Estimation of the sound emergence of wind turbines by semi-supervised learning technique

Jean-Rémy Gloaguen, David Ecotiere, Benoit Gauvreau, Arthur Finez, Arthur Petit, Colin Lebourdat

INCEEUROPE

9th International Conference
on
Wind Turbine Noise
Remote from Europe – 18th to 21st May 2021

Estimation of the sound emergence of wind turbines by semi-supervised learning technique

Jean-Rémy Gloaguen, UMRAE, Cerema, Univ. Gustave Eiffel, Ifsttar, Strasbourg, 67035, France: jean-remy.gloaguen@cerema.fr
David Ecotière, UMRAE, Cerema, Univ. Gustave Eiffel, Ifsttar, Strasbourg, 67035, France
Benoit Gauvreau, UMRAE, Univ. Gustave Eiffel, Ifsttar, Cerema, Bouguenais, 44344, France
Arthur Finez, ENGIE Green, Nantes, 44000 France
Arthur Petit, ENGIE Green, Nantes, 44000 France
Colin Lebourdat, ENGIE Green, Nantes, 44000 France.

Summary

The sound emergence is the main regulatory estimator for wind turbine noise in France. This criterion aims to limit their noise impact on local residents and is highly dependent on the variation of residual noise over time. Therefore, initially defined curtailment plans can sometimes become inadequate, in which case they cannot easily be updated without leading to significant production losses. Machine learning techniques allow today to consider the continuous estimation by measurements of the sound contribution of wind turbine noise in the ambient noise and thus its noise emergence, without needing to stop the wind farm. This operation makes it possible not only to regularly adapt these reduction plans, thus optimizing electricity production, but also limiting the possible noise annoyance for local residents.
For this purpose, semi-supervised Non-negative Matrix Factorization method is considered, enhanced by a temporal regularity constraint. This approach combines a wind turbine dictionary designed on a learning basis and a free dictionary that allows the adaptation of the method to the variability of residual noise. Tests conducted on simulated measurements reveal satisfactory performances with mean estimation errors lower than 2 dBA for wind noise emergences lower than 5 dBA. Finally, the presence of these two types of dictionaries makes it possible to estimate the wind noise emergence according to one or the other depending on the predominance of the estimated wind turbine noise.

1. Introduction

In some European countries, the sound impact of wind farms is assessed by their sound emergences. This indicator is estimated by the difference between the A-statistical sound

pressure level (SPL) of the ambient noise at the receiving point $L_{A,50,amb.}$ (i.e. when the wind farm is operating) and the residual noise $L_{A,50,res.}$ (i.e. when the wind turbine is shutdown):

$$E = L_{A,50,amb.} - L_{A,50,res.} \quad (1)$$

The statistical SPL presents the advantage to be more suited for the wind turbine (WT) noise, which remains quite constant during time, while equivalent SPL is more sensitive to extreme sound pressure levels that might result from emergences of residual noises.
These sound emergences are then limited to regulation thresholds to not exceed. To respect these criteria, the most common solution is the application of a curtailment plan on WTs to limit their operation and thus their noise emission (Rogers, 2020). However, this plan affects strongly the production capacity of the machines.
Furthermore, this solution has the main drawback to depend on the variability of the residual noise, which is a time varying component that changes over the meteorological conditions, the ground characteristics and the surrounding infrastructures. Consequently, this may result in two situations where the curtailment plan might be too strong (i.e. the electrical production is unnecessarily limited) or too low (i.e. the sound emergences exceed the regulation thresholds).
Currently, to update this plan, *on/off* cycle measurements are made. These measurements consist in assessing alternatively the SPLs of the ambient noise and the residual noise by stopping the wind farm periodically. If needed, a new curtailment plan can then be formulated.

