On the Effectiveness of Radio Environmental Map Prediction Through Image-Based Deep Learning

Mohammad Hossein Zadeh, Marina Barbiroli, Enrico Maria Vitucci, *Senior Member, IEEE*, Vittorio Degli-Esposti, *Senior Member, IEEE*, and Franco Fuschini

***Abstract*—This article investigates the use of deep learning (DL) techniques for predicting 2-D radio environment maps in urban microcellular scenarios. Starting from the common framework established by prior works, a comprehensive and systematic analysis is conducted to evaluate the impact of multiple factors—including neural network (NN) architecture, auxiliary input data, data augmentation strategies, and training dataset fidelity—on the prediction performance. A modified U-Net architecture, referred to as DeepUNet, is proposed, incorporating hybrid loss functions and deep supervision to enhance learning effectiveness. Extensive experiments are carried out by training the model via synthetically generated data using standard ray tracing (RT), as well as multiple types of auxiliary data, including empirical path-loss (PL) predictions, line-of-sight (LoS) predictions, and sparsified or interpolated radio environment maps. Results demonstrate that, by properly combining the aforementioned techniques, the proposed DL model achieves high prediction accuracy, even in challenging configurations. Furthermore, generalization capabilities are evaluated by training and testing on different RT datasets. A detailed performance analysis is provided to offer practical insights and design guidelines for future radio environment map prediction tasks.**

***Index Terms*—Convolutional neural networks (CNNs), deep learning (DL), radio environment map, ray tracing (RT), U-Net.**

I. Introduction

RADIO environmental maps (REMs) represent the spatial distribution of radio propagation parameters—typically received power (P_r) or path loss (PL) over a target service area. Both P_r and PL are key quantities in wireless propagation analysis, and on a logarithmic scale (dB), they are essentially equivalent up to a constant offset. P_r is affected by various factors such as free-space propagation and wave interaction with obstacles such as reflections, diffraction, and scattering, resulting in large- and small-scale fading effects [1].

REM can support wireless network optimization and planning, enabling efficient spectrum management and reliable connectivity. In fact, they help identifying wireless coverage holes, and therefore, network operators can enhance the quality of service by strategically deploying additional infrastructure [2]. In dense urban environments, REMs are used to optimize resource allocation, ensuring efficient use of bandwidth and power for cellular and IoT networks [3]. They can also play a critical role in localization, where accurate REMs improve positioning accuracy for GPS-denied environments [4]. In the framework of emerging technologies and applications, REMs are beneficial in path planning for unmanned aerial vehicles (UAVs), ensuring continuous connectivity during autonomous flights [5], [6]. Moreover, they are increasingly relevant in physical layer security, where knowledge of RF propagation assists in detecting potential eavesdropping or jamming threats [7]. Finally, autonomous driving can rely on accurate REM prediction to maintain stable vehicle-to-everything (V2X) communication, ensuring real-time data exchange for traffic management and safety-critical applications [8]. In the context of the ongoing evolution of wireless communication systems, the importance of REMs extends beyond traditional mobile networks, making them a valuable asset in next-generation technologies such as 5G, 6G, and smart city deployments.

In order to effectively support such a broad set of wireless services and applications, REMs must be point-specific, i.e., they should track the difference in propagation over the multitude of spatial locations spread all over the maps. To this aim, propagation data to fill the REMs can be collected either through extensive measurement campaigns or deterministic field prediction models, like ray tracing (RT) or ray launching (RL). Conversely, statistical/empirical propagation models based on simple, analytical formulas (e.g., the Okumura–Hata model [9]) are not well fit for the task, as they primarily describe the average range dependence of the target propagation marker (P_r or PL), i.e., they account for neither shadowing nor multipath effects.

Unfortunately, field measurements often require expensive equipment, while RT/RL accuracy can be heavily limited by imprecision in the environment database. Moreover, they both can be time-consuming, especially for large measurement/simulation areas. These constraints drive the need for more efficient and flexible prediction methods.

Recent advances in machine learning (ML) have enabled data-driven prediction of RF coverage [10], [11]. Unlike conventional methods, ML-based models learn propagation

Received 5 September 2025; revised 4 February 2026; accepted 19 February 2026. Date of publication 26 March 2026; date of current version 9 July 2026. This work was supported in part by European Union (EU)–Next Generation EU through Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, Partnership on "Telecommunications of the Future" (PE00000001–Program "RESTART") under Grant CUP J33C22002880001. *(Corresponding author: Mohammad Hossein Zadeh.)*

The authors are with the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi" (DEI), CNIT, University of Bologna, 40126 Bologna, Italy (e-mail: mohammad.hossein3@unibo.it; marina.barbiroli@unibo.it; enricomaria.vitucci@unibo.it; v.degliesposti@unibo.it; franco.fuschini@unibo.it).

Digital Object Identifier 10.1109/TAP.2026.3676109

TABLE I
RELATED WORKS ON IMAGE-BASED, DL-DRIVEN REMS PREDICTION

Feature	FadeNet [19]	RadioUNet [18]	PMNet [20]	RME-GAN [23]	REM-UNet [22]	DC-Net [21]
Dataset Generation	RT (No details on RT setting)	RT SW *WinProp* [25] (DPM - Dominant Path Model, RT2 - up to 2 bounces, RT4 - up to 4 bounces)	RT SW *Wirel. Insite* [26] (No details on RT setting)	RT SW *WinProp* (No details on RT setting)	RT SW *WinProp* (RT2 - up to 2 bounces)	RT SW *WinProp* (No details on RT setting)
Target Fading Effects	LSF	LSF	LSF	LSF	LSF	LSF
Maps Number	2500	700 (RadioMapSeer DB [24])	N.A.	700 (RadioMapSeer DB)	700 (RadioMapSeer DB)	700 (RadioMapSeer DB)
Map Size	512m × 512m	256m × 256m	243m × 243m	256m × 256m	256m × 256m	256m × 256m
RX Spatial Res.	2m × 2m	1m × 1m	N.A.	1m × 1m	1m × 1m	1m × 1m
TX & RX	1 TX/map h_{TX}=7m h_{RX}=1.5m	DPM/RT2: 80 TX/map IRT4: 2 TX/map, h_{TX}=h_{RX}=1.5m	N.A.	80 TX/map h_{TX}=h_{RX}=1.5m	TX on rooftop h_{RX}=1.5m	80 TX/map h_{TX}=h_{RX}=1.5m
Antennas	N.A. (most likely: isotropic)	Isotropic	Isotropic	Isotropic	Isotropic	Isotropic
Frequency	Not provided	5.9 GHz	2.5 GHz	5.9 GHz	3.5 GHz	5.9 GHz
DL Architecture	UNet-based (28 conv. layers)	2x UNet (cascade, 18 conv. layers each)	Dilated convolutional autoencoder network	Generative Adversarial Network (includes a UNet with 18 conv. layers)	UNet	Double UNet + Distant-range Content Module
Auxiliary Data	YES (LoS map)	YES (300 samples/map from RT4)	NO	YES (sparse map)	YES (LoS map)	N.A.
Data Augmentation	YES (flip/rotation, 20000 final maps)	N.A.	YES (cropping/rotation)	N.A.	YES (flip/rotation ×8)	NO
Performance	RMSE=7.1 dB w/o aux. data RMSE=5.8 dB with aux. data	RMSE = 0.027-0.038 (2.7–3.8 dB) w/o aux. data RMSE = 0.016-0.03 (1.6–3 dB) with aux. data	NDL (Normalized Depth Loss) = 0.01 Normalized MSE = 0.0003	RMSE = 0.014	RMSE = 0.034 (1.2 dB) with aux. data	RMSE = 0.0095 - 0.019

patterns from large-scale datasets and can generalize across different environments [12], [13]. Deep learning (DL), a subset of ML, has further enhanced this capability by using neural networks (NNs) with multiple layers to detect complex patterns and relationships inside data. DL approaches have been applied to RF propagation analysis, demonstrating improved accuracy and efficiency [14], [15]. In the framework of DL techniques, convolutional NNs (CNNs) have been particularly effective in image processing tasks, including image recognition and object detection [16]. Among CNN-based architectures, U-Net has gained significant attention for its encoder–decoder structure, which allows for efficient feature extraction and spatial reconstruction. Originally developed for biomedical image segmentation [17], U-Net has since been applied to different image-related tasks, leveraging skip connections to preserve fine details while capturing high-level contextual information.

Some recent works have competitively addressed REM prediction in urban environments using DL, usually relying on image-based inputs (e.g., city maps) and U-Net-style architectures [18], [19], [20], [21], [22], [23]. A comparative overview of these recent methods is presented in Table I. In order to achieve reliable prediction accuracy, they all simultaneously work/act on the multiple factors affecting the final performance with the highest likelihood, like the U-Net structure and its (hyper-)parameters, the size of the training dataset, the resort to data augmentation techniques, and auxiliary input data. In spite of the effort, the final, overall setting of the U-Net sometimes resulted in a marginal performance improvement compared to other, similar studies.

