From Simulation to Prediction: Wind Turbine Noise Estimation Using Convolutional Recurrent Neural Networks

Abdelazyz Rkhiss[1*], *Gabriel* Vasile[1], *Arthur* Finez[2], *Jean-Remy* Gloaguen[2], and *Julien* Maillard[3]

[1]GIPSA-Lab, Grenoble-INP, Grenoble, France Engie
[2]Green, Acoustic Division, Lyon France
[3]CSTB, Acoustics, Vibration, Lighting and Electromagnetism Division, Grenoble, France

Abstract. Wind turbine noise (WTN) assessment is a critical issue for environmental impact studies and regulatory compliance, particularly for residents near wind farms. This paper investigates the use of convolutional recurrent neural networks (CRNNs) to extract the wind turbine noise from the total noise. A dedicated learning dataset is constructed by combining measured background noise in the environment with simulated WTN signals, yielding representative total acoustic noise time series. The simulations rely on a physics-based framework using an aeroacoustics source and an outdoor sound propagation model to build a site-specific time series of noise levels in 1/3-octave bands, accounting for wind speed, wind direction, turbine operating modes, and day/night atmospheric conditions. The dataset covers multiple wind farms and receiver locations, with wind directions selected based on long-term meteorological statistics. The proposed CRNN-based approach is trained to capture local patterns and the temporal dynamics of WTN, with the objective of accurately estimate overall sound pressure level (OASPL) of WTN in various acoustic conditions.

1 Introduction

Energy supply has always played a central role in geopolitical dynamics and in the coexistence of societies. Since the end of the eighteenth century, fossil fuels have been the dominant source of energy worldwide. However, the energy crisis of the early 1970s, together with the growing awareness of environmental impacts associated with the rapid increase in global average temperature, has progressively driven technological research toward clean and renewable energy sources, notably wind and photovoltaic power [1].

*Corresponding author: abdelazyz.rkhiss@grenoble-inp.fr

Despite their environmental benefits, wind turbines generate acoustic noise that remains a relevant issue for residents near wind farms, as elevated sound levels may affect residents' quality of life, particularly through sleep disturbance and potential impacts on overall health. Freiberg et al. [2] provide a comprehensive review of the potential health effects associated with wind turbine noise exposure in residential areas.

A curtailment plan is implemented to limit the acoustic impact of wind farms and ensure regulatory compliance. During the operational phase, this plan is typically verified through noise measurements performed during repeated on/off cycles of the turbines. However, such measurement-based approaches suffer from several limitations. First, the limited duration of the measurements does not always capture the full variability of the background noise. Second, these procedures are costly, as they require shutting down turbines and thus interrupting energy production.

For efficient monitoring and the implementation of more reliable and dynamic curtailment strategies, an automatic tool capable of estimating WTN without shutting down the wind turbines would be highly beneficial. A method based on non-negative matrix factorization (NMF) was previously proposed by Gloaguen et al. [3]. Although this approach yields promising results in specific scenarios, the associated uncertainty currently limits its deployment at an industrial scale. Consequently, recent research efforts have shifted toward deep neural network–based approaches, given their demonstrated success in related audio signal processing tasks such as source separation [4], sound classification [5], and sound event detection [6]. In addition, a study based on support vector regression (SVR) by Anicic et al. [7] highlights the importance of incorporating auxiliary information, such as wind speed, to improve the accuracy of wind turbine noise prediction models. A previous study employed recurrent neural networks (RNNs), to estimate WTN levels from acoustic data, while also incorporating environmental variables such as wind speed and turbine power output as additional features [8].

The method proposed in this paper relies on convolutional recurrent neural networks (CRNNs) for the estimation of WTN levels from total noise. Recurrent neural networks have demonstrated their effectiveness in modeling acoustic scenes and capturing temporal patterns in audio signals [9], particularly for handling long-term dependencies in sequential data. Combined with convolutional layers, which enable efficient extraction of local time–frequency features, CRNN architectures are well suited to model both the spectral and temporal characteristics of acoustic signals. CRNN architectures have shown strong performance in complex acoustic tasks such as sound source localization in noisy and reverberant environments [10].

The remainder of this paper is organized as follows. Section 2 describes the construction of the dataset, including the WTN simulation and the background measurements. Section 3 presents the proposed CRNN architecture and the associated input features. Section 4 details the experimental setup and discusses the obtained results, and Section 5 concludes the paper and outlines perspectives for future work.

2 Dataset

2.1 Wind Turbine Noise Simulation

Training a supervised tool such CRNN to extract the WTN contribution from total noise times eries requires labelled sound scenes: for every total noise sample the corresponding WTN (the target) must be known. Directly measuring WTN level is challenging due to the presence of environmental noise contributions from multiple sources that cannot be fully isolated for far field receivers (in living people area). To overcome this limitation, a hybrid approach combining numerical simulation and acoustic measurements is adopted.

The wind turbine noise time series is generated using a physics-based simulation framework that models both noise emission and propagation. The emission is described using semianalytical aeroacoustic formulations accounting for dominant mechanisms such as trailingedge and inflow turbulence noise [11], while sound propagation from each source element to the receivers is computed using the Harmonoise model [12]. Scintillation effects are not considered in this study.

The simulations are performed for four onshore wind farms, using the turbine models installed at each location. Two turbine types are considered, corresponding to the Senvion MM92 and Vestas V126 machines. The simulated data consist of time series with a temporal resolution of 4 s, represented as time–frequency features in one-third-octave bands.

