m} + d_{j,m} < d_l + \lambda_{\text{ellipse}} \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where $d_{i,m}$ and $d_{j,m}$ are the distances from the center of pixel m to the two foci i and j, d_l is the link distance, and λ_{ellipse} is the ellipse width parameter.

The Cassini oval model, as shown in Fig. 3b, has the same transmitter and receiver of link l as the foci and uses λ_{Cassini} to decide the oval shape. The weights of the Cassini oval model are constructed as:

$$w_{lm,\text{Cassini}} = \frac{1}{\sqrt{d_l}} \begin{cases} 1, & \text{if } 4d_{i,m}d_{j,m}/d_l^2 < \lambda_{\text{Cassini}} \\ 0, & \text{otherwise} \end{cases}, \tag{7}$$



(a) The ellipse model

(b) The Cassini oval model

Figure 3: The two weight model for selecting pixels (•) that contribute to the shadowing loss of link $l = (i, j)$. Pixels in white have zero weight; pixels in light green have constant positive weight as given in (6) and (7).

The total modeling error z. The total modeling error is the sum of errors from multiple sources: 1) shadow fading, 2) small-scale fading, and 3) measurement error due to thermal noise, which are independent of each other. The distribution for shadow fading, according to the linear additive loss field modeling, is spatially correlated Gaussian in dB. We can also assume the measurement error due to thermal noise to be *i.i.d.* normally distributed in dB scale. For the last small-scale fading, while Rayleigh and Rician are the common models, the sum of all these errors can presumably be Gaussian in dB, according to the Central Limit Theorem [48].

3.3. Loss field learning

Bayesian linear regression. Given the linear joint link model in (4) and the Gaussian loss field prior, we reconstruct the loss field p via Bayesian linear regression. We note the likelihood function of the total modeling error vector is,

$$f(z|W, p, \sigma_\eta^2) = \mathcal{N}(Wp, \sigma_\eta^2 I_L), \tag{8}$$

where $\mathcal{N}(Wp, \sigma_\eta^2 I_L)$ is a Gaussian distribution with a mean of Wp and a covariance of $\sigma_\eta^2 I_L$. Next, the loss field prior is modeled as a Gaussian distribution,

$$f(p) = \mathcal{N}(0, C_p), \tag{9}$$

where C_p is the covariance matrix formed by (5). Therefore the posterior pdf of p is multivariate Gaussian as

$$f(p|z, W, \sigma_\eta^2) = \mathcal{N}(\mu_{p|z}, C_{p|z}), \tag{10}$$

where

$$\begin{aligned} \mu_{p|z} &= \sigma_\eta^{-2}(\sigma_\eta^{-2} W^T W + C_p^{-1})^{-1} W^T z, \\ C_{p|z} &= (\sigma_\eta^{-2} W^T W + C_p^{-1})^{-1}. \end{aligned} \tag{11}$$

As a result, we can acquire the maximum a posterior (MAP) estimator $\hat{p}$ as the posterior mean $\mu_{p|z}$ in (11).

Solution stability. The linear regression problem posed in (4), however, is an ill-posed problem and/or (11) may not exist due to two main factors:

1. $L < M$: there are more pixels to be estimated than link measurements, thus the problem is underdetermined;

2. $L > M$ but with a sparse W: only a few pixels are assigned non-zero weights for each link and thus W is rank-deficient regardless of the number of link samples.

For a stable solution, the regularization constant α is introduced such that the estimator $\hat{p}$ is expressed as:

$$\begin{aligned} \hat{p} &= \Pi_1 z \\ \Pi_1 &= (W^T W + \alpha C_p^{-1})^{-1} W^T, \end{aligned} \tag{12}$$

where σ_η is considered in α. In doing so, the estimator is robust to rank deficiency in the weight matrix, and the inverse term in the operator Π_1 always exists.

Solution efficiency. Latency can be the other concern given large datasets and wide area estimation, and thus requires efficiency improvement. If $L < M$, we can review the problem as sparse linear regression and adopt the minimum norm estimator (MNE) as:

$$\begin{aligned} \hat{p} &= \Pi_2 z \\ \Pi_2 &= C_p W^T (W C_p W^T + \alpha I)^{-1}, \end{aligned} \tag{13}$$

which calculates an inverse of only a smaller matrix of size $\mathcal{R}^{L \times L}$ rather than $\mathcal{R}^{M \times M}$. Note that the inverse of a larger matrix is more computationally complex, i.e., requiring $\mathcal{O}(M^3)$ work [49].

If $L > M$, we leverage the Cholesky decomposition [50] to lower the latency. It is based on the fact that $(W^T W + \alpha C_p^{-1})$ in Π_1 is symmetric and positive definite. Let $A = W^T W + \alpha C_p^{-1}$, and $b = W^T z$. We first calculate the triangular matrix S via the Cholesky factorization:

$$SS^T = A, \quad S = \text{chol}A. \tag{14}$$

By reformulating the problem as $SS^T p = b$, the loss field estimate $\hat{p}$ can be obtained via forward-backward substitution. According to [51], the Cholesky decomposition can be twice as efficient as the general LU decomposition.

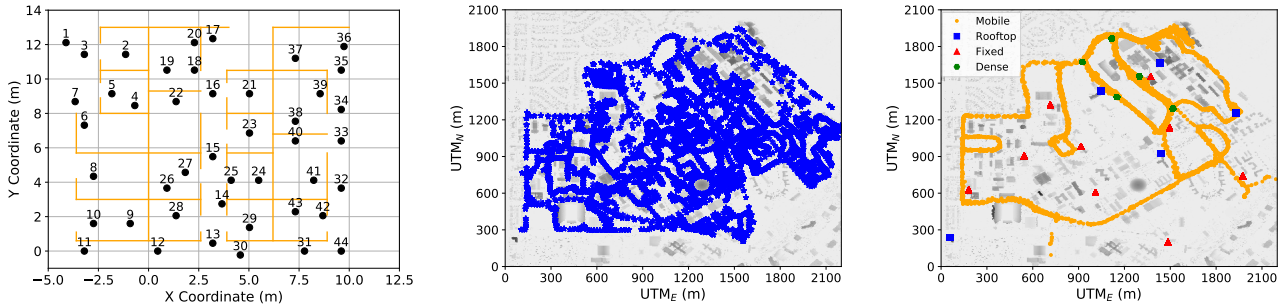
3.4. Explainable learning Method

Explainability is a significant feature of CELF as it provides reasoning for the learning model construction and contextualization for the learning goal, loss field, both in a human-understandable manner. In contrast, ML techniques especially deep learning models are often black-box solutions that can easily suffer from problem diagnosis [52] and adversarial attacks [53]. As a result, CELF as an explainable learning approach can greatly enhance engineers' and regulators' trust in the channel model [54].

CELF embraces explainability in two aspects: model design and model output. For the model design, we first adopt an inherently explainable linear model to describe the additive relation between link modeling errors and the spatial loss field [54]. Second, a Gaussian random field with an exponential kernel as the loss field prior can be justified by the fact that (1) shadow fading is commonly modeled and experimentally verified as Gaussian in dB [9, 24], and (2) shadowing loss due to obstructions across space experiences a distance-based decay [46]. For the model output, we show in Section 5.2 that the spatial loss field can be explained via the physical mechanism of shadowing due to obstacles and can further be numerically validated when the locations of obstructions in the area are known. Such explainability enables a closer tie between a digital spectrum twin and the real environment it is meant to represent [55].