However, these measurements require the shutdown of the machines, which also significantly affects the electrical production. Furthermore, they are only carried out during a couple of weeks and are not sufficiently representative over a long period. Consequently, this updated plan does not guarantee to be still valid few months later. To repeat this process regularly will then increase again the loss of electrical production and is not a suitable solution.
Consequently, there is an interest to be able to estimate the sound emergence of wind farms continuously without stopping them. Thus, it would be possible to control regularly the noise annoyance for neighboring inhabitants, while optimizing the electrical production.
Instead of seeking to improve the sound source emission of WT and the models for outdoor sound propagation (Cotté, 2019), this emergence could be estimated by measurements made *in situ*. Recently, (Gloaguen et al., 2020) propose a first tool to estimate the sound emergence of WTs. This tool is based on the Non-negative Matrix Factorization (NMF) used as a source separation method to first extract the WT component from simulated ambient scenes and then to estimate the sound emergences. This preliminary study has been extended in (Gloaguen et al., 2021) on a larger number of situations where the WT component has been propagated at different distances in different propagation conditions. Supervised learning associated with a temporal constraint reached satisfactory results with a mean error inferior to 2 dBA on many cases. In this study, only WT noise was considered in NMF. It might be then helpful to better consider the residual noise in NMF to bring more flexibility but also robustness to this method.
Thus, this paper extends these works and considers a semi-supervised learning, which allows NMF to consider labelled data (i.e. WT noise) and unknown data that can be adapted to the diversity of the residual noise. Section 2 introduces Semi-supervised NMF, Section 3 presents the different corpora used in this experiment and Section 4 summarizes the main results and details the behavior of the proposed method.

2. Semi-supervised Non-negative Matrix Factorization

In audio field, Non-negative Matrix Factorization (NMF) is usually introduced as a linear approximation of an audio spectrogram $V_{F\times N}$ by the product of two matrices, W and H, such as $V \approx \tilde{V} = WH$. $W_{F\times K}$, called *dictionary*, includes audio spectra and $H_{K\times N}$, called the *activation matrix*, corresponds to the temporal evolution of each of the spectra such as $\hat{V} = WH$ (Dikmen & Mesaros, 2013; Lee & Seung, 1999). As only amplitude (or power) spectra are considered in V and W, only additive combinations are possible. The dimensions of the matrices are chosen

such as $K < \min(F, N)$ to avoid overcomplete representation. In supervised learning, W is learned on labelled samples and H is the unknown to estimate. In the present case, W includes WT noise spectra.
For semi-supervised learning (Kitamura et al., 2014), a second dictionary and activation matrix are considered, respectively $Y_{F\times J}$ and $Z_{J\times N}$. The dimension J corresponds to the number of considered spectra, which is often set such as $J < K$ in order to allow NMF to stay focused on the data learned in W.
These two matrices are learned on each spectrogram to represent, the best as possible, the residual noise. This second part makes it possible to increase the degree of freedom of NMF and to include residual noise directly learn on the spectrogram $\tilde{V}$. In supervised learning, it would have been necessary to learn a residual dictionary on labelled data, which is more complex to do, as it is a noise composed of multiple different kind of sound sources not easy to define precisely.
Furthermore, instead of a linear sum between $[WH]$ and $[YZ]$, the energetic sum between the two components is considered here:

$$V \approx \tilde{V} = \sqrt{[WH]^2 + [YZ]^2}. \tag{2}$$

This modification of the NMF problem is proposed because the available data are expressed in Pa, thus the energetic sum is more appropriate (see Sec. 3). The approximation between V and $\tilde{V}$ is performed by minimizing the cost function $D_\beta(V||\tilde{V})$:

$$\min D_\beta(V||\tilde{V}) = \min\left(\sum_{f=1,n=1}^{F,N} d_\beta(V_{fn}||\tilde{V}_{fn}) + \alpha_{sm}\sum_{n=1}^{N}\left(h_{k,n} - h_{k,n-1}\right)^2\right) \tag{3}$$

$D_\beta(V||\tilde{V})$ belongs to the β-divergence family where only the Euclidean distance is considered such as $d_2(x|y) = \frac{1}{2}(x-y)^2$. The added regularization term, weighted by the coefficient α_{sm}, corresponds to a temporal regularity constraint applied only on H (Févotte et al., 2018). This constraint forces the shape of the activations in H to adapt a smoother behaviour and thus to better corresponds to the time signature of WT noise, which is quite constant in practice. As the NMF formulation has been modified on Eq. 2, the update algorithms of H, Y and Z are changed. The new update algorithms are estimated as in (Lee & Seung, 2000) with a descent gradient. Equation 4 presents the update algorithms for the Euclidean distance for this new formulation:

$$Y \to Y.\frac{\left[[YZ].V\right]Z^T}{\left[[YZ].\tilde{V}\right]Z^T} \tag{4.1}$$

$$Z \to Z.\frac{Y^T\left[[YZ].V\right]}{Y^T\left[[YZ].\tilde{V}\right]} \tag{4.2}$$

