Bridging Data Gaps: Enhancing Wireless Localization with Physics-Informed Data Augmentation

ANONYMOUS AUTHOR(S)

Machine learning (ML) models have emerged as the state-of-the-art approach for wireless applications such as transmitter localization. One of the challenges for ML models, however, is their reliance on the abundance of high quality data. Specifically for localization, a significant challenge is to obtain training data that "covers" the entire landscape, in order to ensure a high accuracy for the ML approaches. This issue is compounded when trying to localize multiple transmitters. To address this problem, we introduce a new data augmentation pipeline, termed Physics-informed Augmentation and RF Modeling (PhARMNet), that can combine existing data with a physics-based simulation model, producing a larger dataset with an improved coverage of the landscape of interest. Our results show that PhARMNet offers significant advantages over traditional path loss models. For localization, we demonstrate that augmenting training data with PhARMNet-produced samples improves localization accuracy, particularly in out-of-distribution regions and multi-transmitter settings.

Additional Key Words and Phrases: data augmentation, propagation modeling, localization, machine learning

ACM Reference Format:
Anonymous Author(s). 2018. Bridging Data Gaps: Enhancing Wireless Localization with Physics-Informed Data Augmentation. *J. ACM* 37, 4, Article 111 (August 2018), 13 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Machine learning (ML) methods have been widely adopted for networking applications such as transmitter localization, where models learn to perform localization in a data-driven manner [12, 21, 22]. In contrast to physics based triangulation methods, data driven methods do not require detailed descriptions of the physical characteristics of the landscape, antennas, etc. Thus, the same model architecture can be trained and used in multiple environments, and potentially be extended to more challenging settings such as multi-transmitter localization.

However, ML methods often come with a significant drawback: models require a large amount of high quality training data in order to generalize (i.e., perform well on unseen inputs). Obtaining large amounts of training data can be expensive, and the data obtained may not be "exhaustive". Data-driven models are also known to struggle when exposed to inputs that differ from the training data, known as out-of-distribution (OOD) samples [2, 9, 11]. Such distribution shifts are common in real-world networking (e.g., due to environmental changes, transmitter/receiver hardware updates, etc.). This creates a critical gap between available training data and the complexity of the real-world scenarios, making it challenging to deploy ML models reliably in localization and also in other networking applications [17].

Data augmentation is a common strategy to address data scarcity in machine learning [3, 6]. In computer vision applications, augmentation relies on techniques such as dropout, blurring, or rotations. However, in our context, a key cause of data gaps is the absence of training data from certain regions of interest, where conventional augmentation methods are insufficient as

ACM 1557-735X/2018/8-ART111
https://doi.org/XXXXXXX.XXXXXXX

they only perturb or *interpolate between existing samples*. To address this, we propose a Physics-informed Augmentation and RF Modeling Network (PhARMNet). PhARMNet integrates insights from training samples with physics-based features to create a neural network model for data augmentation for wireless applications. As illustrated in Fig. 1, PhARMNet uses environment maps and the TIREM propagation model [4, 15, 18] to generate new data. This data is then used to train an improved localization model. Our results are outlined below, in Section 1.1.

Beyond localization, we expect data-driven propagation modeling and received signal strength (RSS) prediction to be essential for other wireless networking applications, including interference estimation, spectrum management, and dynamic spectrum allocation [13]. In these applications, reliable RSS prediction supports tasks such as link budget estimation, interference mapping, and real-time decisions for channel access and power control in dynamic and mobile networks [15, 20]. RSS prediction is a key component in emerging technologies like Digital Spectrum Twins (DSTs), which use accurate propagation modeling to optimize spectrum utilization in shared environments such as the Citizens Broadband Radio Service (CBRS) band [15]. Exploring these downstream applications is an interesting avenue for future research.

1.1 Our results

We present a systematic study of augmentation methods for wireless measurement data, using transmitter localization as the main application. There are two natural approaches for augmentation: (a) physical simulation, and (b) data-driven interpolation. Our main result is to show that a *combination* of the two approaches gives the best results. We evaluate our claims both by measuring the accuracy of augmentation directly (using hold-out data), and by measuring the performance of a localization algorithm trained using the augmented dataset. Our main contributions are as follows:

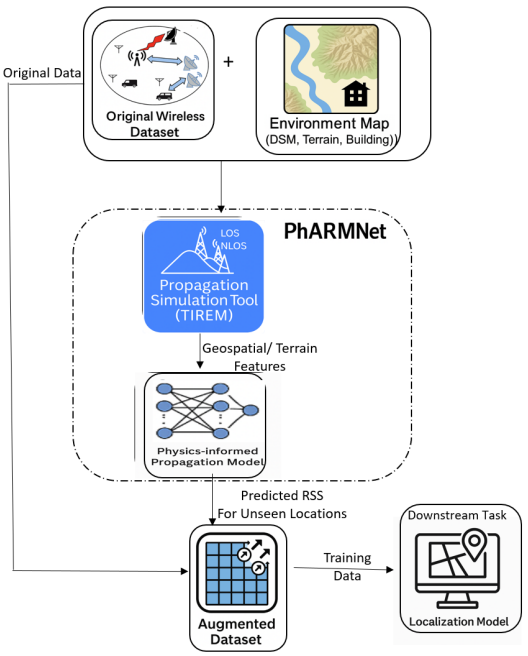


Fig. 1. Overview of PhARMNet

PhARMNet and Localization. We develop an augmentation method that takes in a dataset of received signal strength (RSS) measurements and environment maps, uses derived environmental features from the physics-based modeling software TIREM [4], to produce an augmented dataset of RSS measurements. Using holdout data for evaluation, we show that the RSS measurements produced via this hybrid method outperform purely physics (TIREM-based) methods, as well as purely data-driven interpolation methods by significant margins (see Section 5). We show that it translates to 20% - 40% improvement (on average) in downstream localization tasks.

