Jammer Source Localization with Federated Learning

Jammer Source Localization with Federated Learning

Mariona Jaramillo-Civill
Northeastern University
Dept. of Electrical & Computer Eng.
Boston, MA (USA)
jaramillocivill.m@northeastern.edu

Peng Wu
Northeastern University
Dept. of Electrical & Computer Eng.
Boston, MA (USA)
wu.p@northeastern.edu

Andrea Nardin
Politecnico di Torino
Dept. of Electronics & Telecommunications
Turin, Italy
andrea.nardin@polito.it

Tales Imbiriba
University of Massachusetts Boston
Dept. of Computer Science
Boston, MA (USA)
tales.imbiriba@umb.edu

Pau Closas
Northeastern University
Dept. of Electrical & Computer Eng.
Boston, MA (USA)
closas@northeastern.edu

***Abstract*—Global Navigation Satellite System (GNSS) signals are increasingly vulnerable to jamming, disrupting critical applications like autonomous navigation and aviation. Traditional jammer localization relies on centralized data processing, raising privacy concerns. This work proposes a federated learning (FL) framework for privacy-preserving jammer localization using crowdsourced received signal strength (RSS) measurements. We explore three models: a neural network (NN) for initial localization, a path-loss model (PL), and an augmented physics-based model (APBM) combining both PL and NN models. Evaluations in open-sky, suburban and urban environments show that PL and APBM outperform a non-FL baseline in open-sky and suburban settings, while urban scenarios remain challenging due to multipath and shadowing. In addition, we analyze the impact of client distribution, observation density, and measurement noise on localization accuracy.**

***Index Terms*—GNSS interference, jammer localization, federated learning, privacy-preserving machine learning.**

I. Introduction

Global Navigation Satellite System (GNSS) has been increasingly incorporated into civilian applications [1], and the growing dependence on it within critical infrastructures such as communications, autonomous vehicles, aviation and precise timing has led to a proportional increase in the vulnerability of these systems to jamming [2]. GNSS jamming occurs due to the transmission of high-power radio signals in the L-band spectrum, on which GNSS receivers operate [3]. Jammer devices overpower GNSS receivers until denying their operation. Even though the use of jammers is illegal in most countries, they are easy and inexpensive to purchase online, posing a significant risk to the normal operation of many systems reliant on GNSS. Additionally, some devices can unintentionally interfere with GNSS, such as continuous wave (CW) interference from damaged electronics and signals emitted by Distance Measurement Equipment (DME) technology designed for aircraft navigation [4], [5].

It is therefore crucial to monitor these interferences by detecting, classifying and localizing the jammer sources that cause them. Currently, this monitoring requires expensive signal processing and significant manpower [6]. A viable, cost-effective and efficient solution is to focus on congested areas, where users are more likely to operate and from which crowdsourced data can be leveraged. Crowdsourcing methods involve a system where users navigate an area and share their measured received signal strength (RSS) at a GNSS frequency band with a central processing unit [7], [8]. Subsequently, inference methods are used to learn the corresponding signal propagation model and to make predictions about the RSS at points without observations. If a large number of users collaborate in this framework, the learned signal propagation model can be used to localize the source of the jamming interferences and, ultimately, take corrective measures.

Prior research on crowdsourced jammer localization has addressed several challenges associated with this approach [7], [9]–[12]. One key challenge is the requirement for an accurate signal propagation model to enable reliable inference. While the nominal path-loss model has been shown to be well-suited for estimating the jamming RSS field in open-sky environments, it falls short in more complex settings, such as urban or indoor scenarios, which are often affected by multipath and shadowing effects. A neural network-only approach has also been explored; however, it underperforms due to its lack of physical explainability and the absence of prior knowledge about the path-loss physical model governing jammer signal propagation. To address these limitations, an augmented physics-based neural network model (APBM) was proposed in [13], [14]. The APBM integrates the path-loss model with a neural network serving as a nonlinear correction term, demonstrating improved performance and adaptability in these challenging environments.

This work has been partially supported by the National Science Foundation under Award ECCS-1845833 and CCF-2326559.

Once a signal propagation model is employed to learn the jamming RSS field, it can subsequently be used to estimate and localize the source of jamming interference. For path-loss and APBM models, the jammer position is inherently included as a parameter within the model. Consequently, by accurately learning the RSS propagation model, the jammer's position is also estimated. Conversely, in the case of a neural network-only model, the jammer position is inferred as the location associated with the highest predicted jamming RSS.

Prior research has identified and analyzed critical factors that influence performance when dealing with degraded or denied positioning in crowdsourced estimation. These factors include user positioning accuracy, observation density and measurement noise, all of which have been studied to better understand the behavior of the aforementioned models [14]. Despite the promising results presented in [13], [14], certain limitations remain. For instance, even the APBM model was not able to completely characterize more complex urban jamming fields, which seems to require more informed models. Another issue concerns the centralized processing adopted by these strategies which poses privacy concerns over users' data.

Federated learning (FL) is a learning technique that enables users to collaboratively benefit from shared models trained on crowdsourced data without the need to centrally store the data [15]. In this approach, each client retains a local training dataset that is never uploaded to the central server. Instead, clients receive the current global model, compute updates based on their local data, and send only these updates to the server. The central server then aggregates these updates to refine the global model and the process is repeated. By leveraging edge devices to perform local training, federated learning effectively reduces latency, increases scalability and optimizes resource usage, in addition to enhancing data privacy and security, as previously mentioned.

Previous research has investigated the application of machine learning [16], [17], and more particularly federated learning [18], in the GNSS domain. For instance, the jammer classification method using machine learning algorithms proposed in [19] was adapted to incorporate user data privacy through federated learning in [20].

In this paper, we tackle both field-characterization and privacy limitations. Section II introduces the jammer localization problem and outlines proposed improvements to the neural network, path-loss and APBM models. Additionally, we adapt these models for use within a FL framework, requiring new approaches to initialize the path-loss and APBM models, as the central server can no longer rely on identifying the highest RSS client's position, as was done in non-federated settings.

Section III leverages federated learning to address privacy concerns inherent to crowdsourced systems, by proposing the integration of information from multiple agents accessing the cloud asynchronously. Data generation and experiments are provided in Sections IV and V respectively, where the three models are evaluated in a) an open-sky environment where the path-loss model is accurate, b) a suburban scenario generated using ray-tracing techniques, and c) a complex urban scenario also generated with ray-tracing. Furthermore, we assess the robustness of jammer localization under varying conditions, including different levels of agent positioning accuracy, observation density, measurement errors and the number of nodes in the clients' distribution. The final section discusses the main conclusions of this work and outlines directions for future research.