$$H \to H.\frac{W^T\left[[WH].V\right] + 2\alpha[H_{+1} + H_{-1}]}{W^T\left[[WH].\tilde{V}\right] + 2\alpha\,[H + H_{12}]} \tag{4.3}$$

with $H_{-1} = [\mathbf{0}\ H_{1:N-1}]$, $H_{+1} = [H_{2:N}\ \mathbf{0}]$ and $H_{12} = [\mathbf{0}\ H_{2:N-1}\ \mathbf{0}]$ where $H_{n:m}$ refers to the selection of columns n to m of H and $\mathbf{0}$ a column of K zeros. More details on NMF and some applications can be found in (Févotte & Idier, 2011) , (Heittola et al., 2011) and (Ludeña-Choez et al., 2017).

In summary, among different parameters, each taking several possible values, one tries to find an optimal NMF formulation: the number of basis in W ($K \in \{5, 10, 20\}$), the number of basis in Y

($J \in \{2, 5, 10\}$) and the weight of the temporal regulation ($\alpha_{sm} \in \{0, 0.5, 1, 2\}$). Finally, NMF is performed for 1000 iterations where H, Y and Z are initialized randomly.
Sound emergences can be estimated in two different ways with semi-supervised NMF (SS NMF). First, one estimates from $[WH]$ the WT noise component in V, deduces the residual component, its estimated statistical SPL $\tilde{L}_{A,50,res.}$ and then the estimated emergence $\tilde{E}_{WH}$,

$$\tilde{E}_{WH} = L_{A,50,amb.} - \tilde{L}_{A,50,res,WH}. \quad (5)$$

In a second way, one considers $[YZ]$ as the exclusively residual noise and directly use it to estimate the sound emergence, called $\tilde{E}_{YZ}$. Figure 1 summarizes this process. In Section 4.1.1, the influence of this choice on the setting on the estimation of the sound emergence is exposed.

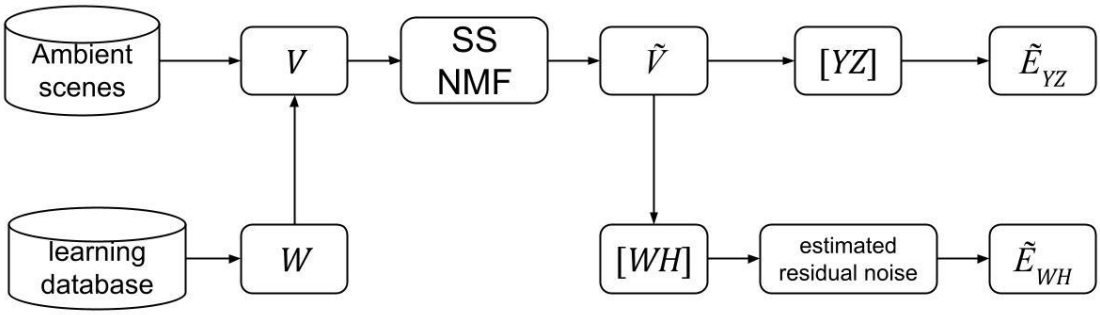


Figure 1: block diagram of the estimation of the sound emergence by NMF.

3. Corpora of environmental sounds

To perform the calculation, two different corpora are used. The first one is a corpus of ambient scenes, built in order to simulate the method based on *on/off* cycle measurements (see Section 3.1). However, instead of considering long periods, one considers short periods of 10 minutes of ambient noise (i.e. *on* cycles when the WTs operate) and only residual noise (i.e. *off* cycles when the WTs are shutdown). This second period simulates the current process used *in situ* that estimates the emergence $E_{on/off}$. This indicator is a baseline method, which will help to compare the performances of NMF. A second corpus composed of exclusively of WT noise spectra is dedicated to the learning of the dictionary W and is presented in section 3.2.
All the following data are expressed in Pa for each third octave band between 20 Hz and 10 kHz.