Rather than contributing to this competitive challenge, a different analysis is proposed in this work, where the goal is not a further (marginal) reduction of the prediction error, but most of all a comparative evaluation of the performance dependence on the following key factors: U-Net complexity, auxiliary input types, augmentation strategies, training dataset size and characteristics, and target label fidelity. Instead of presenting yet another accuracy benchmark, this contribution investigates the sensitivity of REM prediction accuracy to various architectural and input data choices. By systematically evaluating these elements, the aim is to derive general insights into what matters most to achieve practical, scalable, and robust REM prediction, especially under realistic constraints where full-scale RT simulations or large paired datasets may not be feasible. In doing so, this work fills a current gap in the literature by providing a roadmap for optimizing such systems beyond pure accuracy performance, offering both theoretical value and practical guidelines.

The remainder of this article is structured as follows. Section II provides an overview of the related work, discussing prior approaches to RF coverage prediction using both conventional and DL-based methods. Section III describes the dataset generation process, including the RT simulation setup and environmental parameters. Section IV details the proposed DL framework, including model architecture, training methodology, and evaluation metrics. Section V presents experimental results and analysis, highlighting the impact of different parameters and techniques on prediction performance. Finally, Section VI concludes this article and outlines potential directions for future research.

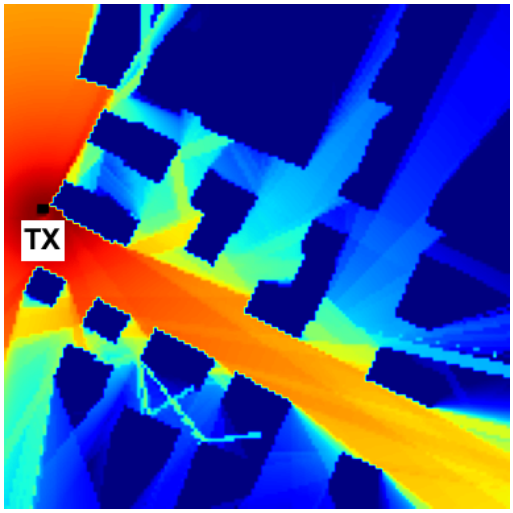


Fig. 1. Example of an urban coverage map.

II. Related Work and Proposed Solution

Starting from the seminal work described in [19], some following studies, shortly summed up in Table I, have investigated DL-assisted urban REM generation, according to the same general approach [18], [20], [21], [22], [23]. Both training and test datasets consist of a number of REMs represented as images, where the strength of the electromagnetic field propagating from a known transmitter (TX) position to a multitude of receiver (RX) locations spread along the streets of an urban map is plotted in a colored scale (see Fig. 1). Propagation data corresponding to the different REMs have always been generated by means of RT simulations. Although not always clearly stated, the evaluation is limited to large-scale fading (LSF), i.e., to distance-dependent path loss and shadowing effect related to possible link obstruction. By contrast, multipath fading interference is likely to be neglected. In order to collect a suitable amount of data, RT simulations have been carried out on a large number of different urban maps.

Part of the images is then used to train an NN, usually based on a U-Net, aimed at REM prediction. The input features to the NN always include the 2-D geometrical layout of the urban scenario and the TX position, while the corresponding generated REM (see Fig. 1), of course, represents the target output label. Sometimes, auxiliary input data have also been considered, as well as augmentation data techniques have been enabled in some cases.

Performance is always evaluated on the test dataset, where basically the REMs predicted by the NN are compared with the ground-truth reference represented by the RT simulation outcome. It is worth highlighting that, before performance assessment, the received power in dB returned by RT is usually normalized as follows [18], [19], [20], [21], [22], [23]:

$$P_{r,norm} = \frac{P_r - P_{r,min}}{P_{r,max} - P_{r,min}} \tag{1}$$

where P_r is the received power value according to RT (in dB units, e.g., dBm or dBW), $P_{r,min}$ is a parameter of the RT simulation corresponding to a minimum power threshold (i.e., $P_r = P_{r,min}$ is automatically set at every RX location, where $P_r \leq P_{r,min}$ is returned by RT), and $P_{r,max}$ is the maximum received power value throughout the whole database (dB units). The normalization in (1) preserves the relative differences in received power while scaling the values to the range [0, 1], which is a more suitable input format for NN [18]. It also improves the model's ability to learn complex patterns by maintaining consistent intensity variations across all samples.

The root-mean-square error (RMSE) between the REM predicted by the NN and the reference, ground-truth image returned by RT, is finally computed. In case *normalized* received power values are considered, then RMSE also belongs to [0, 1], and it can be regarded as a sort of relative percentage expression of the prediction error with respect to the $P_{r,max}$ - $P_{r,min}$ range. For example, if RMSE = 0.1 is achieved with $P_{r,min} = -130$ dBm and $P_{r,max} = -80$ dBm, then the absolute, average error is equal to 5 dB. Although the RMSE evaluation should be, of course, limited to the pixels corresponding to the RX locations, it is sometimes computed all over the image (i.e., buildings included) [18], [23].

Despite this general and common framework, different choices have been made in the different previous studies under several aspects, e.g., in terms of the structure of the U-Net, size of the training/test dataset, and resort to auxiliary input data and data augmentation strategies (see Table I). Anyway, no strong motivations are usually provided about the reasons why each investigation uses a different strategy from the others, which in the end can convince the reader that NN can effectively drive the REM generation process, but without providing them with any physical insight, nor giving them any clear guideline to effectively handle and deploy DL resources and techniques for radio map estimation. Furthermore, information is sometimes lacking or incomplete, as also outlined in Table I.

This is the gap this work aims at filling up, as it is not just a further benchmark to test the performance of DL-driven REM prediction, but rather a more comprehensive investigation on the sensitivity of the prediction accuracy to those factors that are actually expected to mostly affect the performance, like U-Net complexity, auxiliary input types, augmentation strategies, training dataset size and characteristics, and target label fidelity. In agreement with previous studies, the analysis is carried out in an urban environment, as it is particularly challenging from the propagation perspective.

To support the clarity and reproducibility of this multifaceted investigation, a flowchart is introduced in Fig. 2. This diagram provides a concise visual summary of the pipeline adopted throughout the work. Without delving into implementation details, it illustrates the sequence of processing steps applied to generate the model inputs, including optional data sources and learning strategies. Starting from the base inputs—namely, the map of the urban environment and the location of the TX—the process may optionally incorporate additional features, such as externally computed auxiliary maps, or exploit data augmentation techniques. The final step involves feeding the processed data into a selected NN model for training and prediction. The reader may refer

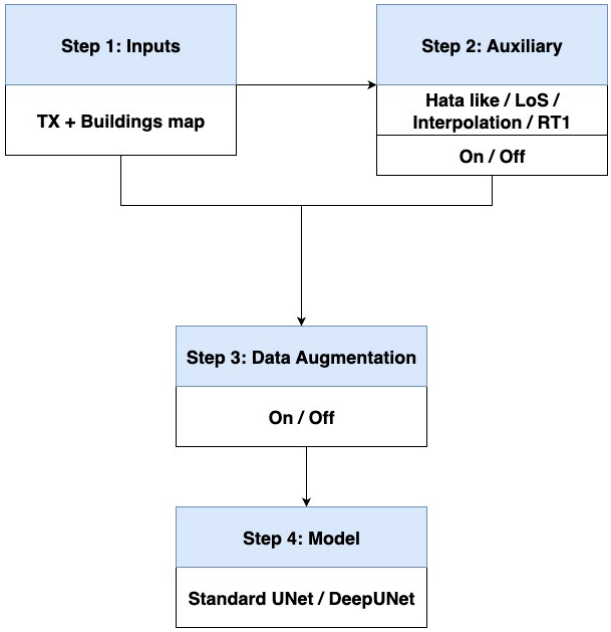


Fig. 2. Block diagram of the proposed REM prediction pipeline.

back to this overview diagram while progressing through Sections III and IV, which provide an in-depth explanation of each component.

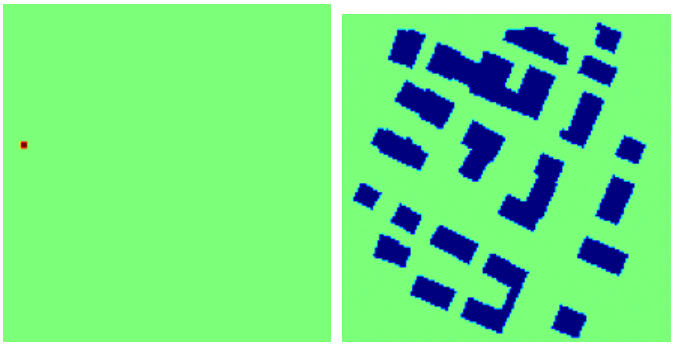
III. Dataset Generation

Input data typically consist of environmental features, such as building maps and TX locations, while the output label describes the corresponding RF coverage characteristics (i.e., the REM). In addition to these primary input features and output labels, auxiliary data can be introduced to further enhance the model's predictive capabilities. Auxiliary data, collected in a faster way with respect to input data, refer to supplementary input information, such as line-of-sight (LoS) maps [19], [22] or (interpolated) sparse coverage maps [18], [23]. They provide context to improve generalization and accuracy. Previous studies have shown that incorporating auxiliary data can actually increase DL-based REM prediction accuracy by reducing uncertainty and improving spatial consistency [30].