Two practical considerations motivated the use of time–frequency energetic features rather than raw audio waveforms. First, privacy and data protection constraints: background recordings are acquired in residential areas (including inside dwellings), where raw audio may contain personally identifiable information; retaining only band-energy time series ensures compliance with privacy regulations (GDPR). Second, storage and computational constraints: the experiments rely on continuous recordings for several weeks to months, making the storage of high-resolution audio impractical, whereas compact third-octave representations significantly reduce storage requirements and data handling costs.

Figure 1 illustrates the spatial configuration of the first wind farm as an example. The wind farm includes three wind turbines (E1–E3) and four noise receptors (R001, R003, R006, and R008). The coordinates of turbines and receptors are expressed in the Lambert-93 projected coordinate system (EPSG:2154), where the X-axis corresponds to the Easting coordinate and the Y-axis to the Northing coordinate, both expressed in meters.

The dashed lines indicate the distances between each wind turbine and the surrounding receptors. These distances are used as input for the acoustic propagation calculations performed with the Harmonoise model.

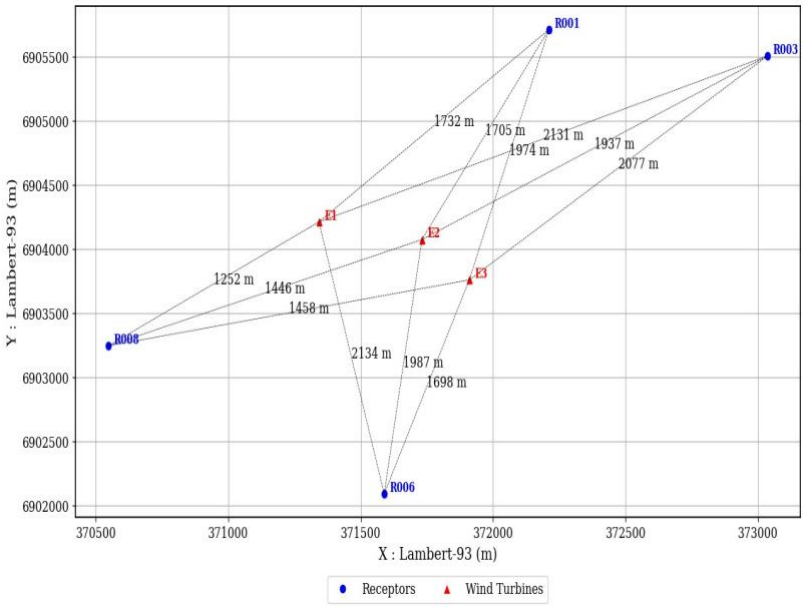


Figure 1. Spatial configuration of the first wind farm used as an illustrative example. Coordinates are expressed in Lambert-93 (Easting X, Northing Y) in meters.

Site-specific meteorological parameters are considered in the simulations, including wind speed class, air temperature, humidity, and atmospheric conditions, selected based on longterm meteorological observations at each wind farm. Wind directions used in the simulations are chosen according to wind rose analyses, with a finer angular resolution applied in the most frequently occurring sectors in order to better represent dominant wind regimes.

Wind direction is defined with respect to the East and increases positively toward the North. Under this convention, a wind direction of 0 corresponds to a wind blowing from East to West, while 90 corresponds to a wind blowing from North to South. The selected wind directions are therefore not uniformly spaced but reflect the statistical distribution of wind occurrences observed at each wind farm. The meteorological configurations retained for the simulations, including wind directions, temperature, and humidity, are summarized for each wind farm in Table 1, and the desired wind speeds for each turbine and operating modes are given in table 2.

Table 1. Atmospheric conditions for wind farms

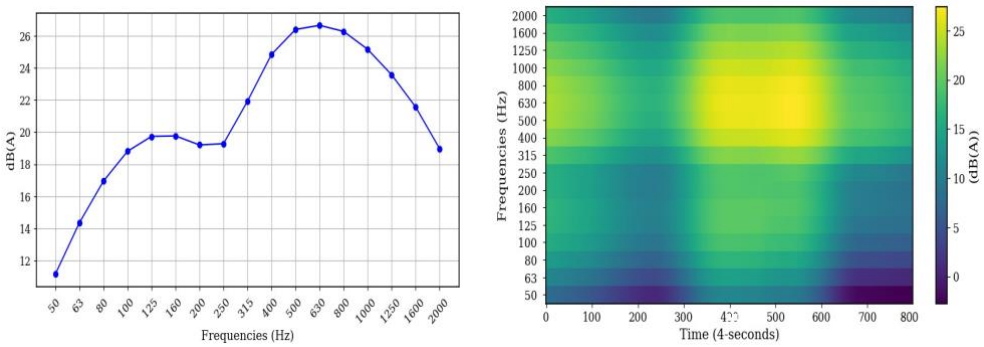
First wind farm	
Wind directions [°]	45, 135, 160, 180, 192, 205, 215, 225, 235, 245, 260, 290, 330
Temperature (day/night) [°C]	15, 5
Humidity [%]	70
Number of wind turbines	3 (Model: MM92)
Second wind farm	
Wind directions [°]	0, 45, 90, 135, 167, 190, 212, 235, 270, 315
Temperature (day/night) [°C]	10, 5
Humidity [%]	70
Number of wind turbines	4 (Model: V126)
Third wind farm	
Wind directions [°]	0, 45, 90, 135, 180, 202, 225, 247, 270, 315
Temperature [°C]	15
Humidity [%]	70
Number of wind turbines	4 (Model: MM92)
Fourth wind farm	
Wind directions [°]	0, 45, 67, 90, 135, 180, 225, 247, 270, 292, 326
Temperature (day/night) [°C]	13.8, 11.3
Humidity (day/night) [%]	79.8, 86
Number of wind turbines	4 (Model: V126)