II. Estimation of Jammer's Position

In the proposed framework, we aim at estimating the jammer's position from N observations of its power, y_n (dBW), measured at agent locations $\mathbf{x}_n = \left(x_n^{(1)}, \ldots, x_n^{(D)}\right)^\top$ in a D-dimensional space. In this contribution, we assume $D = 2$. The measured RSS signal y_n is a function of the jammer position $\boldsymbol{\theta}$ and the location, $\mathbf{x}_n$, where the measurement took place. That is

$$y_n = f^{\mathrm{TRUE}}(\mathbf{x}_n; \boldsymbol{\theta}) + \xi_n, \quad n = 1, \ldots, N \tag{1}$$

where $f^{\mathrm{TRUE}}(\cdot)$ represents the true underlying model governing the measurements. The literature [13], [14] presents different parametric models for $f^{\mathrm{TRUE}}(\cdot)$ for different scenarios (e.g., open sky and urban), which may include different levels of non-ideal effects such as multipath. To estimate the jammer's position, we resort to parametric models of the interference field which may, or may not, include the jammer's position as part of its parametrization, as will be discussed in the next sections. Thus, the observed RSS is modeled as a parametric function, given by:

$$y_n = f_m(\mathbf{x}_n; \Psi_m) + \xi_n, \quad n = 1, \ldots, N. \tag{2}$$

The family of functions $\{f_m\}_{m=1}^3$ represents the specific model being used, with m indexing the neural network, path-loss, or augmented physics-based model respectively. The parameter set Ψ_m is defined based on the corresponding model. The noise term ξ_n accounts for additive measurement noise and is assumed to be independent of $\mathbf{x}_n$ and Ψ_m; however, in practice, spatially correlated effects such as multipath or shadowing may arise, which can be implicitly captured by the NN and APBM models.

This modeling approach is particularly well-suited to GNSS interference scenarios, where signals typically fall below the noise floor, making large signal powers strong indicators of jamming activity.

Our practical experience modeling RSS fields shows that accurately characterizing the power measurements constitutes a modeling challenge due to the peaky nature of the RSS field. To prioritize the accuracy of RSS peak detection while ensuring overall field accuracy, a weighted loss function is employed.

We define $\mathbf{y} = [y_1, y_2, \ldots, y_N]^\top$ as the vector containing the observed RSS values at different locations, and $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N]^\top$ as the corresponding predicted RSS values, where $\hat{y}_n$ is the predicted RSS value at position $\mathbf{x}_n$.

The weight for each observation is computed as:

$$\tilde{w}_n = \left(\frac{y_n - \min(y)}{\max(y) - \min(y) + \epsilon}\right)^2, \tag{3}$$

where $\epsilon = 10^{-6}$ prevents division by zero. These weights are then normalized and applied to the weighted mean squared error loss function, defined as:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \sum_{n=1}^{N} w_n \left(y_n - \hat{y}_n\right)^2, \tag{4}$$

$$w_n = \frac{\tilde{w}_n}{\sum_{m=1}^{N} \tilde{w}_m} \quad \text{for } n = 1, \ldots, N. \tag{5}$$

The loss or cost function can be equivalently expressed in terms of Ψ_m:

$$\mathcal{L}(\Psi_m; \mathcal{D}) = \sum_{n=1}^{N} w_n \left(y_n - f_m(\mathbf{x}_n; \Psi_m)\right)^2 \tag{6}$$

where the training dataset is defined as $\mathcal{D} = \{y_n, \mathbf{x}_n\}_{n=1}^{N}$ in a centralized architecture. We assumed that ξ_n is zero-mean such that $\hat{y}_n = \mathbb{E}\{y_n|\mathcal{D}\} = f_m(\mathbf{x}_n; \Psi_m)$.

The estimation problem is then formulated as an expected loss optimization problem:

$$\hat{\Psi}_m = \arg\min_{\Psi_m} \mathbb{E}_{\hat{p}(\mathcal{D})} \left\{\mathcal{L}(\Psi_m; \mathcal{D})\right\}, \tag{7}$$

where $\mathbb{E}_{\hat{p}(\mathcal{D})}$ represents the expectation over the empirical data distribution, available through the training dataset $\mathcal{D}$.

Another challenge in estimating the RSS field concerns the imprecise knowledge of agent locations. In practice, agent positions are not perfectly known and are instead represented by estimates $\hat{\boldsymbol{x}}_n$, which are modeled as noisy observations of the true positions $\boldsymbol{x}_n$. Here, we assume these estimates follow a Gaussian distribution with an error characterized by a standard deviation $\boldsymbol{\sigma}_{\text{pos}}$ along the D dimensions:

$$\hat{\boldsymbol{x}}_n \sim \mathcal{N}(\boldsymbol{x}_n, \boldsymbol{\sigma}_{\text{pos}}^2 \mathbf{I}). \tag{8}$$

Additionally, in real-world deployments, differences in sensor hardware (e.g., phone models) may introduce systematic biases in RSS or position estimates due to varying antenna gains or calibration. While not explicitly modeled in this work, such biases could be mitigated through calibration procedures or domain adaptation methods in future extensions.

A. Neural Network-Based Initialization of Jammer's Position

A purely data-driven model of the power values over space, that is a neural network (NN), is not parametrized by the jammer's location. As a consequence, its effectiveness in estimating the location of the jamming source is limited and it is only considered here for initialization purposes. This NN-based initialization of the estimated field is seen to be critical when direct access to client data is restricted, as required in federated learning. It corresponds to $f_1(\mathbf{x}_n; \boldsymbol{\phi})$, where the parameter set $\Psi_1 = \boldsymbol{\phi}$ represents the weights and biases of the NN. The NN is implemented as a feed-forward architecture with three hidden layers of 500, 256 and 128 neurons, using Leaky ReLU activation functions to avoid gradient vanishing, and linear output layer, allowing negative dB as output. Xavier initialization [21] is employed to initialize the weights, ensuring proper gradient scaling and stable convergence. To enhance training stability and prevent over-smoothing and overfitting of the RSS field, batch normalization and dropout layers are incorporated. Training is performed using the Adam optimizer [22] with a learning rate of 0.001.

The NN predicts the RSS field over a grid and the jammer's position is identified as the grid point with the maximum predicted RSS. Although the NN alone does not achieve the lowest RMSE, as shown in Section V, it provides a robust initialization of the jammer position for subsequent models.

