

Learned Elevation Models as a Lightweight Alternative to LiDAR for Radio Environment Map Estimation

Ljupcho Milosheski*, Fedja Močnik, Mihael Mohorčič, Carolina Fortuna
Department of Communication Systems, Jožef Stefan Institute, Ljubljana, Slovenia
*ljupcho.milosheski@ijs.si

Abstract—Next-generation wireless systems such as 6G operate at higher frequency bands, making signal propagation highly sensitive to environmental factors such as buildings and vegetation. Accurate Radio Environment Map (REM) estimation is therefore increasingly important for effective network planning and operation. Existing methods, from ray-tracing simulators to deep learning generative models, achieve promising results but require detailed 3D environment data such as LiDAR-derived point clouds, which are costly to acquire, several gigabytes per km^2 in size, and costly to keep updated in dynamic environments. We propose a two-stage framework that eliminates the need for 3D data at inference time, confining LiDAR use to a one-time training. In the first stage, a learned estimator predicts elevation maps directly from aerial RGB imagery, which are then fed alongside antenna parameters into the REM estimator in the second stage. Across existing CNN-based REM estimation architectures, the proposed approach improves RMSE by up to 6.8% over image-only baselines, reducing deployment storage from 294 MB/ km^2 to a fixed 606 MB regardless of coverage and lowering re-acquisition carbon cost by 17% compared to LiDAR-dependent approaches.

Index Terms—Radio Environment Map, Large-Scale Pathloss Modeling, Deep Learning, Aerial Imagery, 3D Environment Models

I. INTRODUCTION

Emerging applications such as Extended Reality (XR), Holographic Communication, and Digital Twins impose increasingly stringent requirements on wireless networks in terms of throughput and latency [1]. To meet these demands, next-generation systems such as 6G are moving towards higher frequency bands [2]. However, higher frequency electromagnetic waves are inherently more sensitive to the environment: they experience greater attenuation, reduced diffraction around obstacles, and more significant blockage from buildings and vegetation. This necessitates far more careful and precise planning of network coverage compared to previous generations. Radio Environment Maps (REMs) are a powerful tool for this purpose, enabling the prediction of signal propagation characteristics, such as pathloss, across a given area based on environment-specific inputs, including 3D geometry and antenna parameters. Approaches for REM estimation range from classical spatial interpolation methods [3] and physics-based ray-tracing simulators such as Wireless InSite [4] and NVIDIA Sionna [5], to recently proposed Deep Learning (DL) generative models based on U-Net [6] and GAN-based archi-

tectures [7]. While these methods achieve remarkable results, they share a critical dependency on fine-grained 3D input data to accurately model the propagation environment, which limits their potential for large-scale deployments and dynamic environments. Acquiring such 3D data is highly demanding in practice, as LiDAR campaigns require specialized equipment, cover limited areas, and produce datasets of tens of gigabytes requiring intensive preprocessing [6]. Open-source alternatives such as OpenStreetMap [8] provide building footprints but omit vegetation, which can significantly affect signal propagation [9]. Moreover, real-world environments change continuously due to construction and temporary structures, making 3D data-dependent estimators ill-suited for dynamic scenarios such as environment-aware network management and control [10]. A promising alternative is to estimate scene geometry directly from aerial imagery using DL models [11], eliminating the need for dedicated LiDAR campaigns at inference time. We propose a two-stage framework that requires 3D data for training, and only aerial images during inference. A DL-based height estimator [11] is trained on paired RGB and LiDAR data to predict elevation maps directly from aerial imagery, which are then used alongside antenna parameters as input to the REM estimator. The proposed framework improves RMSE by up to 6.8% over the image-only baseline, reduces deployment storage from a coverage-dependent 294 MB/ km^2 to a fixed 606 MB, and cuts re-acquisition carbon cost by 17% compared to LiDAR-dependent approaches. The contributions of this paper are as follows:

- A two-stage REM estimation framework replacing 3D data at inference time with a learned elevation estimator, achieving up to 6.8% RMSE improvement over image-only baselines while closing approximately one fifth of the gap to the LiDAR upper bound.
- A memory scalability and carbon footprint analysis showing that acquisition storage scales from 294 MB/ km^2 (LiDAR) to a fixed 606 MB deployment footprint, and that subsequent RGB re-acquisition requires 17% less carbon than LiDAR campaigns.
- Model-agnostic validation across three CNN-based architectures on the RMDirectionalBerlin benchmark, confirming the generalisability of the proposed approach.