Table 2. Selected wind speeds and operating modes

Model	Mode	Wind speeds at HH [m/s]
MM92	Standard	3, 5, 7, 8, 9, 11, 13, 15
MM92	SMII C	5, 7, 9
MM92	SMII D	9, 11, 13
V126	Standard	3, 5, 7, 9, 10, 11, 13, 15
V126	SO12	5, 7, 9, 11, 13, 15

The simulated WTN signals exhibit characteristic spectral and temporal patterns. Figure 2a presents an example of the simulated WTN spectrum at receiver R01 of the first wind farm. The one-third-octave-band spectrum of the synthesized WTN shows dominant contributions in the frequency range between 315 and 1600 Hz. Two main spectral components can be identified: a contribution centered around 630 Hz, associated with trailing-edge noise, and a lower-frequency contribution around 125 Hz, related to inflow turbulence noise at the leading edge of the blades.

Figure 2b illustrates the corresponding time–frequency representation over a time window during daytime conditions, highlighting the temporal evolution of these spectral components in the simulated WTN signal.



(a) Sample of the simulated wind turbine noise spectrum.

(b) Time–frequency representation of a simulated WTN signal.

Figure 2. Results of wind turbine noise simulation: (a) spectrum and (b) spectrogram representation.

2.2 Background Noise Measurements

The background noise level L_{BN} is measured using a class I sound level meter (Svantek SV277C) equipped with windshields and full outdoor weather protection kits. Each microphone is installed at a height of 1.5 m above ground in free-field conditions. The acoustic data are recorded in one-third-octave bands covering the frequency range from 50 Hz to 2000 Hz, with a temporal resolution of 4 s. Fig. 5 shows a sample of the time–frequency representation of a measured background noise signal.

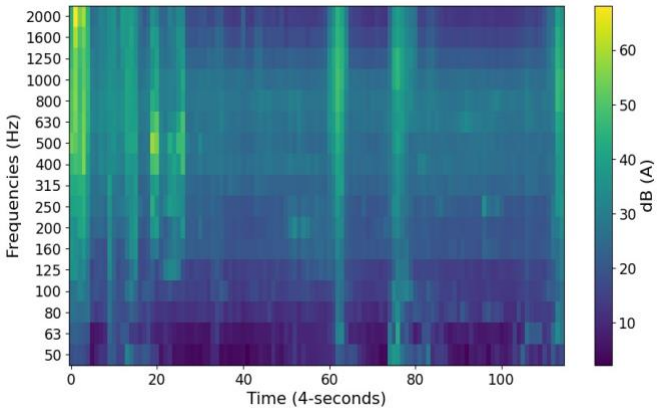


Figure 3. Time–frequency representation of the background noise measurement.

2.3 Dataset Construction

The total noise level L_{TN} is obtained by combining the synthesized wind turbine noise L_{WTN} with the measured background noise L_{BN} The total noise is computed through an energetic summation of sound pressure levels, expressed as:

$$L_{\mathrm{TN}} = L_{\mathrm{WTN}} \oplus L_{\mathrm{BN}}, \tag{1}$$

where the operator $\oplus$ denotes the energetic summation of decibel values. The resulting dataset consists of several months of time series data collected across four wind farms, each characterized by multiple receiver locations. For each receiver, paired time series of total noise L_{TN} and corresponding wind turbine noise L_{WTN} are generated with a temporal resolution of 4 s ($L_{Aeq,4s}$). These paired signals constitute the input–output samples used for supervised training and evaluation of the proposed model.

3 Neural Network Architecture

3.1 Proposed CRNN Architecture

To estimate WTN level from total noise, a convolutional recurrent neural networks (CRNNs) architecture is adopted. CRNNs have demonstrated strong performance in audio signal processing tasks by combining convolutional with recurrent layers. This hybrid architecture has been applied to sound event detection [13], motivating its use in the present study. These models excel by leveraging their ability to learn complex patterns and non-linear relationships in acoustic data.

The proposed architecture, illustrated in Fig. 5, processes the total noise time series. Each input sample consists of 115 consecutive time steps, with 17 features per time step, corresponding to one-third-octave-band sound levels. The network outputs an estimate of the WTN OASPL for each input sequence.

The sequence length of 115 time steps was determined based on an analysis of the temporal correlation of wind turbine noise. To estimate the characteristic correlation time, a time series of L_{WTN} was recorded at a 1 s resolution using a sound level meter located approximately 400 m from an industrial wind turbine (110 m rotor diameter, 2.2 MW capacity, 90 m hub height). Although this dataset is not sufficiently long to train and evaluate the CRNN model, it is suitable for estimating the temporal dynamics of the signal. The correlation time is derived from the autocorrelation function (ACF), following standard approaches used in turbulence and environmental noise analysis [14]. The procedure is summarized as follows:

- Compute the normalized autocorrelation function $\rho(\tau)$ of the measured L_{WTN} time series.
- Evaluate its logarithm, $\ln(\rho(\tau))$.
- Identify the asymptotic region corresponding to the exponential decay of the ACF. • Fit a linear model of the form:

$$\ln(\rho(\tau)) = \alpha\tau + b \tag{2}$$

over this region.