B. Path-Loss Based Estimation of Jammer's Position

For the path-loss model [23], $f_2(\mathbf{x}_n; \boldsymbol{\theta}, P_0, \gamma)$ is used, where the parameter set $\Psi_2 = \{\boldsymbol{\theta}, P_0, \gamma\}$. The function is expressed as:

$$f_2(\mathbf{x}_n; \boldsymbol{\theta}, P_0, \gamma) = P_0 - \gamma 10 \log_{10} d(\mathbf{x}_n, \boldsymbol{\theta}), \tag{9}$$

where $\boldsymbol{\theta}$ represents the jammer's position, P_0 is the jammer's transmission power in dBW at the reference distance of 1 meter and γ is the path-loss exponent. The distance $d(\mathbf{x}_n, \boldsymbol{\theta})$ is the Euclidean distance between the observer at position $\mathbf{x}_n$ and the jammer at position $\boldsymbol{\theta}$.

In the considered path-loss model, no assumptions about the model parameters are made a priori. Instead, the jammer's position $\boldsymbol{\theta}$ is initialized using the neural network's estimate, and the parameters P_0 and γ are learned directly from the data during the optimization process. This makes the path-loss model autonomous and adaptable to diverse environments, as it does not rely on pre-defined parameter values.

The parameters $\boldsymbol{\theta}$, P_0 and γ are estimated by solving (7), using learning rates of 0.5 for the jammer position $\boldsymbol{\theta}$ and 0.01 for both the transmission power P_0 and the path-loss exponent γ. The jammer's position is retrieved from the optimized parameter $\boldsymbol{\theta}$.

As discussed in [24], [25], the Gaussian assumption on the noise term ξ_n is reasonable in a log-distance path-loss model. In addition, when the weighting is uniform, the optimization of the cost function (6) corresponds to finding the maximum likelihood (ML) estimator for the parameters of the path-loss model, as was derived in [13]. This is because minimizing the squared error term in the cost function is equivalent to maximizing the likelihood under the assumption of Gaussian noise.

C. Augmented Physics-Based Estimation of Jammer's Position

The augmented physics-based model combines into a single function the path-loss model and a neural network correction term to account for complex propagation phenomena, such as multipath and shadowing expressed as:

$$\begin{aligned} f_3(\mathbf{x}_n; \boldsymbol{\theta}, P_0, \gamma, \boldsymbol{\phi}, \lambda) = \lambda f_2(\mathbf{x}_n; \boldsymbol{\theta}, P_0, \gamma) \\ + (1 - \lambda) f_1(\mathbf{x}_n; \boldsymbol{\phi}), \end{aligned} \tag{10}$$

where $f_2(\mathbf{x}_n; \boldsymbol{\theta}, P_0, \gamma)$ represents the path-loss model and $f_1(\mathbf{x}_n; \boldsymbol{\phi})$ represents the NN correction term. The parameter λ is a learnable weight that controls the contribution of each model and is constrained within the range $\lambda \in [0, 1]$, ensuring a balance between the two terms.

Initially, λ is set to 0.8, favoring the path-loss model and physical interpretability. The parameter λ is optimized during training with a small learning rate of 0.001 to ensure dynamic adjustment to different environments.

The parameters $\boldsymbol{\theta}$, P_0, γ and $\boldsymbol{\phi}$ are jointly optimized by solving (7) and the same learning rates and optimizers as in the PL and NN approach. Again, in this approach, the jammer's position is inferred from the optimized parameter $\boldsymbol{\theta}$. By dynamically adjusting the contributions of the two models, the APBM should ensure robust and accurate RSS field predictions and jammer localization.

III. Federated Learning Methodology

We assume that the dataset $\mathcal{D}$ is composed of M IID disjoint subsets $\mathcal{D}_i = \{y_n^{(i)}, \mathbf{x}_n^{(i)}\}_{n\in\mathcal{I}_i}$, $i \in \{1, \ldots, M\}$, corresponding to M collaborative nodes or clients. Here, $\mathcal{I}_i \subset \{1, \ldots, N\}$ is the set of indices corresponding to the N_i samples allocated to client i, with the subsets satisfying $\bigcup_{i=1}^{M} \mathcal{I}_i = \{1, \ldots, N\}$ and $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset$ for $i \neq j$. This ensures that all data points are uniquely assigned to exactly one client and the total dataset size satisfies $\sum_{i=1}^{M} N_i = N$.

Under IID data settings, the standard FL approach is Federated Averaging (FedAvg) [15], where local models are trained independently by clients using their respective datasets and subsequently uploaded to a central cloud server. The server aggregates these models by averaging their parameters to compute a global model. For the three aforementioned models, where $y_n = f_m(\mathbf{x}_n; \Psi_m) + \xi_n$, the parameter set Ψ_m is the one averaged across clients to produce the global parameters.

In FedAvg, the training objective is to minimize a single global loss function $\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}})$, which is obtained as a weighted combination of the local loss functions defined for each client in Eq. (4), $\mathcal{L}_i(\mathbf{y}_i, \hat{\mathbf{y}}_i) = \frac{1}{N_i} \sum_{n\in\mathcal{I}_i} w_n^{(i)} (y_n - \hat{y}_n)^2$:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \sum_{i=1}^{M} \frac{N_i}{N} \mathcal{L}_i(\mathbf{y}_i, \hat{\mathbf{y}}_i), \tag{11}$$

where $w_n^{(i)}$ ($n \in \mathcal{I}_i$) represents the weights for individual samples in training node i.

The learning problem is then an optimization of (11), which is solved iteratively through multiple rounds of local optimization at the clients, followed by model fusion at the central server. At each communication round t, every client i initializes its local model with the fused global parameters $\Psi_{m,t}$ from the previous iteration and updates it by minimizing the local loss function over its dataset $\mathcal{D}_i$ for a certain number of local epochs:

$$\Psi_{m,t+1}^{(i)} = \arg\min_{\Psi_m} \mathcal{L}_{i,t}(\mathbf{y}_i(\Psi_m), \hat{\mathbf{y}}_i), \tag{12}$$

where t in the loss function indicates that the local parameters were initialized with the fused global parameters, $\Psi_{m,t}$.

Once all participating clients complete their local training, the central server aggregates the locally updated models using a weighted averaging strategy to obtain the new global model:

$$\Psi_{m,t+1} = \sum_{i=1}^{M} \frac{N_i}{N} \Psi_{m,t+1}^{(i)}, \tag{13}$$

where $\Psi_{m,t+1}$ represents the updated global model parameters at iteration $t+1$ and $\Psi_{m,t+1}^{(i)}$ corresponds to the locally updated model parameters of client i.

This iterative procedure is repeated until convergence, effectively learning a global model that optimally combines the individual local models while preserving data privacy.

