Urban Sound Propagation: a Benchmark for 1-Step Generative Modeling of Complex Physical Systems

Martin Spitznagel   MARTIN.SPITZNAGEL@HS-OFFENBURG.DE
Institute for Machine Learning and Analytics (IMLA)
Offenburg University
Germany

Janis Keuper   KEUPER@IMLA.AI
Institute for Machine Learning and Analytics (IMLA)
Offenburg University
Germany

Reviewed on OpenReview: *https: // openreview. net/ forum? id= 1vCAi53AVj*

Abstract

Data-driven modeling of complex physical systems is receiving a growing amount of attention in the simulation and machine learning communities. Since most physical simulations are based on compute-intensive, iterative implementations of differential equation systems, a (partial) replacement with learned, 1-step inference models has the potential for significant speedups in a wide range of application areas. In this context, we present a novel benchmark for the evaluation of 1-step generative learning models in terms of speed and physical correctness.

Our *Urban Sound Propagation* benchmark is based on the physically complex and practically relevant, yet intuitively easy to grasp task of modeling the 2d propagation of waves from a sound source in an urban environment. We provide a dataset with $100k$ samples, where each sample consists of pairs of real 2d building maps drawn from *OpenStreetmap*, a parameterized sound source, and a simulated *ground truth* sound propagation for the given scene. The dataset provides four different simulation tasks with increasing complexity regarding reflection, diffraction and source variance. A first baseline evaluation of common generative *U-Net, GAN* and *Diffusion* models shows, that while these models are very well capable of modeling sound propagations in simple cases, the approximation of sub-systems represented by higher order equations systematically fails.

Information about the dataset, download instructions and source codes are provided on our website: `https://www.urban-sound-data.org`.

Keywords: Generative Models, 1-step Physic Simulation, Sound Propagation

1 Introduction

Critical for urban planning and noise pollution management, traditional sound mapping methods are resource-heavy [Dutilleux et al. (2009)]. Our approach uses a dataset from OpenStreetMap, annotated with simulated sound maps to mirror diverse urban scenarios [Śliwińska Kowalska and Zaborowski (2017)]. The research evaluates the effectiveness of U-Net, Generative Adversarial Networks (GAN) [Goodfellow et al. (2014)], and Denoising Diffusion Probabilistic Models [Ho et al. (2020b)] in adhering to the complexities of ur-

ban soundscapes. We specifically investigate these models' capabilities in capturing sound reflections and diffractions.

In the realm of generative model development for physical phenomena, the availability and specificity of datasets play a crucial role. While datasets are catering to a variety of environmental and structural scenarios, there remains a noticeable gap in resources specifically designed for urban sound propagation. This gap is particularly significant given the complex interplay of variables in urban environments that affect sound dynamics, such as building layouts, material properties, and ambient conditions.

Our research fills this void by introducing a novel dataset for studying urban sound propagation. This dataset is distinct in its focus on the intricate patterns and behaviors of sound waves as they navigate through urban landscapes. The introduction of this dataset marks a pivotal advancement in the field. It provides a much-needed resource for precise simulation and analysis of urban soundscapes and sets a new standard for future explorations in this area. By offering a dataset specifically focused on urban sound dynamics, we aim to catalyze further research and innovation in sound mapping using advanced machine learning techniques, thereby contributing to the evolution of urban planning and noise mitigation strategies.

1.1 Related Work

The integration of physical principles within generative models represents an active and burgeoning area of research, particularly in fields such as image and sound processing. Models like PUGAN [Cong et al. (2023)] and FEM-GAN [Argilaga (2023)] have showcased the potential of combining GANs with physical modeling to enhance performance in environments with complex physical laws. Likewise, progress in fluid dynamics or structural system identification via the PG-GAN approach has highlighted the improvements in both efficiency and precision achievable when generative models incorporate physics-based loss functions and simulations [Kim et al. (2018); Yu and Liu (2024)].

These studies collectively underscore a crucial trend: the integration of physical laws into the training of generative models not only enhances the fidelity and reliability of these models but also significantly improves their performance in tasks involving complex physical phenomena [Zhang et al. (2021); Karpatne et al. (2022); Chen et al. (2023)]. While significant advancements have been made in the realm of generative models across diverse disciplines, the domain of urban sound propagation presents a unique set of challenges that remain largely unaddressed. Recognizing this deficiency, our research underscores the necessity of incorporating physics-guided principles into generative modeling. By doing so, we aim to establish a new baseline for urban sound propagation research, marking an essential first step towards 1-Step generative modeling of complex physical systems.

2 Physics of Sound Propagation

Mathematically, the propagation of sound over time is described via partial differential wave equations. Due to space constrains and the practical nature of our problem setting, we will neglect the derivation from continuous wave equations and directly focus an the discrete and iterative implementations of sound propagation which have been applied for our *ground-truth* simulations.

Following [Vorländer (2007)], for a discrete set of receivers R, the amplitude $L_{R_k}^j$ of receiver R_k at frequency j is computed via iterative differences:

$$L_{R_k}^j \quad = \quad L_W^j - A_{div_{R_k}} - A_{atm_{R_k}}^j - A_{dif_{R_k}}^j - A_{grd_{R_k}}^j \quad , \text{ where} \tag{1}$$

$$L_W^j \quad \text{models the source,} \tag{2}$$

$$A_{div_{R_k}} \quad \text{captures the geometrical spreading,} \tag{3}$$

$$A_{atm_{R_k}}^j \quad \text{represents the atmospheric absorption,} \tag{4}$$

$$A_{dif_{R_k}}^j \quad \text{models diffraction,} \tag{5}$$

$$A_{grd_{R_k}}^j \quad \text{the ground effect - which is neglected in our study.} \tag{6}$$

Additionally, the model accounts for reflections by adjusting the power level L_W^j based on the absorption coefficient α_{vert} of the surfaces involved. This adjustment is performed using the equation

$$L_W^{(n_{ref})} = L_W^{(n_{ref}-1)} + n_{ref} \times 10 \log_{10}(1 - \alpha_{vert}) \tag{7}$$

where n_{ref} indicates the number of reflections considered. Specular reflections are modeled using the image receiver method, which provides a computationally efficient way to account for the angle of incidence being equal to the angle of reflection [Vorländer (2007)]. The sound level at each receiver reflects the cumulative effect of direct, diffracted, and reflected sound paths.

As the number of reflections increases, the complexity of calculating the power level L_W^j also increases. This is because each reflection path introduces additional calculations that are not strictly linear due to the logarithmic nature of the decibel scale and the multiplicative effect of each reflection's absorption.