This paper is organised as follows. Section II summarises

related work, Sections III and IV elaborate on the proposed framework and provide methodological details, while Section V presents the results. Section VI concludes the paper.

II. RELATED WORK

REM estimation has attracted growing attention in the wireless community in recent years, particularly with the emergence of Digital Twins, where REMs serve as a core component, capturing spatial distributions of radio propagation for network planning and optimization. **Ray tracing and datasets.** Advances in ray-tracing methods have progressed from primarily commercial packages [4] to optimized open-source libraries such as Sionna [5]. Alongside these simulation tools, several high-quality datasets have been released to benchmark and train data-driven methods. OpenStreetMap [8] provides 3D building models for major urban areas but lacks vegetation data. RadioMapSeer [12] offers widely used simulated radio maps with estimated building heights. More detailed datasets have recently appeared, including the RMDirectionalBerlin dataset with paired LiDAR and aerial imagery [6], physics-informed radio maps [13], and the multiband 3D SpectrumNet benchmark [14]. **Deep learning approaches.** Generative DL models have emerged as particularly promising for REM estimation due to their ability to capture complex propagation patterns. Wang et al. [15] deploy a generative diffusion model to achieve accurate REM estimation from sparse measurements and environment geometry. Chen et al. [16] employ dilated convolutional layers with an enlarged receptive field alongside a high-resolution feature-preserving architecture. GAN-based approaches such as ACT-GAN [7] further demonstrate the effectiveness of generative methods. Collectively, these works highlight the strong potential of generative architectures for capturing the spatial complexity of radio propagation. **Positioning of this work.** To the best of our knowledge, only two published works [6], [17] address REM estimation from aerial RGB imagery, both via end-to-end CNNs mapping RGB images and antenna parameters directly to path loss. In contrast, we decouple the problem into two independently optimised stages: elevation estimation and REM prediction. At inference time, only aerial RGB imagery is required, yet the proposed approach achieves up to 6.8% RMSE improvement over single-stage RGB-only approaches. Compared to LiDAR-dependent methods, the framework also reduces deployment storage and carbon footprint, and allows the elevation estimator to be updated independently as environments change.

III. PROPOSED APPROACH

Let $I \in \mathbb{R}^{H \times W \times C}$ denote a satellite RGB image, $\mathbf{E} \in \mathbb{R}^{H \times W}$ a LiDAR-derived elevation map, $G \in \mathbb{R}^d$ a vector of transmitter metadata, including 3D coordinates, orientation angles, and antenna pattern identifiers, which are spatially projected as input features, and $R \in \mathbb{R}^{H \times W}$ a radio environment map (REM). Conventional learning-based REM estimation models compute:

$$R = f_\theta(\mathbf{E}, I, G) \quad (1)$$

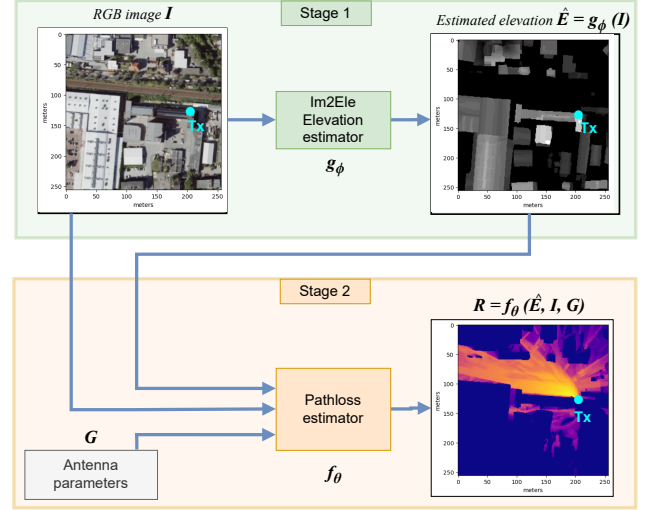


Fig. 1: System diagram of the two-stage framework. Data and pathloss estimators [6], elevation estimator Im2Ele model [11].