- Estimate the correlation time as:

$$T_c = -\frac{1}{\alpha}. \tag{3}$$

Figure 4 illustrates the logarithm of the normalized autocorrelation function together with the linear regression fitted over the exponential decay region. The estimated slope α yields a correlation time of:

$$T_c = 460 \text{ s}. \tag{4}$$

Given the temporal resolution of the dataset ($\Delta t = 4$ s), the corresponding sequence length in discrete time steps is:

$$L = \frac{T_c}{\Delta t} = \frac{460}{4} = 115. \tag{5}$$

This sequence length is adopted as the input window of the CRNN to capture the effective temporal memory of the wind turbine noise process.

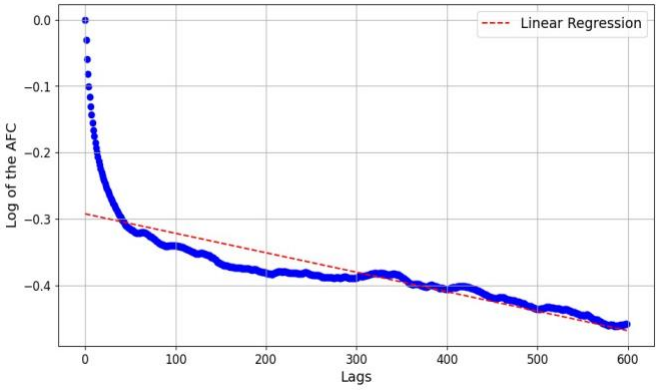


Figure 4. Logarithm of the normalized autocorrelation function of the wind turbine noise time series and linear regression over the exponential decay region used to estimate the correlation time.

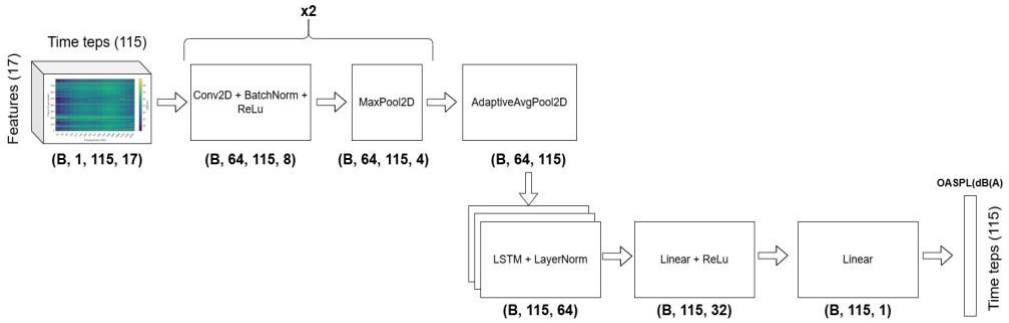


Figure 5. Proposed CRNN architecture illustrating the forward propagation from the input time–frequency representation to the predicted OASPL sequence.

The first stage of the network consists of two 2D convolutional layers with 32 and 64 filters, respectively, applied jointly along the temporal and frequency dimensions. These layers aim to extract local spectro-temporal patterns from the input signals while preserving the temporal resolution of the sequence. Frequency-domain dimensionality is progressively reduced using max-pooling operations applied exclusively along the frequency axis.

The resulting convolutional feature maps are subsequently averaged along the frequency dimension through global frequency pooling, yielding a sequence of learned spectral embeddings for each time step. Temporal dependencies over longer time scales are then modeled using two stacked long short-term memory (LSTM) [15] layers with 64 hidden units each. A layer normalization is applied to the LSTM outputs in order to stabilize hidden state dynamics during training.

Finally, a time-distributed linear output layer produces an estimate of the OASPL associated with the turbine noise at each time step of the input sequence.

3.2 Data representation

The input data are time–frequency representations of acoustic signals in third-octave bands. Each sample is segmented into fixed-length temporal sequences to allow sequential modeling:

• Input spectrogram (total noise):

$L_{TN} \in \mathrm{R}^{T\times F}$, with $T = 115$ time steps and $F = 17$ third-octave frequency bands.

• Target overall sound pressure level of the WTN: $L_{WTN} \in \mathrm{R}^{T}$.

The problem addressed in this work corresponds to a regression task: the model learns a mapping

$$f_\theta : \mathrm{R}^{T\times F} \rightarrow \mathrm{R}^{T},$$

where f_θ denotes a neural network parameterized by θ that predicts OASPL value per time step.

3.3 Loss function and evaluation metrics

Evaluation metrics.

Model performance is evaluated using the Mean Absolute Error (MAE), computed directly in dB(A). Let N denote the total number of predicted samples across the dataset. The MAE is defined as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|L_{\mathrm{WTN},i} - \hat{L}_{\mathrm{WTN},i}\right|, \tag{6}$$

where $L_{\mathrm{WTN},i}$ and $\hat{L}_{\mathrm{WTN},i}$ denote the ground-truth and predicted wind turbine noise levels at sample i, respectively, expressed in dB(A). The MAE provides a physically interpretable measure of the average absolute prediction error in decibels. In addition to MAE, the prediction bias is also evaluated in order to quantify potential systematic overestimation or underestimation of the WTN level. The bias is defined as:

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{L}_{\text{WTN},i} - L_{\text{WTN},i}\right), \tag{7}$$

where a positive value indicates an average overestimation of the wind turbine noise level, while a negative value indicates an underestimation.

Loss function

The model is trained using the Huber loss to improve robustness against outliers and highly variable acoustic conditions. For a residual $e_i = L_{\text{WTN},i} - \hat{L}_{\text{WTN},i}$, the Huber loss is defined as:

$$\mathcal{L}_\delta(e_i) = \begin{cases} \frac{1}{2}e_i^2, & \text{if } |e_i| \le \delta, \\ \delta\left(|e_i| - \frac{1}{2}\delta\right), & \text{otherwise,} \end{cases} \tag{8}$$

where $\delta > 0$ controls the transition between quadratic and linear behavior. For small residuals, the loss behaves as a Mean Squared Error (MSE), promoting stable gradient updates. For larger residuals, it transitions to a linear penalty similar to MAE, reducing the influence of large deviations and improving robustness during optimization.