Dataset	Indoor	SLC1	SLC2	ANTW
Freq. (MHz)	2443	462.7	3543	868
Area	17.5×15 (m^2)	2.2×2.1 (km^2)	2.2×2.1 (km^2)	4.1×5.6 (km^2)
Sample Size	9,460	59,323	84,682	162,568
TX Power	10 mW	1 W	1 W	unknown

Table 1: Specifications for the indoor and three outdoor datasets. Note that the outdoor SLC1 dataset, due to uncalibrated receivers, is treated as 4 different subdatasets.



(a) The locations of 44 indoor nodes (b) The SLC1 outdoor transmitter locations (c) The SLC1 outdoor receiver locations

(d) The SLC2 transceiver locations (e) The ANTW transceiver locations

Figure 4: The transmitter and receiver locations of the indoor and outdoor received power datasets. The cubicles (–) in the indoor office and the outdoor environment are also plotted as a reference.

4. Evaluation Methodology

In this section, we describe one indoor and three outdoor real-world received power datasets. The system-related details of each dataset are provided in Table 1. Commonly used methods for channel estimation are presented next including one empirical and three ML-based methods. We describe in the end two evaluation metrics for assessing the performance of the CELF algorithm.

4.1. Real-world received power datasets

Indoor Dataset. This dataset [56] is from in an indoor office area, a 17.5×15 m^2 space surrounded by 1.8 m high cubicle walls, as shown in Fig. 4a. Channels between all pairs of 44 device locations are measured in sequence by transmitting a pseudo-noise code with a 40 MHz chip rate at 2443 MHz. The transmit power is 10 mW. By repeating sample collection for each link 5 times, this indoor dataset has in total $44 \times 43 \times 5 = 9460$ measurements, as described in [57].

Dataset	Receiver	Count	SDR	Antenna Height (m)	Samples
SLC1	Rooftop	5	X310	20–40	13114
	Fixed	12	B210	1.5	24253
	Mobile	7	B210	2.0	8688
	Dense	5	B210	8.8	13268

Table 2: The 4 subdataset under the SLC1 outdoor datasets due to uncalibrated receivers.

SLC1 Outdoor Dataset. The first set of measurements [58] is collected from a 2.2 × 2.1 km^2 university campus area in Salt Lake City (SLC). A portable commercial radio is used as the transmitter, and the receivers are 25 software-defined radio (SDR) nodes with omnidirectional antennas deployed on POWDER, an open wireless experimental testbed [59]. The carrier frequency is 462.7 MHz and the transmit power is 1W. The receivers are one of 4 types, *Rooftop, Fixed, Mobile, and Dense*, according to the radio-antenna-placement differentiation. Table 2 gives specifications for each receiver type. Fig. 4b and 4c show the GPS coordinates of the transmitter and all the receivers on the campus map. As the four types of receivers are heterogeneous and uncalibrated, this work treats the data collected by each type as a separate dataset.

SLC2 Outdoor Dataset. The second dataset [60] is collected on the same 2.2 × 2.1 km^2 University of Utah campus in SLC. However, it differs from the SLC1 outdoor dataset in that the center frequency is 3534 MHz which is in the CBRS band for shared spectrum use. Including this dataset helps evaluate how precisely and efficiently CELF can perform in the real-world dynamic spectrum sharing scenario. According to [60], the SLC2 outdoor dataset uses 5 *Dense* nodes on POWDER to transmit a continuous wave (CW) signal at 1W transmit power while a portable SDR receiver is carried by walk and driving for sample collection. The transmitter and receiver locations are shown in Fig. 4d.

ANTW Outdoor Dataset. The last outdoor dataset [61] is the largest in terms of data size and coverage area. The measurements are taken in the city center of Antwerp (ANTW), Belgium by stationary cell towers. The mobile transmitters are carried by Antwerp's postal service vehicles while transmitting LoRaWAN messages at 868 MHz. The intention of adding this dataset is to evaluate CELF in a large urban area. Fig. 4e presents the 11 stationary receivers and the GPS coordinates of the mobile transmitter.

Train-Test Split. Each dataset needs to be split without overlapping for loss field estimation (training) and shadowing loss prediction (testing) purposes. We choose the link index as the criterion to partition the datasets. Each dataset is split with a 7:3 ratio. Each data point is randomly assigned for training or testing.

4.2. Methods for comparison

We adopt three general-purpose ML models, *Random Forest, SVR, and MLP-ANN*, in this work for performance comparisons. The rationale behind such choices is that they require neither site-specific terrain information nor large-scale datasets, unlike complex deep learning models such as RadioUNet [12] and PL-GAN [62]. They have also been widely used as benchmarks for path loss prediction [24, 63, 37, 38].

- *Random Forest* [64]: it is an ensemble learning approach that first constructs multiple decision trees on random subsets of the dataset and then combines them to improve the accuracy and robustness of the model.

- *SVR* [65]: it is a variation of support vector machines used for regression. Unlike traditional squared error minimization, SVR fits a curve by maximizing the margin between different kernels.

- *MLP-ANN* [38]: MLP-ANN is a feedforward neural network that consists of an input layer, an output layer, and multiple hidden layers. It is trained iteratively using algorithms like stochastic gradient descent for squared error minimization.

4.3. Evaluation metrics

We adopt two evaluation metrics, *variance reduction* and *running time*, to quantify the performance of the tested algorithms. To specify, *variance reduction* is defined as the percentage decrease of the modeling error variance, i.e.,

$$\epsilon = \frac{\sigma^2_{z_\mathcal{T}} - \sigma^2_{\text{CELF}}}{\sigma^2_{z_\mathcal{T}}} \cdot 100\%, \tag{15}$$

where $\sigma^2_{z_\mathcal{T}}$ is the modeling error variance of a dataset $\mathcal{T}$ after the first-order channel estimation in Section 3.1, and σ^2_{CELF} is the final variance after applying the learned loss field via CELF for shadowing prediction, which is computed as the mean-squared error (MSE):

$$\sigma^2_{\text{CELF}} = \frac{\|z_\mathcal{T} - W_\mathcal{T} \cdot \hat{p}\|^2}{N_\mathcal{T}}, \tag{16}$$

where $\hat{p}$ is the attenuation image learned from Section 3.3, $W_\mathcal{T}$ is the weight matrix model, and $N_\mathcal{T} = |\mathcal{T}|$ is the size of the dataset $\mathcal{T}$.

Note that the variance reduction metric in (15) is considered in this work a measure of accuracy although, by definition, it describes precision. The rationale is that: 1) the ground truth for the loss field p is unavailable so we cannot directly quantify its accuracy, and 2) the datasets for training and testing are assumed to be sampled from identical distributions, and thus no sample bias is involved.

The other metric, *running time*, is a measure of the computational efficiency of the proposed CELF algorithm. It has been crucial in time-sensitive applications such as real-time spectrum access and management systems [66]. This metric includes the execution time for loss field learning and shadowing loss prediction. Note that the terms "learning" and "training", "prediction" and "testing" are used interchangeably for comparing CELF to the selected approaches in Section 5.

We take the following three steps to ensure result comparability. First, all the models are trained and tested on the same partitioned datasets. Second, the inputs of these ML models are the 2D coordinates of transmitters and receivers to be consistent with CELF. Lastly, all the results are obtained by running the algorithm on the same Linux system with a 16-core Intel Xeon Gold 6130 processor.