Two-transmitter localization. A particular challenge for ML-based localization is the multi-transmitter setting. Here, the issue of training data "covering" the space becomes more severe; to see this, suppose have n grid cells where transmitters may be located. There are $\binom{n}{2}$ different 2-TX pairs, each of which may cause distinct interference patterns at the receivers, so unless they are all *covered* by the training data, ML methods may perform poorly. Furthermore, in our datasets, the

number of 2-TX samples is much smaller (only around 350 samples). Here, we design a variant of the PhARMNet model, and show that augmentation improves the 2-TX identification accuracy from under 50% to over 70%. We also observe that PhARMNet captures interference patterns more accurately than simple superposition.

We conducted all our experiments on three publicly available city-scale localization datasets [1, 8, 14], which provide diverse urban environments and reliable RSS measurements. In addition to the results above, we also demonstrate that augmentation allows us to adapt localization models to more dynamic settings, e.g., when new sensors are deployed. We also study the question of *how much* of synthetic data to add in order to obtain optimal results.

2 Background and Related Work

Data augmentation is a classic technique in Computer Vision and ML, often applied to improve model accuracy by increasing training diversity, helping models generalize to unseen data [5, 16]. In the context of wireless localization, very few works have applied data augmentations. In [10], Mitchell et al. apply adversarial training to a CNN-based localization model to improve robustness to attack. LLOCUS [12] interpolates RSS values to a set of fixed locations, but this was done to enable localization using mobile sensors rather than to improve model generalization. Several works use synthesized RSS propagation data to train localization models [22], [23], but these works also evaluate the localization model on the same synthetic data, failing to address the question of augmentation and improving model generalization. Recent works such as RadioUNet [7] have used CNNs for path loss prediction to train another CNN model on synthetic data. However, the practical application of this approach remains entirely untested due to the lack of real-world datasets for evaluation and the oversimplification of the training environments. Our work addresses whether a localization model can be accurate in real-world scenarios when trained using synthetic data.

Terrain Integrated Rough Earth Model (TIREM): TIREM is a widely used RF propagation model for estimating large-scale path loss by considering terrain profiles, surface roughness, atmospheric conditions, and antenna heights [4]. Supporting frequencies from 20 MHz to 1 THz, TIREM models both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios. TIREM simplifies the environment to a two-dimensional elevation profile, ignoring multipath reflections, scattering, and antenna directionality. These assumptions limit its accuracy in complex propagation phenomena. Nevertheless, TIREM remains efficient for large-scale modeling where ray tracing is impractical.

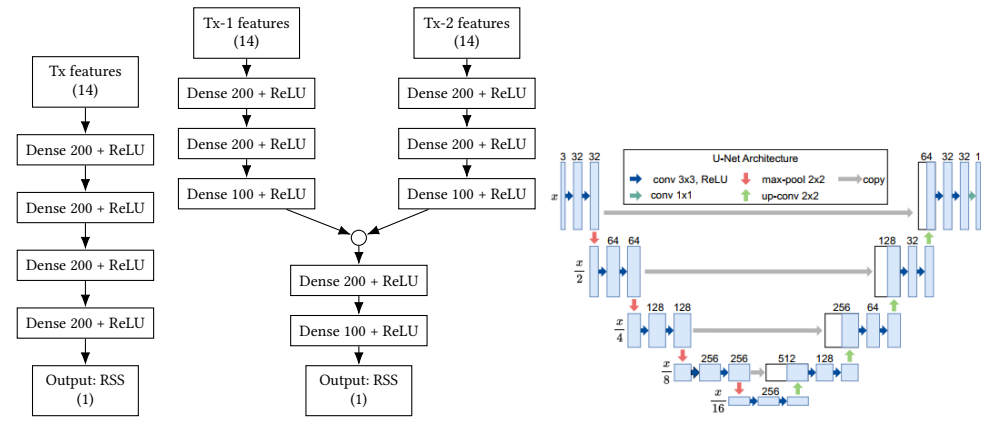
3 PhARMNet: Augmentation Framework and Localization

This section presents our PhARMNet framework, at the architecture level. The architecture has three key components: (i) the feature extraction step that utilizes the DSM (terrain map) of the domain of interest and computes environment and physics-aware features for each transmitter-receiver pair, (ii) the PhARMNet propagation model, a neural network that learns to predict RSS values using the collected data and the physics-based features, (iii) the dataset augmentation procedure and the downstream localization model. PhARMNet takes as input the base station (receiver) coordinates, a DSM terrain map of the propagation area, training data with transmitter coordinates and receiver RSS values, and outputs an augmented dataset. This dataset is used to train an improved localization NN model.

(i) Feature Extraction using TIREM. Given DSM data, we use the TIREM [4] to model our propagation environment and calculate a set of 14 geometric and propagation features. Specifically, given a pair (tx, rx), the generated features using TIREM capture the propagation environment between tx and rx. These features consider both line-of-sight (LOS) and non-line-of-sight (NLOS) conditions and include: diffraction parameters (3), diffraction and elevation angles (5), LOS/NLOS indicator

(1), number of "knife edges" (1), number of obstacles (1), and shadowing angles (2). A detailed breakdown of each extracted feature and their expression is provided in Appendix A. These features are computed for each transmitter-receiver pair in the training data, as well as for candidate receiver locations used during synthetic augmentation. Since this is computationally intensive for city-scale domains, we develop a parallel implementation by dividing the domain into sub-grids and focusing only on regions where augmentations are placed. We could thus reduce the computation time from hours to under 45 seconds on a 12-core AMD Ryzen 9 3900*X* CPU on all of our datasets.

(ii) PhARMNet Propagation Model. The next step in our augmentation framework takes the vector of features obtained from step (i) (for a given tx, rx pair) and produces an RSS value. For this, we use a simple shallow NN architecture, shown in Figure 2a. It consists of four intermediate layers of 200 neurons each. For the case of two-transmitter localization, we use a "multi head" architecture (Fig. 2b). The choice of hyperparameters (number of intermediate neurons and number of layers) was made based on manual tuning given the data and domain size, and developing auto-tuning methods is an interesting direction for further research. We use the given data to train the model, using a standard Adam optimizer.