A. Input Map Generation

In order to cope with the computational effort required by RT simulations to generate the propagation dataset to train and test the DL model, some expedients have been enforced in this study to properly restrict the REM size. In particular, the analysis has been limited to microcellular urban layouts, i.e., with the TX and the RXs placed at a height, respectively, equal to 5 and 1.5 m, i.e., well below the average height of the buildings included inside the urban area, equal to 14 m. With this measure, given that propagation primarily occurs in the horizontal plane between the TX and RX, we can simplify the simulations by using a 2-D urban layout, significantly reducing their complexity.

A 5 $\times$ 5 km high-resolution map of the city of Bologna (Italy) was randomly divided into 1000 smaller maps with a size of 150 $\times$ 150 m, corresponding to suitable input samples for RT simulations. To ensure a meaningful urban environment, each tile contains a minimum of 25 buildings, serving as a threshold to filter out uninformative regions.



(a) (b)

Fig. 3. Example of input maps: TX map and buildings map. (a) TX location. (b) Buildings map.

In particular, two different maps have been considered as input features for the DL model.

1) *TX Location Map:* A binary matrix accounting for the TX position in the map. The TX is randomly placed over the map region not covered by buildings, and its location corresponds to the only active pixel in the input matrix [see Fig. 3(a)].
2) *Building Map:* A binary matrix describing the position, size, and shape of buildings within the urban map. Buildings are marked as active pixels in the matrix, while the others are left inactive [see Fig. 3(b)].

As clearly shown in Fig. 3, the information conveyed by the two input maps is limited to 2-D data, i.e., height values are not included. Although electromagnetic propagation is inherently a 3-D process, it is likely to occur below the rooftop level in microcellular scenarios, i.e., in the streets surrounding the buildings rather than above them. In this regard, 2-D information is assumed to be sufficient for a reliable REM prediction.

B. RT Simulation

In agreement with previous investigations, the propagation data to fill the database for training and testing the DL model have been collected by means of RT simulations carried out over the 1000 available urban maps. While a single TX has been placed in each map, a multitude of RXs have been spread along the streets between buildings, at a height of 1.5 m and with a spatial resolution of 1 m. Simulations have been run at the frequency f_0 = 3.5 GHz and assuming an isotropic radiation pattern for both the transmitting and the receiving antennas. In order for the simulation to better mimic real-world propagation, diffuse scattering (S) can be enabled in the considered RT tool [31], in addition to reflection (R) and diffraction (D) considered in previous studies. The maximum number of bounces experienced by each ray, as well as the type of electromagnetic interactions, has been set as summed up in Table II, which represents an effective tradeoff between the overall simulation time and the need for realistic results. The minimum power threshold was set to $P_{r,min}$ = $-$110 dBm,

TABLE II
RT SETTINGS AND CORRESPONDING COMPUTATION TIME FOR 1000 SIMULATIONS

Type	Max. no. of int.	R	D	R and D	S	Time
LoS	0	0	0	0	Off	Fast
RT1	1	1	1	1	Off	12h
RT2	2	2	1	2	Off	1 week
RT3	3	3	1	2	On	2 weeks

TABLE III
MAJOR PARAMETERS OF THE MICROCELLULAR SCENARIOS

Parameter	**Value**
Map Size	150 m X 150 m
Maps Number	1000
Antenna Type	Isotropic
Transmit Power	0 dBm
Transmitter Height	5 m
Receiver Height	1.5 m
(Average) Building Height	14 m
Carrier frequency	3.5 GHz
Image Resolution	1 RX/m
Minimum Threshold Power	-110 dBm

corresponding to a maximum path loss equal to 110 dB, in agreement with previous works [18], [22]. Still in compliance with previous investigations, the field intensities carried by the different rays have been commonly summed up, neglecting any phase relationship, i.e., multipath interference was not taken into account. This means that the total received power at any RX location has been computed as follows:

$$P_r = 10\log_{10}\left(\sum_i P_{r,i}\right) \tag{2}$$

$$P_{r,i} = \frac{\lambda^2}{8\pi\eta}\, g_r(\theta_i,\varphi_i)\cdot|\vec{p}_r(\theta_i,\varphi_i)\cdot\vec{E}_i(Rx)|^2 \tag{3}$$

where N is the number of rays tracked by the RT tool, (θ_i,ϕ_i) is the direction of arrival of the ith ray, g_{rx} is the directional gain of the receiving antenna, $\vec{p}_{rx}$ is the polarization vector of the receiving antenna, and $\vec{E}_i$ is the electric field arriving at the RX through the ith ray. Of course, λ is the wavelength and $\eta = 120\pi$ the free-space impedance.

Just for the RT3 case in Table II, an additional dataset was also generated, enabling coherent simulation mode. The interference among the fields conveyed to the same RX along the corresponding rays has been taken into account in the computation of the overall received power according to the following expression:

$$P_{r,\text{coh}} = 10\log_{10}\left(\frac{\lambda^2}{8\pi\eta}\left|\sum_{i=1}^{N_r}\sqrt{g_r(\theta_i,\varphi_i)}\cdot\vec{p}_r(\theta_i,\varphi_i)\cdot\vec{E}_i(Rx)\right|^2\right). \tag{4}$$

To further speed up the simulation process, parallel computing was also enforced. The simulations were executed on a 12-core CPU, enabling concurrent processing of different maps. Of course, the computation time T_c reported in Table II refers to the parallel computation on the 12 cores. The values can then be easily scaled to a different number of cores n_c as $T_c \cdot 12/n_c$. The major parameters related to the simulated microcellular scenarios are reported in Table III. An example of REM returned by RT enabling three maximum interactions (RT3 in Table II) and scattering included is reported in Fig. 1.

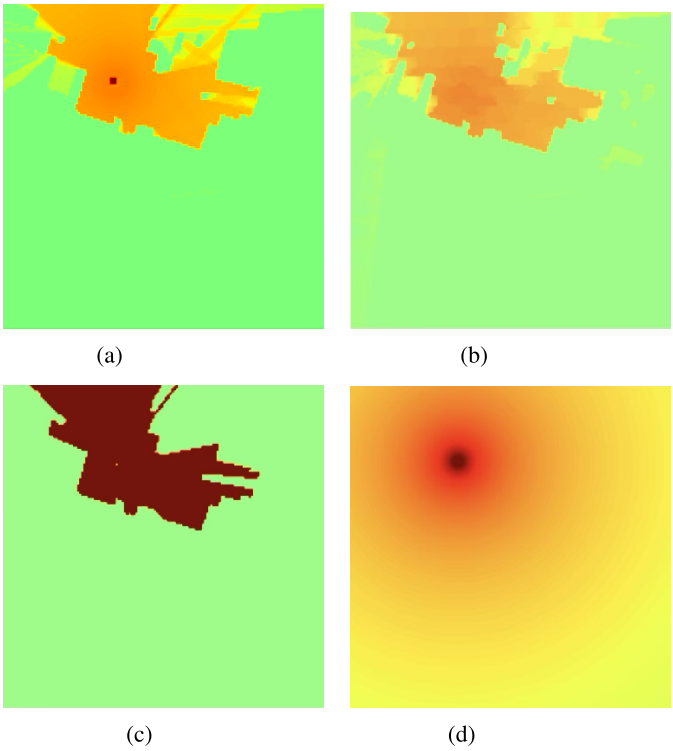


Fig. 4. Different types of auxiliary input maps. (a) Simplified. (b) Sparse. (c) LoS. (d) Hata-like.

C. Auxiliary Input Data

In addition to the primary input maps (TX position and building layout), the dataset can include auxiliary data to enhance the learning process. Auxiliary data provide additional input information that further helps the model improving its insight into the complex relationship between the primary input files and the corresponding output label (i.e., the REM). In this study, different types of auxiliary data are considered, as briefly described in the following.

1) *Simplified Coverage Map:* It simply consists of the REM achieved through RT simulations carried out with a reduced number of interactions compared to the REMs generated and used for training and test [see Fig. 4(a)]. Although these simulations are run enabling a lower number of reflections/diffractions/scattering, they provide additional propagation insights while significantly saving computation time compared to full-scale simulations. In the following, REMs corresponding to the RT setting marked as RT1 in Table II are considered as possible auxiliary input data.
2) *Interpolated Sparse Map:* It is a coarse coverage map based on propagation data collected from a lower number of RXs, i.e., corresponding to a reduced spatial density of RXs compared to the REMs generated and

used for training and testing. Then, the sparse coverage map can be interpolated to get a full-resolution image [see Fig. 4(b)]. In the following, sparse maps achieved with 20-m spacing between the RXs are considered.