3 Dataset Creation and Properties

The proposed dataset has been generated from $25k$ real geolocations, spread across 10 different pre-selected cities. For each of these cities, our dataset provides 2,500 samples from distinct locations, thereby offering a wide range of urban environments for our analysis. The associated image samples originate from pre-processed satellite images that capture the urban landscape within a rectangular 500m² area, depicting buildings as black pixels, while open spaces are represented by white pixels. All of this geodata has been collected from open sources via the *Overpass API* [1] and has been processed with *GeoPandas* [Jordahl et al. (2020)] to ensure a consistent coverage of 500m² per sample. Further, we applied a location selection heuristic to steer the sample positioning relative to building structures and required each site to be encircled by a minimum of 10 buildings within a 200-meter radius. Additionally, a constraint of a 50-meter minimum distance from any building to the sample location was imposed to model realistic urban scenarios.

Ground-Truth Simulation. We applied the open source *NoiseModelling v4* simulation framework [Bocher et al. (2019)], which fulfills the *CNOSSOS-EU* standard for noise emission [Kephalopoulos et al. (2012)] in order to generate *ground-truth* propagation maps for

---

1. OSM-data for this study was collected using the Overpass API, available at `http://overpass-api.de/api/map`.

the collected samples according to eq. 1. We extended the existing regular 5m simulation grid by placing additional sound receivers at every 5 meters along the edges and at corners of buildings, while systematically removing any receivers positioned inside the buildings. After the simulation step, we linearly interpolated the receiver values onto a 512x512 or 256x256 pixel maps, where each pixel indicates the decibel level at that specific location. The decibel values are normalized to a grayscale range, mapping a 0-100 dB range to 0-255 grayscale values. Fig. 10 gives an overview of the full process. More details on our scalable implementation of the generation pipeline are given in Appendix A.

| (a) Sattelite image | (b) Receiver grid | (c) Urban layout | (d) Sound propagation |
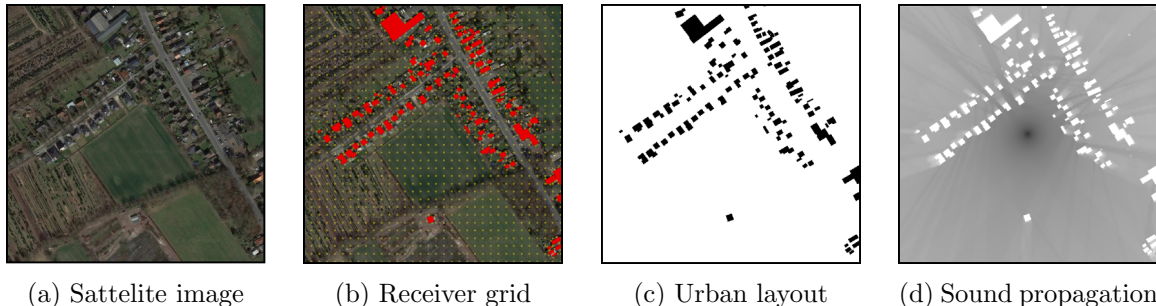
Figure 1: Starting with the selection of a $500m^2$ area (a), buildings are identified, followed by placing a receiver grid (b). The urban layout (c) and the corresponding sound propagation, simulated using the NoiseModelling Framework from a central signal source, are then used in the dataset (d).

Limitation to 2d Propagations. A significant limitation of our dataset is the absence of building height information, which restricts our ability to accurately model vertical sound diffraction. Consequently, our current dataset focuses only on horizontal diffraction. Vertical diffraction is disabled during the simulation step.

3.1 Prediction Tasks

We simulate the *ground-truth* propagation maps at all locations for four different prediction tasks with increasing complexity, resulting in a total of $100k$ samples in the final dataset.
Baseline Task: In the most simplistic baseline setup, a steady noise source with a level of 95 dB at 500 Hz has been simulated without diffraction and reflection. This allows to test generative models under stable conditions, focusing on changes in the environment around the sound source.
Diffraction Task: To isolate the effect of sound wave diffraction around obstacles this setup is nearly identical to the baseline task, except for enabling horizontal diffraction during the simulation process. This task allows to assess the models' precision in predicting how sound waves bend and spread upon encountering buildings.
Reflection Task: Maintaining the same sound level and frequency, this setting focused on modeling sound reflections off surfaces. The constant power level $L_{W,i}^j$ and a standardized absorption coefficient α_{vert} (set to 0.1) were used to simulate paths with multiple reflections.
Combined Task: Incorporating variance in source sound levels (60 to 115 dB) and environmental variables like humidity and temperature, this setting added complexity. The power level L_W^j and atmospheric absorption coefficient A_{atm}^j varied per sample. Reflection and

diffraction were included to replicate realistic sound propagation, with each phenomenon adding complexity by altering the sound paths and power levels at receivers.

4 Evaluation & Baseline Results

4.1 Experimental Setup

Generative Models. For a first baseline evaluation, we benchmarked three widely used architectures for image generation on all four tasks: a standard U-Net [Ronneberger et al. (2015)], a Pix2Pix-based GAN [Isola et al. (2016)], and a Denoising Diffusion Probabilistic Model [Ho et al. (2020a)]. Each model was constructed upon a unified U-Net backbone, scaling from 64 to 1028 channels, and then reconverging to 64. All models were trained with a consistent batch size of 18 for a maximum of 50 epochs. An early stopping criterion with a patience of 3 epochs was implemented to curtail training when no improvement in validation loss was detected, thereby preventing overfitting. All additional hyperparameters can be found in Appendix C.

Evaluation Metrics. The primary interest of our baseline evaluation is to investigate the capabilities of standard image-to-image generative models to predict physically correct sound propagations in our defined tasks. The pixel difference at a given location between true and prediction is used to determine the quality of this prediction. While Mean Absolute Error (MAE) quantifies the average magnitude of prediction errors, the implementation of Weighted Mean Absolute Percentage Error (wMAPE) enhances this evaluation by particularly penalizing inaccuracies in instances where predictions gives high values in areas that physically should have low amplitudes, such as regions behind buildings. This approach assigns a maximal error rate of 100% to these errors. Additionally, to assess how well each model captured the reflections or diffractions we specifically measured both metrics in areas not in direct line of sight to the central signal source in the OSM images. This assessment was conducted by performing ray tracing from the sound source, allowing us to evaluate the models' effectiveness in prediction propagations in Line-of-Sight (LoS) and outside Line-of-Sight (NLoS) by reflection and diffraction.

4.2 Baseline Performance

The initial assessment of U-Net, GAN, and Diffusion models using MAE and MAPE established a baseline for comparison (see Table1). The baseline provides a controlled environment to assess the generative models' fundamental capabilities. With a fixed sound level and frequency, the models' performance in predicting $L_{R_k}^{j}$ could be evaluated without the additional complexities of environmental interactions. This setup confirmed that even under stable conditions, there is some variance in the models ability to capture the subtle changes around the noise source.

Reflection and Diffraction Tasks. For the reflection task, each model displayed distinct performance characteristics regarding their ability to model reflective patterns. A sample from the reflection dataset, as illustrated in Figure 2, showcases the unique approaches of each model architecture in handling the higher-order complexities of reflections (see eq. 7). More qualitative samples are provided in Appendix B, quantitative results are shown in table 2.