(a) PhARMNet for 1-TX (b) PhARMNet for 2-TX (c) UNet model for localization

Fig. 2. PhARMNet architectures for single and two transmitters, and the UNet localization model.

(iii) Synthetic Data Augmentation and Localization NN. Once the PhARMNet propagation model is obtained, the dataset is augmented with RSS values corresponding to (tx, rx) pairs for receivers of interest, and tx locations that are not close to a sufficient number of transmitters in the training data. For the two transmitter case, a similar set of synthetic data points is generated and a random sample of desired size is used for augmentation.

We adopt the state-of-the-art UNet-based localization model proposed in [9] as the downstream localization model for our PhARMNet augmented dataset. As an aside, we were able to improve the model performance by introducing non-linear calibration using bspline (as opposed to a simple linear scaling used in [9]). We also modified the output calculation of the model by introducing clustering instead of just peak selection to make it suitable for multi-transmitter case. We describe the modifications

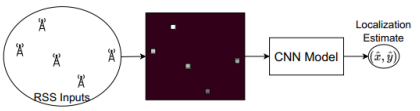


Fig. 3. CNN pipeline: RSS measurements are converted to an image and input to CNN for localization.

in Appendix B. The localization pipeline follows the standard paradigm used in recent works [7, 11, 21, 22], where RSS measurements are first converted into a 2D image representation such as Fig. 3, and the output is an image showing the heatmap for the localization. (This enables the use of an image-to-image architecture such as UNet.) In training, we use the center-of-mass (CoM) approach from [21] to obtain the predicted device coordinates from the UNet's output heatmap.

4 Evaluation Methodology

We begin with an overview of our evaluation framework. We will first introduce the three large-scale outdoor datasets that we use for our experiments. Our evaluation will have two parts. First, we evaluate the quality of the data augmentation, and second, we will evaluate the effect of augmentation on the downstream task of transmitter localization.

4.1 Evaluation Datasets

We use three real-world datasets (summarized in Table 1). Each dataset is associated with a high-resolution DSM (i.e., a terrain map), with resolution in the range 0.5–1m.

Table 1. Dataset parameters. *Num. Inputs* denotes the number of active receivers in the samples.

	DS1 [8]	**DS2 [14]**	**DS3 [1]**
Freq. [MHz]	462.7	3543	868
Area [km^2]	2.2×2.4	1.4×1.4	6.1×5.6
Receiver Sensors	23	6	15
Mobile Sensors	0-4 (Rx)	0	0
Num. Inputs	9-23	4-5	5-10
One-Tx Samples	4,577	15,613	10,032
Two-Tx Samples	346	0	0
RSS Samples	95,517	77,639	57,777

DS1 (POWDER-FRS [8]): Campus-scale FRS-band (462.7 MHz) signals from mobile and stationary sensors. The data includes RSS measurements at multiple receivers, and it contains samples with one (1-TX) and two (2-TX) transmitters. The terrain is complex with significant elevation differences and a campus environment with buildings.
DS2 (POWDER-CBRS [14]): CBRS-band (3.5 GHz) from fixed transmitters with sparse RSS measurements (4–5 active receivers). This dataset also covers scenarios where the active receivers are *all* far from transmitter.
DS3 (Antwerp [1]): Urban LoRaWAN traces at 868 MHz from a mobile transmitter. This dataset is the most complex due to occlusions and urban canyons. We use the DSM provided by [19].

4.2 Evaluation of the Augmentation

We compare PhARMNet with three established methods for RSS estimation, that we first describe briefly. The first two methods are data-driven, using interpolation techniques to predict the RSS value at a certain location. The last method is physics based, as discussed earlier.

(1) RBF Interpolation: Uses SciPy's `RBFInterpolator` with a linear kernel ($S = 1$). To extend beyond the convex hull of training points, RSS values are first predicted at the region's corners using a log-distance model. For the 2-TX case, individual predictions are obtained as above, and combined linearly in the power domain.

(2) CELF [20]: Applies Bayesian linear regression to learn a spatial loss field over residuals from a log-distance model. Path loss is computed by integrating the learned field along direct paths. CELF requires only 2D locations, enabling flexible mobility settings without DSM input.

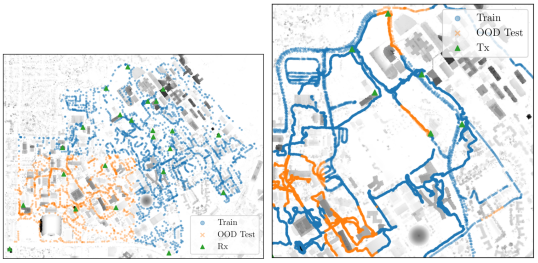
(3) TIREM [4]: Classical model using knife-edge diffraction over terrain profiles. The model uses high-resolution DSM to extract features and leverages them for prediction. Care is taken to align the locations from the RSS data with the DSM before feeding them to TIREM.

Evaluation. The accuracy of the augmentation is evaluated using a natural hold-out approach: a subset of the training data is hidden from the models during training, and the accuracy is obtained

by comparing model's output at these points to the true RSS measurements. Our experiments evaluate different choices for the holdout set, as we now explain. The full results are presented in Section 5.

4.3 Selecting Holdout Data

As we discussed earlier, the difficulty in data augmentation comes from the fact that certain regions of the domain may not be well-represented in the training data. To emulate this in our evaluation, we partition the domain of our training data into an $N \times N$ grid of regions, and we hold out all transmitter data from a subset of the regions (roughly 20% of them, similar to [9, 21]). For the purposes of the ML training process, data from these regions can be viewed as being "out-of-distribution" (OOD). Figure 4 shows examples with $N = 2$ and $N = 5$ for two of our datasets.