Hyperparameter	Description	Indoor	*Rooftop*
δ_p	Pixel width (m)	0.35	25
σ_S^2/σ_Z^2	Shadowing variance ratio	0.30	0.58
δ	Space constant (m)	2.5	35
λ_{ellipse}	Excess length (m)	0.18	105
α	Regularization	41	0.3

Table 3: Model hyperparameters of CELF for the indoor and SLC1-*Rooftop* dataset.

5. Results

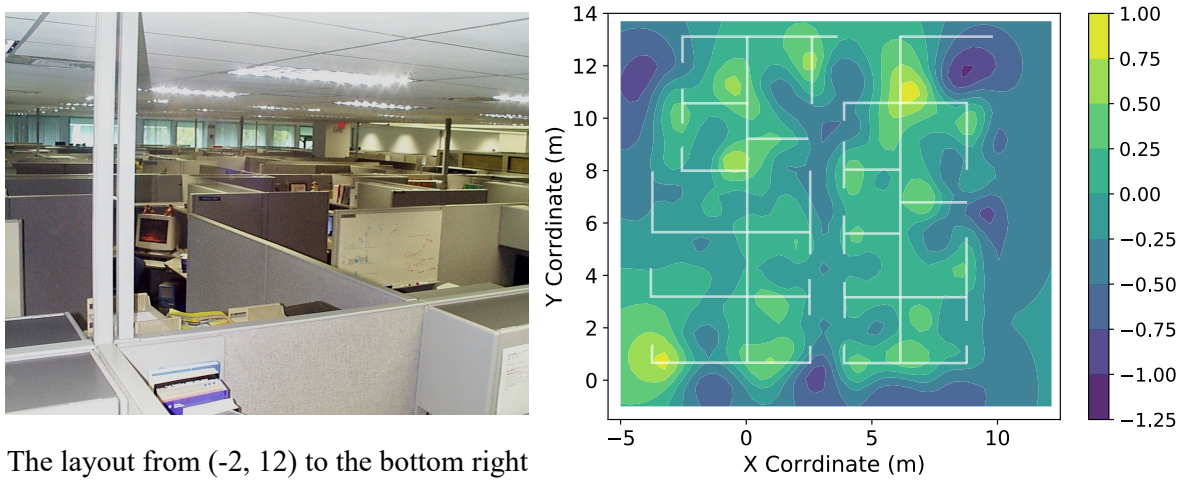
Experimental results of the proposed CELF algorithm are given in this section. We first present three loss field image examples which are learned from the indoor and SLC1 outdoor datasets in Section 4.1. We then compare CELF with the chosen methods via *variance reduction* and *running time* from Section 4.3. We further explore CELF's accuracy improvement and robustness when TIREM is used as the channel base model. A discussion of the ellipse and Cassini oval weight matrix models is presented next. The impact of the hyperparameters on accuracy is discussed in the end.

5.1. Example loss field images

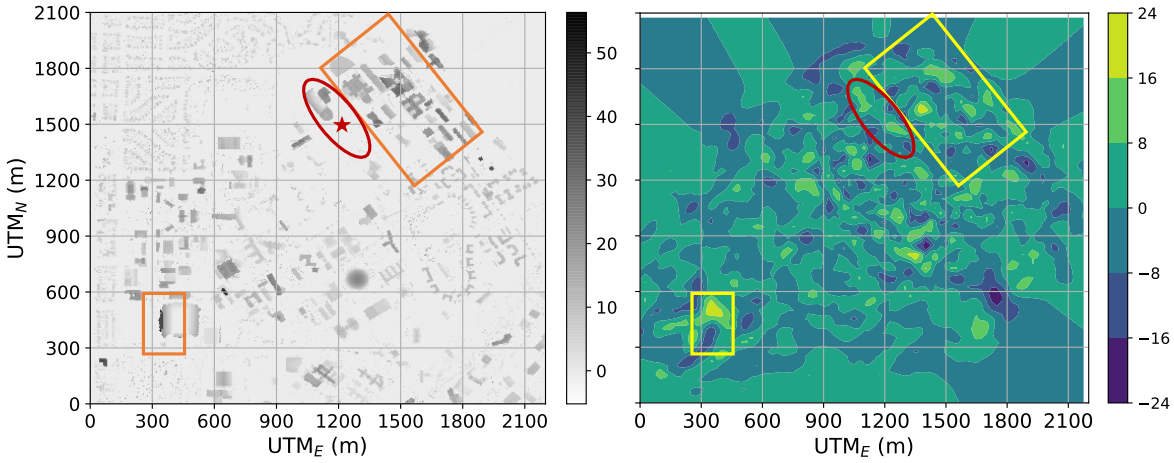
This subsection presents two example loss field images using the *log-distance path loss* model as the channel base model in Section 3.1 and the *ellipse* weight matrix model in Section 3.2. The rationale behind the *Rooftop* dataset choice is that these receivers, as deployed high above the ground, give better coverage of the campus area. We select both indoor and outdoor datasets to discuss CELF's practical use in various types of environments. The image boundaries are the same as Fig. 4a and Fig. 4b.

The statistical analysis follows the next four steps. First, we determine the path loss exponent n_p and the reference loss $P_T - \Pi_0$ in (1) via linear regression. The reference distance Δ_0 is set to be 1 m across the datasets. The results of the two examples are (1) indoor: $n_p = 2.26$ and $P_T - \Pi_0 = -37.04$ dBm, and (2) SLC1-*Rooftop*: $n_p = 2.73$ and $P_T - \Pi_0 = -1.25$ dBm.

Second, we tune hyperparameters for CELF and interpret their values. The model hyperparameters are selected via 5-fold cross-validation. This procedure is to randomly sample 1/5 data out of the training dataset for hyperparameter validation and overfitting prevention. Their descriptions and values are given in Table 3. The first hyperparameter, δ_p, denotes the attenuation image resolution and impacts both computation time and prediction accuracy. The second shadowing variance ratio, σ_S^2/σ_Z^2, represents the contribution of shadowing loss to the total modeling error. In comparison to outdoor environments, indoor surroundings have more multipath components as indoor obstacles that obstruct radio wave propagation are relatively uniformly placed throughout the area. Therefore indoor environments have more significant small-scale fading [46]. The third space constant δ indicates the obstruction size in the environment [46]. We expect that obstacles will be smaller for the indoor area. In this case, the δ for the SLC1-*Rooftop* dataset is 35m, larger



(a) The indoor training dataset. Correlation can be seen between the cubic wall locations and high spatial losses.



(b) The SLC1-*Rooftop* training dataset. The rectangles mark high obstructions and high spatial losses. The red oval highlights a missing new building at the star ($\star$) location which is correctly captured by CELF.

Figure 5: Example loss images learned via the proposed CELF algorithm and the site maps as a reference.

than 2.5m for the indoor dataset. The next hyperparameter λ is introduced by the ellipse weight model to select valid pixels for each link. It is determined by the area size and the pixel width. The last hyperparameter α balances the loss field prior and the data from the area of interest. We notice that α of the indoor dataset is about 100 times larger than that of the outdoor case. This can be explained by the $1/\sqrt{d_l}$ weight in (6). The path lengths d_l of the indoor measurements are 100 times smaller, which makes α 100 times larger to balance the $1/d_l$ discrepancy in (12).