Table 1: Performance metrics including MAE and MAPE across different tasks for all architectures. Note: Due to varying decibel value ranges, the combined task metrics cannot be directly compared with the other three tasks.

Model	Baseline		Diffraction		Reflections		Combined	
	MAE	wMAPE	MAE	wMAPE	MAE	wMAPE	MAE	wMAPE
UNet	2.08	19.45	**1.65**	9.75	3.22	31.87	1.77	20.59
GAN	**1.52**	**8.21**	1.66	**8.03**	**2.88**	**16.57**	1.76	**19.12**
Diffusion	2.57	25.21	2.12	11.85	4.14	35.20	**1.57**	21.45
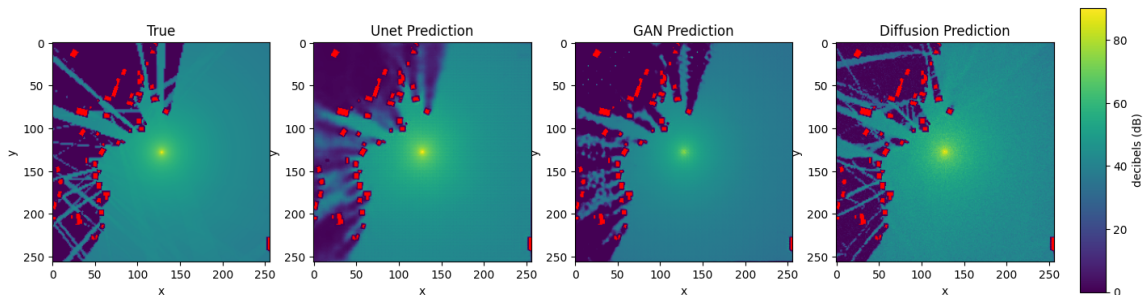
Figure 2: Comparing the ground-truth simulation with the predictions from U-Net, GAN, and diffusion model for a single sample within the **reflection** task.

The U-Net model tends to blur most areas, indicating a general approach to sound mapping that prioritizes broad patterns over specific details. On the other hand, the GAN model attempts to imprint specific patterns within these areas, reflecting a more detailed-oriented strategy. Notably, the Diffusion model, despite having a high MAE and MAPE in NLoS areas, makes an apparent effort to visually replicate reflection patterns.

In the reflection task, the recursive nature of sound interactions (see eq. 7) introduces non-linear complexities. The nonlinearity arises because the adjustment for each reflection depends on the cumulative effect of all previous reflections, each potentially altering the sound level in a unique manner. The generative models must, therefore, be capable not just of modeling the initial interaction of sound with the environment, but also of capturing the compounded effect of multiple reflections.

Combined Task. The final comprehensive evaluation using variance in all terms revealed the adaptability of the Diffusion model (see table 2), showing a marked improvement in its performance. Meanwhile, U-Net and GAN exhibited a stable performance across different tasks, highlighting their consistency.

4.3 Analysis

By separating the loss in LoS and NLoS conditions, each model exhibited distinct strengths and weaknesses. When models encounter tasks with reflections or diffractions, we observe an increase in MAE for NLoS regions across all tested architectures. Both the UNet and GAN models show a moderate rise in NLoS error, but it is most significant in the Diffu-
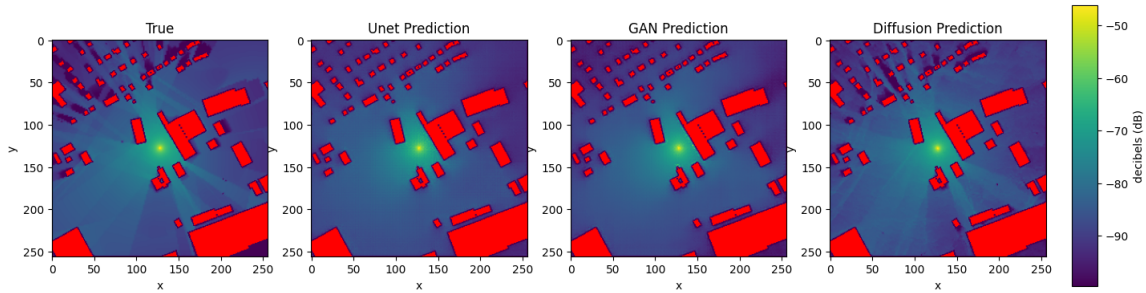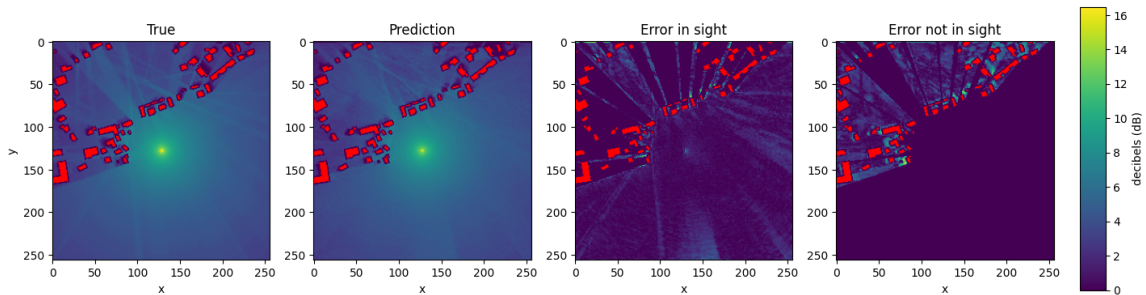
Figure 3: Comparing the ground-truth simulation with the prediction of the diffusion model for a single sample within the **combined task**, distinguishing between the MAE in LoS and NLoS.

sion model. Despite its ability to visually replicate sound reflection patterns, there is a notable discrepancy between its visual outputs and actual acoustic precision. This discrepancy suggests that the Diffusion model, while visually detailed, does not fully capture the complexities of sound physics.

Figure 4: Comparing the ground-truth simulation with the prediction of the diffusion model for a single sample within the **reflection** task, distinguishing between the MAE in LoS and NLoS.

An illustrative example from the final setting incorporating a complex mix of direct sound paths, reflections, and diffractions, with varying environmental conditions is depicted in Figure 4. This sample visualizes the nuanced challenges faced by the diffusion model. More qualitative results are provided in Appendix B, quantitative results are shown in table 2.

Runtime Analysis The comparison of performance metrics in Table 2 reveals differences between generative models and the traditional sound propagation simulation in terms of processing speed. The generative models, especially GAN and U-Net, show a significant improvement in runtime over the conventional simulation (up to factor $20k$). The Diffusion model has a slightly higher runtime than GAN and U-Net. This comparison underscores the advantages of GAN and U-Net in terms of speed, highlighting their effectiveness as

Table 2: Consolidated performance metrics across tasks for all architectures. The "Runtime per Sample" is reported in seconds and represents an average computed over 100 samples.