We propose to replace the physical elevation map with a learned estimator:

$$\hat{\mathbf{E}} = g_\phi(I) \quad (2)$$

where g_ϕ is a neural image-to-elevation model trained on paired RGB and 3D data. The final REM prediction then becomes:

$$\hat{R} = f_\theta(\hat{\mathbf{E}}, I, G) \quad (3)$$

Training proceeds in two independent stages. First, g_ϕ is optimized to minimize the elevation estimation error:

$$\min_{\phi} \mathcal{L}_{\text{elev}}(g_\phi(I), \mathbf{E}) \quad (4)$$

With ϕ fixed, f_θ is then trained to minimize the REM prediction error:

$$\min_{\theta} \mathcal{L}_{\text{REM}}(f_\theta(g_\phi(I), I, G), R) \quad (5)$$

To realize $\hat{R}$ we propose a modular two-stage framework illustrated in Fig. 1 as follows.

1) *Stage 1 — Elevation Estimation:* In the first stage, to obtain $\hat{\mathbf{E}}$ from Eq. 2, we retrain the Im2Ele model [11] to learn $g_\phi(I)$ as depicted in the upper part of Fig. 1. Im2Ele is a SNet-based architecture that takes a single aerial RGB image as input and predicts a normalized Digital Surface Model (nDSM) representing the height of above-ground structures. We focus specifically on buildings, as these constitute the primary obstacles influencing radio propagation in urban environments. This focus also reduces scene variability, which is particularly important given that the RMDirectionalBerlin dataset provides only 424 image-elevation pairs, considerably fewer than the several thousand samples used in the original Im2Ele training [11], making generalization to geometrically inconsistent elements unreliable. The elevation model is trained independently, and its weights are frozen prior to Stage 2 training, ensuring that the REM estimation model receives consistent elevation inputs throughout its optimization.

2) *Stage 2 — Radio Environment Map Estimation*: In the second stage, depicted in the lower part of Fig. 1, the pathloss estimator f_θ follows the experimental setup of Jaensch et al. [6]. The model receives as input the aerial RGB image I , the estimated elevation map $\hat{E}$ produced by Stage 1, a one-hot encoded transmitter location map, and antenna gain parameters G . These inputs are concatenated as separate feature channels and passed through the network, which outputs a spatial pathloss map $\hat{R} \in \mathbb{R}^{H \times W}$ representing radio signal propagation across the environment.

A. Reference implementation and baseline selection

To validate the proposed framework from Fig. 1, we propose a reference implementation as follows. Stage 1 is realized with Im2Ele as the elevation estimator. Stage 2 is realized via the model-agnostic pathloss estimator evaluated with three neural architectures: LitRadioUNet, LitPMNet, and LitUNetDCN [6]. We refer to this proposed instantiation as **Im2Ele-predicted**→**nDSM+images**. We consider the following two baselines for evaluation purposes. Both baselines also include evaluations with the LitRadioUNet, LitPMNet and LitUNetDCN architectures on the RMDirectionalBerlin dataset. We refer to the first baseline as **LiDAR**→**nDSM+images**. It uses LiDAR-derived elevation alongside aerial RGB imagery and antenna parameters, representing the best achievable performance given the richest available input data [6]. The second baseline, **Image-only**, uses aerial RGB imagery and antenna parameters without any elevation input, providing an end-to-end single-stage baseline with the least input information. This configuration is equivalent to the prior single-stage approaches of [6], [17], and serves as the benchmark against which the direct benefit of the proposed two-stage decoupling is measured.

IV. METHODOLOGY

In this section, we describe the dataset, training protocol, evaluation strategy, and baseline configurations used to assess the proposed two-stage pipeline.

A. Training and Evaluation Data

All experiments were conducted using the RMDirectionalBerlin dataset, which contains 74,515 radio map samples across 424 RGB aerial images paired with corresponding normalized Digital Surface Models (nDSM) and transmitter metadata, including antenna gain parameters and location. The data covers both urban and vegetated areas, enabling separation of structural and natural height components. While vegetation is known to affect signal propagation [9], preliminary experiments incorporating vegetation into the elevation estimator degraded REM prediction performance, which we attribute to the high geometric variability of vegetation combined with the limited dataset size of 424 samples. Restricting the estimator to buildings reduces input variability and improves generalization under this data constraint. We adopted the training-validation-test splits of 80%-10%-10%, accordingly.