3) *LoS Map:* It is a binary map providing information about the existence of indicated LoS between the TX and each RX location spread on the map [see Fig. 4(c)].
4) *Hata-Like Map:* It is generated by means of the following simple analytical formula:

$$P_r\,[\mathrm{dB}] = P_t\,[\mathrm{dB}] - PL\,[\mathrm{dB}]$$
$$PL\,[\mathrm{dB}] = 10 \cdot \alpha \cdot \log_{10}\left(\frac{d}{d_0}\right) + PL_0\,(d_0) \qquad (5)$$

where α is the *path-loss exponent* (PLE or *propagation factor*) and $PL_0(d_0)$ is the free-space attenuation at a reference distance d_0, here set equal to 1 m. Equation (5) describes the average range dependence of received power, and its slope is related to the value of α. In an urban environment, α usually belongs to the range [3, 3.5] [1]. The result is a smooth, linear decay of received signal strength with distance (in a log-to-log scale) [see Fig. 4(d)]. Although this map does not account for specific building obstructions, it nevertheless provides additional information about the expected (average) rate of power reduction with distance, which can anyway support the prediction of the output target REM to some extent. In the following, $\alpha = 3.5$ is considered.

Of course, the different auxiliary input maps are expected to improve the accuracy of the REM prediction to a different extent (as addressed in Section V). At the same time, they pose different requirements on the final user of the trained model for REM prediction. In order to generate simplified coverage maps, the final user must have access to both the urban maps and an RT tool, which might not always be obvious. Interpolated sparse maps can also be achieved by means of RT simulations, but they might also consist of real propagation data collected from working end-user equipment (e.g., sensor nodes) distributed over the urban area [23]. With reference to the LoS maps, they certainly require the availability of the urban layout, but they can be computed without an RT model, as the presence of LoS is a geometrical condition that can be checked in some different ways [22]. Finally, the computation of Hata-like maps just requires the value of the PLE, which can be fixed by means of simulations, but also by resorting to ad hoc experimental data, experience, or suggestions from existing literature.

D. Data Augmentation

Data augmentation is widely used in DL and refers to any technique aimed at artificially increasing the size of the training dataset by creating a modified representation of existing data. This approach helps preventing overfitting and improves the model's robustness by exposing it to diverse scenarios. In the context of RF coverage prediction, data augmentation can be particularly beneficial as it mitigates the need for extensive data collection or computationally expensive RT simulations to generate additional labeled samples.

To maximize the benefits of data augmentation while preserving the original spatial resolution of 150 × 150 m tiles, a comprehensive transformation strategy was employed. First, all possible rotations (90°, 180°, and 270°) were applied to each image, corresponding to three additional images. Then, horizontal, vertical, and diagonal flips were also enforced, thus introducing four new representations. Since some combinations lead to the same final result, the size of the dataset extracted from RT is finally expanded by a factor of eight. This dataset extension exposes the model to a broader range of spatial patterns and propagation scenarios, improving its ability to generalize across different urban layouts and TX placements. The impact of data augmentation on the performance of the DL model for REM prediction is discussed in Section V.

E. Output Map Generation

The final output maps, returned by RT simulations performed with high resolution, represent the ground truth for training and testing the DL model described in Section IV. These maps provide the spatial distribution of received power levels across the outdoor urban environment, accounting for the RF signal strength spatial fluctuations due to building obstruction and/or multipath propagation (see Fig. 1). To make these maps effective input for NNs, they are usually converted into grayscale images with values ranging from 0 to 1, according to the normalization procedure described in Section II.

IV. DL Model

ML is a field of artificial intelligence that enables computers to learn from data and make predictions or take decisions without being explicitly programmed. ML techniques have transformed numerous disciplines by automating complex tasks, identifying patterns in large datasets, and optimizing decision-making processes. In recent years, ML has demonstrated remarkable success in fields such as computer vision (object detection or recognition), natural language processing, and biomedical analysis [32], [33], [34]. These advancements have opened new frontiers in various fields, including wireless communication.

DL, a subset of ML, has further advanced predictive modeling by utilizing artificial NNs (ANNs) with multiple layers. Unlike traditional ML approaches that rely on handcrafted features, DL models learn hierarchical representations from raw data, making them highly effective for tasks that require intricate pattern recognition. These advancements have expanded the applications of ML and DL to domains such as wireless communication and RF coverage prediction, where data-driven models can compete with traditional propagation models and improve network design and optimization.

A. Convolutional Neural Networks

At the core of DL lie ANNs, which are computational models structured to emulate the layered processing mechanisms of the human brain. These models consist of interconnected layers of processing units (or neurons) that enable hierarchical learning of data patterns. Among the various types of ANNs, CNNs [35], [36] have proven

particularly effective for handling tasks involving spatially organized data, such as images and videos. Unlike traditional ANNs that typically rely on separate feature extraction and classification phases, CNNs seamlessly integrate both into a unified learning framework [37], [38].

In traditional ML pipelines, features must be manually engineered—a process that is not only time-consuming but may also result in suboptimal performance. CNNs, by contrast, learn hierarchical feature representations directly from raw input data, effectively identifying the most relevant patterns for the task at hand. This built-in ability to automatically extract and refine features during training significantly enhances both prediction accuracy and processing efficiency [37], [38].

CNNs are especially well-suited for 2-D input formats, as they apply learnable convolutional filters across the spatial dimensions of the data. This approach enables the model to detect and exploit spatial structures such as edges, contours, textures, and other localized patterns that are essential for high-level interpretation.

B. U-Net

U-Net [17] is a specialized CNN architecture designed for pixel-level predictions, making it particularly suitable for tasks such as image segmentation [39], [40], [41], video predicting [42], super-resolution [43], inverse problems in imaging [44], image-to-image translation [45], and medical image analysis [46]. It is structured as an encoder–decoder network.

1) *The encoder* gradually downsamples the input image while increasing the number of feature channels. This path efficiently extracts hierarchical features, capturing both local and global spatial dependencies.
2) *The decoder* progressively upsamples the feature maps, reconstructing the spatial resolution of the input while reducing the number of feature channels. This pathway synthesizes the learned high-level concepts back into a detailed output image.
3) *Skip connections* are used to concatenate feature maps from corresponding encoder layers to decoder layers of the same resolution. This mechanism preserves high-resolution information, ensuring spatial accuracy and structural consistency in the output.
4) Unlike traditional CNNs, U-Net does not contain fully connected layers, enabling it to maintain spatial coherence throughout the network. This design makes it robust for a wide range of applications, including image segmentation, super-resolution, image-to-image translation, and RF coverage prediction.

C. DeepUNet for RF Coverage Map Prediction

The proposed DeepUNet architecture is designed to effectively capture both low- and high-level spatial features necessary for accurate RF coverage prediction. As illustrated in Fig. 5, the model follows a symmetric encoder–decoder structure with five levels of depth, enabling the extraction of hierarchical representations from the input maps, each followed by a nonlinear activation function known as the rectified linear unit (ReLU) and a max-pooling operation for spatial downsampling (red arrows). The use of a nonlinear activation function is essential in deep NNs to allow the model to learn and represent complex, nonlinear relationships in the data. In particular, ReLU is one of the most commonly used activation functions due to its computational efficiency and ability to mitigate the vanishing gradient problem. ReLU operates by outputting zero for any negative input and passing positive inputs unchanged, thereby introducing nonlinearity without saturating the output. The max-pooling operation, on the other hand, is employed to progressively reduce the spatial dimensions of the feature maps. By retaining only the maximum value within small local regions (typically 2×2), max pooling helps in reducing computational complexity and achieving spatial invariance while also preserving the most salient features in the data. Together, convolution, activation, and pooling layers enable the encoder to extract meaningful hierarchical features that are crucial for accurate RF coverage prediction.

The decoder mirrors the encoder using transpose convolutions for upsampling. Blue boxes indicate the number of feature maps (filters) at each stage, growing deeper in the encoder and symmetrically reducing in the decoder. Skip connections between corresponding encoder and decoder layers (gray arrows) help preserve spatial details, which are crucial in high-resolution prediction tasks. The network outputs the final prediction map via a 1×1 convolutional layer. In addition to the main output, auxiliary outputs are extracted (yellow arrows) at each decoder stage to implement deep supervision, which strengthens multiscale feature learning and stabilizes training. The considered architecture is summarized in Table IV.

To guide the training process, a hybrid loss function is employed, combining mean squared error (MSE) and structural similarity index (SSIM). In the context of ML, a loss function is a quantitative measure of the difference between the model's predicted output and the actual result (known as the ground truth—in this case, the RT-generated coverage map). During training, the model parameters are iteratively adjusted to minimize this loss function, thereby improving the accuracy of predictions. The two components of the hybrid loss serve complementary purposes: MSE focuses on pixel-wise accuracy, while SSIM emphasizes structural similarity and perceptual fidelity. The hybrid formulation aims to ensure both accurate value prediction and preservation of spatial patterns, which are important in radio signal mapping.