Optimization.

Model parameters are optimized using the AdamW algorithm [16, 17], an adaptive gradientbased optimization method with decoupled weight decay. Unlike standard stochastic gradient descent, AdamW maintains exponential moving averages of both the gradients and their squared values, enabling parameter-specific adaptive learning rates. At iteration t, given the gradient

$$g_t = \nabla_\theta \mathcal{L}_\delta(\theta_t), \tag{9}$$

the first and second moment estimates are updated as

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)g_t, \tag{10}$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)g^2{}_t, \tag{11}$$

where β_1 and β_2 are exponential decay rates controlling the momentum and variance estimates, respectively. To correct the initialization bias, the estimates are bias-corrected as:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \hat{v}_t = \frac{v_t}{1 - \beta_2^t}. \tag{12}$$

The parameter update rule in AdamW is then given by:

$$\theta_{t+1} = \theta_t - \eta\left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda\theta_t\right), \tag{13}$$

where η denotes the learning rate, ϵ is a small constant ensuring numerical stability, and λ is the weight decay coefficient. Gradient norm clipping with a maximum norm of 1.0 was applied during training to prevent exploding gradients in the recurrent layers.

The learning rate η is a critical hyperparameter that controls the magnitude of parameter updates during optimization. To determine a suitable value, a logarithmic learning-rate range test was conducted over the interval $[10^{-7},10^{-1}]$, following a standard empirical approach for deep neural network training. During this procedure, the learning rate was progressively increased on a logarithmic scale while monitoring the validation MAE. As shown in Fig. 6, the MAE initially decreases within a stable region, indicating efficient learning, before increasing sharply due to training instability at higher learning rates. The optimal learning rate is selected at the point corresponding to the minimum validation MAE. Based on this analysis, an initial learning rate of $\eta = 10^{-1}$ was retained. Although this value is relatively high compared to commonly used settings, it was found to provide stable and efficient convergence in our case. This behavior can be attributed to the use of the AdamW optimizer with gradient norm clipping, which stabilizes the updates even at higher learning rates, and to the scale of the input features expressed in dB(A), which leads to moderate gradient magnitudes.

4 Experimental Setup and Results

4.1 Training and Validation Procedure

The acoustic data are organized as continuous time series. To preserve temporal continuity while generating training samples, the signals are segmented into fixed-length sequences of 115 time steps using an overlapping window strategy. An overlap of 35 time steps (approximately 30% of the sequence length) is applied between consecutive sequences.

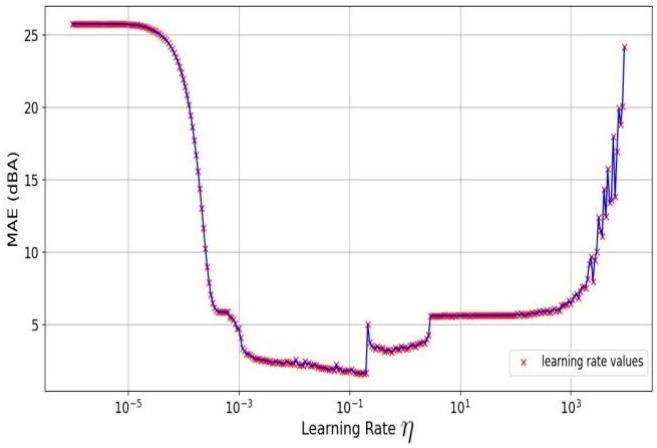


Figure 6. Learning rate search results for the CRNN model. The optimal η, associated with the minimum MAE is around 10^{-1}.

This approach allows adjacent sequences to share part of the signal, thereby preserving temporal dependencies and reducing boundary effects between consecutive segments. The choice of a 30% overlap results from empirical evaluation. Larger overlap values were tested during preliminary experiments; with, no significant improvement was observed in terms of temporal continuity of the estimated WTN during inference. Increasing the overlap mainly led to higher computational cost without noticeable performance gain. Therefore, a 30% overlap was selected as a trade-off between performance and computational efficiency.

The dataset is divided into training and test subsets. From the training portion, 20% of the sequences are further reserved for validation. The validation set is constructed by randomly selecting sequences from different segments of the training time series. The test set consists of independent time series selected from each wind farm in order to evaluate the model under different scenarios.

4.2 Dataset Description

The Signal-to-Noise Ratio (SNR) is defined as the difference between the overall sound pressure level of the wind turbine noise (L_{WTN}) and that of the background noise (L_{BN}), i.e.,

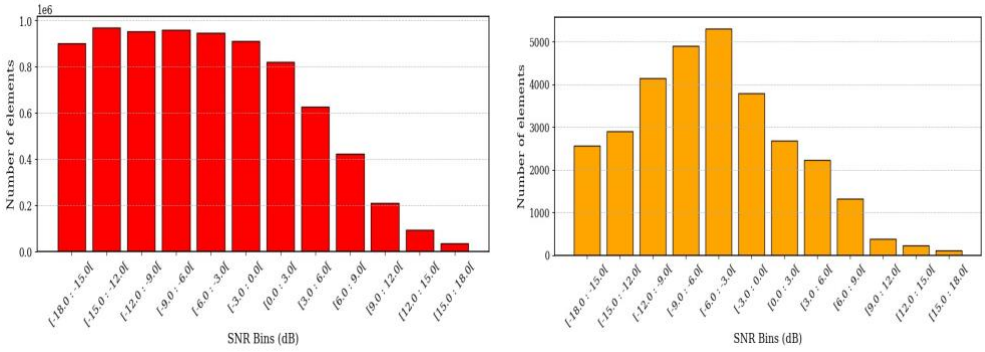
$$\mathrm{SNR} = L_{WTN} - L_{BN}. \tag{14}$$

The SNR provides an indication of the relative contribution of wind turbine noise compared to the ambient background noise. Negative SNR values correspond to situations where the background noise level exceeds the wind turbine noise level, while positive SNR values indicate conditions where the turbine noise dominates.