IV. Data Generation

Synthetic data is widely used in GNSS studies, as real-world data collection can be challenging in certain scenarios. For instance, constructing a real-world scenario with active jamming is difficult, particularly when testing different jammer types, received power levels and propagation conditions. Additionally, accurately simulating effects such as multipath reflections or shadowing adds further complexity.

Therefore, in our analysis, we adopted similar data generation techniques to those reported in [14], while expanding to additional scenarios and incorporating specific factors relevant to this research.

We considered three propagation scenarios for the electromagnetic field produced by the jammer:

a) Open-sky path-loss scenario: an open-sky environment entirely described by the path-loss formula in (9).
b) Suburban ray-tracing scenario: generated using ray-tracing techniques in Chicago's Grant Park, representing a suburban or partially open-sky environment.
c) Urban ray-tracing scenario: also generated using ray-tracing in Downtown Chicago, where complex effects such as multipath reflections and shadowing are prevalent.

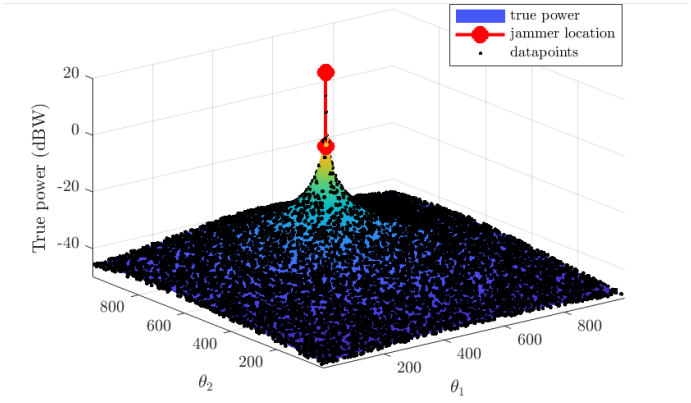
Table I reports the values used for generating the open-sky path-loss scenario (Fig. 1a), while Table II describes the parameters for the ray-tracing simulation [26], [27] in both suburban (Fig. 1b) and urban scenarios (Fig. 1c). In all scenarios, the jammer is assumed to transmit with an omni-directional gain pattern, simplifying the propagation model and reflecting typical assumptions in interference modeling.

TABLE I: Open-sky path-loss scenario simulated parameters.

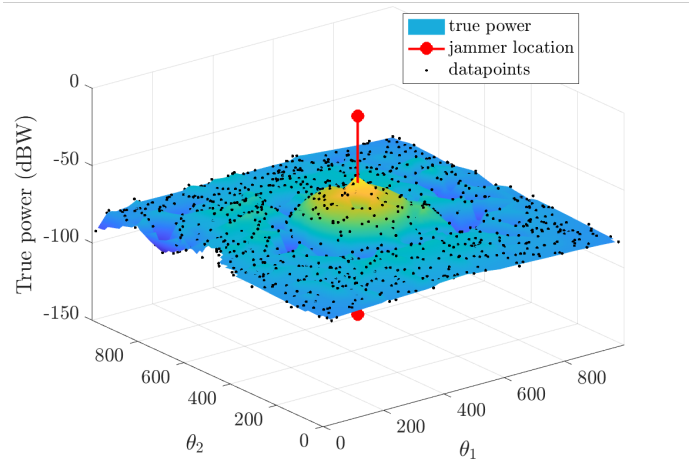
Parameter	Value
P_0	10 dBW
γ	2
Number of samples	10000
Observed area	1 km^2

As our objective is to perform crowdsourced estimation, where users navigate an area and share their measured RSS with a central unit, we utilized building information from OpenStreetMap [28] to ensure measurements are taken only on the exterior of buildings. This filters out negligible power measurements, which correspond to $-\infty$ in the logarithmic dBW scale.

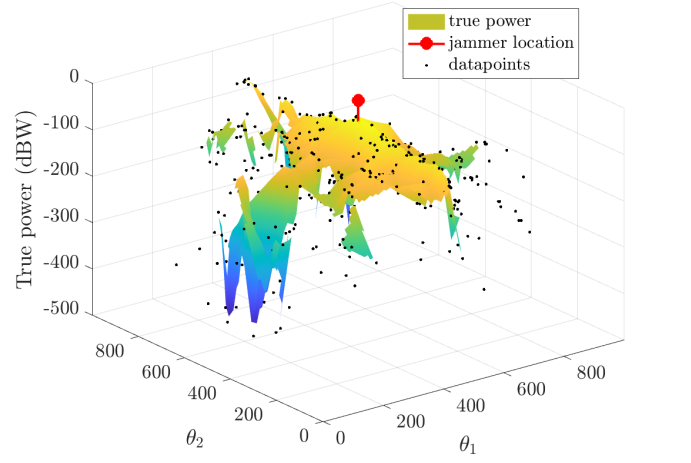
To further refine the data processing for the three models used to predict the jammer field, we applied a threshold at -150 dBW. Measurements below this threshold were excluded to prevent the models from being influenced by extremely



(a) Open-sky path-loss scenario.



(b) Suburban ray-tracing scenario.



(c) Urban ray-tracing scenario.

Fig. 1: Simulated jammer power field under open-sky path-loss and suburban and urban ray-tracing.

TABLE II: Ray-tracing scenarios simulated parameters.

Parameter	**Value**
P_0	10 dBW
Jammer frequency	1575.42 MHz (L1)
Suburban location (Chicago's Grant Park)	41.876442 -87.618971
Urban location (Downtown Chicago)	41.880011 -87.629338
Max. no. of reflections	100
Ray-tracing method	Shooting and bouncing rays (SBR)
Launched rays average separation	0.5391°
Number of samples	1000
Observed area	1 km^2

low values, which are not informative for predictions. Instead, we focus on the meaningful measurements that contribute significantly to RSS field prediction.

In the open-sky path-loss and suburban ray-trace scenarios, all measurements in our datasets are above this threshold. However, in the urban ray-trace scenario, the dataset is significantly reduced due to multipath and shadowing effects, with only 183 of the original 1000 samples remaining above the threshold. This highlights the challenge posed by the urban environment, where the number of informative observations is considerably lower.

V. Experiments

To evaluate the model's performance, we partitioned the dataset into 80% training and 20% testing points. The training and testing losses for jammer field prediction were analyzed and visualized to ensure convergence across all models and to apply early stopping, preventing overfitting to the training points. Furthermore, we performed 10 Monte Carlo (MC) simulations to account for variations in data distribution across nodes and to assess the model's robustness.