B. Model Training

Training was performed independently for the elevation estimation model (Im2Ele) and the radio propagation models (LitRadioUNet, LitPMNet and LitUNetDCN), ensuring optimization stability and preventing gradient interference between geometric reconstruction and radio propagation losses.

1) *Elevation Estimation*: The Im2Ele model was retrained on the RMDirectionalBerlin dataset to predict nDSM maps from monocular aerial imagery, following the normalization and augmentation strategy described in [11], with one modification: color jitter augmentation was removed as color perturbations increased prediction instability in scenes containing large industrial roofs with high reflectance. Removing color jitter improved robustness in such failure cases without degrading overall performance. Thus, augmentation includes linear normalization of height values to range $[0, 1]$, random geometric scaling, and random 8-symmetry transformations (rotations and flips).

2) *Pathloss Estimation*: The radio propagation model receives as input the aerial RGB image, an optional elevation map (ground-truth nDSM or estimated $\hat{H}$, clipped to the maximum height), a one-hot encoded transmitter location, and antenna gain parameters as per Eqs. 1 and 3. All configurations were trained with Adam optimizer at a learning rate of 10^{-4} , batch size 32, and FP16-mixed precision.

C. Evaluation

1) *Elevation and Pathloss*: Both stages of the pipeline are regression problems and are thus evaluated using the same standard metrics: the Mean Absolute Error is $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ and the Root Mean Square Error is $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$. For elevation estimation, y_i and $\hat{y}_i$ denote ground-truth and predicted height values, and metrics are reported in meters. For radio map estimation, the same expressions apply with y_i representing path loss values, scaled by the normalization factor PL_{scale} to express results in dB. Beyond aggregate metrics, we also perform an error distribution analysis that examines the per-sample RMSE distribution across configurations to assess systematic shifts in prediction quality and the frequency of severe prediction failures.

2) *Memory Scalability*: To quantify the practical storage implications of replacing LiDAR with a learned elevation model, we compare the memory footprint of each approach related to the geographic coverage. We distinguish between the transient memory demands of the acquisition and preprocessing pipeline, which requires access to the full LiDAR point cloud regardless of how compact the final derived artefacts are. Since the transient LiDAR cost grows with the area covered while the learned model maintains a fixed deployment footprint regardless of scale, we report theoretical and empirical storage figures for both approaches on the RMDirectionalBerlin dataset, alongside peak training and inference memory consumption.

3) *Energy and Carbon Footprint*: To quantify the environmental impact of each pipeline, we analyze the carbon footprint of the proposed framework relative to the image-based and 3D-dependent baselines, using commercially available drone platforms as a reference for data acquisition costs. Energy consumption of the Im2Ele model was measured using the eCAL library [18]. Carbon emissions are estimated using the German grid carbon intensity of $363 \text{ gCO}_2\text{e/kWh}$ [19], equivalent to $1.008 \times 10^{-4} \text{ gCO}_2/\text{J}$, reflecting the geographic context of the RMDirectionalBerlin dataset. For data acquisition costs, we use the DJI Matrice 350 RTK as a reference platform, equipped with either the Zenmuse L2 for LiDAR or the Zenmuse P1 for RGB imagery, allowing a direct comparison under identical flight conditions with only the sensor payload exchanged. The reported LiDAR figures cover flight energy only and therefore represent a conservative lower bound, as the energy cost of preprocessing the raw point cloud is not accounted for and would further increase the carbon footprint of the 3D-dependent pipeline.

V. EXPERIMENTAL RESULTS

We present results across four dimensions: elevation estimator accuracy, model-agnostic REM estimation performance across three pathloss estimators, memory scalability, and energy and carbon footprint. All results are reproducible via the code available at our repository.¹

A. Elevation Estimation Performance

In Table I, the height estimation accuracy of the retrained Im2Ele model on the RMDirectionalBerlin dataset is compared with the originally reported results [11]. As can be seen from the table, the retrained model successfully minimizes average errors but struggles with localized extreme outliers compared to the baseline. Specifically, our retrained model achieves an MAE of 1.02 m, improving upon the baseline’s 1.19 m. However, the RMSE increases slightly from 2.88 m in the baseline to 3.11 m in our retrained version. A contributing factor is that only buildings were used for ground-truth, as discussed in Section IV-A, while in the RGB both vegetation and buildings are present. As the two configurations are evaluated on different datasets, this comparison is indicative rather than a direct measure of the retraining strategy. Figure 2 presents an example of an input RGB image, with the corresponding generated height maps and ground-truth LiDAR-derived elevation maps. It can be observed that, while the generated maps reliably capture the general urban morphology, they also introduce height variation in the vegetated areas, a known failure mode discussed further in Section V-B1.