Figures 7a and 7b present the distribution of SNR values for the training and test sets, respectively. The SNR values in the dataset span a wide range from approximately −18 to 18 dB, allowing the models to be trained and evaluated under a variety of acoustic conditions.

As observed in both figures, the majority of samples correspond to negative SNR values, indicating that in most simulated situations the background noise level exceeds the wind turbine noise level.

Conversely, positive SNR values correspond to situations where the wind turbine noise becomes dominant over the background noise. These conditions typically occur at shorter turbine–receiver distances, or low background noise, or under meteorological conditions favoring sound propagation of WTN.



(a) SNR distribution of the training set. (b) SNR distribution of the test set.

Figure 7. Distribution of SNRs in the training and testing data. The SNR values are divided into bins, showing the number of elements per bin.

The training set contains a large number of samples in order to expose the learning model to diverse acoustic scenarios. The test set follows a similar distribution but with different

samples that cover multiple scenarios, ensuring an independent evaluation of the model performance across SNR.

4.3 Results and Performance Evaluation

This subsection presents the performance of the trained model evaluated on the independent test set. During training the validation MAE reaches a minimum value of 1.96 dB(A), while the test MAE is 1.98 dB(A), confirming the ability of the model to generalize from training to unseen samples within the dataset. For more details the evaluation focuses on the estimation MAE of the wind turbine noise level under different SNR conditions.

Figure 8 illustrates the Mean Absolute Error obtained for the OASPL of wind turbine noise estimation as a function of the SNR bins. Each bar represents the average MAE computed over the corresponding SNR interval, while the error bars indicate the variability of the estimation error within each bin.

The largest errors are observed for low SNR values, particularly for the intervals below −12 dB. In these conditions, the background noise level is significantly higher than the wind turbine noise, which makes the estimation problem more challenging since the turbine contribution is strongly masked by background noise. As a result, the local spectral patterns associated with WTN become less distinguishable, limiting the ability of the convolutional layers to extract relevant features.

For SNR values close to 0 dB, where the turbine noise and background noise have comparable levels, the MAE falls below approximately 0.75 dB(A). When the wind turbine noise becomes dominant (SNR > 0 dB), the model achieves its best performance with MAE values typically below 0.5 dB(A).

Figure 9 shows the prediction bias of the model for SNR bins on the test set. The bias for each SNR interval is defined as the mean difference between the predicted and the simulated labels of the OASPL wind trubine noise. A positive bias indicates an overestimation of the wind turbine noise level, while a negative bias corresponds to an underestimation.

The vertical error bars represent the uncertainty associated with the mean bias and correspond to the standard error of the mean (SEM), computed as:

$$\mathrm{SEM} = \frac{s}{\sqrt{N_{\mathrm{bin}}}}, \tag{15}$$

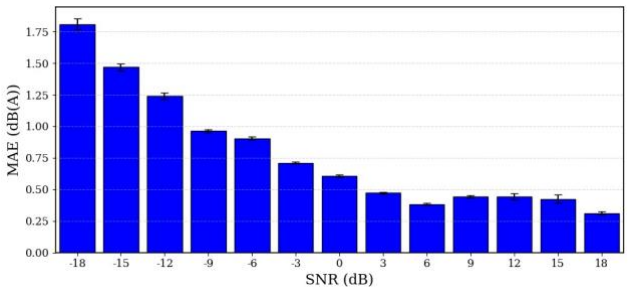


Figure 8. MAE of the wind turbine noise level estimation as a function of the SNR bins on the test dataset.

where s is the standard deviation of the prediction errors within the SNR bin. The SEM provides an estimate of the confidence in the mean bias value, with smaller values indicating a more stable estimate across the samples of the corresponding SNR interval.

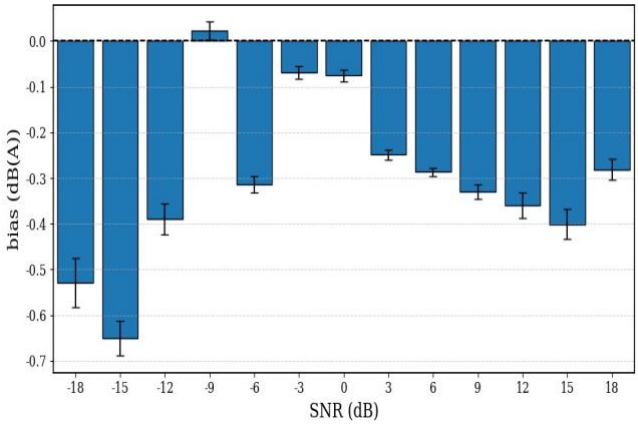


Figure 9. Bias of wind turbine OASPL predictions on the test set.