MSE is defined as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{6}$$

where y_i and $\hat{y}_i$ represent the ground-truth and predicted pixel values, respectively, and N is the number of valid pixels in the map.

SSIM evaluates the similarity between predicted and ground-truth images in terms of luminance, contrast, and structure. Its formula is

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{7}$$

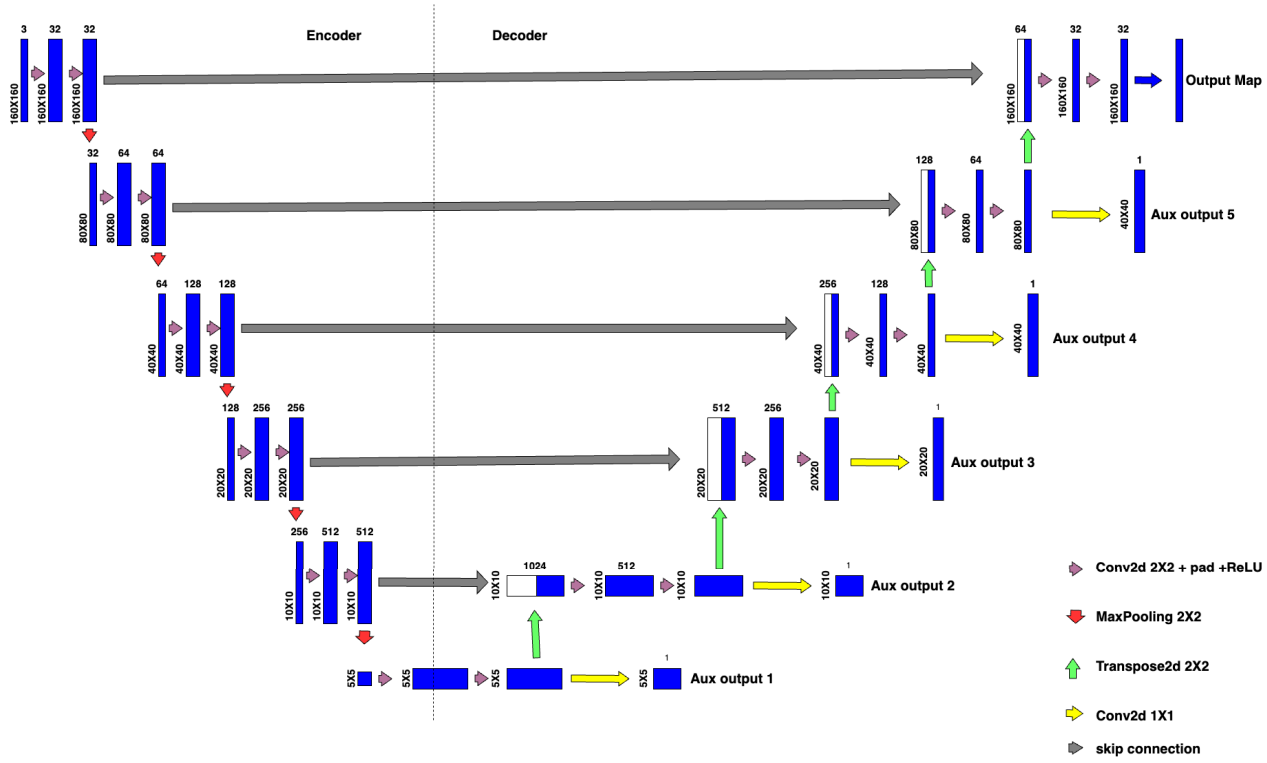


Fig. 5. DeepUNet architecture.

TABLE IV
DEEPUNET ARCHITECTURE

Layer Type	# of Layers	Filter Size	Activa Func.
Convolutional Layer	22	2×2	ReLU
Pooling Layer	5	2×2 (MaxPool)	-
Upsampling Layer	5	2×2 (Transpose)	-
Skip Connections	5	-	-
Final Output Layer	1	1×1	-

where μ_x and μ_y are the mean values, σ_x^2 and σ_y^2 are variances, and σ_{xy} is the covariance. Constants C_1 and C_2 stabilize the division.

Deep supervision is applied by introducing auxiliary losses at intermediate decoder stages, each using the same hybrid loss. The total training loss is computed as

$$L_{total} = L_{main} + \alpha_1 L_{aux_1} + \alpha_2 L_{aux_2} + \ldots + \alpha_n L_{aux_n} \quad (8)$$

where L_{main} is the loss at the final output, L_{aux_i} is the auxiliary loss, and α_i is the corresponding weight. In the deep supervision design, the network produces five auxiliary outputs (one per decoder level). When auxiliary weights are not explicitly provided, uniform weights are automatically assigned through $\alpha_i = \frac{1}{n}$, where n = 5. Thus, each auxiliary loss is multiplied by 0.2 and contributes equally to the total loss in addition to the main output loss. This choice stabilizes multiscale learning and improves convergence during training.

It is important to clearly distinguish between the training and evaluation metrics. The training process relies on the hybrid loss function (MSE + SSIM), while the model's performance is evaluated using the RMSE on the test dataset. This separation aligns with established practice and avoids confusion between optimization and evaluation objectives.

Since the output coverage maps are normalized to grayscale values between 0 and 1, RMSE is computed as

$$RMSE_{GS} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} \left(P_{r,norm,true} - P_{r,norm,pred}\right)^2}. \quad (9)$$

Since the grayscale values are derived from the received power levels using min–max normalization, the RMSE in grayscale can be approximately converted to RMSE in decibels using the following relationship:

$$RMSE_{dB} \approx RMSE_{GS} \cdot \left(P_{r,max} - P_{r,min}\right) \quad (10)$$

where $RMSE_{GS}$ denotes the RMSE computed on the normalized grayscale maps. In this work, the received power values are normalized using $P_{r,\max} = -77$ dBm and $P_{r,\min} = -130$ dBm, resulting in a dynamic range of 53 dB. Consequently, the RMSE expressed in decibels can be roughly computed as $RMSE_{dB} \approx 53 \cdot RMSE_{GS}$.

In agreement with [19], the RMSE is calculated only over the map area where valid RT-based ground-truth values exist, e.g., excluding the area covered with buildings. This ensures that the evaluation is limited to areas relevant to RF propagation, providing a more reliable assessment of model performance. Although this may sound obvious, it is worth pointing out that RMSE is instead computed over the whole map (i.e., buildings included) in some other works [21], [22], [23], which is expected to sweeten the performance evaluation.

The training process of the DeepUNet was conducted using the Adam optimizer, a widely used stochastic optimization

TABLE V
STEP-BY-STEP PERFORMANCE ANALYSIS

Model Type	Aux. Data	Data Aug.	RMSE	RMSE (dB)
Standard UNet	None	No	0.25	13.25
		Yes	0.16	8.48
DeepUNet	None	No	0.14	7.42
Standard UNet	Hata-like	No	0.13	6.89
	LoS		0.07	3.71
	Sparse Map		0.06	3.18
	RT1 Map		0.04	2.12
DeepUNet	RT1 Map	Yes	0.03	1.59

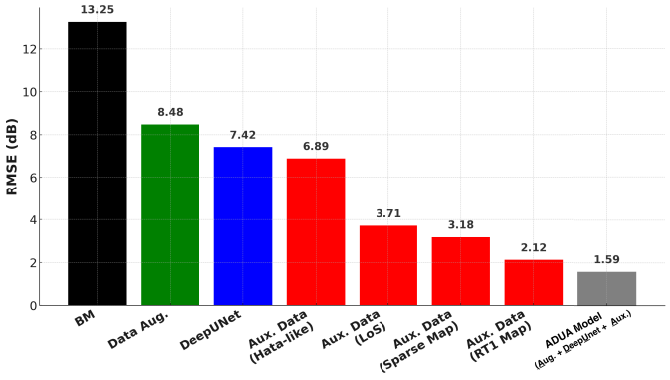


Fig. 6. Bar plot of RMSE for different models.

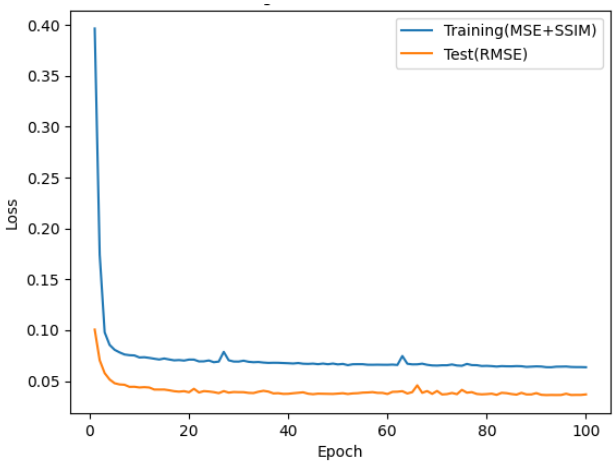


Fig. 7. Train and test loss.

method that combines momentum and adaptive learning rates. The initial learning rate was set to 0.001 and decreased by a factor of 0.1 after 60 epochs to promote convergence. Training was performed for 100 epochs using a batch size of 16. These values, often referred to as hyperparameters, control how the learning algorithm updates the model weights. All models were trained using a hybrid loss function with deep supervision applied at each decoder level.