Next, we derive the weight matrix and estimate the loss image via Bayesian linear regression. Fig. 5 demonstrates the two trained loss images and the site maps as a reference. It can be observed that they have spatial loss ranges of -1.25 – 1.00 dB and -24 – 24 dB, respectively. Higher losses can be seen at higher obstructions such as near cubicle walls in Fig. 5a and the marked rectangle areas in Fig. 5b.

The red oval area of Fig. 5b highlights a mismatch between the estimated two high-loss regions and one high obstruction of the site map. The loss image estimate is in fact more accurate because

the terrain profile is outdated; a new building recently constructed at the star ($\star$) location was not in the database used to generate the left image in Fig. 5b. Note that CELF does not use any terrain or building information. Collecting and maintaining the site-specific terrain dataset could be time-consuming and expensive, but CELF can use channel loss measurements for accurate and cost-effective loss field estimation.

The last step for loss field learning is to quantitatively assess the training accuracy of the learned loss image via variance reduction. For the *Rooftop* training dataset, the modeling error variance after the *log-distance path loss model* is 58.4 dB^2. The shadowing loss estimates via CELF decreases the channel loss variance to 30.7 dB^2 which is 47.4% less than that of the base model. For the indoor training dataset, the modeling error variance reduces from 19.8 dB^2 to 10.1 dB^2, which corresponds to a 49.3% reduction.

5.2. Loss fields' explainability

The estimated images are called shadowing loss fields as they can be explained as site maps of additional radio shadowing due to obstacles in the environment. As an example, consider the loss image of the indoor office in Fig. 5a. We can see from the photo of the area that desks, computers, and bookcases are generally positioned close to the cubicle walls, and our site data includes the wall locations. We can see in Fig. 5a that the estimated loss field is lower in the middle of each cubicle and higher close to the cubicle walls. Similarly, the vertical corridor region at $x \approx 3.2$ m experiences lower loss than inside the cubicles. It is intuitively clear that the shadowing loss field can be explained by the locations of the obstructions in the area.

We further quantify this argument by showing that the increase in the loss field near known obstructions is, in fact, statistically significant, in both our indoor and outdoor environments. We apply the two-sample t-test to verify numerically that the loss field is higher near known obstructions. Let the shadowing loss field near the known obstructions be p_O and the field values at the remaining positions be p_R. Our hypotheses to test are: $H_0 : \mu_{p_O} = \mu_{p_R}$ and $H_1 : \mu_{p_O} > \mu_{p_R}$, where μ_{p_O} and μ_{p_R} are the sample mean of each population. To identify obstructions for the SLC1-*Rooftop* building map, we use any location where the building height is over 30m, which is the average antenna height of the *Rooftop* receivers. For the indoor set, we use the cubicle walls as the obstruction locations. Note that the obstruction locations are not only the exact coordinates, but also any neighboring pixel within 3% of the field width. This distance accounts for any potential error between our "ground truth" obstruction location and the location of high attenuation in the loss field estimate. This 3% of the field width translates to 1.4 and 2.4 times the pixel width for the indoor and SLC1-*Rooftop* cases, respectively. The two-sample t-test for the loss field estimates shown in Fig. 5a and Fig. 5b result in p-values of 8×10^{-133} and 3×10^{-3}, respectively. Therefore, we reject H_0 in both cases at a significance level of 0.01. We can further conclude that loss field values are in fact statistically higher near actual attenuating obstructions in the environment. This validates that the loss field via the proposed CELF is explainable in part by the locations of significant obstructions in the area, providing a key feature to its users.

5.3. Accuracy analysis

Upon obtaining the loss field, we evaluate CELF's performance on the test datasets. The first is accuracy analysis using the variance reduction metric. Fig. 6 demonstrates the modeling error variance reductions on the indoor and outdoor test datasets using three ML methods and CELF. It can be seen from Fig. 6 that all ML methods can lower the modeling error variance

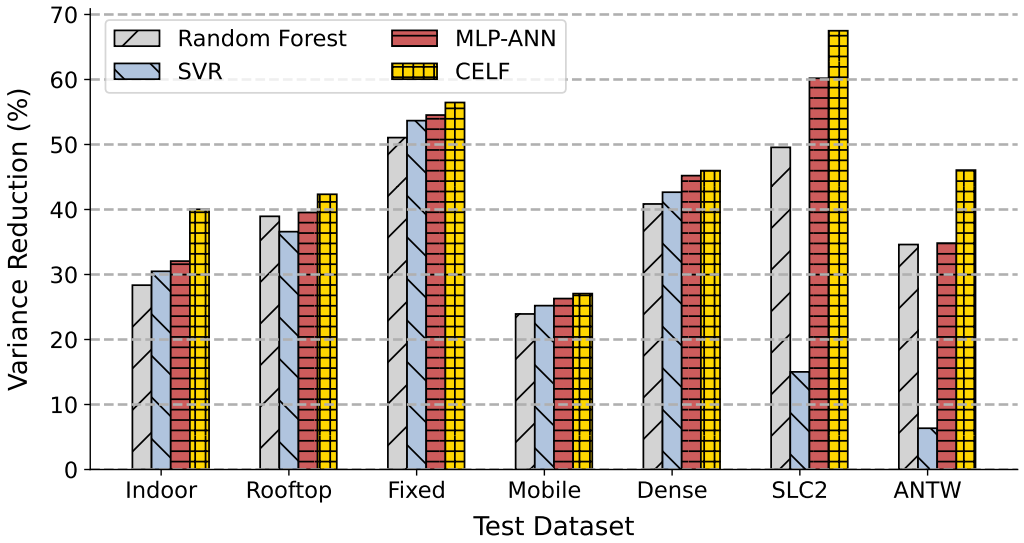


Figure 6: Modeling error variance reductions on the indoor and outdoor test datasets via three ML methods and CELF.

SLC1	Data Variance (dB^2)			Maximum
Dataset	Stationary	Radius$\leq 1\lambda_f$	Sum	Reduction
Rooftop	3.9	20.3	24.3	58.4%
Fixed	2.9	11.0	13.9	76.9%
Dense	2.6	8.2	10.8	58.8%

Table 4: Data variance when the portable transmitter is stationary or rotating with a radius $\leq 1\lambda_f$ and the maximum variance reduction space for CELF. The sum of the data variance approximates the lowest possible modeling error variance.

to a certain degree. MLP-ANN gives the largest variance reduction among the three ML-based methods. However, CELF outperforms all the ML models across the test datasets. Take the indoor dataset for instance. CELF can achieve 40.0% variance reduction which is higher than MLP-ANN's 32.1%. It can be further seen that CELF reduces the most variance of the SLC2 outdoor dataset, which validates CELF for channel modeling in the real-world spectrum sharing scenario. To summarize, we are able to show that the CELF algorithm outperforms the three ML methods in terms of variance reduction.

To further understand the lowest possible modeling error variance that CELF can reduce to, we analyze a subset of the SLC1 outdoor dataset which is collected when a portable transmitter is either stationary or rotating with a radius less than or equal to 1 wavelength (λ_f). The subset has 14,026 received power observations. Variation in stationary data approximates the measurement noise variance, and the data for link distances changing on the order of the signal wavelength can estimate the small-scale fading loss [24]. Hence the sum of the two is an estimate of the lowest possible variance of the modeling errors σ_Z^2. Table 4 illustrates the variance of the two measurement sets. Note that the SLC1-*Mobile* dataset is not applicable as the receivers are constantly moving.