Model	Task	LoS MAE	NLoS MAE	LoS wMAPE	NLoS wMAPE	Runtime per Sample (s)
Sim.	Baseline	0.00	0.00	0.00	0.00	20.4717 ±1.4885
UNet	Baseline	2.29	1.73	12.91	37.57	0.0126 ±0.0012
GAN	Baseline	**1.73**	**1.19**	**9.36**	**6.75**	**0.0095 ±0.0012**
Diffusion	Baseline	2.42	3.26	15.57	51.08	4.1560 ±0.0061
Sim.	Diffraction	0.00	0.00	0.00	0.00	20.6027 ±0.7953
UNet	Diffraction	0.94	**3.27**	4.22	22.36	0.0126 ±0.0012
GAN	Diffraction	**0.91**	3.36	**3.51**	**18.06**	**0.0095 ±0.0012**
Diffusion	Diffraction	1.59	**3.27**	8.25	20.30	4.1560 ±0.0061
Sim.	Reflection	0.00	0.00	0.00	0.00	25.0973 ±3.2928
UNet	Reflection	2.29	5.72	12.75	80.46	0.0126 ±0.0012
GAN	Reflection	**2.14**	**4.79**	**11.30**	**30.67**	**0.0095 ±0.0012**
Diffusion	Reflection	2.74	7.93	17.85	80.38	4.1560 ±0.0061
Sim.	Combined	0.00	0.00	0.00	0.00	29.2395 ±4.7059
UNet	Combined	1.39	2.63	10.10	45.15	0.0126 ±0.0012
GAN	Combined	1.37	2.67	**9.80**	40.68	**0.0095 ±0.0012**
Diffusion	Combined	**1.26**	**2.21**	13.07	**40.38**	4.1560 ±0.0061

quicker alternatives to traditional simulation methods. For additional analysis, please refer to Appendix 6.

5 Discussion & Future Work

Our first baseline results showed that the proposed dataset provides a suitable proxy-problem for further research and development of data driven models for the prediction of complex physical tasks. The provided 2d sound propagation tasks have a manageable compute complexity, both on the simulation side as well for the training of state-of-the-art generative models, while providing different levels of difficulty. The evaluation of current image-to-image generative models shows, that speedups of up to factor $20k$ compared to the simulation are realistic, while the physical correctness still needs further improvement. The initial analysis of the prediction results points towards two very interesting phenomena, which need further investigation: I) the very different error pattern produced by different generative approaches, and II) the eminent disability of all model to capture higher order dependencies.

Broader Impact Statement. The aim of the presented dataset is to foster research towards fast and physically correct 1-step generative models. Such models have the potential to drastically speed-up complex simulation of environmental (climate) or engineering problems, resulting in a wide range of positive effects in various applications. As for any simulation technique, the authors can not entirely rule out malignant applications.

Reproducibility Statement

Detailed information about the dataset and the download link are provided under the website: `https://www.urban-sound-data.org/`. The code necessary to replicate the experiments discussed in this paper has been made publicly available on the GitHub website: `https://github.com/urban-sound-data/urban-sound-data`. Additionally, a comprehensive description of the dataset is given as a Datacard in Section C.

Funding Acknowledgement

References

A. Argilaga. Fem-gan: A physics-supervised deep learning generative model for elastic porous materials. *Materials*, 2023. doi: 10.3390/ma16134740.

Erwan Bocher, Gwenaël Guillaume, Judicaël Picaut, Gwendall Petit, and Nicolas Fortin. Noisemodelling: An open source gis based tool to produce environmental noise maps. *ISPRS International Journal of Geo-Information*, 8(3):130, 2019. doi: 10.3390/ijgi8030130.

Yilun Chen, Qi Wu, Yujie Sui, Yizhou Wang, and Song-Chun Zhu. Physically grounded vision-language models for robotic manipulation. *arXiv preprint arXiv:2309.02561*, 2023.

Runmin Cong, Wenyu Yang, Chongyi Zhang, Weiand Li, Chunle Guo, and Sam Huang, Qingmingand Kwong. Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Transactions on Image Processing*, 2023. doi: 10.1109/TIP.2023.3286263.

Guillaume Dutilleux, Jérôme Defrance, Fabrice Junker, David Ecotière, Benoit Gauvreau, Dirk van maercke, Emmanuel Duc, Michel Bérengier, and Francis Besnard. *Road noise prediction 2 - Noise propagation computation method including meteorological effects (NMPB 2008)*. 06 2009.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *CoRR*, abs/2006.11239, 2020a. URL `https://arxiv.org/abs/2006.11239`.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020b.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. URL `http://arxiv.org/abs/1611.07004`.

Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Adrian Garcia Badaracco, Carson Farmer, Geir Arne Hjelle, Alan D. Snow, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Nick Eubank, maxalbert, Aleksey Bilogur, Sergio Rey, Christopher Ren, Dani Arribas-Bel, Leah Wasser, Levi John Wolf, Martin Journois, Joshua Wilson, Adam Greenhall, Chris Holdgraf, Filipe, and François Leblanc. geopandas/geopandas: v0.8.1, July 2020. URL https://doi.org/10.5281/zenodo.3946761.

Anuj Karpatne, Imme Ebert-Uphoff, Sai Ravela, Hassan A. Babaie, and Vipin Kumar. Physics-guided neural networks (pgnn): An application in lake temperature modeling. *arXiv preprint arXiv:2212.02500*, 2022.

S Kephalopoulos, M Paviotti, and F Anfosso-Lédée. Common noise assessment methods in europe (cnossos-eu). Technical Report EUR 25379 EN, Publications Office of the European Union, Luxembourg (Luxembourg), 2012. JRC72550.

Byungsoo Kim, V. C. Azevedo, N. Thürey, Theodore Kim, M. Gross, and B. Solenthaler. Deep fluids: A generative network for parameterized fluid simulations. *Computer Graphics Forum*, 38, 2018. doi: 10.1111/cgf.13619.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL http://arxiv.org/abs/1505.04597.

Michael Vorländer. *Auralization: Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality.* Springer Science & Business Media, 2007.

Yang Yu and Yongming Liu. Physics-guided generative adversarial network for probabilistic structural system identification. *Expert Systems with Applications*, 239:122339, 2024. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2023.122339. URL https://www.sciencedirect.com/science/article/pii/S0957417423028415.

Yufeng Zhang, Wei Li, Yangang Gao, and Shihong Li. Physdiff: Physics-guided human motion diffusion model. *arXiv preprint arXiv:2106.03704*, 2021.

Mariola Śliwińska Kowalska and Kamil Zaborowski. Who environmental noise guidelines for the european region: A systematic review on environmental noise and permanent hearing loss and tinnitus. *International Journal of Environmental Research and Public Health*, 14(10), 2017. ISSN 1660-4601. doi: 10.3390/ijerph14101139. URL https://www.mdpi.com/1660-4601/14/10/1139.