B. Radio Map Estimation Performance

Table II details the radio map prediction accuracy, presenting the RMSE and MAE for three elevation input configurations: *Image-only*, *Im2Ele-predicted*→*nDSM+image*,

TABLE I: Evaluation of Stage 1: comparison of performance of our retrained model with the results reported in Im2Ele [11].

Configuration	Dataset	MAE (m)	RMSE (m)
Im2Ele [11]	IEEE DFC2018	1.19	2.88
Im2Ele	RMDirectionalBerlin	1.02	3.11

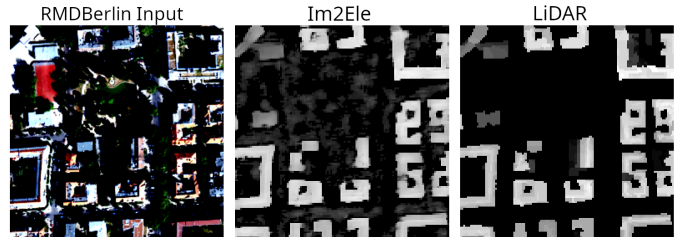


Fig. 2: Qualitative examples comparing generated height maps with ground-truth LiDAR derived elevation maps.

and *LiDAR*→*nDSM+image*. The results show that incorporating learned elevation information provides a consistent and measurable improvement over the image-only baseline. The proposed *Im2Ele-predicted*→*nDSM+image* configuration achieves a Test RMSE of 0.0901 with LitRadioUNet (−4.3%), 0.0847 with LitUNetDCN (−4.3%), and 0.0856 with LitPM-Net (−6.8%), respectively. Expressed relative to the performance ceiling set by the LiDAR upper bound, the proposed method recovers approximately one fifth of the gap between the image-only baseline and the LiDAR-derived elevation input, without requiring any 3D data at inference time. A similar pattern is observed for the MAE values. In practice, such error reductions mean fewer misclassified coverage and outage areas during network planning. It is important to note that neither the image-only nor the proposed baseline eliminates LiDAR from the training pipeline: all configurations rely on LiDAR-derived ray-tracing simulations to generate the ground-truth REM maps used for supervision. These simulations also include vegetation effects, whereas the learned elevation model reconstructs buildings only. This means part of the remaining gap to the LiDAR upper bound is therefore irreducible under the buildings-only design. The remaining gap to the LiDAR upper bound is expected to narrow with larger training datasets.

1) *Error Distribution Analysis*: The results of the error distribution study are presented in Figure 3a, which plots the probability density function (PDF) of the per-sample RMSE across the three evaluated configurations, where Image-only is colored blue, *Im2Ele*→*nDSM+Images* is red and *LiDAR*→*nDSM+Images* is green. We observe that *Im2Ele*→*nDSM+image* provides a systematic and measurable pathloss estimation accuracy improvement over the image-only baseline, successfully shifting the overall error distribution toward lower values and reducing the frequency of severe prediction failures. The vertical dashed lines marking the mean RMSE confirm this shift, consistent with Table II. The *Im2Ele*→*nDSM+image* curve also exhibits a lower den-