Dataset	Training Time (s)				Testing Time (μs/link)			
	Random Forest	SVR	MLP-ANN	CELF	Random Forest	SVR	MLP-ANN	CELF
Indoor	0.1	0.0	2.4	0.1	25	21	7	18
SLC1-*Rooftop*	1.2	7.4	26.8	8.2	5	184	3	102
SLC1-*Fixed*	1.6	28.7	59.7	15.8	4	323	3	88
SLC1-*Mobile*	1.6	4.2	13.6	4.8	5	119	2	93
SLC1-*Dense*	0.7	7.1	25.5	6.0	4	190	1	90
SLC2	7.0	21.8	72.5	10.8	4	136	1	46
ANTW	13.6	41.5	177.0	64.6	4	139	1	93

Table 5: Running time comparison for training and testing among the ML models, and the CELF algorithm. We use the training time for the entire training set to analyze CELF's loss field learning efficiency. Instead, the testing time in μs/link is used. It is to evaluate how fast CELF can predict channel loss for one unseen arbitrary link.

We can learn that for the SLC1-*Dense* dataset, the variance reduction upper limit is 58.8% which, based on Fig. 6, is 12.9% higher than the result of CELF. By comparing Table 4 and Fig. 6, we can conclude that there is still room to lower the shadowing loss variance, but the proposed CELF method has shown results closer to the limits.

5.4. Efficiency analysis

We compare the training and testing efficiency of the methods via running time. Their results are shown in Table 5. First, MLP-ANN, among the remaining methods, is the most efficient for shadowing loss prediction but the most computationally expensive for training. Second, the slowest model for testing is SVR except for the indoor dataset. Last, CELF is approximately 3 times faster than MLP-ANN for shadowing loss field learning. As a result, it can update the model with new measurements or learn the spatial loss of a new environment with much less computational cost.

Contrary to the training time which is for the entire training set, the prediction time is given in μs/link. This measure can directly describe, given the learned shadowing loss field, how efficiently CELF predicts channel loss for a single unseen link. We can see from Table 5 that CELF only needs on average 76 μs for one link prediction. However, it is slower than MLP-ANN across all the datasets. This is due to the time-expensive weight matrix computation for each link. Optimization of the weight model is needed for prediction efficiency improvement and remains future work.

5.5. TIREM enhancement

We use *TIREM*, a hybrid empirical/physical model, as another channel base model to discuss whether CELF can robustly enhance the channel estimation performance of a base model that requires no measurements from the area of deployment. A known problem of TIREM is its over-prediction of shadowing loss as it does not consider signal reflection and diffraction around obstacles [29]. In this section, we train a newly learned loss field from TIREM's modeling errors and present CELF's accuracy improvement.

Fig. 7 shows an example of TIREM's channel loss estimation map. A transmitter located at (1048, 1435) emits signals through the University of Utah campus. The value at each pixel presents the loss estimate given by TIREM. We can see from Fig. 7 several radiating rays outward

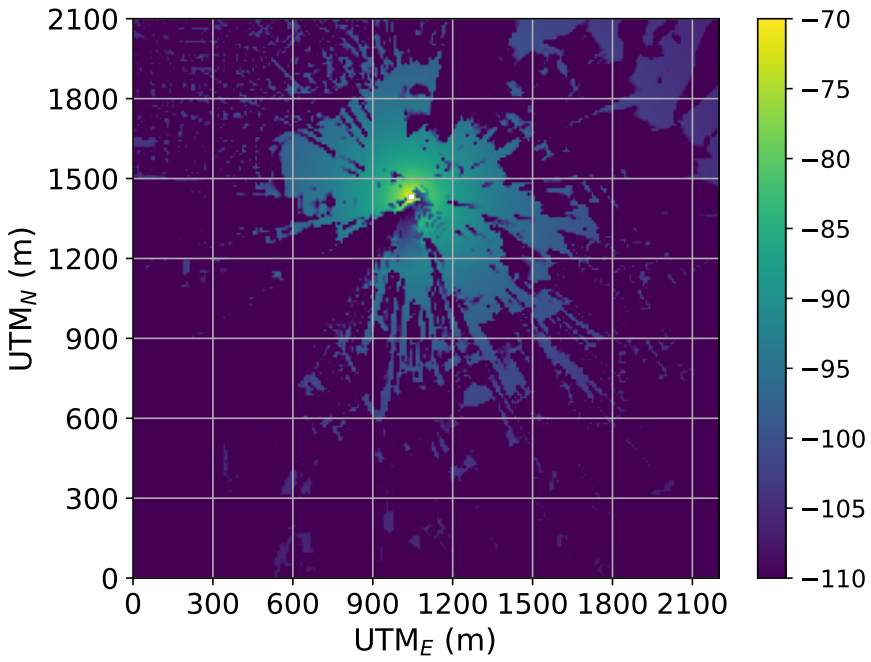


Figure 7: An example of the channel loss estimation map via TIREM for the University of Utah campus.

from the transmitter due to the LOS paths at different angles. However, regions at a -270 degree angle relative to the transmitter, e.g., at (900, 1200), show a sharp decline in channel loss. This corresponds to the lack of considering signal reflection and diffraction.

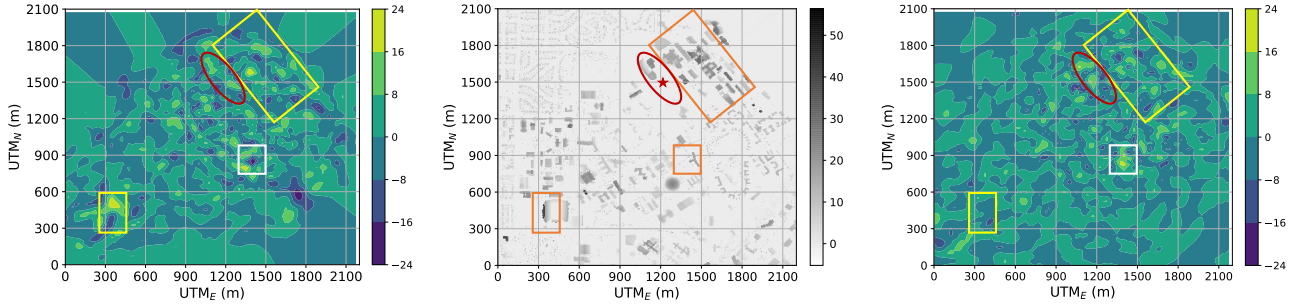
Table 6 shows the data size, the modeling error variances of the two base models, and the variance reduction results, all on the outdoor test datasets. The indoor dataset is not included as the terrain features are unavailable for TIREM. First, we can see higher variances using the TIREM base model. This is because the terrain profile available for TIREM is outdated and TIREM does not rely on the measurements from the area of interest. Therefore, TIREM shows larger variations in the modeling errors than the fitting-based log-distance path loss model. Second, CELF can decrease the variances across all the datasets, up to 67.5% for the log-distance path loss model and up to 63.1% for TIREM. Third, we observe smaller variance reductions using TIREM as the base model for all test datasets except ANTW. It can be that, unlike the log-distance path loss model, TIREM does characterize shadowing loss but has the over-prediction problem. As a result, the modeling errors in TIREM contain less shadowing loss and are not as highly spatially correlated. In summary, it verifies that CELF can robustly improve different base models using the explainable shadowing loss field.