The DeepUNet model consists of approximately 31.1 million trainable parameters across 22 convolutional layers. Training the best-performing configuration for 100 epochs on a dataset of 900 image pairs required approximately 15 min using an NVIDIA Tesla P100-PCIE-12 GB GPU (12-GB memory).

V. RESULTS AND DISCUSSION

This section analyzes the prediction results by organizing the discussion into three logical parts: sensitivity analysis, fast fading tracking capability, and generalization potential toward real-world propagation.

A. Sensitivity Analysis: Effect of Model and Data Enhancements

Standard U-Net, trained on RT3 targets, without auxiliary data and without data augmentation, is here considered as the baseline model (BM). As reported in Table V, it yields a normalized RMSE of 0.25 (13.25 dB), which is actually not particularly good. Enabling data augmentation alone reduces the RMSE to 0.16 (8.48 dB), proving its regularization effect and contribution to improved generalization. Replacing the Standard U-Net with the DeepUNet architecture, still without auxiliary inputs and data augmentation, results in an RMSE of 0.14 (7.42 dB), highlighting the benefit of deeper feature extraction and hybrid loss supervision.

Finally, Table V also shows that auxiliary input information turns out to be beneficial to the greatest extent, with RMSE dropped to less than 7 dB regardless of the type of auxiliary data map. This result confirms that low-complexity simulated REMs can serve as effective training support for predicting higher fidelity coverage maps. Interestingly, the RMSE reduction brought by auxiliary input maps is directly related to the complexity of their generation process. Hata-like maps can be quite easily achieved, but also correspond to the lowest gain, whereas simplified coverage maps computation requires heavier effort but also produces the most remarkable advantage. Not surprisingly, the most accurate REM prediction, corresponding to an RMSE equal to 0.03 (1.59 dB), is achieved when the DeepUNet architecture, data augmentation, and auxiliary input (RT1 Map) are simultaneously enforced, as also shown in Fig. 8. The RMSE improvement triggered by the different cases considered in Table V is also represented in Fig. 6.

Fig. 7 shows the training loss and test RMSE over 100 epochs for the best-performing model configuration (DeepUNet + RT1 auxiliary input). The model exhibits stable and monotonic convergence, with the training loss decreasing consistently and the test RMSE stabilizing at a low value. The minimal gap between training and test curves confirms strong generalization and the absence of overfitting. This behavior reinforces the reliability of the selected configuration and suggests that the learning process is both stable and efficient.

The qualitative results for the best-performing configuration—DeepUNet with RT1 Map as auxiliary input—are shown in Fig. 8, which includes the ground-truth and predicted coverage maps for both the test samples with minimum and maximum RMSE. In addition, the figure reports the corresponding absolute error maps, computed as the pointwise difference between the predicted and ground-truth REMs, to explicitly highlight local prediction inaccuracies. While the minimum RMSE case illustrates excellent agreement between the prediction [see Fig. 8(b)] and ground truth [see Fig. 8(a)], the maximum RMSE sample

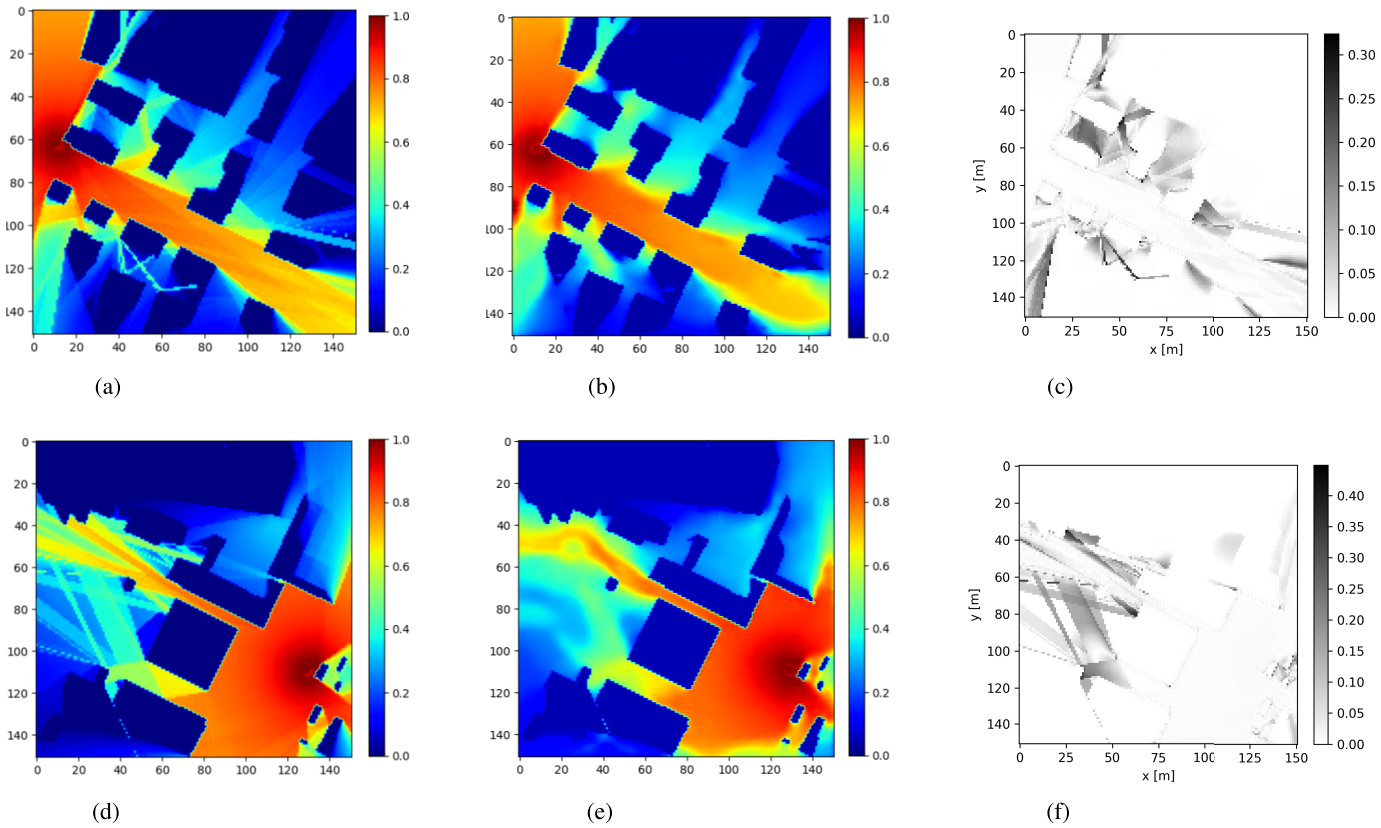


Fig. 8. Qualitative comparison between ground-truth REMs, DeepUNet predictions, and absolute error maps for the best-case (RMSE = 0.005, top row) and worst case (RMSE = 0.089, bottom row) test samples. Error maps are visualized using a grayscale representation, where brighter regions indicate lower prediction error. (a) Ground truth (min RMSE). (b) Prediction (min RMSE). (c) Absolute error (min RMSE). (d) Ground truth (max RMSE). (e) Prediction (max RMSE). (f) Absolute error (max RMSE).

[see Fig. 8(d) and (e)] also shows a strong qualitative match. Although some deviation is expected, especially in complex propagation zones, the model still captures the dominant structures and gradients accurately. This consistency across both extremes reflects the model's overall robustness and stability, as confirmed by the very low standard deviation of RMSE reported for this configuration. It highlights that even in relatively harder scenarios, the model maintains reliable predictive quality. The error maps [see Fig. 8(c)–(f)] show that discrepancies are mostly localized in regions where more complex propagation conditions occur, as around (multiple) building corners or at the furthest NLoS receiving locations. This is actually not fully surprising, as it is where the learning task is harder.

B. Fast Fading Tracking Capability Under Coherent Propagation

To assess the model's ability to handle more complex propagation effects, simulations were repeated in the coherent case, where multipath components arriving at the RX are combined, taking into account phase information, therefore capturing the interference patterns related to fast fading. Unlike the noncoherent scenario, where power contributions from different rays are simply summed up, neglecting any phase relationship, the coherent model introduces stronger local fluctuations and fine-grained spatial variations that might represent a further challenge for the model learning process. In this context, the received power should ideally be modeled according to (4), which accounts for the vectorial (i.e., phase-aware) summation of the field components, as opposed to (2), which assumes incoherent power addition.