Appendix A. Scalable Simulation Pipeline.

The dataset generation pipeline visualized in Figure 5 is a crucial component of our study, developed to efficiently process sound propagation data in diverse urban settings. Utilizing the NoiseModelling framework Bocher et al. (2019), we have automated the data input and simulation processes within a Docker-containerized environment. The pipeline commences with the automatic download of a 500m² area map from OpenStreetMap for each location using the Overpass API, followed by their import into the NoiseModelling framework alongside the signal source.
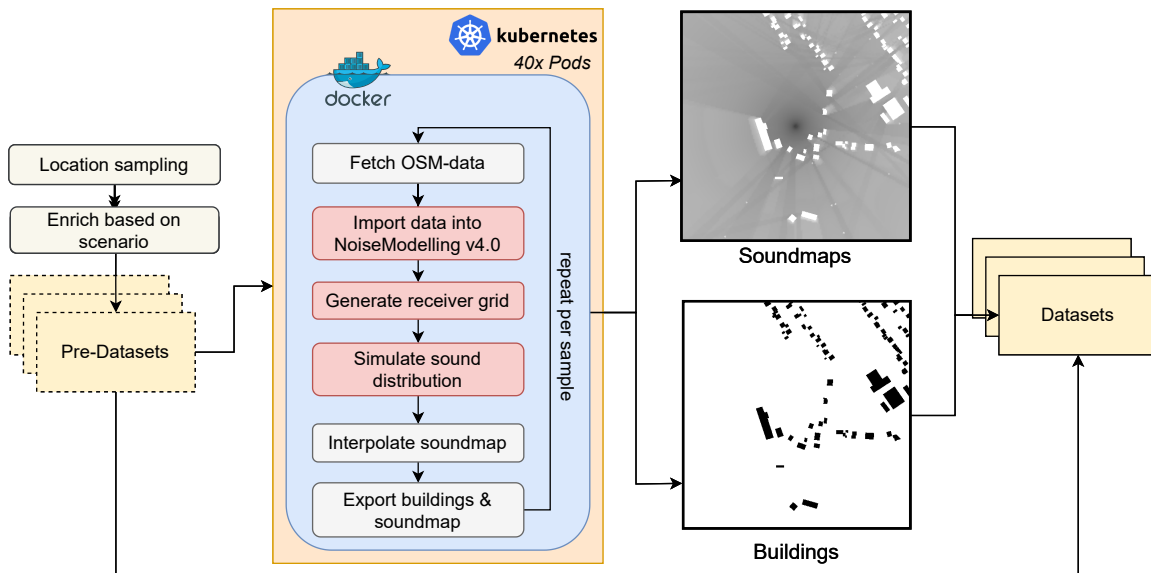
Figure 5: Detailed visualization of the dataset generation pipeline.

Considering the computational intensity of this process, with an average duration of 30 seconds per sample, our pipeline is structured for scalability. It operates on a Kubernetes cluster with 40 pods, enabling us to complete the generation of the entire dataset, encompassing 25,000 data points for each complexity level, in approximately 20 hours.