5.6. The Cassini oval weight matrix models

Different weight matrix models formulate distinct spatial patterns that contribute to the variations in channel losses due to multipath propagation. That is, obstacles in an environment, e.g., outdoor buildings and indoor walls, can reflect, diffract, or scatter a transmitted signal wave, which causes the channel loss to be dependent on spatial locations in the field [24]. Here we compare the Cassini oval weight model to the ellipse and discuss their impact on CELF's accuracy improvement across datasets.

Table 6: The modeling error variances of TIREM and the log-distance path loss as the base models and variance reductions on outdoor test datasets via CELF.

Dataset	Size	Base Model: Log-distance Path Loss		Base Model: TIREM	
		Modeling Error Var (dB^2)	CELF Var/ Reduction (dB^2/%)	Modeling Error Var (dB^2)	CELF Var/ Reduction (dB^2/%)
SLC1-*Rooftop*	3,935	58.6	33.8 (42.3%)	176.2	118.0 (33.1%)
SLC1-*Fixed*	7,276	60.0	26.1 (56.5%)	209.4	129.6 (38.1)%
SLC1-*Mobile*	2,607	40.1	29.3 (27.1%)	175.6	135.2 (23.0%)
SLC1-*Dense*	3,981	26.0	14.0 (46.0%)	163.7	108.9 (33.5%)
SLC2	25,405	67.6	22.0 (67.5%)	166.8	65.6 (60.7%)
ANTW	48,771	34.3	18.5 (46.1%)	237.7	87.8 (63.1%)



(a) The loss field via the ellipse model. (b) The obstruction map as a reference. (c) The loss field via the Cassini oval model.

Figure 8: The learned loss fields on the SLC1-*Rooftop* dataset using the Cassini oval and the ellipse weight models. The site map is added as a reference but not used by CELF.

We first present the learned loss fields using the SLC1-*Rooftop* dataset using the Cassini oval weight model. As seen in Fig. 8, both loss fields share the same spatial loss range of -24 – 24 dB. Comparing the marked rectangle areas between the obstruction map in Fig. 8b and the loss field by the Cassini oval in Fig. 8c, we can still see the correlation, i.e., higher losses at higher buildings.

A new white rectangle region in Fig. 8a and Fig. 8c highlights a difference in the learned spatial loss given by the weight models. It can be clearly seen that the loss field via the Cassini oval has a higher correlation with the reference map, which demonstrates the potential of using Cassini ovals as the weight matrix model.

We next describe the accuracy improvement results on the test datasets in Fig. 9. It can be seen that CELF with the Cassini oval model performs better on the SLC1-*Rooftop*, SLC1-*Mobile*, SLC1-*Dense*, and ANTW test datasets. Counterintuitively, CELF with the Cassini oval reduces more variance in the SCL1-*Dense* dataset but less in the SLC2 dataset, even though both are collected using the same POWDER nodes in the same geographic area. This could be due to the coverage difference between the datasets or different carrier frequencies. Despite the difference, CELF with the two weight models result in similar variance reduction. To summarize, we find no systematic improvement in the Cassini oval model compared to the ellipse model. However, there may be environments and deployment types in which the Cassini oval is a better fit. Regardless,

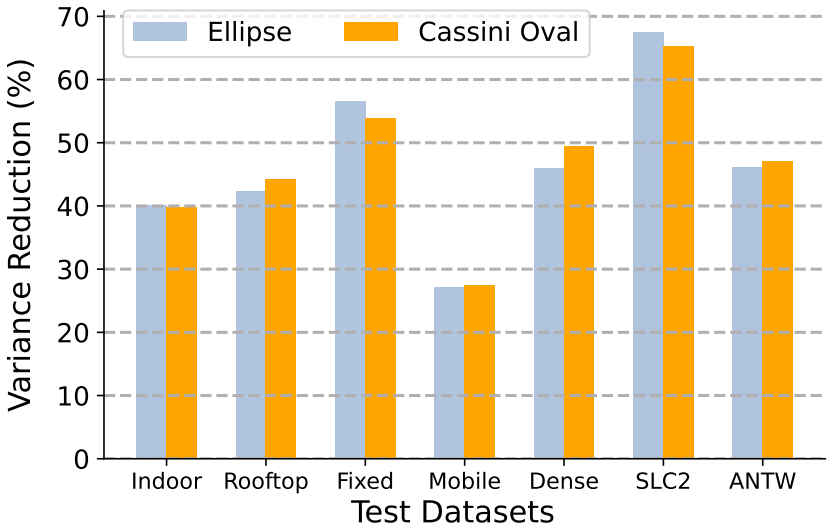
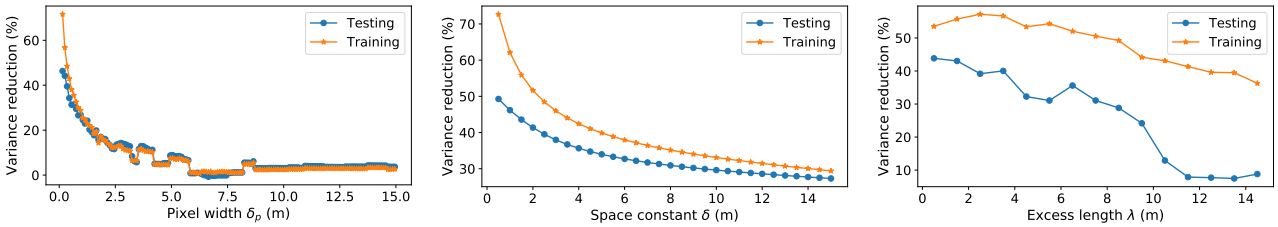


Figure 9: Comparison of CELF's weight models via variance reductions on the indoor and outdoor test datasets.



(a) The effect of the pixel width δ_p. (b) The effect of the space constant δ. (c) The effect of the excess length λ_{ellipse}.

Figure 10: Variance reduction vs. CELF's hyperparameters on the indoor dataset.

we can report that both the ellipse and Cassini oval models are valid for the weight matrix and can lower the test variance along with CELF.

5.7. Effect of hyperparameters

CELF's hyperparameters play a significant role in its performance. We here present variance reduction as a function of CELF's three major hyperparameters on the indoor dataset.

Fig. 10a shows that the variances for both training and test datasets mostly reduce less as the pixel width δ_p increases from 0.15m to 15m. Fluctuations occur at near 2.5 m, 3.5 m, 5 m, and 8 m. While the lower the pixel width, the higher the variance reduction, it comes with training time sacrifice. Fig. 10b discusses the reduction variation vs. the space constant δ. Reductions for testing and training decrease as δ increase from 0.5 m to 15 m. As δ approximates the obstruction size, unreasonablely large space constants give lower variance reduction for training and testing. Fig. 10c presents the effect of the excess length λ_{ellipse} on variance reduction. It can be seen that too large of the excess length includes too many pixels for loss field estimation and thus leads to lower variance reduction.

6. Conclusion

This paper proposes CELF, which learns a spatial loss field and uses it to predict shadowing loss on any new links in a deployment area. It formulates total fading loss via a discretized linear model and applies Bayesian linear regression and optimization for the loss image estimation. CELF provides an explainable learning approach for fast and precise site-specific channel loss estimation.