(a) DS1 with 2×2 grid (b) DS2 with 5×5 grid

Fig. 4. Device and base station locations with grid-based holdout. Training (blue), OOD (orange), base stations (green).

5 Evaluation of Data Augmentation

We evaluate PhARMNet's RSS prediction performance against the three baselines discussed above: RBF interpolation, CELF [20], and TIREM [4].

5.1 Qualitative Comparison for a Fixed Transmitter

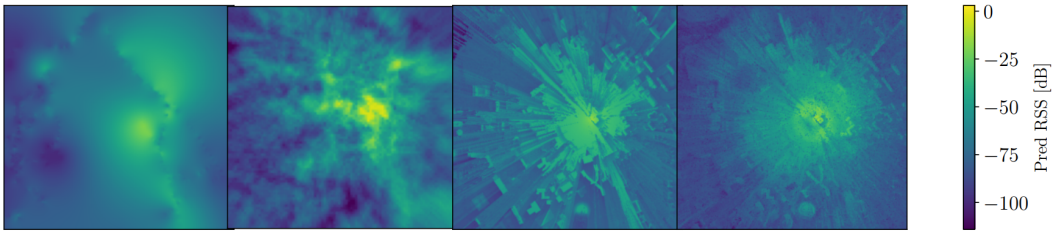
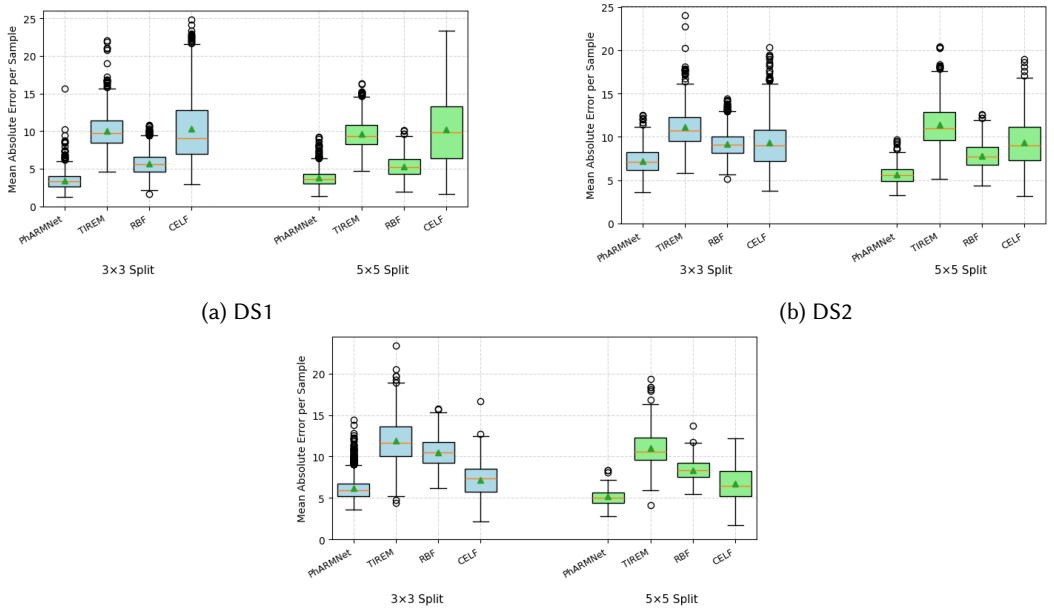


Fig. 5. Predicted RSS over DS2 (5×5 grid, 1-TX). Left to right: RBF, CELF, TIREM, PhARMNet.

Before delving into quantitative results, we show in Figure 5 the predicted RSS values at various receiver locations, from a fixed transmitter. RBF and CELF are data-driven, but do not use the DSM terrain data, while TIREM uses only DSM. Note that RBF yields smooth surfaces but ignores obstacles. CELF provides a better structure but lacks DSM awareness. TIREM models occlusions but underpredicts in shadowed zones. PhARMNet balances obstacle-sensitivity and spatial smoothness, preserving edges while reducing artifacts. Its environmental features and training on real measurements enable strong generalization in unseen regions, that as we will see quantitatively, outperform purely geometric or data-driven methods.



(a) DS1

(b) DS2

(c) DS3

Fig. 6. Mean absolute RSS error (in dB) on OOD test samples across three datasets.

5.2 RSS Prediction Results (1 TX)

We evaluate the generalization ability of PhARMNet against three baselines—RBF interpolation, CELF, and TIREM—on hold-out (or out-of-distribution (OOD)) test samples generated via spatial splits (3×3, 5×5; see Section 4.3) in the 1-TX setting. The boxplots in Fig. 6 present the distribution of absolute RSS prediction errors per sample across all test configurations. PhARMNet consistently achieves the lowest median and lower spread of errors, indicating both accuracy and robustness. The classical TIREM model, while physics-based and stable, lacks real-world adaptability and tends to over-simplify the RSS prediction. CELF, though modeling a spatial shadowing field using Bayesian regression, suffers in DS1 and DS2 due to limited training density. Surprisingly, RBF interpolation performs competitively on average, benefiting from the high spatial continuity in densely measured regions.

However, RBF fails sharply in spatially isolated regions. As shown in Fig. 7, for DS3's Sensor 15, the RSS values are dramatically underpredicted by RBF—often exceeding 80 dB in error—due to lack of nearby ground-truth points. PhARMNet, in contrast, leverages geospatial features such as building obstruction, terrain angle, and LOS/NLOS indicators to generalize even in these low-coverage zones. Together, these results confirm a clear trend: while RBF can interpolate well within the dense training regions, PhARMNet provides consistent and realistic predictions in all our evaluation settings.