The bias analysis indicates that the model slightly underestimates the turbine noise for most SNR conditions, with a maximum underestimation of approximately 0.7 dB(A). The magnitude of the bias is small, indicating that the model predictions are stable across different samples within the same SNR bins.

The largest negative biases occur in the lowest SNR bins, where the turbine noise is most difficult to detect. The relatively small SEM values show that the estimated bias is stable across samples within each interval.

This systematic underestimation likely induced by the dominance of low-SNR conditions in the dataset and the difficulty of extracting weak WTN signal from background noise. An exception is observed around SNR = −9 dB, where the model slightly overestimates the turbine noise, resulting in a small positive bias. This behavior may be influenced by the training data distribution and the model parameterization.

To further illustrate the test performance of the proposed model, Fig. 10 presents an example of OASPL WTN estimation over a time window from the test set. The figure compares the predicted WTN level with the corresponding reference signal.

The predicted background noise level (predict BN) is not directly predicted by the model but is deduced from the total noise and the estimated WTN, while the real background noise is the measured environmental noise.

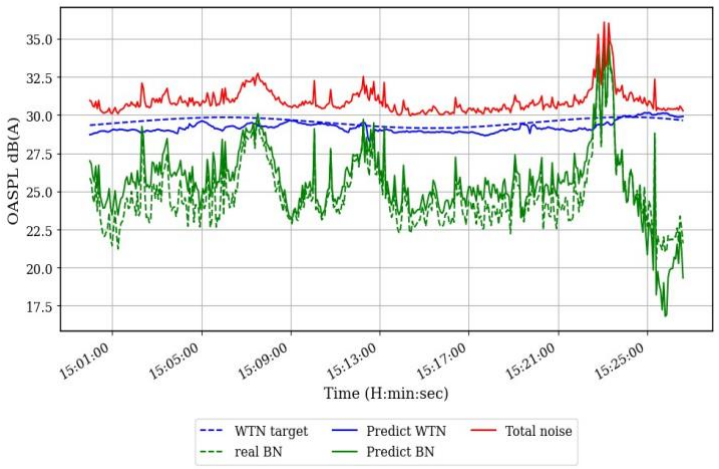


Figure 10. Example of wind turbine noise estimation from total noise on a test time series. reference signal over time, with small deviations and a slight tendency toward underestimation.

A more detailed observation shows that the model captures well the temporal dynamics of the WTN, including its variations. In particular, for time periods where the SNR is positive, this behavior is particularly relevant, as regulatory assessments of wind turbine noise focus on conditions where the turbine noise is impactful.

4.4 Comparison with baseline method

To assess the effectiveness of the proposed CRNN model, a comparison is conducted with non-negative matrix factorization (NMF) approach, previously introduced for wind turbine noise estimation [3]. The NMF model used in this study was trained on a previously published dataset, following the methodology described in [3]. It is then evaluated on the dataset constructed in the present work, which combines measured background noise with simulated wind turbine noise. This setup allows evaluating the robustness and generalization capability of both CRNN and NMF

Figure 11 presents the MAE of both models as a function of the SNR bins on a test set. The results show a significant performance gap between the two approaches across all SNR conditions. The NMF method exhibits substantially higher estimation errors, particularly in low SNR regimes, where the wind turbine noise is strongly masked by background noise. This behavior reflects the limitations of the NMF in capturing WTN from different scenarios and complex acoustic patterns.

CRNN model achieves consistently lower MAE values across all SNR bins. The improvement is especially pronounced in moderate and high SNR conditions, where the model is able to accurately track the wind turbine noise contribution. This performance gain highlights the ability of the CRNN architecture to learn discriminative time–frequency representations and exploit temporal dependencies in the data. Overall, this comparison demonstrates the clear advantage of the proposed data-driven approach over traditional signal decomposition methods for wind turbine noise estimation.

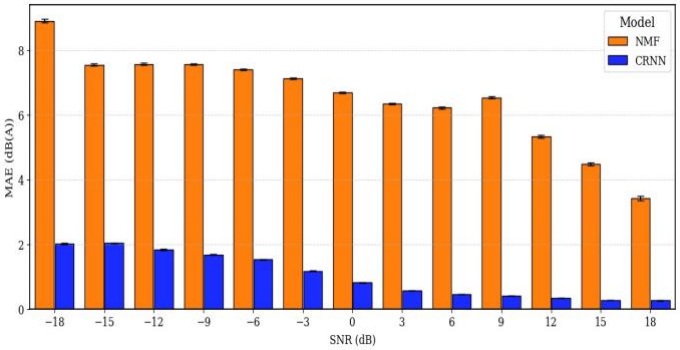


Figure 11. Comparison of MAE between NMF and CRNN models as a function of SNR on the test dataset.

5 Conclusion

This work presented a data-driven approach for estimating the overall sound pressure level of wind turbines from third-octave band time–frequency sequences. Given the absence of experimental methods capable of isolating wind turbine noise from total noise spectrograms, a supervised learning framework was adopted. To enable training, hybrid acoustic scenes were generated by combining measured background noise with simulated wind turbine signals, providing reliable ground truth labels for each generated total noise sample. It should be noted that scintillation effects and wind fluctuations were not considered; therefore, the simulations capture only a subset of real acoustic phenomena and remain limited in their representativeness. Nevertheless, they provide a dataset to assess the neural network's ability to identify and learn relevant acoustic characteristics.

The proposed CRNN architecture, integrating 2D convolutional layers for joint time–frequency feature extraction with stacked LSTM layers for temporal modeling, demonstrated its effectiveness in the test set. Stable and efficient training was achieved through the use of the Huber loss function, AdamW optimization with systematic learning-rate tuning, and gradient norm clipping.