The proposed method has been validated with two evaluation metrics, variance reduction and running time, for training and prediction. It is tested on one indoor and three outdoor real-world datasets. Three ML-based methods, SVR, random forest, and MLP-ANN, are used for performance comparison. Experimental results demonstrate that CELF presents larger variance reductions than all the other methods and can also estimate the loss field more efficiently than the most accurate MLP-ANN model. Numerical results via two-sample t-tests show a statistically higher loss field at locations near known attenuating obstructions in the area, which validates the explainability feature of CELF.

CELF is further tested on a different channel base model, TIREM, and a different weight matrix model, Cassini oval, for comprehensive discussion. Numerical results show that, with TIREM, CELF can robustly reduce the test modeling error variance by up to 63%; and with the Cassini oval, CELF shows similar accuracy improvement as the ellipse weight model.

References

[1] W. S. H. M. W. Ahmad, N. A. M. Radzi, F. Samidi, A. Ismail, F. Abdullah, M. Z. Jamaludin, M. Zakaria, 5G technology: Towards dynamic spectrum sharing using cognitive radio networks, IEEE Access 8 (2020) 14460–14488.

[2] S. Bhattarai, J.-M. J. Park, B. Gao, K. Bian, W. Lehr, An overview of dynamic spectrum sharing: Ongoing initiatives, challenges, and a roadmap for future research, IEEE Transactions on Cognitive Communications and Networking 2 (2) (2016) 110–128.

[3] M. M. Sohul, M. Yao, T. Yang, J. H. Reed, Spectrum access system for the citizen broadband radio service, IEEE Communications Magazine 53 (7) (2015) 18–25.

[4] T. Kidd, National radio quiet and dynamic zones, CHIPS–The Department of Navy's Information Technology Magazine (2018).

[5] M. Zheleva, C. R. Anderson, M. Aksoy, J. T. Johnson, H. Affinnih, C. G. DePree, Radio dynamic zones: Motivations, challenges, and opportunities to catalyze spectrum coexistence, IEEE Communications Magazine 61 (6) (2023) 156–162. doi:10.1109/MCOM.005.2200389.

[6] D. Eppink, W. Kuebler, TIREM/SEM handbook, Defense Technical Information Center (1994).

[7] Z. Yun, M. F. Iskander, Ray tracing for radio propagation modeling: Principles and applications, IEEE Access 3 (2015) 1089–1100.

[8] M. Hata, Empirical formula for propagation loss in land mobile radio services, IEEE Transactions on Vehicular Technology 29 (3) (1980) 317–325.

[9] T. S. Rappaport, Wireless communications: Principles and practice, 2/E, Pearson, 2010.

[10] A. Clegg, Google and CBRS, https://www.youtube.com/watch?v=2vxjFDieP1Q, Speech at NEDAS 2019 NYC Summit (Sept. 2019).

[11] A. Seretis, C. D. Sarris, An overview of machine learning techniques for radiowave propagation modeling, IEEE Transactions on Antennas and Propagation 70 (6) (2021) 3970–3985.

[12] R. Levie, Ç. Yapar, G. Kutyniok, G. Caire, RadioUNet: Fast radio map estimation with convolutional neural networks, IEEE Transactions on Wireless Communications 20 (6) (2021) 4001–4015.

[13] J. Gerlings, A. Shollo, I. Constantiou, Reviewing the need for explainable artificial intelligence (xAI), Proceedings of the Hawaii Intl. Conf. System Sciences (HICSS) 54ArXiv preprint arXiv:2012.01007 (2020).

[14] A. Holzinger, C. Biemann, C. S. Pattichis, D. B. Kell, What do we need to build explainable AI systems for the medical domain?, arXiv preprint arXiv:1712.09923 (2017).

[15] C. R. Karanam, Y. Mostofi, 3D through-wall imaging with unmanned aerial vehicles using WiFi, in: Proceedings of the 16th ACM/IEEE Intl. Conf. Information Processing in Sensor Networks, 2017, pp. 131–142.

[16] J. Wilson, N. Patwari, Radio tomographic imaging with wireless networks, IEEE Transactions on Mobile Computing 9 (5) (2010) 621–632.

[17] J. Wang, N. Patwari, Git repository of the proposed channel estimation via loss field (CELF) model, https://gitlab.flux.utah.edu/Jie_Wang/channel-estimation-via-loss-field (2024).

[18] N. Patwari, J. Wilson, RF sensor networks for device-free localization: Measurements, models, and algorithms, Proceedings of the IEEE 98 (11) (2010) 1961–1973.

[19] E. Paolini, A. Giorgetti, M. Chiani, R. Minutolo, M. Montanari, Localization capability of cooperative anti-intruder radar systems, EURASIP Journal on Advances in Signal Processing 2008 (2008) 1–14.

[20] O. Norklit, J. Andersen, Diffuse channel model and experimental results for array antennas in mobile environments, IEEE Transactions on Antennas and Propagation 46 (6) (1998) 834–840. doi:10.1109/8.686770.

[21] H. Hashemi, The indoor radio propagation channel, Proceedings of the IEEE 81 (7) (1993) 943–968.

[22] J. Wang, M. G. Weldegebriel, N. Patwari, Channel estimation via loss field: Accurate site-trained modeling for shadowing prediction, in: 2024 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), 2024, pp. 312–321. doi:10.1109/DySPAN60163.2024.10632818.

[23] C. Phillips, D. Sicker, D. Grunwald, A survey of wireless path loss prediction and coverage mapping methods, IEEE Communications Surveys & Tutorials 15 (1) (2012) 255–270.

[24] A. Goldsmith, Wireless communications, Cambridge Univ. Press, 2005.

[25] A. Hrovat, G. Kandus, T. Javornik, A survey of radio propagation modeling for tunnels, IEEE Communications Surveys & Tutorials 16 (2) (2013) 658–669.

[26] J. S. Seybold, Introduction to RF propagation, John Wiley & Sons, 2005.

[27] G. A. Hufford, A. G. Longley, W. A. Kissick, et al., A guide to the use of the ITS irregular terrain model in the area prediction mode, US Department of Commerce, National Telecommunications and Information Administration, 1982.

[28] M. R. Souryal, T. T. Nguyen, Effect of federal incumbent activity on CBRS commercial service, in: 2019 IEEE Intl. Symposium on Dynamic Spectrum Access Networks (DySPAN), 2019, pp. 1–5.

[29] M. A. Varner, F. Mitchell, J. Wang, K. Webb, G. D. Durgin, Enhanced RF modeling accuracy using simple minimum mean-squared error correction factors, in: 2022 IEEE 2nd Intl. Conf. Digital Twins and Parallel Intelligence (DTPI), 2022, pp. 1–5.

[30] P. Series, Recommendation ITU-R P.1812-6: A path-specific propagation prediction method for point-to-area terrestrial services in the frequency range 30 MHz to 6000 MHz, International Telecommunication Union, 2021.

[31] S. S. Szyszkowicz, H. Yanikomeroglu, J. S. Thompson, On the feasibility of wireless shadowing correlation models, IEEE Transactions on Vehicular Technology 59 (9) (2010) 4222–4236.

[32] J. Lee, F. Baccelli, On the effect of shadowing correlation on wireless network performance, in: IEEE INFOCOM, 2018, pp. 1601–1609.