A. *Quantitative Analysis of Jammer Localization Performance*

Table III provides a detailed analysis of the root mean square error (RMSE) in jammer localization for the neural network, path-loss and augmented physics-based models across three propagation scenarios: open-sky, suburban and urban. The experiments are conducted within a federated learning framework, using 10 nodes for the open-sky and suburban scenarios and 5 nodes for the urban scenario. The interference-to-noise ratio (INR) is set to 10 dB and the observation density is $0.001\,\text{obs/m}^2$ (equivalent to 1000 observations in $1\,\text{km}^2$ for open-sky and suburban settings). In the urban scenario, the observation density is lower (1.83×10^{-4}) due to the aforementioned multipath effects and shadowing.

A key reference point for evaluating the models is the error associated with the closest training point to the jammer's location. In a non-federated setup, where all data is centrally accessible, the closest training point could be directly used to estimate the jammer's location. In the open-sky and suburban scenarios, the PL and APBM models successfully achieve a lower error, even within a FL framework, demonstrating the potential of these models to provide accurate jammer

TABLE III: Jammer Localization and Prediction Errors Across Different Scenarios.

Metric for Jammer Localization Error (m)	Open-Sky Path-Loss			Suburban Ray Tracing			Urban Ray Tracing		
	NN	PL	APBM	NN	PL	APBM	NN	PL	APBM
Average	33.63	1.90	2.17	25.57	4.39	7.96	50.54	36.48	22.09
Standard Deviation	14.13	0.97	1.09	8.25	2.60	3.03	19.09	83.22	44.71
Median	33.61	1.67	1.86	27.86	3.45	7.51	53.72	10.72	4.55
Minimum	15.1	0.49	1.09	10.96	1.3	3.8	17.67	2.17	2.22
Maximum	57.25	3.92	4.16	36.6	8.9	12.98	88.61	285.79	155.07
Closest Training Point to Jammer	18.13 ± 9.04			13.69 ± 0.00			14.63 ± 1.59		

localization while adhering to the privacy-preserving principles of FL.

An essential advantage of the NN model lies in its ability to initialize the jammer's position in the FL framework. Unlike previous centralized approaches in past research, where the jammer's position was initialized as the point with the highest received signal strength (RSS), the decentralized nature of FL prohibits direct access to such global data. The NN model addresses this challenge by providing an initial estimate of the jammer's location in a fully decentralized manner. This initialization is crucial for enabling the PL and APBM models to refine the jammer localization, as it allows FL to function effectively while preserving data privacy.

Open-Sky Path-Loss Scenario: In the open-sky scenario, the PL model achieves the best performance, with an average localization error of $1.90\,\text{m}$, followed closely by the APBM with $2.17\,\text{m}$. Both models outperform the baseline error associated with the closest training point ($18.13 \pm 9.04\text{m}$), which highlights their ability to accurately capture the RSS field and localize the jammer under ideal propagation conditions. The NN model, while less accurate with an average error of $33.63\,\text{m}$, provides a robust initial estimate that enables the subsequent refinement by the PL and APBM models.

Suburban Ray-Tracing Scenario: In the suburban scenario, which includes partial multipath effects, the PL model again achieves the lowest average localization error at $4.39\,\text{m}$, followed by the APBM at $7.96\,\text{m}$. Even though the error is higher than in the open-sky scenario, both models still outperform the non-FL baseline error of $13.69\,\text{m}$ associated with the closest training point.

Urban Ray-Tracing Scenario: The urban scenario presents the most challenging conditions, with significant multipath effects and shadowing reducing the number of useful observations to 183. In this scenario, the NN model achieves an average localization error of $50.54\,\text{m}$, while the PL and APBM models achieve $36.48\,\text{m}$ and $22.09\,\text{m}$, respectively. Unlike the open-sky and suburban cases, none of the models outperform the baseline error associated with the closest training point in a non-FL framework.

However, the median localization error remains lower than that of the closest training point, indicating that while the models often achieve accurate jammer localization, large outliers occasionally occur. This suggests that the primary challenge with the urban ray-tracing scenario is how data distributes across nodes: some Monte Carlo (MC) simulations result in precise localization, whereas others produce large errors.

B. Qualitative Analysis of Jammer Field Prediction and Localization

Fig. 2 provides a qualitative visualization of the jammer field prediction and localization for the three scenarios: open-sky, suburban and urban. Each scenario includes four subplots: (1) the training points distributed across multiple nodes, (2) the NN-predicted RSS field with the NN-predicted jammer location, (3) the PL-predicted RSS field with the PL-predicted jammer location, and (4) the APBM-predicted RSS field with the APBM-predicted jammer location. The true jammer location is indicated for reference in all subplots.

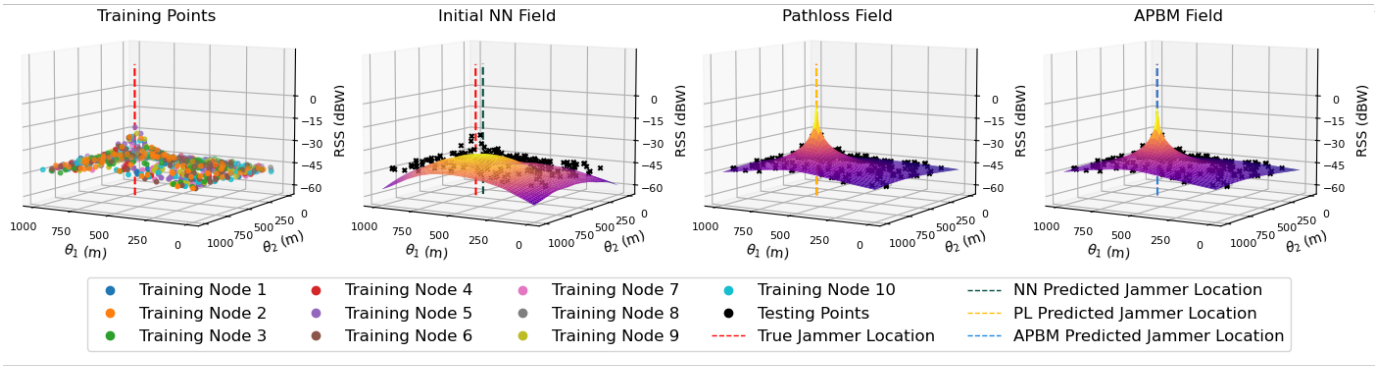
In the open-sky and suburban scenarios, the NN model fails to accurately capture the peak RSS values, but it provides a predicted jammer location (green marker) reasonably close to the true location (red marker). The PL and APBM models effectively fit the RSS field and achieve near-perfect localization, as indicated by their yellow and blue markers, closely overlapping the true jammer location.