Test experiments conducted on the dataset, covering SNR range from approximately −18 dB to +18 dB, demonstrate that the proposed model can estimate turbine noise levels under diverse acoustic conditions. The results reaching the best performance when the turbine noise becomes dominant (SNR > 0 dB), with errors typically below ∼0.5 dB(A).

The bias analysis further indicates a moderate underestimation of the turbine noise in strongly masked conditions (negative SNRs, down to approximately −0.7 dB(A)), while a underestimation (about −0.4 dB(A)) is observed at high SNR values. The small confidence intervals obtained for each SNR bin confirm the stability of the model predictions on the test.

As a perspective, future work will focus on evaluating the proposed CRNN on real-world data and compare its performance against the conventional on/off measurement method. This reference approach estimates wind turbine noise level by measuring background noise during wind turbine shutdown periods and total noise during operation, with the WTN level derived from the difference between these two phases assuming temporally stable background noise during this on/off phases. Future developments will also focus on incorporating more realistic environmental fluctuations and acoustic scintillation into the WTN simulations.

Author Contributions

A.R. generated the WTN signals from meteorological parameters, developed the deep learning methodology, performed the model training and testing, and prepared the manuscript.
J. Maillard generated specific condtions of the wind turbine noise data used in the hybrid dataset construction. J.-R. Gloaguen contributed to the non-negative matrix factorization (NMF) evaluation. A. Finez and G. Vasile supervised the work and contributed to the review and revision of the manuscript. All authors read and approved the final version of the paper.

Data Availability Statement

The data used in this study are not publicly available due to privacy constraints related to background noise measurements in residential environments. Processed data may be made available from the corresponding author upon reasonable request.

Acknowledgements

The authors would like to thank the partners involved in the collection of background noise measurements and the development of the WTN simulation framework used in this work.

References

1. P.A. Owusu, S. Asumadu-Sarkodie, A review of renewable energy sources, sustainability issues and climate change mitigation, Cogent Engineering 3, 1167990 (2016). 10.1080/23311916.2016.1167990
2. A. Freiberg, C. Schefter, M. Girbig, V.C. Murta, A. Seidler, Health effects of wind turbines on humans in residential settings: Results of a scoping review, Environmental Research 169, 446 (2019). 10.1016/j.envres.2018.11.032
3. J.R. Gloaguen, D. Ecotière, B. Gauvreau, A. Finez, A. Petit, C. Bourdat, Automatic estimation of the sound emergence of wind turbine noise with nonnegative matrix factorization, The Journal of the Acoustical Society of America 150, 3127 (2021). 10.1121/10.0006782
4. P.S. Huang, M. Kim, M. Hasegawa-Johnson, P. Smaragdis, Deep learning for monaural speech separation, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2014), pp. 1562–1566
5. C. Zhang, H. Zhan, Z. Hao, X. Gao, Classification of Complicated Urban Forest Acoustic Scenes with Deep Learning Models, Forests 14, 206 (2023). 10.3390/f14020206
6. A. Gorin, N. Makhazhanov, N. Shmyrev, DCASE 2016 Sound Event Detection System Based on Convolutional Neural Network (2016)
7. O. Anicic, D. Petkovic, S. Cvetkovi´ c, Evaluation of wind turbine noise by soft comput- ´ ing methodologies: A comparative study, Renewable and Sustainable Energy Reviews 56, 1122 (2016). 10.1016/j.rser.2015.12.024
8. A. Rkhiss, A. Finez, J.R. Gloaguen, G. Vasiele, J. Maillard, Impact of environmental data on wind turbine noise level estimation, Journal of Environmental Science and Agricultural Research 3, 1 (2025). 10.61440/JESAR.2025.v3.65
9. S. Mun, S. Shon, W. KIM, D. Han, H. Ko, A Novel Discriminative Feature Extraction for Acoustic Scene Classification Using RNN Based Source Separation, IEICE

Transactions on Information and Systems E100.D, 3041 (2017). 10.1587/transinf.2017EDL8132

10. A. Pérotin, R. Serizel, E. Vincent, A. Guérin, CRNN-Based Multiple DoA Estimation Using Acoustic Intensity Features for Ambisonics Recordings, in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (2019)
11. A.P.C. Bresciani, J. Maillard, A. Finez, Wind farm noise prediction and auralization, Acta Acustica 8, 15 (2024). 10.1051/aacus/2024007
12. J. Defrance, E. Salomons, I. Noordhoek, D. Heimann, B. Plovsing, G. Watts, B. Rasmussen, D. van Maercke, Outdoor sound propagation reference model developed in the european harmonoise project, Acta Acustica united with Acustica 93, 213 (2007). 10.3813/AAA.918086
13. E. Cakir, G. Parascandolo, T. Heittola, H. Huttunen, T. Virtanen, Convolutional recurrent neural networks for polyphonic sound event detection, IEEE/ACM Transactions on Audio, Speech, and Language Processing 25, 1291 (2017). 10.1109/TASLP.2017.2670575
14. A.S. Monin, A.M. Yaglom, Statistical Fluid Mechanics: Mechanics of Turbulence (MIT Press, 1975)
15. S. Hochreiter, J. Schmidhuber, Long Short-term Memory, Neural computation 9, 1735 (1997). 10.1162/neco.1997.9.8.1735
16. D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, International Conference on Learning Representations (ICLR) (2015). 10.48550/arXiv.1412.6980
17. I. Loshchilov, F. Hutter, Decoupled weight decay regularization, International Conference on Learning Representations (ICLR) (2019). 10.48550/arXiv.1711.05101