¹<https://github.com/sensorlab/REM-estimate>

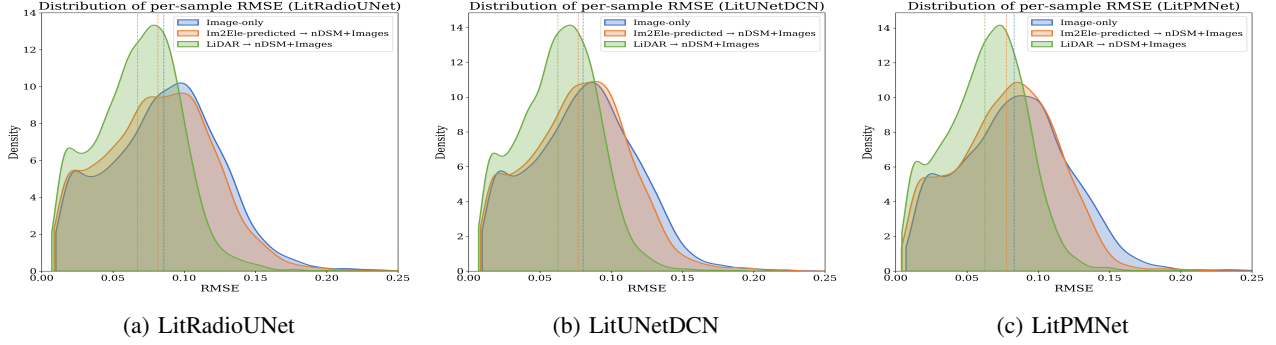


Fig. 3: Per-sample RMSE distributions for all three elevation input configurations across the evaluated architectures.

TABLE II: Evaluation of Stage 2: comparison of training configurations and results (RMSE, MAE). Our results are in bold.

Configuration	Architecture	RMSE	MAE
Image-only	LitRadioUNet	0.0942	0.0417
LiDAR→nDSM+images	LitRadioUNet	0.0735	0.0313
Im2Ele-predicted→nDSM+images	LitRadioUNet	0.0901	0.0401
Image-only	LitUNetDCN	0.0885	0.0400
LiDAR→nDSM+images	LitUNetDCN	0.0684	0.0304
Im2Ele-predicted→nDSM+images	LitUNetDCN	0.0847	0.0380
Image-only	LitPMNet	0.0918	0.0408
LiDAR→nDSM+images	LitPMNet	0.0686	0.0295
Im2Ele-predicted→nDSM+images	LitPMNet	0.0856	0.0372

sity in the high-error region where RMSE exceeds 0.125. These observations indicate that while the learned elevation model does not reach the upper bound performance of the *LiDAR→nDSM+Images* (green), it effectively mitigates extreme prediction errors and provides a more reliable estimation than imagery alone. To further characterise the relationship between the two pipeline stages, Figure 4 plots the per-tile elevation RMSE against the corresponding RML RMSE across all test tiles, with colours indicating five training checkpoints of the elevation estimator. The bulk of tiles show no strong correlation between the two error sources, indicating that the REM estimator is robust to typical levels of height estimation noise. However, tiles with extreme elevation errors ($\text{RMSE} > 0.05$) consistently exhibit elevated RML RMSE, confirming that only severe height estimation failures propagate into pathloss prediction errors.

C. Memory and Scalability: Data vs. Model

Beyond predictive performance, an important consideration is memory scalability. LiDAR-based approaches impose a large transient memory cost during acquisition and preprocessing that grows linearly with geographic coverage, although the final preprocessed artifacts are compact. In contrast, the learned elevation model has a fixed storage footprint regardless of the area covered. On the RMDirectionalBerlin dataset (424 tiles, $256\text{ m} \times 256\text{ m}$, 27.79 km^2), the LiDAR point cloud at 9.8 points/m^2 [6] corresponds to 294 MB/km^2 uncompressed, i.e., 8.17 GB for the tile footprint alone, while the full ac-

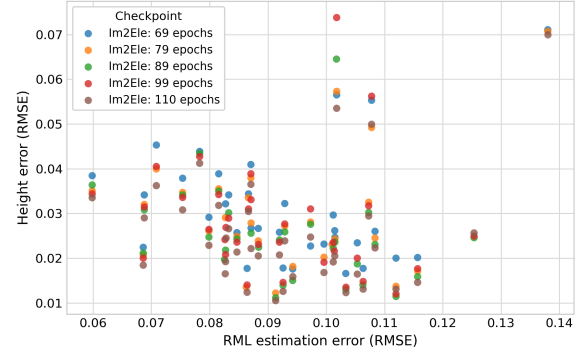


Fig. 4: Per-tile elevation RMSE vs. RML RMSE. Colours indicate different training checkpoints of the elevation estimator.