The urban scenario illustrates the challenges of multipath and shadowing. The NN model provides an approximate jammer location, but the RSS field lacks detail. The PL model peaks at the jammer's true location but fails to capture high RSS points in the main street where the jammer is placed and surrounding areas. The APBM model, however, provides a more detailed field, accurately capturing the main street and peaking at the jammer's location.

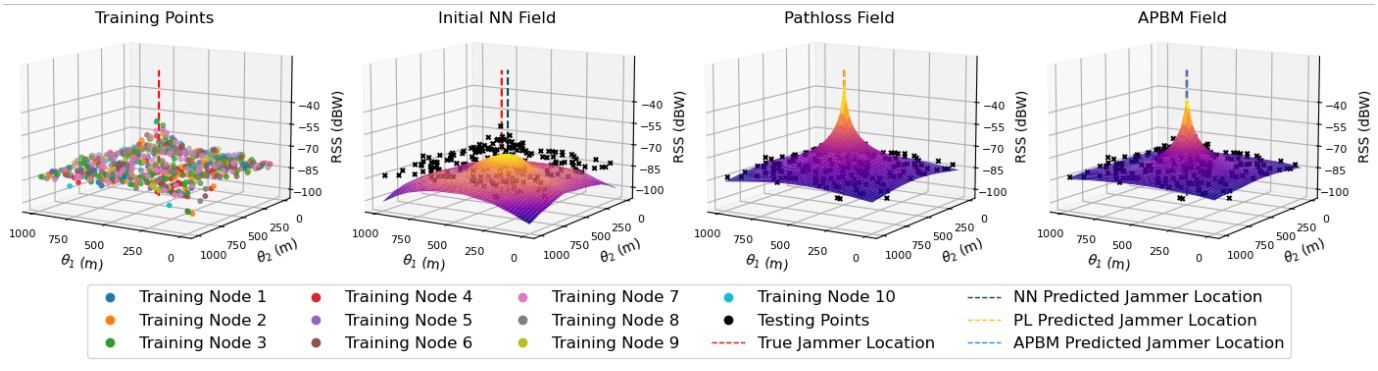
C. Factors Affecting Jammer Localization

To evaluate how different factors influence jammer localization error within the FL framework, we conducted a series of experiments by varying key parameters related to data availability and FL settings. For the open-sky and suburban scenarios, we established a baseline configuration with 10 nodes, an INR of $10\,\text{dB}$ and an observation density of $0.001\,\text{obs/m}^2$. Given the limited number of useful observations in the urban scenario, we used a reduced setting with 5 nodes and 183 significant observations, maintaining the same INR of $10\,\text{dB}$.

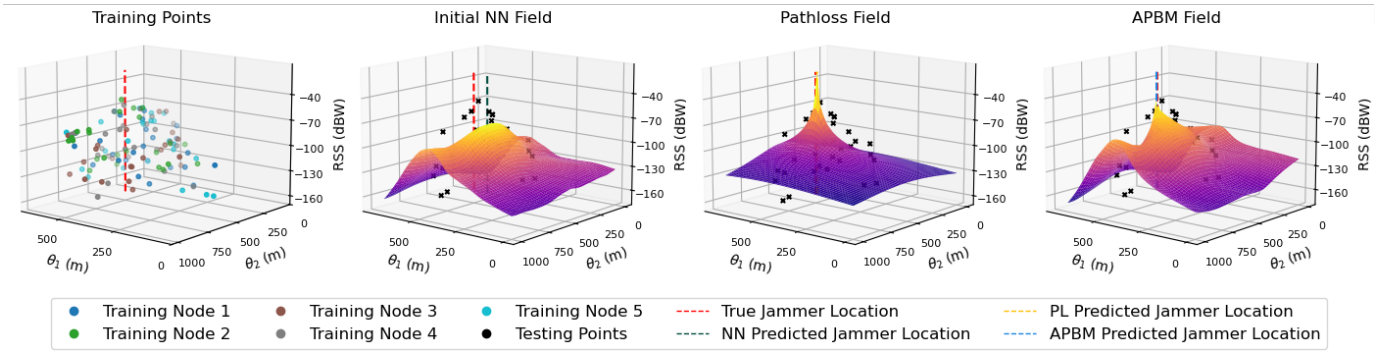
We examined the impact of the number of clients per node, observation density and measurement noise on jammer localization error, presenting the results in boxplots. In each



(a) Open-sky path-loss scenario.



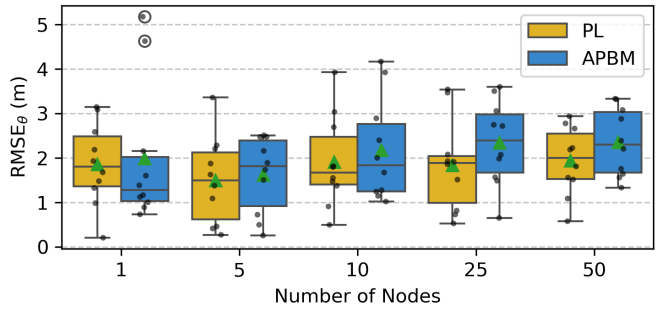
(b) Suburban ray-tracing scenario.



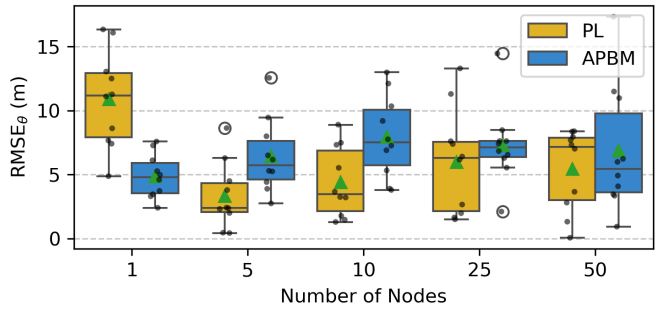
(c) Urban ray-tracing scenario.

Fig. 2: Jammer field prediction and localization across different scenarios: (a) open-sky path-loss, (b) suburban ray-tracing, and (c) urban ray-tracing. Each scenario contains four subplots (from left to right): 1) the training points distributed across multiple nodes used for model training, 2) the predicted NN field with the NN-predicted jammer location, 3) the predicted PL field with the PL-predicted jammer location, and 4) the predicted APBM field with the APBM-predicted jammer location. The true jammer location is indicated for reference in each subplot.