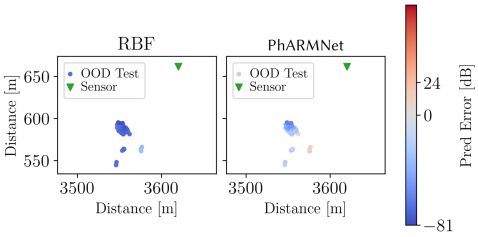
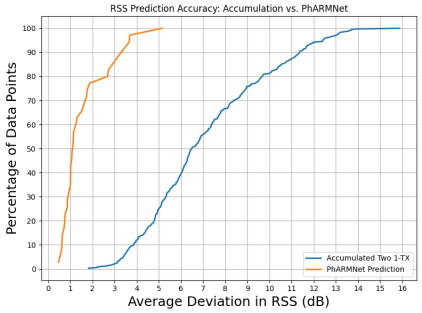
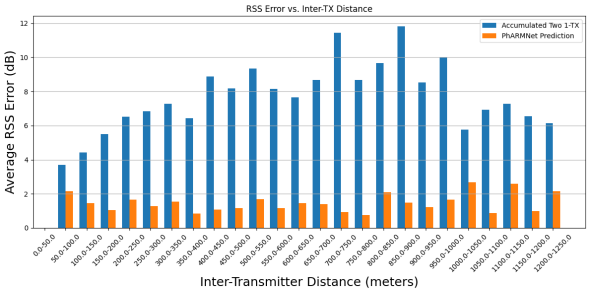


Fig. 7. RBF and PhARMNet prediction error for OOD test samples for Sensor 15 in DS3 (5×5 split). In the absence of nearby training samples, RBF underpredicts severely, while PhARMNet maintains lower error.

5.3 Modeling Superposition



(a) CDF of RSS prediction error: PhARMNet vs. accumulated 1-TX.



(b) RSS error vs. inter-TX distance.

Fig. 8. PhARMNet vs. naive accumulation on 2-TX RSS prediction.

We assess PhARMNet's abitlity to model complex multi-transmitter propagation, by predicting RSS in 2-TX scenarios and comparing it against a naive baseline: log-domain accumulation of two spatially matched 1-TX samples from the corresponding single-transmitter samples: $\text{RSS}_{\text{combined}} = 10\log_{10}\left(10^{\frac{\text{RSS}_1}{10}} + 10^{\frac{\text{RSS}_2}{10}}\right)$. Figure 8a shows that PhARMNet yields significantly lower prediction error, with over 90% of samples under 3 dB error. In contrast, the accumulation baseline (which is perfect if we assume the additivity of power) has a median error over 6 dB, with large deviations in many cases. As illustrated in Figure 8b, PhARMNet maintains consistent accuracy regardless of transmitter separation, even in high-interference cases. Meanwhile, the naive method degrades as the distance increases, unable to capture complex signal interactions. These results confirm PhARMNet's effectiveness in modeling real-world multi-transmitter propagation beyond linear signal superposition.

5.4 Use Cases and Trade-offs of RSS Estimation Models

Each RSS model presents distinct strengths depending on deployment conditions. RBF performs best with dense, in-distribution data, and requires no DSM, but has high inference cost. CELF offers efficient DSM-free predictions with moderate generalization in low-data settings. TIREM provides fast, low-data estimates based on terrain profiles but lacks adaptability. PhARMNet excels in OOD regions, especially near isolated base stations, using minimal data and DSM-derived features. The pros/cons of different methods are shown in Table 2; method selection should align with environmental constraints and computational and data resources.

Table 2. Use-case summary of RSS prediction methods.

Method	Runtime [s]		Needs DSM	Data Needs
	Train	Infer		
RBF	9.8	65	No	High
CELF	2.8	1.1	No	Medium
TIREM	< 300	< 1	Yes	Minimal
PhARMNet	< 800	3.8	Yes	Low

6 Evaluation of Localization with PhARMNet Augmentation

We now evaluate how data augmentation using PhARMNet improves localization accuracy. We use the localization model from [9] for our experiments.

Table 3. Mean localization error (MLE) in meters, across five trials using different training settings and spatial splits for the hold-out. For settings with 2TX samples, both MLE and the percentage of samples where both transmitters were correctly identified (2TX-Acc) are reported, latter in parentheses.

Dataset	Training Setting	Split	NoAug	PhARMNet	CELF	RBF	TIREM-FS
DS1	True: 1TX + Aug.: 1TX	3×3	379	265	347	328	348
		5×5	185	162	177	174	275
	True: 1TX + 2TX + Aug.: 1TX + 2TX	3×3	255 (47.08%)	175 (70.13%)	–	209 (59.17%)	233 (41.03%)
		5×5	172 (53.73%)	146 (73.44%)	–	177 (66.18%)	179 (46.07%)
	True: 1TX + 2TX + Aug.: 2TX Only	3×3	255 (47.08%)	181 (72.58%)	–	229 (62.81%)	241 (42.16%)
		5×5	172 (53.73%)	138 (81%)	–	189 (69.63%)	186 (47.04%)
DS2	True: 1TX + Aug.: 1TX	3×3	353	235	292	295	282
		5×5	216	205	252	256	239
DS3	True: 1TX + Aug.: 1TX	3×3	1120	935	1065	1030	1109
		5×5	644	602	609	670	593

6.1 Improvement in Localization Accuracy

Table 3 shows that PhARMNet consistently improves mean localization error (MLE), especially in grid-based OOD splits. For instance, in DS1 (3×3), MLE drops from 379 m to 265 m. As a technical detail, for DS1, sensor-specific models are used as there were three categories of sensors; DS2 and DS3 employ a single model.

In samples with two transmitters(2TX), augmentation using PhARMNet improves both the estimation error, and the 2TX identification accuracy (2TX-Acc). The latter rises from 47% to over 80% in some cases. RBF based augmentation shows a comparable MLE but lower 2TX-Acc, highlighting its weakness in modeling interference. TIREM performs poorly, presumably because it is purely physics based and not data-driven. CELF is excluded as it does not support multi-TX input. Two augmentation strategies were tested: full (1TX+2TX) and 2TX-only. The latter improves 2TX-Acc slightly by reducing 1TX–2TX imbalance in the dataset, though it can increase MLE in unseen 1TX regions. This trade-off illustrates the need to balance synthetic coverage. Another spatial error improvement analysis is discussed in Appendix C.