quisition campaign from which the tiles are extracted measures 249 GB in LAZ format, exceeding 1 TB uncompressed. Although the derived nDSM height maps occupy only a few megabytes, the complete point cloud must be retained and processed to produce them, imposing an unavoidable $\mathcal{O}(A)$ preprocessing demand. In contrast, Image-only and $\text{nDSM+image} \rightarrow \text{Im2Ele}$ require no LiDAR data at inference time. The deployed Im2Ele model occupies 606 MB , and the height reference input $\mathbf{H}$ at inference requires only 11 MB , yielding a total deployment footprint that remains fixed as the number of covered regions grows. At the raw point-cloud footprint of 294 MB/km^2 , the learned model thus becomes the more storage-efficient option for any deployment area beyond approximately 2 km^2 . The peak memory consumption during training was 20 GB (including framework and training overhead), while inference required approximately 600 MB .

D. Energy and Carbon Footprint

For LiDAR acquisition, the manufacturer-stated coverage of 2.5 km^2 per flight at 150 m requires 12 flights for the full 27.79 km^2 dataset, consuming $22,741\text{ kJ}$ ($2,293\text{ g CO}_2$) based on the onboard battery capacity of 526.4 Wh ($2 \times \text{TB65}$). For RGB acquisition, the Zenmuse P1 (35 mm lens) at its manufacturer-specified 240 m operating altitude achieves a ground sampling distance of 3 cm/px , nearly seven times finer than the 20 cm/px imagery used as training input, with a coverage of 3 km^2 per flight, requiring 10 flights and consum-

ing 18,950 kJ (1,911 g CO₂). All figures cover flight energy only and represent lower bounds, excluding transportation, operator costs, and preprocessing compute, which is expected to be much more significant for the LiDAR data. Table III summarises the estimated energy and carbon footprint of each approach. The one-time training of Im2Ele requires 411.6 kJ (41.50 g CO₂), after which each inference consumes only 0.16 kJ (0.02 g CO₂) per tile. For the full 424-tile dataset, the total inference energy amounts to 67.84 kJ and 6.84 g CO₂. In the case of repeated surveys over the same region due to environmental changes, the RGB campaign requires approximately 17% less energy than LiDAR. The re-acquisition frequency needed to maintain accuracy is dictated by the pace of environmental change and is the same for both modalities. The advantage of the RGB pipeline lies in the lower cost of each repeated survey.

TABLE III: Estimated energy and carbon footprint (German grid, 363 gCO₂e/kWh [19]).

Method	Energy (kJ)	CO ₂ (g)
Drone LiDAR survey	22741	2293
Drone RGB survey	18950	1911
Im2Ele training (one-time)	411.6	41.50
Im2Ele inference per tile	0.16	0.02
Im2Ele inference full dataset	67.84	6.84
LitRadioUNet inference per tile	0.004	<0.01

VI. CONCLUSIONS AND FUTURE WORK

This paper presented a two-stage framework for Radio Environment Map estimation that eliminates the need for LiDAR data at inference time by replacing it with a learned elevation model. LiDAR remains required during training for both elevation supervision and ground-truth REM generation, a dependency shared by all learning-based approaches on this dataset. Evaluated on the RMDirectionalBerlin dataset, the proposed approach improves RMSE by up to 6.8% over the image-only baseline, recovering approximately one fifth of the gap to the LiDAR upper bound while using the same input features at inference time. Beyond predictive accuracy, the learned elevation model requires only 606 MB of storage regardless of geographic coverage, compared to LiDAR's 294 MB/km² that scales linearly with area, and incurs no additional data acquisition cost per new deployment region. There are two main limitations. First, the evaluation uses a single dataset representative of German urban architecture. Second, the buildings-only elevation design leaves vegetation-induced propagation effects unmodeled. Future work will address these limitations through broader geographic validation, larger training datasets, and satellite imagery as an even more cost-efficient deployment data source.