[33] N. Patwari, P. Agrawal, Nesh: A joint shadowing model for links in a multi-hop network, in: 2008 IEEE Intl. Conf. Acoustics, Speech and Signal Processing, 2008, pp. 2873–2876.

[34] A. Abdi, M. Kaveh, On the utility of gamma PDF in modeling shadow fading (slow fading), in: 1999 IEEE 49th Vehicular Technology Conference, Vol. 3, 1999, pp. 2308–2312.

[35] P. Ramírez-Espinosa, F. J. Lopez-Martinez, Composite fading models based on inverse gamma shadowing: Theory and validation, IEEE Transactions on Wireless Communications 20 (8) (2021) 5034–5045.

[36] H.-S. Jo, C. Park, E. Lee, H. K. Choi, J. Park, Path loss prediction based on machine learning techniques: Principal component analysis, artificial neural network, and Gaussian process, Sensors 20 (7) (2020) 1927.

[37] Y. Zhang, J. Wen, G. Yang, Z. He, J. Wang, Path loss prediction based on machine learning: Principle, method, and data expansion, Applied Sciences 9 (9) (2019) 1908.

[38] L. Wu, D. He, B. Ai, J. Wang, H. Qi, K. Guan, Z. Zhong, Artificial neural network based path loss prediction for wireless communication network, IEEE Access 8 (2020) 199523–199538.

[39] E. Ostlin, H.-J. Zepernick, H. Suzuki, Macrocell path-loss prediction using artificial neural networks, IEEE Transactions on Vehicular Technology 59 (6) (2010) 2735–2747.

[40] S. Ojo, A. Imoize, D. Alienyi, Radial basis function neural network path loss prediction model for LTE networks in multitransmitter signal propagation environments, Intl. Journal of Communication Systems 34 (3) (2021) e4680.

[41] J. Thrane, D. Zibar, H. L. Christiansen, Model-aided deep learning method for path loss prediction in mobile communication systems at 2.6 GHz, IEEE Access 8 (2020) 7925–7936.

[42] C. Bettstetter, C. Hartmann, Connectivity of wireless multihop networks in a shadow fading environment, in: Proc. 6th ACM Workshop on Modeling Analysis & Simulation of Wireless & Mobile Systems, 2003, pp. 28–32.

[43] R. Hekmat, P. Van Mieghem, Connectivity in wireless ad-hoc networks with a log-normal radio model, Mobile Networks & Applications 11 (2006) 351–360.

[44] Y. Chen, A. Terzis, On the implications of the log-normal path loss model: an efficient method to deploy and move sensor motes, in: ACM SenSys, 2011, pp. 26–39.

[45] M. Gudmundson, Correlation model for shadow fading in mobile radio systems, Electronics letters 23 (27) (1991) 2145–2146.

[46] P. Agrawal, N. Patwari, Correlated link shadow fading in multi-hop wireless networks, IEEE Transactions on Wireless Communications 8 (8) (2009) 4024–4036.

[47] N. Patwari, P. Agrawal, Effects of correlated shadowing: Connectivity, localization, and RF tomography, in: 2008 Intl. Conf. Information Processing in Sensor Networks (IPSN 2008), 2008, pp. 82–93.

[48] W. A. Shewhart, S. S. Wilks, Wiley series in probability and mathematical statistics, Wiley, 1984.

[49] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, M. O'Neil, Fast direct methods for Gaussian processes, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (2) (2016) 252–265.

[50] G. H. Golub, C. F. Van Loan, Matrix computations, JHU Press, 2013.

[51] R. Garnett, Bayesian optimization, Cambridge Univ. Press, 2023.

[52] D. Gunning, D. Aha, Darpa's explainable artificial intelligence (xai) program, AI magazine 40 (2) (2019) 44–58.

[53] C. Fiandrino, L. Bonati, S. D'Oro, M. Polese, T. Melodia, J. Widmer, Explora: Ai/ml explainability for the open ran, Proceedings of the ACM on Networking 1 (CoNEXT3) (2023) 1–26.

[54] S. Wang, M. A. Qureshi, L. Miralles-Pechuan, T. Huynh-The, T. R. Gadekallu, M. Liyanage, Applications of explainable ai for 6g: Technical aspects, use cases, and research challenges, arXiv preprint arXiv:2112.04698 (2021).

[55] S. Tadik, K. M. Graves, M. A. Varner, C. R. Anderson, D. Johnson, S. K. Kasera, N. Patwari, J. Van der Merwe, G. D. Durgin, Digital spectrum twins for enhanced spectrum sharing and other radio applications, IEEE Journal of Radio Frequency Identification (2023).

[56] N. Patwari, Indoor channel impulse response dataset, `https://crawdad.org/utah/CIR/20070910/matlab/` (2022).

[57] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, R. J. O'dea, Relative location estimation in wireless sensor networks, IEEE Transactions on Signal Processing 51 (8) (2003) 2137–2148.

[58] F. Mitchell, The FRS/GMRS outdoor received power dataset, `https://zenodo.org/record/7259895#.ZDbPiezML0p` (2022).

[59] J. Breen, A. Buffmire, J. Duerig, K. Dutt, E. Eide, M. Hibler, D. Johnson, S. K. Kasera, E. Lewis, D. Maas, A. Orange, N. Patwari, D. Reading, R. Ricci, D. Schurig, L. B. Stoller, J. Van der Merwe, K. Webb, G. Wong, POWDER: Platform for open wireless data-driven experimental research, in: Proc. 14th Intl. Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization (WiNTECH), 2020.

[60] S. Tadik, A. Singh, F. Mitchell, Y. Hu, X. Yao, K. Webb, A. Sarbhai, D. Maas, A. Orange, J. Van der Merwe, N. Patwari, M. Ji, S. K. Kasera, A. Bhaskara, G. D. Durgin, Salt lake city 3534 MHz multi-transmitter measurement campaign, `https://github.com/serhatadik/slc-3534MHz-meas` (March 2024).

[61] M. Aernouts, R. Berkvens, K. Van Vlaenderen, M. Weyn, Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas, `https://doi.org/10.5281/zenodo.3904158` (Mar. 2018).

[62] A. Marey, M. Bal, H. F. Ates, B. K. Gunturk, PL-GAN: Path loss prediction using generative adversarial networks, IEEE Access 10 (2022) 90474–90480.

[63] J. Wen, Y. Zhang, G. Yang, Z. He, W. Zhang, Path loss prediction based on machine learning methods for aircraft cabin environments, IEEE Access 7 (2019) 159251–159261.

[64] C. A. Oroza, Z. Zhang, T. Watteyne, S. D. Glaser, A machine-learning-based connectivity model for complex terrain large-scale low-power wireless deployments, IEEE Transactions on Cognitive Communications and Networking 3 (4) (2017) 576–584.

[65] N. Moraitis, L. Tsipi, D. Vouyioukas, Machine learning-based methods for path loss prediction in urban environment for LTE networks, in: 2020 16th Intl. Conf. Wireless and Mobile Computing, Networking and Communications (WiMob), 2020, pp. 1–6.

[66] G. D. Durgin, M. A. Varner, N. Patwari, S. K. Kasera, J. Van der Merwe, Digital spectrum twinning for next-generation spectrum management and metering, in: 2022 IEEE 2nd Intl. Conf. Digital Twins and Parallel Intelligence (DTPI), 2022, pp. 1–6.