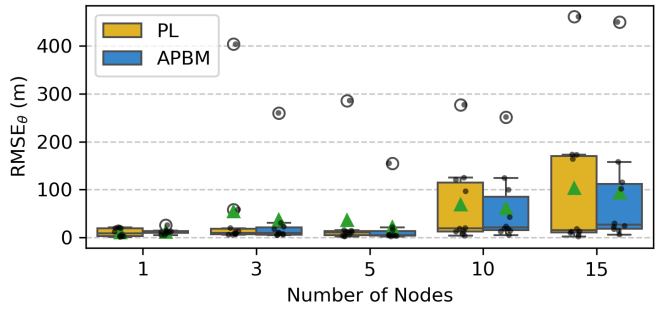
experiment, one factor was varied while the others remained fixed at the baseline values.



(a) Open-sky path-loss scenario.



(b) Suburban ray-tracing scenario.

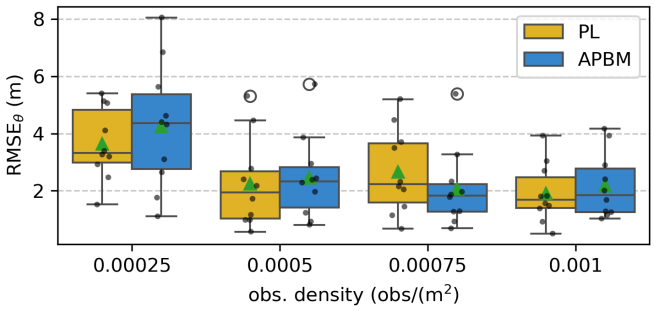


(c) Urban ray-tracing scenario.

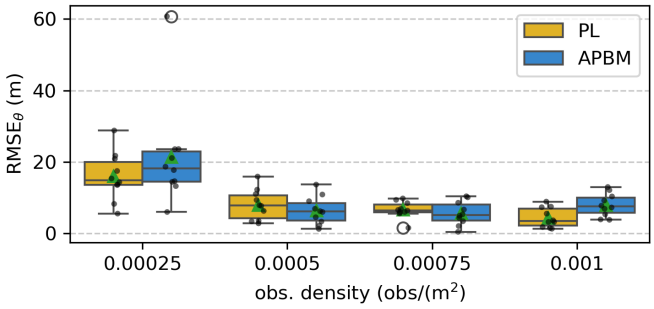
Fig. 3: Impact of the number of nodes on jammer localization performance across different scenarios.

1) Impact of Number of Clients per Node: In open-sky and suburban scenarios, increasing the number of nodes does not reveal a clear trend in performance, as the RMSE remains low across all cases, see Fig. 3. The maximum number of nodes considered was 50, which corresponds to each client having approximately 40 training observations. The lack of a strong correlation between node count and localization error suggests that the model is robust to data partitioning, meaning that an increased number of clients does not necessarily degrade performance.

In the urban scenario, however, a different trend emerges. Since a lower maximum number of nodes (15) was used due to the limited number of observations, each client only had around 10 significant observations. This led to a significant performance drop, suggesting that urban environments pose greater challenges for jammer localization. In these scenarios, the limited number of meaningful observations due to shadowing, rapid signal attenuation and multipath makes it more difficult to obtain accurate jammer predictions under FL.



(a) Open-sky path-loss scenario.



(b) Suburban ray-tracing scenario.

Fig. 4: Impact of observation density on jammer localization performance across different scenarios.

2) Impact of Observation Density: As expected, reducing the observation density leads to an increase in RMSE, see Fig. 4, indicating a degradation in performance. However, this effect is not drastic. Even in the lowest-density case (where each node had approximately 20 training observations), the performance drop was moderate. In the open-sky scenario, the localization error increased to only about 4 m for both the PL and APBM models, while in the suburban scenario, the error increased to approximately 20 m. These results suggest that while higher observation densities improve performance, the models can still achieve reasonable localization accuracy with fewer observations.

3) Impact of Measurement Noise: The measurement noise directly influences the Interference-to-Noise Ratio (INR), defined as:

$$\mathrm{INR} = 10 \log_{10} \frac{P_0}{\sigma^2}, \tag{14}$$

where P_0 is the jammer's transmission power (set to 10 dBW in the considered datasets) and σ^2 represents the measurement noise variance.

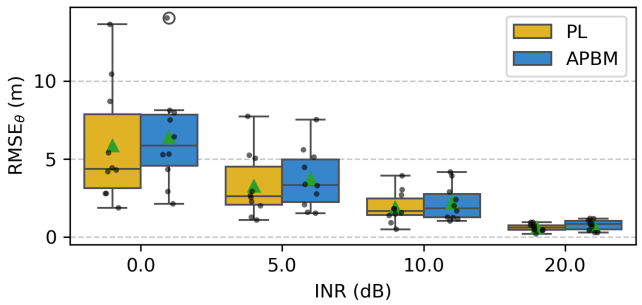


Fig. 5: Impact of Interference-to-Noise Ratio (INR) on jammer localization performance in an open-sky path-loss scenario.

As shown in Fig. 5, under ideal open-sky conditions, a clear trend is observed: higher INR values correlate with lower localization errors. This aligns with expectations, as improved signal quality typically enhances localization accuracy. The suburban and urban scenarios are excluded from the figure, as no consistent correlation between INR and localization error was found in those cases, suggesting that other factors—such as multipath and shadowing—may overshadow its effect.

VI. Conclusion

This work presented a federated learning (FL) framework for jammer localization using crowdsourced received signal strength (RSS) measurements. By preserving data privacy, FL overcomes limitations of centralized methods that require direct access to user data. We evaluated three models: a neural network (NN) for initialization, a path-loss (PL) model and an augmented physics-based model (APBM), across open-sky, suburban and urban scenarios.

In the open-sky scenario, the PL and APBM models achieved localization errors of $1.90\,\text{m}$ and $2.17\,\text{m}$, respectively, outperforming the non-FL baseline of $18.13\,\text{m}$. In the suburban scenario, errors remained low at $4.39\,\text{m}$ and $7.96\,\text{m}$. However, urban environments remained challenging due to multipath and shadowing. The APBM achieved the lowest error ($22.09\,\text{m}$) but did not surpass the closest training point to the jammer distance and exhibited high variability.

Further analysis of key factors affecting jammer localization revealed that the number of clients per node does not significantly impact performance in open-sky and suburban scenarios. However, in urban environments, obtaining significant observations is more challenging due to factors such as shadowing and multipath. This results in a reduced number of meaningful observations per client, which significantly hinders the ability to estimate the jammer's position, leading to performance degradation. Observation density has a moderate effect, with the lowest density increasing errors to around $4\,\text{m}$ in open-sky and $20\,\text{m}$ in suburban settings. Measurement noise has a limited impact in complex environments but reducing it improves localization in open-sky conditions.