6.2 Localization in Isolated Regions

We assess PhARMNet's utility in "isolated" locations —-test points that are over 200 m from any training sample– where interpolation methods fail. Table 4 shows that PhARMNet augmentation significantly improves localization, reducing error by up to 192 m (42%) in these regions. While traditional models falter in data-sparse zones, PhARMNet's terrain-informed predictions effectively bridge gaps, enabling accurate inference even in unobserved locations.

Table 4. Mean localization error (m) over five trials with and without PhARMNet augmentation. "200m Avg" refers to isolated test points that are more than 200m from any training samples.

Setting	No Aug		PhARMNet Augmented	
	Avg	200m Avg	Avg	200m Avg
DS1 3×3	379	471	265	302
DS1 5×5	185	258	162	116
DS2 3×3	353	385	235	182
DS3 3×3	1120	861	935	652
DS3 Rand.	294	720	278	205

6.3 Augmentation for Localization in Dynamic Environments

PhARMNet supports improved localization in *adaptive* or dynamic settings, enabling models to incorporate newly deployed base stations without additional data collection. We randomly select 20% of the base stations in each (1-TX) dataset, treat them as newly deployed, and compare three models: (i) **non-adaptive**—trained without any RSS inputs from those receivers, (ii) **adaptive**—augmented using PhARMNet-generated RSS, and (iii) **full-data**—trained with all ground-truth.

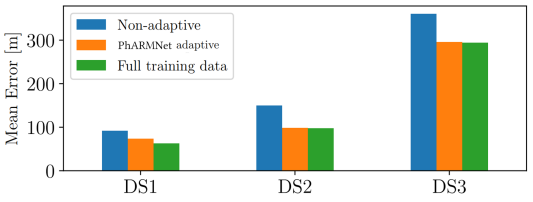


Fig. 9. Mean localization error across five trials: non-adaptive, PhARMNet-adaptive, and full-data models.

As shown in Figure 9, PhARMNet augmentation recovers much of the performance gap. In DS2 and DS3, the adaptive model is within 2 m of the full-data model. In DS1, it yields a 19 m (21%) improvement over the non-adaptive baseline. Across 15 trials, PhARMNet improved accuracy in 14, and was within 1 m of the baseline in the remaining case. Gains are especially prominent in DS2, which has only 5 RSS inputs per sample, making each augmentation more impactful. These results highlight PhARMNet's ability to synthesize the measurements, supporting fast adaptation to new infrastructure without retraining on real data.

6.4 How much Synthetic Data?

Using the 2-TX setting, we also conduct a simple experiment where we increase the number of synthetic 2-TX samples used in training, while keeping 80% of the real 2-TX training data (around 300 samples; using the other 20% for validation). All the 1-TX samples are used during training. As seen in Figure 10, without augmentation, 2-TX identification has accuracy under 50%. Accuracy improves rapidly with added PhARMNet samples (up to ~ 600 samples), saturating around 80–85%.

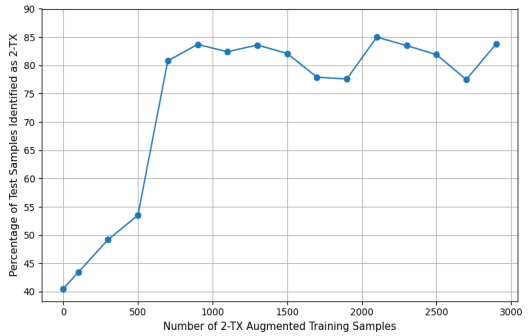


Fig. 10. Varying number of synthetic samples (fixed 80% true 2-TX).

7 Conclusion

We presented PhARMNet, a physics-informed signal propagation modeling framework that combines terrain-aware simulation with data-driven learning. PhARMNet not only predicts RSS in real-world environments but also serves as an effective data augmentation engine to enhance localization models. Our experiments demonstrate that augmentations generated by PhARMNet significantly improve localization accuracy, particularly in out-of-distribution (OOD) and multi-transmitter settings. We also show that PhARMNet enables adaptive localization by synthesizing RSS data for newly added base stations, achieving near-parity with fully supervised models without requiring additional measurements. By bridging classical propagation theory with modern machine learning, PhARMNet addresses key challenges in scalable, realistic localization and opens the door for broader use of physics-augmented learning in wireless systems.

References

[1] Michiel Aernouts, Rafael Berkvens, Koen Van Vlaenderen, and Maarten Weyn. 2018. Sigfox and LoRaWAN datasets for fingerprint localization in large urban and rural areas. *Data* 3, 2 (2018), 13.

[2] Maximilian Arnold and Mohammed Alloulah. 2022. Benchmarking learnt radio localisation under distribution shift. *arXiv preprint arXiv:2210.01930* (2022).

[3] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. 2019. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 113–123.

[4] D. Eppink and W. Kuebler. 1994. "Tirem/sem handbook , tech. rep ELECTROMAGNETIC COMPATIBILITY ANALYSIS CENTER ANNAPOLIS MD5. https://apps.dtic.mil/sti/tr/pdf/ADA296913.pdf

[5] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*. Vol. 1. MIT press Cambridge.

[6] Sven Gowal, Sylvestre-Alvise Rebuffi, Olivia Wiles, Florian Stimberg, Dan Andrei Calian, and Timothy A Mann. 2021. Improving robustness using generated data. *Advances in neural information processing systems* 34 (2021), 4218–4233.

[7] Ron Levie, Çağkan Yapar, Gitta Kutyniok, and Giuseppe Caire. 2021. RadioUNet: Fast radio map estimation with convolutional neural networks. *IEEE Transactions on Wireless Communications* 20, 6 (2021), 4001–4015.