VII. ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency (ARIS) under grant P2-0016.

REFERENCES

- [1] I. F. Akyildiz and H. Guo, "Wireless communication research challenges for extended reality (xr)," *ITU Journal on Future and Evolving Technologies*, vol. 3, pp. 1–15, 2022.
- [2] R. Schwarz, "The metaverse and extended reality - implications for wireless communications," 2023, accessed: 2025-05-14. [Online]. Available: https://www.rohde-schwarz.com/us/solutions/wireless-communications-testing/wireless-standards/5g-nr/extended-reality-xr-testing/white-paper-the-metaverse-and-extended-reality_257899.html
- [3] M. Pesko, T. Javornik, A. Košir, M. Štular, and M. Mohorčič, "Radio environment maps: The survey of construction methods," *KSII Transactions on Internet and Information Systems (TIIIS)*, vol. 8, pp. 3789–3809, 2014.
- [4] Remcom Inc., "Wireless insite propagation software," 2025, accessed: June 17, 2025. [Online]. Available: <https://www.remcom.com/wireless-insite-propagation-software>
- [5] J. Hoydis *et al.*, "Sionna," 2022. [Online]. Available: <https://nvlabs.github.io/sionna/>
- [6] F. Jaensch, G. Caire, and B. Demir, "Radio map prediction from aerial images and application to coverage optimization," *IEEE Transactions on Wireless Communications*, vol. 25, pp. 308–320, 2026.
- [7] Q. Chen, J. Yang, M. Huang, and Q. Zhou, "Act-gan: Radio map construction based on generative adversarial networks with act blocks," *IET Communications*, vol. 18, no. 19, p. 1541–1550, Oct. 2024.
- [8] OpenStreetMap contributors, "Planet dump retrieved from <https://planet.osm.org/>," 2026, accessed: 2026-03-09. [Online]. Available: <https://www.openstreetmap.org>
- [9] J. Schampheleer, A. Huss, and M. Deruyck, "Network-scale impact of vegetation loss on coverage and exposure for 5g networks," *IEEE Access*, vol. 13, pp. 23 902–23 912, 2025.
- [10] Y. Zeng and X. Xu, "Toward environment-aware 6g communications via channel knowledge map," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 84–91, 2021.
- [11] C.-J. Liu, V. A. Krylov, P. Kane, G. Kavanagh, and R. Dahyot, "Im2elevation: Building height estimation from single-view aerial imagery," *Remote Sensing*, vol. 12, no. 17, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/17/2719>
- [12] Ç. Yapar, R. Levie, G. Kutyniok, and G. Caire, "Dataset of pathloss and ToA radio maps with localization application," *arXiv preprint:2212.11777*, 2022.
- [13] S. Zhang, B. Choi, F. Ouyang, and Z. Ding, "Physics-inspired machine learning for radiomap estimation: Integration of radio propagation models and artificial intelligence," *IEEE Communications Magazine*, vol. 62, no. 8, pp. 155–161, Aug. 2024.
- [14] S. Zhang, S. Jiang, W. Lin, Z. Fang, K. Liu, H. Zhang, and K. Chen, "Generative ai on spectrumnet: An open benchmark of multiband 3-d radio maps," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 2, pp. 886–901, 2025.
- [15] X. Wang, K. Tao, N. Cheng, Z. Yin, Z. Li, Y. Zhang, and X. Shen, "Radiodiff: An effective generative diffusion model for sampling-free dynamic radio map construction," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 2, pp. 738–750, 2025.
- [16] Q. Chen, M. Huang, and J. Yang, "Rem-net+: Quantified 3d radio environment map construction guided by radio propagation model," *IEEE Transactions on Vehicular Technology*, pp. 1–12, 2025.
- [17] L. Wu, D. He, B. Ai, and K. Guan, "Electromagnetic-geometric coupled environmental features-based intelligent path loss model in urban scenarios," *IEEE Antennas and Wireless Propagation Letters*, vol. 25, no. 3, pp. 1259–1263, 2026.
- [18] S.-K. Chou, J. Hribar, V. Hanžel, M. Mohorčič, and C. Fortuna, "The energy cost of artificial intelligence lifecycle in communication networks," *IEEE Journal on Selected Areas in Communications*, vol. 44, pp. 2427–2443, 2026.
- [19] P. Icha and T. Lauf, "Development of the specific greenhouse gas emissions of the german electricity mix, 1990–2024," Umweltbundesamt, Tech. Rep. 13/2025, Apr. 2025. [Online]. Available: https://www.umweltbundesamt.de/system/files/medien/11850/publikationen/13_2025_cc.pdf