Future work will enhance FL robustness by exploring alternative data distributions, including non-IID settings and improving jammer field estimation and localization in urban scenarios by integrating geographic information and city layout to model shadowing effects. Additionally, refining FL aggregation techniques and incorporating more advanced physics-informed models could further enhance localization accuracy while preserving privacy.

References

[1] Y. J. Morton, F. van Diggelen, J. J. Spilker Jr, B. W. Parkinson, S. Lo, and G. Gao, *Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications*. John Wiley & Sons, 2021.

[2] M. G. Amin, P. Closas, A. Broumandan, and J. L. Volakis, "Vulnerabilities, threats, and authentication in satellite-based navigation systems [scanning the issue]," *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1169–1173, Jun. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7471534/

[3] K. Radoš, M. Brkić, and D. Begušić, "Recent Advances on Jamming and Spoofing Detection in GNSS," *Sensors*, vol. 24, no. 13, p. 4210, Jun. 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/13/4210

[4] D. Borio, C. O'Driscoll, and J. Fortuny, "GNSS Jammers: Effects and countermeasures," in *2012 6th ESA Workshop on Satellite Navigation Technologies (Navitec 2012) & European Workshop on GNSS Signals and Signal Processing*. Noordwijk, Netherlands: IEEE, Dec. 2012, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/document/6423048/

[5] H. Li, D. Borio, and P. Closas, "Dual-Domain Robust GNSS Interference Mitigation," in *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, Miami, Florida, Oct. 2019, pp. 991–1002. [Online]. Available: https://www.ion.org/publications/abstract.cfm?articleID=16991

[6] I. E. Mehr, A. Minetto, and F. Dovis, "A Navigation Signals Monitoring, Analysis and Recording Tool: Application to Real-Time Interference Detection and Classification," in *Proceedings of the 36th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2023)*, Denver, Colorado, Oct. 2023, pp. 3878–3887. [Online]. Available: https://www.ion.org/publications/abstract.cfm?articleID=19391

[7] D. Borio, C. Gioia, A. Štern, F. Dimc, and G. Baldini, "Jammer localization: From crowdsourcing to synthetic detection," in *Proceedings of the 29th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2016)*, 2016, pp. 3107–3116.

[8] E. Arias-de Reyna, P. Closas, D. Dardari, and P. M. Djuric, "Crowd-Based Learning of Spatial Fields for the Internet of Things: From Harvesting of Data to Inference," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 130–139, Sep. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8454394/

[9] L. Strizic, D. M. Akos, and S. Lo, "Crowdsourcing GNSS Jammer Detection and Localization," Reston, Virginia, Feb. 2018, pp. 626–641. [Online]. Available: https://www.ion.org/publications/abstract.cfm?articleID=15546

[10] P. Wang and Y. T. Morton, "Efficient Weighted Centroid Technique for Crowdsourcing GNSS RFI Localization Using Differential RSS," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 3, pp. 2471–2477, Jun. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8717611/

[11] R. C. Blay and D. M. Akos, "GNSS RFI Localization using a Hybrid TDOA/PDOA Approach," Reston, Virginia, Feb. 2018, pp. 703–712. [Online]. Available: https://www.ion.org/publications/abstract.cfm?articleID=15554

[12] G. K. Olsson, E. Axell, E. G. Larsson, and P. Papadimitratos, "Participatory Sensing for Localization of a GNSS Jammer," in *2022 International Conference on Localization and GNSS (ICL-GNSS)*. Tampere, Finland: IEEE, Jun. 2022, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/9797031/

[13] A. Nardin, T. Imbiriba, and P. Closas, "Jamming Source Localization Using Augmented Physics-Based Model," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island, Greece: IEEE, Jun. 2023, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/10095731/

[14] ——, "Crowdsourced Jammer Localization Using APBMs: Performance Analysis Considering Observations Disruption," in *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Apr. 2023, pp. 511–519. [Online]. Available: http://www.ion.org/publications/abstract.cfm?jp=p&articleID=18775

[15] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, Florida, 2017. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a/mcmahan17a.pdf

[16] A. Siemuri, K. Selvan, H. Kuusniemi, P. Valisuo, and M. S. Elmusrati, "A systematic review of machine learning techniques for GNSS use cases," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 6, pp. 5043–5077, 2022.

[17] P. Closas, L. Ortega, J. Lesouple, and P. M. Djurić, "Emerging trends in signal processing and machine learning for positioning, navigation and timing information: special issue editorial," *EURASIP Journal on Advances in Signal Processing*, vol. 2024, no. 1, p. 84, 2024.

[18] P. Wu, "Bayesian data fusion for distributed learning," PhD Thesis, Northeastern University, 2024.

[19] R. Morales Ferre, A. de la Fuente, and E. S. Lohan, "Jammer Classification in GNSS Bands Via Machine Learning Algorithms," *Sensors (Basel, Switzerland)*, vol. 19, no. 22, p. 4841, Nov. 2019.

[20] P. Wu, H. Calatrava, T. Imbiriba, and P. Closas, "Jammer classification with Federated Learning," in *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. Monterey, CA, USA: IEEE, Apr. 2023, pp. 228–234. [Online]. Available: https://ieeexplore.ieee.org/document/10140124/

[21] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Sardinia, Italy, 2010. [Online]. Available: https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf

[22] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA, 2015, version Number: 9. [Online]. Available: https://arxiv.org/abs/1412.6980

[23] A. Goldsmith, *Wireless communications*. Cambridge ; New York: Cambridge University Press, 2005.

[24] R. Bernhardt, "Macroscopic Diversity in Frequency Reuse Radio Systems," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 862–870, Jun. 1987. [Online]. Available: http://ieeexplore.ieee.org/document/1146594/

[25] D. C. Cox, R. R. Murray, and A. W. Norris, "800-MHz Attenuation Measured In and Around Suburban Houses," *AT&T Bell Laboratories Technical Journal*, vol. 63, no. 6, pp. 921–954, Jul. 1984. [Online]. Available: https://ieeexplore.ieee.org/document/6771659

[26] K. Schaubach, N. Davis, and T. Rappaport, "A ray tracing method for predicting path loss and delay spread in microcellular environments," in *[1992 Proceedings] Vehicular Technology Society 42nd VTS Conference - Frontiers of Technology*. Denver, CO, USA: IEEE, 1992, pp. 932–935. [Online]. Available: http://ieeexplore.ieee.org/document/245274/

[27] Z. Yun and M. F. Iskander, "Ray Tracing for Radio Propagation Modeling: Principles and Applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7152831/

[28] O. Contributors, "OpenStreetMap Data," 2025, data licensed under the Open Data Commons Open Database License (ODbL), https://opendatacommons.org/licenses/odbl/. [Online]. Available: https://www.openstreetmap.org