[8] Frost Mitchell, Aniqua Baset, Sneha Kumar Kasera, and Aditya Bhaskara. 2022. A dataset of outdoor RSS measurements for localization. *Zenodo* (2022).

[9] Frost Mitchell, Neal Patwari, Aditya Bhaskara, and Sneha Kumar Kasera. 2023. Learning-based techniques for transmitter localization: A case study on model robustness. In *2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 133–141.

[10] Frost Mitchell, Phillip Smith, Aditya Bhaskara, and Sneha Kumar Kasera. 2023. Exploring adversarial attacks on learning-based localization. In *Proceedings of the 2023 ACM workshop on wireless security and machine learning*. 15–20.

[11] Frost Mitchell, Jie Wang, Aditya Bhaskara, and Sneha Kumar Kasera. 2024. Utilizing Confidence in Localization Predictions for Improved Spectrum Management. In *2024 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 483–492.

[12] Shamik Sarkar, Aniqua Baset, Harsimran Singh, Phillip Smith, Neal Patwari, Sneha Kasera, Kurt Derr, and Samuel Ramirez. 2020. Llocus: learning-based localization using crowdsourcing. In *Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*. 201–210.

[13] Serhat Tadik, Kaitlyn M Graves, Michael A Varner, Christopher R Anderson, David M Johnson, Sneha Kumar Kasera, Neal Patwari, Jacobus Van der Merwe, and Gregory D Durgin. 2023. Digital spectrum twins for enhanced spectrum sharing and other radio applications. *IEEE Journal of Radio Frequency Identification* 8 (2023), 376–391.

[14] S Tadik, A Singh, F Mitchell, Y Hu, X Yao, K Webb, A Sarbhai, D Maas, A Orange, J Van der Merwe, et al. 2024. Salt Lake City 3534 MHz multi-transmitter measurement campaign.

[15] Serhat Tadik, Michael A Varner, Frost Mitchell, and Gregory D Durgin. 2023. Augmented rf propagation modeling. *IEEE Journal of Radio Frequency Identification* 7 (2023), 211–221.

[16] Rohan Taori, Achal Dave, Vaishaal Shankar, Nicholas Carlini, Benjamin Recht, and Ludwig Schmidt. 2020. Measuring robustness to natural distribution shifts in image classification. *Advances in Neural Information Processing Systems* 33 (2020), 18583–18599.

[17] Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. 2018. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 969–977.

[18] Michael A Varner, Thomas A Rodriguez, and Gregory D Durgin. 2022. Rf coverage mapping of bistatic radio links using the terrain integrated rough earth model (tirem). In *2022 16th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 1–5.

[19] Agentschap Digitaal Vlaanderen. 2014. Digitaal Hoogtemodel Vlaanderen II, DTM, raster, 5 m.

[20] Jie Wang, Meles G Weldegebriel, and Neal Patwari. 2024. Channel estimation via loss field: Accurate site-trained modeling for shadowing prediction. In *2024 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 312–321.

[21] Çağkan Yapar, Ron Levie, Gitta Kutyniok, and Giuseppe Caire. 2023. Real-time outdoor localization using radio maps: A deep learning approach. *IEEE Transactions on Wireless Communications* 22, 12 (2023), 9703–9717.

[22] Caitao Zhan, Mohammad Ghaderibaneh, Pranjal Sahu, and Himanshu Gupta. 2022. Deepmtl pro: Deep learning based multiple transmitter localization and power estimation. *Pervasive and Mobile Computing* 82 (2022), 101582.

[23] Anatolij Zubow, Suzan Bayhan, Piotr Gawłowicz, and Falko Dressler. 2020. Deeptxfinder: Multiple transmitter localization by deep learning in crowdsourced spectrum sensing. In *2020 29th International Conference on Computer Communications and Networks (ICCCN)*. IEEE, 1–8.

A Terrain and Environmental Features

This appendix provides formal definitions and mathematical expressions for the 14 terrain-aware and physics-based propagation features extracted using the TIREM engine, used to inform both real and synthetic training data within the PhARMNet framework.

1. Diffraction Parameters (2): These parameters quantify signal obstruction due to knife-edge diffraction at terrain or building discontinuities using normalized clearance: $v = \frac{h-h_l}{\sqrt{\frac{\lambda d_1 d_2}{d_1+d_2}}}$ where: v is the Fresnel diffraction parameter, h is the height of the obstacle apex, h_l is the line-of-sight height at the same horizontal position, d_1 and d_2 are distances from transmitter and receiver to the obstacle, λ is the signal wavelength. From v, the diffraction loss L_d is approximated via:

$$L_d(v) = \begin{cases} 0 & v \le -0.78 \\ 6.9 + 20\log_{10}\left[\sqrt{(v-0.1)^2+1} + v - 0.1\right] & v > -0.78 \end{cases}$$

We extract two features per link: (i) maximum v value, and (ii) total cumulative diffraction loss L_d across all obstacles.

2. Max Obstacle Diffraction (1): This feature measures the normalized clearance v at the most significant obstacle: $v_{\max} = \max_i v_i$,where i indexes the obstacles along the Tx-Rx path.

3. Diffraction Angles (2): We compute the angular spread at obstacles as: $\theta_{d,i} = \tan^{-1}\left(\frac{h_i - h_t}{d_{ti}}\right) + \tan^{-1}\left(\frac{h_i - h_r}{d_{ri}}\right)$,where: h_i is the obstacle height, h_t and h_r are transmitter and receiver heights, d_{ti} and d_{ri} are distances from tx/rx to the obstacle. We report: (i) average $\theta_{d,i}$ and (ii) max $\theta_{d,i}$.