It is worth noting that the model architecture adopted is the proposed DeepUNet. The LoS map is here selected as auxiliary input as it provides a favorable tradeoff between prediction accuracy and auxiliary data generation complexity as illustrated in Fig. 6. Although RT1 maps yield the best overall performance, their generation requires additional RT simulations, which were not available for the coherent case considered. For this reason, LoS-based auxiliary information is adopted in this case.

Surprisingly, this added complexity does not seem to be affecting the model's performance to a significant extent. When tested on coherent REMs using LoS maps as auxiliary input, the final RMSE was found equal to 0.07 (3.7 dB), which is nearly identical to the value obtained under noncoherent conditions. This result suggests that the model generalizes fairly well to more complex propagation phenomena like multipath spatial interference.

Fig. 9 provides a qualitative comparison between the ground-truth coherent REM and the predicted one. Owing to the enforced coherent simulation mode, the ground-truth REM coming from RT now clearly includes interference-induced multipath ripples. Although the DeepUNet cannot

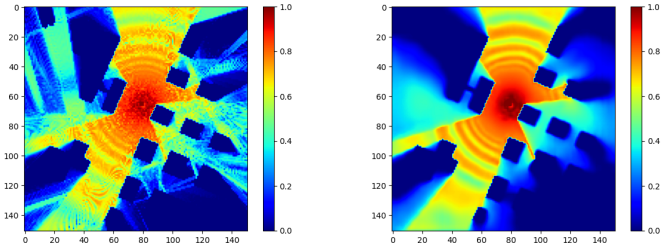
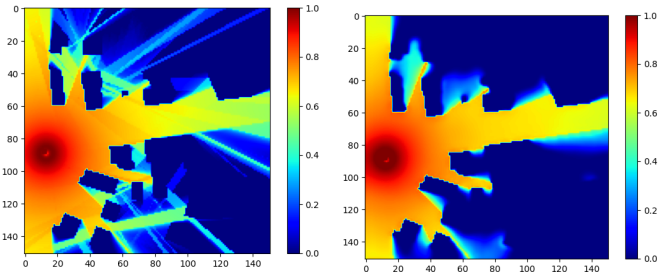


Fig. 9. Prediction results under coherent propagation conditions. Ground-truth REM (left). Predicted REM (right).



(a) (b)

Fig. 10. Prediction of RT3 map based on training on the RT1 Map. (a) Ground truth. (b) Prediction.

actually track them at the point-specific level, it anyway can, nonetheless, fairly well predict the macro structure of the interference pattern (especially present in the LoS region in Fig. 9).

C. Glance to Real-World Propagation

In practical scenarios, the ultimate goal is to predict REMs as they actually are in the real world. However, collecting comprehensive real-world measurement data for training DL models is infeasible due to time, cost, and logistic limitations [50]. To simulate such conditions to some fair extent, a realistic evaluation strategy involves training the DeepUNet model on synthetic data generated by a simplified RT simulation (RT1) and testing it on data from a more accurate and detailed simulation (RT3), treated as a proxy for real-world measurements.

This approach allows for assessing the model's generalization ability when faced with increased physical complexity and previously unseen propagation behaviors. When the model is both trained and tested on RT1 data, it achieves an RMSE of 0.05 (2.65 dB). After training on RT1 and testing on RT3, the RMSE rises to 0.08 (4.24 dB). As clearly shown in Fig. 10, the main reason for this increase lies in the poor agreement corresponding to the furthest NLoS RX locations. This is actually not surprising, as the single bounce enabled in RT1 can, of course, spread rays only to the RX positions closer to the TX. Therefore, the training stage could just feed the model with information about short-range propagation, unless LoS conditions occur. Although image-based U-Net can quite reliably grasp information on the propagation process from the training dataset (as discussed in Section V-A), any limitations or flaws in the database, making it not well representative of the real-world target process to be modeled, automatically reduce the final prediction skill, regardless of the degree of refinement of the model. Although this should sound quite clear, it is worth pointing out.

VI. Conclusion

This work presents a detailed investigation of image-based DL techniques for radio environment maps prediction, going beyond performance benchmarking to provide an in-depth understanding of key design factors. The proposed DeepUNet model, leveraging hybrid loss and deep supervision, increases prediction accuracy and structural fidelity of the generated maps with respect to standard, general-purpose U-Net architectures. In line with common sense, enabling data augmentation strategies also improves the model performance. Through systematic evaluation, it is shown that the resort to carefully selected auxiliary input turns out to be beneficial to the greatest extent. Results also highlight that the model is still reliable if fast fading effects are included in the training/test datasets. In case of limitations or inaccuracies in the synthetic propagation database with respect to the real propagation process, some reduction in the model performance must be expected, irrespective of its degree of refinement. The proposed analysis and design framework offers practical guidance for applying DL to wireless propagation modeling and sets the stage for future work involving real-world measurements or transfer learning techniques.

Acknowledgment

The authors are grateful to Dr. Nicola Di Cicco for his fruitful contribution to the starting stage of this research activity. The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

[1] R. L. Bertoni, *Radio Propagation for Modern Wireless Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[2] M. M. Azari et al., "Evolution of non-terrestrial networks from 5G to 6G: A survey," *IEEE Commun. Surveys Tut.*, vol. 24, no. 4, pp. 2633–2672, 4th Quart., 2022.

[3] D. F. Külzer, S. Stanczak, and M. Botsov, "CDI maps: Dynamic estimation of the radio environment for predictive resource allocation," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 892–898.

[4] S. Sarkar et al., "LLOCUS: Learning-based localization using crowdsourcing," in *Proc. 21st Int. Symp. Theory, Algorithmic Found., Protocol Design Mobile Netw. Mobile Comput.*, Oct. 2020, pp. 201–210.

[5] E. Krijestorac, S. Hanna, and D. Cabric, "Spatial signal strength prediction using 3D maps and deep learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–6.

[6] S. Mignardi, M. J. Arpaio, C. Buratti, E. M. Vitucci, F. Fuschini, and R. Verdone, "Performance evaluation of UAV-aided mobile networks by means of ray launching generated REMs," in *Proc. 30th Int. Telecommun. Netw. Appl. Conf. (ITNAC)*, Melbourne, VIC, Australia, Nov. 2020, pp. 1–6, doi: 10.1109/ITNAC50341.2020.9315177.

[7] Z. Utkovski, P. Agostini, M. Frey, I. Bjelakovic, and S. Stanczak, "Learning radio maps for physical-layer security in the radio access," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, France, Jul. 2019, pp. 1–5.

[8] B. P. Nayak, L. Hota, A. Kumar, A. K. Turuk, and P. H. J. Chong, "Autonomous vehicles: Resource allocation, security, and data privacy," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 117–131, Mar. 2022, doi: 10.1109/TGCN.2021.3110822.