Appendix B. Additional Qualitative Results
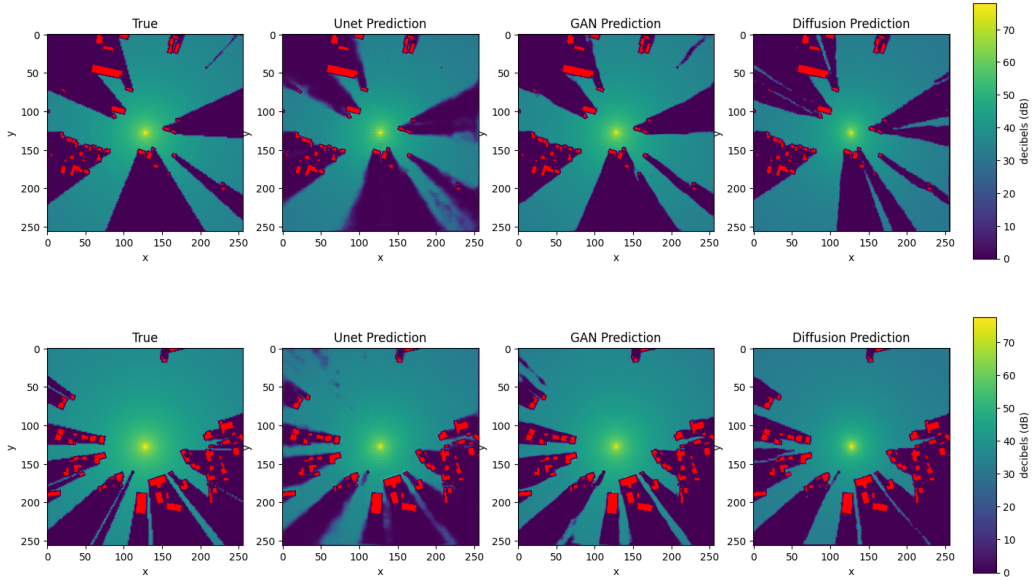
Figure 6: Comparing the output of the physical simulation with the predictions from U-Net, GAN, and diffusion model for a single sample within the **baseline** task dataset.
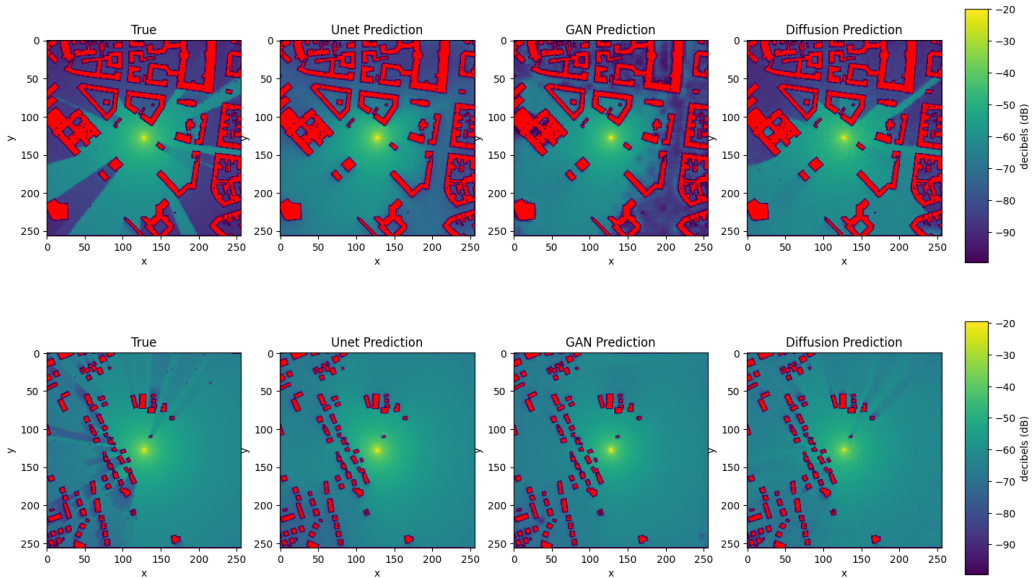
Figure 7: Comparing the output of the physical simulation with the predictions from U-Net, GAN, and diffusion model for a single sample within the **diffraction** task dataset.
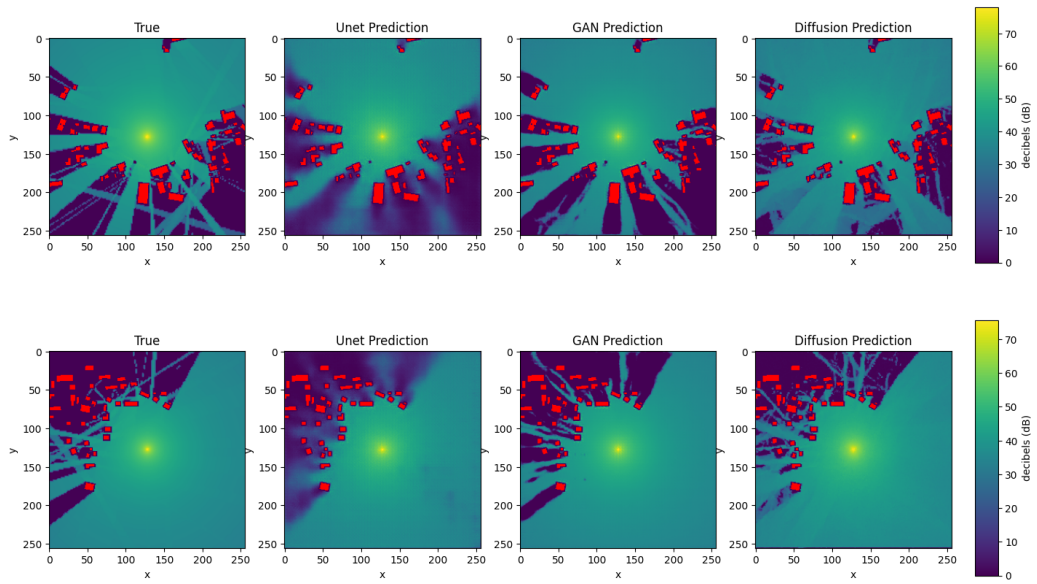
Figure 8: Comparing the output of the physical simulation with the predictions from U-Net, GAN, and diffusion model for a single sample within the **reflection** task dataset.
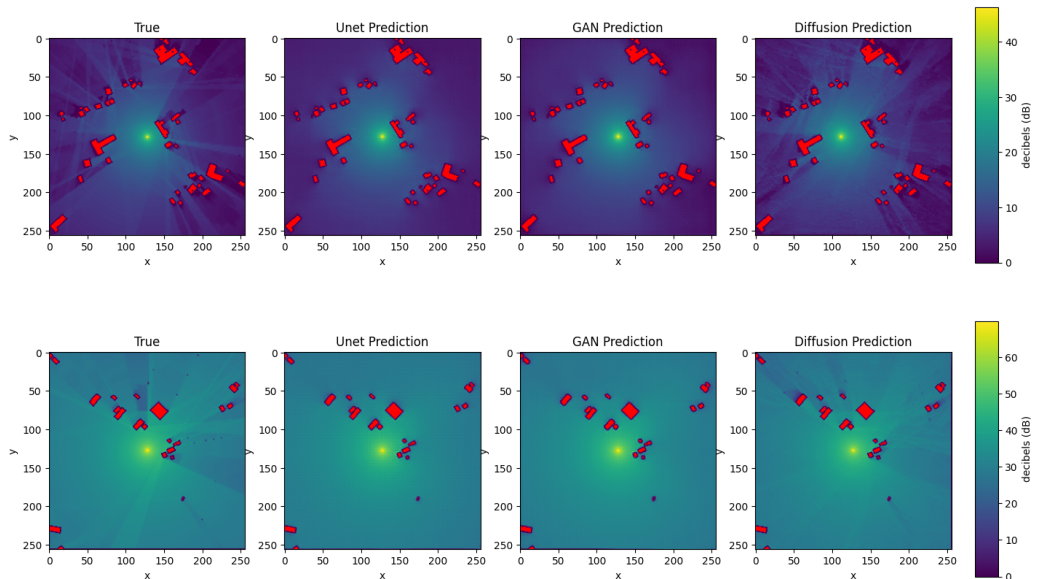
Figure 9: Comparing the output of the physical simulation with the predictions from U-Net, GAN, and diffusion model for a single sample within the **combined** task dataset.
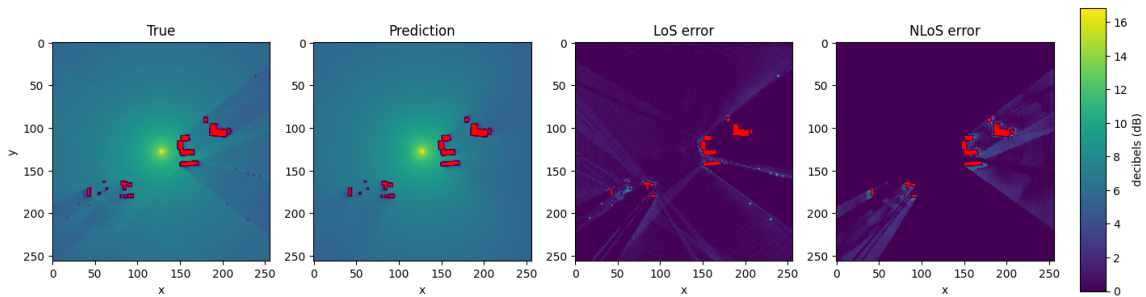
(a) U-Net
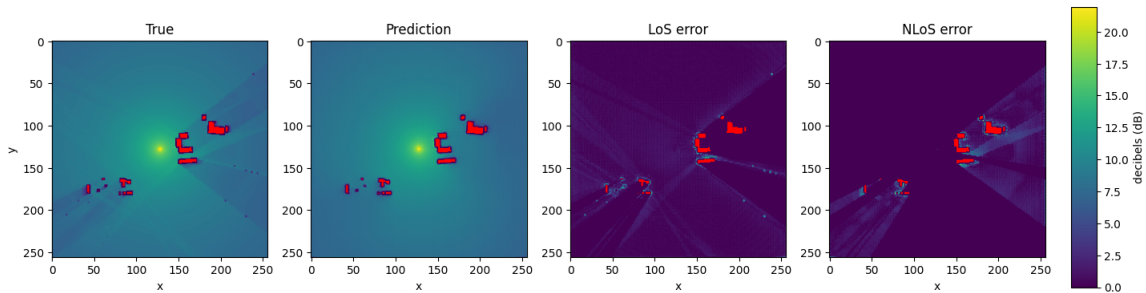
(b) GAN
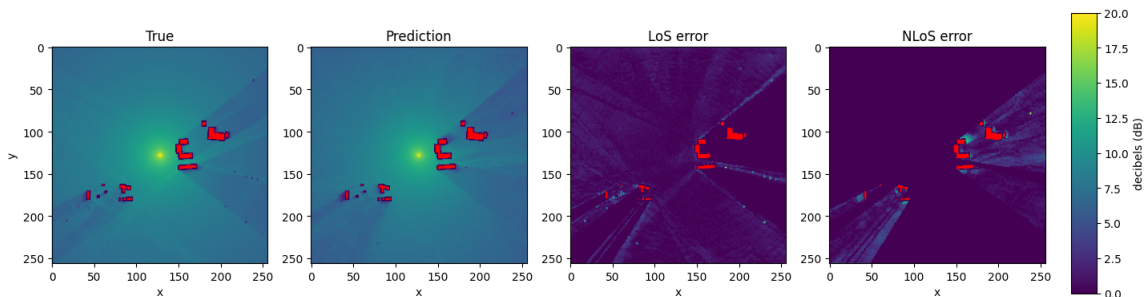
(c) Diffusion model