4. Elevation Angles in NLOS (2): For NLOS paths, we define the elevation angles between the transmitter/receiver and the first obstructing object: $\theta_t^{\text{NLOS}} = \tan^{-1}\left(\frac{h_{\text{obs}} - h_t}{d_t}\right)$, $\theta_r^{\text{NLOS}} = \tan^{-1}\left(\frac{h_{\text{obs}} - h_r}{d_r}\right)$,where h_{obs} is the height of the first blocking object, and d_t, d_r are distances from the transmitter and receiver to the object.

5. Tx-Rx Elevation Angle (1): The direct elevation angle θ_{tr} between the transmitter and receiver is: $\theta_{tr} = \tan^{-1}\left(\frac{h_r - h_t}{d_{tr}}\right)$, where d_{tr} is the horizontal distance between the Tx and Rx.

6. LOS/NLOS Indicator (1): This binary flag is computed by performing terrain and obstacle ray-intersection analysis. The path is considered line-of-sight (LOS) if: $\forall x \in [x_t, x_r] : h(x) < h_t + \frac{(h_r - h_t)}{d_{tr}}(x - x_t)$. Otherwise, it is non-line-of-sight (NLOS).

7. Number of Knife-Edges (1): A terrain profile is segmented into a series of triangle peaks. Each sharp peak with convex shape and height gradient above a threshold is counted as a knife-edge obstacle: Knife-Edges $= \sum_i \mathbb{1}\left[\frac{h_{i+1} - h_i}{d_{i+1} - d_i} > \tau\right]$ with threshold τ defined empirically (e.g., 15° slope).

8. Number of Obstacles (1): Each terrain or building structure intersecting the direct line-of-sight ray is considered an obstacle. We count: $N_{\text{obs}} = \sum_i \mathbb{1}[h_i > h_l(x_i)]$, where h_i is obstacle height at point x_i and $h_l(x_i)$ is height of the direct LOS ray at x_i.

9. Shadowing Angles (2): These angles capture lateral shadowing of the signal path due to topography: $\theta_{s,L} = \tan^{-1}\left(\frac{h_{L,\max} - h_t}{d_L}\right)$, $\theta_{s,R} = \tan^{-1}\left(\frac{h_{R,\max} - h_t}{d_R}\right)$, where $h_{L,\max}$ and $h_{R,\max}$ are maximum obstacle heights within a lateral window to the left and right of the Tx-Rx path.

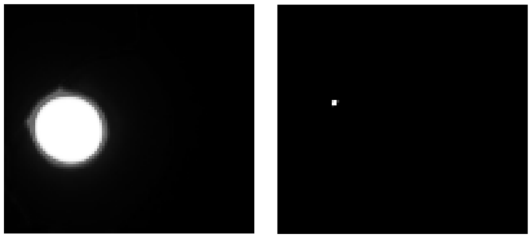
B Improving UNet Localization Model using BSpline and Clustering

Non-Linear Sensor Calibration with B-Splines: In prior work [10] described in section 3, UNet-based RSS localization models used *linear* calibration layers to compensate for hardware differences across sensors by learning per-sensor scale and shift parameters: $\hat{r}_{ij} = w_j \cdot r_{ij} + b_j$. While effective for reducing bias, this approach cannot model non-linear effects such as antenna gain variation

or hardware filtering. To address this, we introduce **B-spline-based** non-linear calibration layers. A B-spline models a smooth, piecewise polynomial transformation: $\hat{r} = \sum_k c_k B_k(r)$, where B_k are fixed basis functions and c_k are trainable weights. Each sensor is assigned an independent calibrator. This architecture allows the model to learn smooth, non-linear corrections, capturing complex real-world distortions.

Sharper Heatmaps with B-Spline Calibration: Figure 11 shows output heatmaps for the same input. Linear calibration produces diffuse, ambiguous regions, whereas B-splines yield sharply localized peaks, improving interpretability and localization precision.

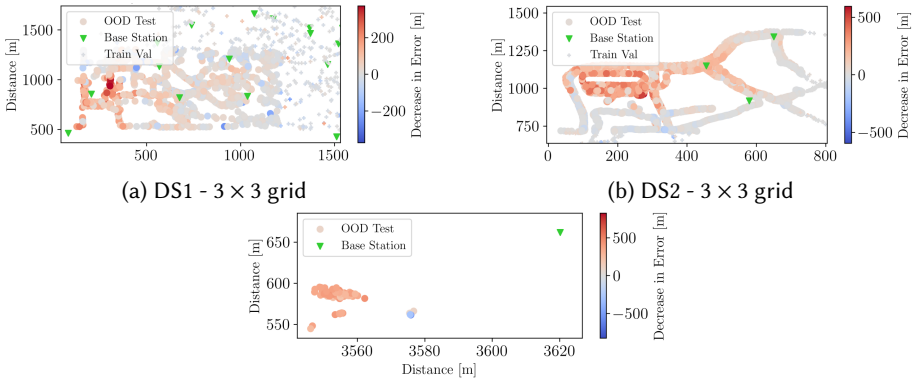
Extension to Multi-TX: Standard UNet uses center-of-mass (CoM) to extract 1-TX locations from heatmaps. In 2-TX settings, this fails due to overlapping peaks. We apply clustering (e.g., DBSCAN) to the output heatmap to detect distinct activation blobs and extract multiple predicted locations. This generalization enables the model to scale to multi-transmitter localization.



(a) Linear calibration (b) B-spline calibration

Fig. 11. Comparison of predicted transmitter heatmaps using linear vs. B-spline calibration. B-spline calibration produces more concentrated and confident outputs.

C Spatial Improvement in Localization Accuracy



(a) DS1 - 3 × 3 grid (b) DS2 - 3 × 3 grid

(c) DS3 - 5 × 5 grid

Fig. 12. Reduction in localization error due to PhARMNet augmentation for OOD test samples across datasets.

Figure 12 shows spatial error improvements. Gains are strongest in regions far from training data using PhARMNet, confirming that physics-aware augmentation compensates where data-driven methods fail. Some degraded samples are attributed to random variation in the trained models.