[9] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. VT-29, no. 3, pp. 317–325, Aug. 1980, doi: 10.1109/T-VT.1980.23859.
[10] D. Romero and S.-J. Kim, "Radio map estimation: A data-driven approach to spectrum cartography," *IEEE Signal Process. Mag.*, vol. 39, no. 6, pp. 53–72, Nov. 2022, doi: 10.1109/MSP.2022.3200175.
[11] M. Yang et al., "AI-enabled data-driven channel modeling for future communications," *IEEE Commun. Mag.*, vol. 62, no. 4, pp. 112–118, Apr. 2024.
[12] M. F. Ahmad Fauzi, R. Nordin, N. F. Abdullah, and H. A. H. Alobaidy, "Mobile network coverage prediction based on supervised machine learning algorithms," *IEEE Access*, vol. 10, pp. 55782–55793, 2022, doi: 10.1109/ACCESS.2022.3176619.
[13] M. Vasudevan and M. Yuksel, "Machine learning for radio propagation modeling: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 5123–5153, 2024, doi: 10.1109/OJCOMS.2024.3446457.
[14] S. Zeng, Y. Ji, W. Chen, L. Yan, and X. Zhao, "Mobile network coverage prediction using multi-modal model based on deep neural networks and semantic segmentation," *Sensors*, vol. 24, no. 16, p. 5178, Aug. 2024, doi: 10.3390/s24165178.
[15] Y. Li, C. Zhang, W. Wang, and Y. Huang, "RMTransformer: Accurate radio map construction and coverage prediction," in *Proc. IEEE 101st Veh. Technol. Conf. (VTC-Spring)*, Jun. 2025, pp. 1–5.
[16] M. Browne and S. S. Ghidary, "Convolutional neural networks for image processing: An application in robot vision," in *Proc. 6th Australas. Joint Conf. Artif. Intell.*, 2003, pp. 641–652.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
[18] R. Levie, Ç. Yapar, G. Kutyniok, and G. Caire, "RadioUNet: Fast radio map estimation with convolutional neural networks," 2019, *arXiv:1911.09002*.
[19] V. V. Ratnam et al., "FadeNet: Deep learning-based mm-wave large-scale channel fading prediction and its applications," *IEEE Access*, vol. 9, pp. 3278–3290, 2021, doi: 10.1109/ACCESS.2020.3048583.
[20] J.-H. Lee, O. G. Serbetci, D. P. Selvam, and A. F. Molisch, "PMNet: Robust pathloss map prediction via supervised learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2023, pp. 4601–4606, doi: 10.1109/GLOBECOM54140.2023.10437562.
[21] Q. Chen, M. Huang, and J. Yang, "DC-Net: A distant-range content interaction network for radio map construction," *ICT Exp.*, vol. 10, no. 5, pp. 1145–1150, Oct. 2024, doi: 10.1016/j.icte.2024.08.008.
[22] H. Sallouha, S. Sarkar, E. Krijestorac, and D. Cabric, "REM-U-Net: Deep learning based agile REM prediction with energy-efficient cell-free use case," *IEEE Open J. Signal Process.*, vol. 5, pp. 750–765, 2024, doi: 10.1109/OJSP.2024.3378591.
[23] S. Zhang, A. Wijesinghe, and Z. Ding, "RME-GAN: A learning framework for radio map estimation based on conditional generative adversarial network," *IEEE Internet Things J.*, vol. 10, no. 20, pp. 18016–18027, Oct. 2023, doi: 10.1109/JIOT.2023.3278235.
[24] Ç. Yapar, R. Levie, G. Kutyniok, and G. Caire, "Dataset of pathloss and ToA radio maps with localization application," 2022, *arXiv:2212.11777*.
[25] R. Hoppe, G. Wölfle, and U. Jakobus, "Wave propagation and radio network planning software WinProp added to the electromagnetic solver package FEKO," in *Proc. Int. Appl. Comput. Electromagn. Soc. Symp.-Italy (ACES)*, Florence, Italy, Mar. 2017, pp. 1–2.
[26] REMCOM. *Wireless Insight*. Accessed: Mar. 23, 2026. [Online]. Available: https://www.remcom.com/wireless-insite-em-propagation-software
[27] E. Krijestorac, S. Hanna, and D. Cabric, "Spatial signal strength prediction using 3D maps and deep learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–6.
[28] Y. Teganya and D. Romero, "Deep completion autoencoders for radio map estimation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1710–1724, Mar. 2022.
[29] K. Li et al., "Model and transfer spatial–temporal knowledge for fine-grained radio map reconstruction," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 2, pp. 828–841, Jun. 2022, doi: 10.1109/TCCN.2021.3133844.
[30] H. Agrawal, A. Hietanen, and S. Särkkä, "Utilizing U-Net architectures with auxiliary information for scatter correction in CBCT across different field-of-view settings," *Proc. SPIE*, vol. 12925, Apr. 2024, Art. no. 1292539, doi: 10.1117/12.3004168.
[31] E. M. Vitucci et al., "Ray tracing RF field prediction: An unforgiving validation," *Int. J. Antennas Propag.*, vol. 2015, pp. 1–11, Aug. 2015.
[32] E. Elyan et al., "Computer vision and machine learning for medical image analysis: Recent advances, challenges, and way forward," *Artif. Intell. Surg.*, vol. 2, no. 1, pp. 24–45, 2022, doi: 10.20517/ais.2021.15.
[33] A. Le Glaz et al., "Machine learning and natural language processing in mental health: Systematic review," *J. Med. Internet Res.*, vol. 23, no. 5, May 2021, Art. no. e15708, doi: 10.2196/15708.
[34] B. Chinnaiyan, S. Balasubaramanian, M. Jeyabalu, and G. S. Warrier, "AI applications–computer vision and natural language processing," in *Model Optimization Methods for Efficient and Edge AI*. Hoboken, NJ, USA: Wiley, 2024, doi: 10.1002/9781394219230.ch2.
[35] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv:1511.08458*.
[36] A. Dhillon and G. K. Verma, "Convolutional neural network: A review of models, methodologies and applications to object detection," *Prog. Artif. Intell.*, vol. 9, no. 2, pp. 85–112, Jun. 2020.
[37] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1–9.
[39] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[40] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[41] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Stanford, CA, USA, Oct. 2016, pp. 565–571.
[42] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Juan, Puerto Rico, 2016, pp. 1–14.
[43] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.
[44] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
[45] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2868–2876.
[46] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[47] L. Jia, A. Huang, X. He, Z. Li, and J. Liang, "A residual multi-scale feature extraction network with hybrid loss for low-dose computed tomography image denoising," *Signal, Image Video Process.*, vol. 18, pp. 1215–1226, Nov. 2024, doi: 10.1007/s11760-023-02809-3.
[48] R. Li et al., "A comprehensive review on deep supervision: Theories and applications," in *Proc. Comput. Vis. Pattern Recognit.*, 2022, pp. 1–23.
[49] Z. Utkovski, P. Agostini, M. Frey, I. Bjelakovic, and S. Stanczak, "Synthetic structure of industrial plastics," in *Plastics* (Polymers of Hexadromicon), vol. 3, J. Peters, Ed., 2nd ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64. [Online]. Available: http://www.bookref.com
[50] S. Bakirtzis, C. Yapar, M. Fiore, J. Zhang, and I. Wassell, "Empowering wireless network applications with deep learning-based radio propagation models," 2024, *arXiv:2408.12193*.

Mohammad Hossein Zadeh received the M.Sc. degree in telecommunications engineering from the University of Bologna, Bologna, Italy, in 2022, where he is currently pursuing the Ph.D. degree in telecommunications engineering.

His activity is focused on the research topic of assessment and development of channel modeling by machine learning. The research activity includes participation in European Research and Cooperation Programs (COST INTERACT).

Open Access funding provided by 'Universita di Bologna' within the CRUI CARE Agreement

Marina Barbiroli received the Laurea degree in electronic engineering and the Ph.D. degree in computer science and electronic engineering from the University of Bologna, Bologna, Italy, in 1995 and 2000, respectively.

She is currently an Associate Professor with the Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi," University of Bologna. She has authored or co-authored more than 40 journal articles on radio propagation and wireless system design. Her research focuses on radio systems designing through theoretical and experimental modeling of the wireless propagation channel.

Enrico Maria Vitucci (Senior Member, IEEE) was a Visiting Scholar at Aalto University, Espoo, Finland, in 2007. In 2015 and 2016, he was an Invited Researcher at Polaris Wireless Inc., Mountain View, CA, USA. He has been appointed as a Teacher in several editions of the courses on mobile radio propagation and short-range propagation at European School of Antennas (ESoA). He is currently an Associate Professor in applied electromagnetics, antennas, and propagation with the Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi," University of Bologna, Bologna, Italy, where he is the Head of Cesena-Forli Unit of the Center for Industrial Research on ICT (CIRI-ICT). He has authored about 140 technical articles in international journals and conferences, and a co-inventor of five international patents. His research interests are in deterministic and hybrid wireless propagation modeling, machine learning applied to radio propagation, macroscopic electromagnetic models for metasurfaces/RIS, and dynamic ray tracing techniques for high-mobility radio channels.

Dr. Vitucci is a Senior Member of the International Union of Radio Science (URSI). He served as an academic editor, an associate editor, and a guest editor for several international journals. He participated to European Research and Cooperation Actions COST 2100, COST IC1004, COST IRACON, and COST INTERACT, and a European Networks of Excellence NEWCOM and NEWCOM++, and in several projects of the Eu FP7, H2020, and Horizon Europe Programs.

Vittorio Degli-Esposti (Senior Member, IEEE) is a Full Professor with the Department of Electrical Engineering, University of Bologna, Bologna, Italy. He has been a Post-Doctoral Researcher at Brooklyn Polytechnic University (now NYU Tandon School of Engineering), Brooklyn, NY, USA; a Visiting Professor at Aalto University, Espoo, Finland, in 2006 and Tongji University, Shanghai, China, in 2013; and the Director of Research at Polaris Wireless Inc., Mountain View, CA, USA, from 2015 to 2016. He has authored or co-authored over 170 peer-reviewed technical articles and is a co-inventor of seven international patents in the fields of applied electromagnetics, radio propagation, and wireless systems.

Dr. Degli-Esposti was a recipient of the "2023 EurAAP Propagation Award." He has been the General Chair of the Conference URSI EMTS 2025, the Vice Chair of EuCAP 2010 and 2011, the Short-Course and Workshops Chair of EuCAP 2015, and the Convened Sessions Chair of EuCAP 2023. He is the Co-Chair of the Working Group WG1 "Radio Channels" of European COST Action 20120 "Interact." He is an Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Franco Fuschini received the M.Sc. degree in telecommunication engineering and the Ph.D. degree in electronics and computer science from the University of Bologna, Bologna, Italy, in March 1999 and July 2003, respectively.

He is currently an Associate Professor with the Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi," University of Bologna. He has authored or co-authored more than 50 journal articles on electromagnetic wave propagation and wireless systems design. His main research interests are in the area of radio systems technologies, radio propagation theoretical modeling, and experimental investigation.

Prof. Fuschini received the "Marconi Foundation Young Scientist Prize" in the context of the XXV Marconi International Fellowship Award in April 1999.