Figure 10: Comparing the output of the physical simulation with the prediction of U-net (a), GAN (b) and diffusion model (c) for a single sample within the **reflection** task dataset, distinguishing between the MAE in LoS and NLoS.

Appendix C. Training Setup

This appendix provides a detailed overview of the architecture, input specifications and hyperparameter used for the generative models utilized in this research. The U-Net model architecture, adopted from [Ronneberger et al. (2015)], and the GAN setup, based on [Isola et al. (2016)], are designed to process grayscale image inputs. For the **combined** task, the input is extended by appending additional parameters as a separate dimension.

For the Diffusion model, we followed the methodology described in [Ho et al. (2020a)] while incorporating conditional inputs as separate dimensions besides the noised input image. Each model was constructed upon a unified U-Net backbone, scaling from 64 to 1028 channels, and then reconverging to 64.

Table 3: UNet Training Hyperparameter

Hyperparameter	Value
Batchsize	18
Learning Rate	1×10^{-4}
Image Size	256
Max Epochs	50

Table 4: GAN Training Hyperparameter

Hyperparameter	Value
Batchsize	18
Learning Rate Discriminator	1×10^{-4}
Learning Rate Generator	2×10^{-4}
Image Size	256
Max Epochs	50
L1 Lambda	100
Lambda GP	10

Table 5: Diffusion Model Training Hyperparameter

Hyperparameter	Value
Batchsize	18
Learning Rate	1×10^{-4}
Image Size	256
Max Epochs	50
Noise Steps	1000

Table 6: Model vs. Simulation Performance Comparison for Single Sample Processing

Model – Condition	Mean Runtime (s)	Std. Dev. (s)
UNet	0.0126	0.0012
GAN	0.0095	0.0012
Diffusion	4.1560	0.0061
Simulation - Baseline	20.4717	1.4885
Simulation - Diffraction	20.6027	0.7953
Simulation - Reflection	25.0973	3.2928
Simulation - Combined	29.2395	4.7059
Simulation - Combined - 3rd Order Reflections	186.2295	16.8491

Urban Sound Propagation

This dataset is assembled for research into urban sound propagation, comprising 25,000 data points across 10 diverse cities. Each city is represented by 2,500 locations, offering a comprehensive analysis of various urban configurations. The dataset utilizes OpenStreetMap (OSM) imagery to detail the urban layout within a 500m x 500m area for each location, where buildings are delineated with black pixels and open spaces with white pixels.

Supplementing the urban structural images, the dataset includes sound distribution maps at resolutions of 512x512 and 256x256. These maps are precisely generated through the NoiseModelling v4.0 framework, an advanced simulation tool engineered for accurate modeling of sound dynamics within urban environments.

For researchers and experts interested in exploring the intricacies of sound simulation, additional insights can be obtained from the NoiseModelling framework documentation.

DATASET LINK	DATA CARD AUTHOR(S)
https://doi.org/10.5281/zenodo.10609793	**Martin Spitznagel, IMLA**: Owner **Janis Keuper, IMLA**: Owner

Authorship

Publishers

PUBLISHING ORGANIZATION(S)	INDUSTRY TYPE(S)	CONTACT DETAIL(S)
Institute for Machine Learning and Analytics (IMLA)	Academic – Tech	**Affiliation:** Offenburg University **Website:** https://imla.hs-offenburg.de/

Funding Sources

INSTITUTION(S)	FUNDING OR GRANT SUMMARY(IES)
German Federal Ministry of Education and Research (BMBF)	This project is part of the "Forschung an Fachhochschulen in Kooperation mit Unternehmen (FH-Kooperativ)" program, within the joint project KI-Bohrer [https://www.ki-bohrer.de/] and is funded under the grant number 13FH525KX1.

Dataset Overview

DATA SUBJECT(S)	DATASET SNAPSHOT	CONTENT DESCRIPTION
Data about natural phenomena Data about places and objects	<table><tr><td>Size of Dataset</td><td>~5 GB</td></tr><tr><td>Number of Instances</td><td>~100,000</td></tr><tr><td>Training</td><td>19908 x 4</td></tr><tr><td>Evaluation</td><td>3732 x 4</td></tr><tr><td>Test</td><td>1244 x 4</td></tr></table> **Additional Notes:** The dataset is segmented into four distinct subsets, each tailored to explore specific aspects of sound propagation in urban environments: Baseline, Reflection, Diffraction, and Combined.	A data point in this dataset consists of two main components: an urban layout image from OpenStreetMap and a corresponding sound distribution map. The urban layout image is a 500m x 500m area depiction where buildings are marked in black and open spaces in white. The sound distribution map generated using NoiseModelling v4.0, illustrates the sound dynamics within that urban environment at resolutions of 512x512 or 256x256. The dataset comprises four subsets: **Baseline** for basic sound behavior, **Reflection** examining sound wave interactions with surfaces, **Diffraction** focusing on sound navigating around objects, and **Combined** which merges reflection, diffraction and changing environmental factors like temperature and humidity.

Dataset Version and Maintenance

MAINTENANCE STATUS	VERSION DETAILS	MAINTENANCE PLAN
Regularly Updated (New versions of the dataset have been or will continue to be made available.)	**Current Version:** 2.0 **Last Updated:** 02/2024 **Release Date:** 02/2024	**Feedback:** martin.spitznagel@hs-offenburg.de

Example of Data Points

PRIMARY DATA MODALITY	SAMPLING OF DATA POINTS	DATA FIELDS
Multimodal - Image Data - Geospatial Data - Tabular Data		

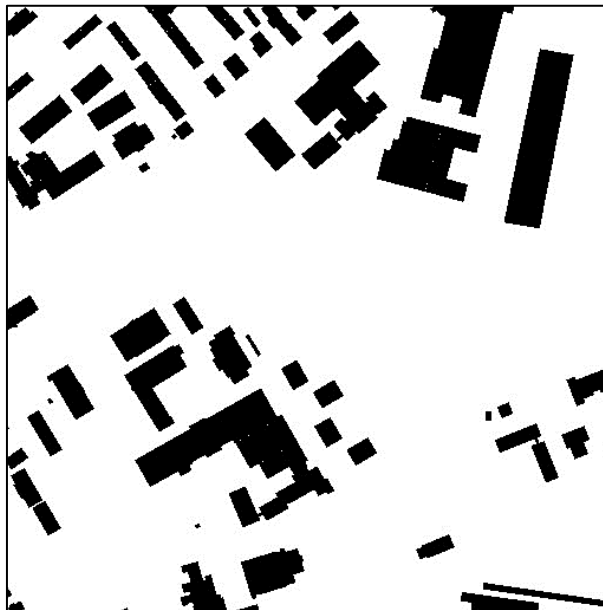| Field Name	Field Value	Description
lat	float	Latitude of the sound measurement location.
long	float	Longitude of the sound measurement location.
db	Object	Key-value pairs of sound levels in decibels for a given frequency (lwd{fqz}).
soundmap	string	Path to 256x256 resolution sound distribution image.
soundmap_512	string	Path to 512x512 resolution sound distribution image.
osm	string	Path to Open Street Map image showing urban layout.
temperature	float	Temperature (°C) at the location.
humidity	float	Humidity (%) at the location.
yaw	float	Orientation of the noise source. Can be empty.
sample_id	int	Unique identifier for the data point.

TYPICAL DATA POINT	EXAMPLE OF DATA POINT
| ```
{
  "lat": 48.030229082138526,
  "long": 11.367773397906852,
  "db": {"lwd500": 69},
  "soundmap":
  "./soundmaps/256/0_LEQ_256.png",
  "soundmap_512":
  "./soundmaps/512/0_LEQ_512.png",
  "osm":
"./buildings/osm_23747.png",
  "temperature": 12,
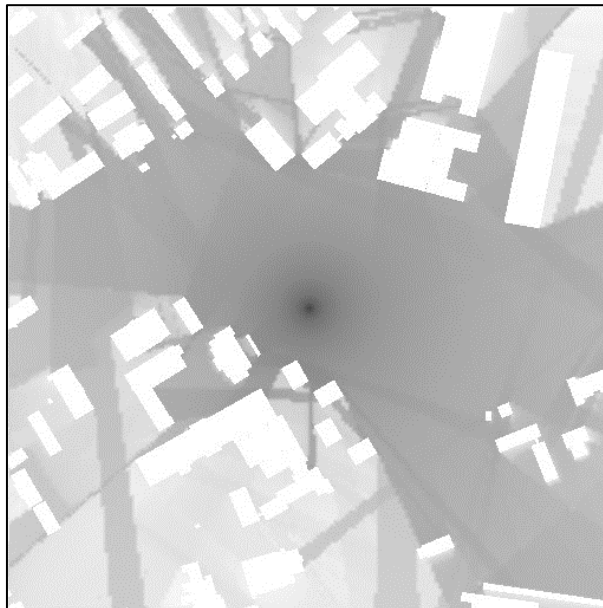  "humidity": 35,
  "yaw": None,
  "sample_id": "23747"
}
``` | Below is an example of an OSM and Simulated Sound Propagation pair:

OSM:

Simulated Sound Propagation:

 |

Motivations & Intentions

Motivations

PURPOSE(S)	DOMAIN(S) OF APPLICATION	MOTIVATING FACTOR(S)
Research	*`Generative Models`, `1-step Physic Simulation`, `Sound Propagation`, `Machine Learning`*	Generative models, through their capacity to learn from complex datasets, hold significant potential in understanding the intricate physics behind sound propagation. By training on data that encompasses various urban layouts and the corresponding sound distribution maps, these models can uncover the underlying patterns and principles governing how sound travels and interacts with different obstacles, such as buildings and open spaces. This capability enables the creation of predictive models that can simulate sound behavior in any urban environment, thereby offering valuable insights for urban planning, acoustic design, and noise mitigation strategies, all rooted in a deep understanding of the physical laws of sound propagation.

Intended Use

DATASET USE(S)	SUITABLE USE CASE(S)	UNSUITABLE USE CASE(S)
Safe for research use	**Sound propagation**: Enhancing models for predicting how sound travels in densely built areas.	**Predicting indoor noise levels**: The dataset is designed for outdoor urban sound distribution, not for indoor environments. **Traffic flow or congestion analysis:** It focuses on sound distribution and ignores vehicle movements or traffic patterns.
	RESEARCH AND PROBLEM SPACE(S)	
	The dataset specifically addresses the problem space of outdoor urban noise propagation. It is intended for developing models that can predict sound distribution around buildings within urban environments.	

Provenance

Collection

METHOD(S) USED	METHODOLOGY DETAIL(S)	SOURCE DESCRIPTION(S)
- API - Physical Simulation Framework	**Overpass API** **Source:** The Overpass API is a read-only API that serves up custom selected parts of the OSM map data. It acts as a database over the web: the client sends a query to the API and gets back the data set that corresponds to the query. **Platform:** https://overpass-api.de/ **Is this source considered sensitive or high-risk?** [Yes / **No**] **Dates of Collection:** [10 2023 - 12 2024] **Primary modality of collected data:** Geospatial Data **NoiseModelling v4.0** **Source:** An advanced simulation tool engineered for accurate modeling of sound dynamics within urban environments. **Platform:** https://github.com/Universite-Gustave-Eiffel/NoiseModelling **Is this source considered sensitive or high-risk?** [Yes / **No**] **Dates of Collection:** [10 2023 - 12 2024] **Primary modality of collected data:** Geospatial Data	**OSM Buildings:** This source provides images from Open Street Map (OSM) that depict urban layouts, specifically focusing on buildings within cities. In these images, black pixels represent buildings, and white pixels indicate open spaces. **Sound Propagation:** This component of the dataset involves simulated sound distribution images around urban centers, where the noise source is placed at the center.

Collection Criteria

DATA SELECTION	DATA INCLUSION	DATA EXCLUSION
Location Sampling: The locations are randomly sampled across 10 cities/areas: ["Hamburg", "Hannover", "Augsburg", "Bonn", "Muenchen", "Schwerin", "Berlin", "Paris", "Stuttgart", "Aachen"]	**Enough Obstacles:** At least 10 Buildings within a circle r=200m around the sound source. **No Obstacle to close:** No Building within a r=50m circle around the sound source. **Additional Notes:** 	**Additional Notes:** If the Data Inclusion criteria is not met, the data is excluded. No additional exclusion criteria are introduced. |