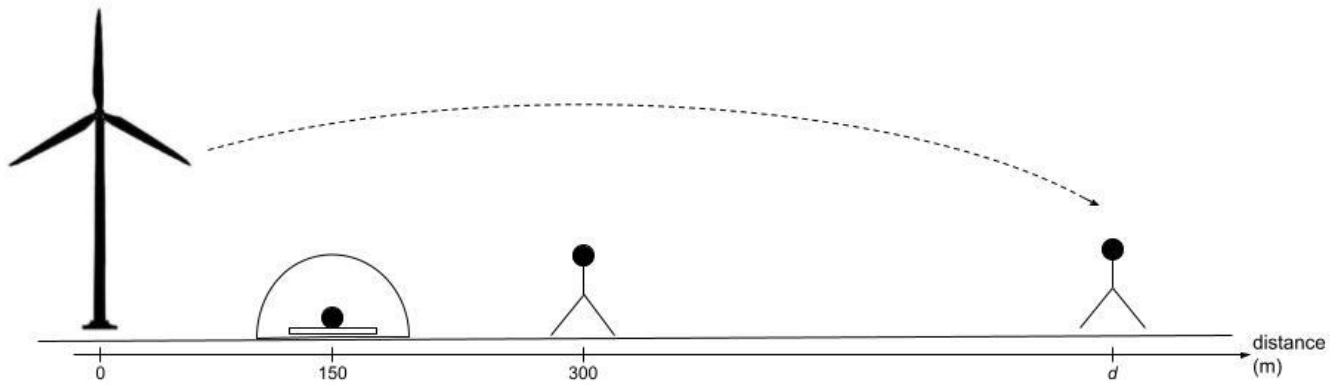


Figure 2: schema of the different measurement points used for the dictionary corpus (150 m) and the corpus of simulated ambient scenes (300 m). The distance d corresponds to the distance where the ambient scene are located (500 m, 1000 m, 1500 m).

3.1 Corpus of simulated ambient scenes

A corpus of 30 simulated ambient scenes is built for this experiment. This simulation process consists in the sum of a WT noise component with a residual noise component. This process makes it possible to know the 'exact' sound emergence $E_{ex.}$(see Eq. 1) and can then be compared to the one estimated by NMF. By using *in situ* measurements, this 'exact' value would have been unknown.

3.1.1. Wind turbine noise samples

The WT noise components are collected during *on/off* cycle measurements at 300 m from the front line of a wind farm at 1.5 m height (Kayser et al., 2018) (see Figure 2). To consider the most accurate samples, the residual noise spectra, estimated with the 30 minutes preceding and following the *off* measurements, filter the *on* measurements. From these cleaned samples, 30 samples of 10 minutes are extracted (see Figure 3). To generate ambient scenes accurately, an attenuation filter must filter the WT noise in order to simulate its propagation at a certain distance through an inhomogeneous media.

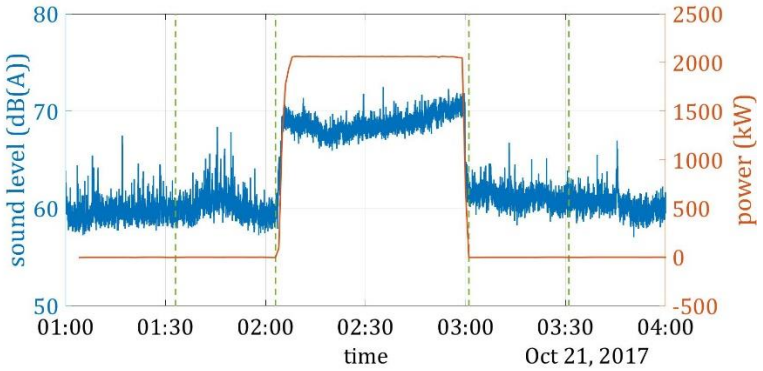


Figure 3: example of the extraction of wind turbine spectra. The emergence on *cycle is considered and filtered by the mean residual spectra estimated from the measurements carried out between the dashed lines.*

3 distances between the WTs and a receiving point are considered ($d \in \{500, 1000, 1500\}$ m) through one inhomogeneous media, which is considered here as a downward propagation condition. Propagation conditions are characterized although the vertical profile of the effective sound celerity, defined as:

$$c_{eff}(z) = \sqrt{\gamma R T(z)} + U(z)\cos(\theta) \tag{6}$$

with γ the heat capacity ratio of dry air at constant pressure and volume (γ = 1.4), R, the specific gas constant for dry air (R = 287 J.kg^{-1}.K^{-1}), θ the angle between the wind direction and the propagation direction and $T(z)$ and $U(z)$ the mean vertical profiles of air and temperature and wind speed respectively. As turbulence is not considered, these profiles are only expressed according to their mean part:

$$T(z) = T_0 + a_t \log\left(\frac{z-d}{z_0}\right), \tag{7.1}$$

$$U(z) = a_u \log\left(\frac{z-d}{z_0}\right) \tag{7.2}$$

where T_0 (K) is the ground surface temperature, d = 0.66h_v is the displacement height accounting for the influence of the vegetation height h_v (m), z_0 = 0.13h_v is the roughness height of flux profiles and a_t and a_u are the coefficients that determine the shape of the temperature and wind vertical profiles, respectively. The relative humidity is set to 70 %. The settings of the downward refraction condition used in the study are T_0 = 15°C, a_u = 100 m.s^{-1}, a_t = 0.15 K.m^{-1}, θ = 45.
Sound propagation is modelled through the use of Parabolic Equation (PE) (Kayser et al., 2020) and resolved by the split-step Padé approach with an average ground floor impedance (airflow

resistivity σ = 500 kN.s.m^{-4}). To improve the estimation of this filter, the sound emission of WT noise is defined by the extended source model proposed by Cotté (Cotté, 2019).
PE is performed for all the third octave bands from 20 Hz to 3150 Hz. This process makes it possible to obtain an accurate estimation of the SPL of WT to ensure the validity of the results of NMF.
From this resolution, the attenuation filters at the distances d are estimated with a reference microphone at 300 m (corresponding to the position where the WT samples in the ambient scenes have been collected) (see Figure 4). Due to computation limit, the attenuations for the third octave bands from 4 kHz to 10 kHz are modelled only with the geometrical divergence and the atmospheric absorption. This limitation is not prohibitive as the WT noise contribution in these bands are very low.

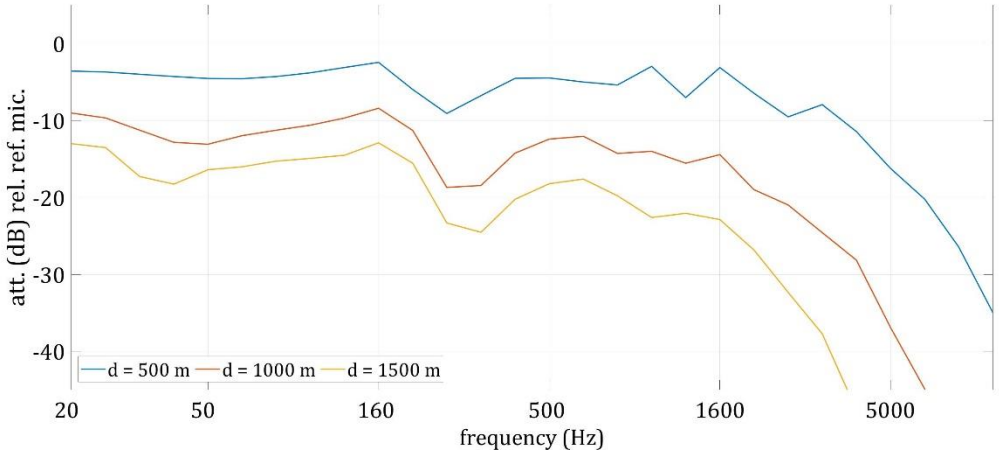


Figure 4: attenuation filters used to propagate the wind turbine noise at distance d *with a reference receiver located at 300 m from the front line of the wind farm and at 1.5 m height above the ground.*

3.1.2. Residual noise samples

The residual noise samples come from two different locations (called WF1 and WF2) and have been collected during *on/off* cycle measurements, but this time during the *off* periods. For each location, 30 samples are collected. The residual noises from WF1 account for weak residual SPL as the wind farm is located far from any anthropogenic noise sources. In the opposite, WF2 is located close to a highly frequently road. The residual noise samples related to this place are then much more dynamic.
Each sample lasts 25 minutes. The first 10 minutes are used to build the ambient noise sample with the WT noise sample. The last 10 minutes are used as only residual noise to simulate the shutdown of the wind farm and to estimate the baseline emergence $E_{on/off}$,

$$E_{on/off} = L_{A,50,amb.} - L_{A,50,res.,off} \quad (8)$$

The last 5 minutes (between the 10th minutes and the 15th minutes) can be seen as the transition time before the WT are shut down and are then discarded.

3.1.3. Building of the ambient scenes

From the WT and residual noise samples, the 30 simulated ambient scenes are built. These scenes length 10 minutes with a 1 second step. The WT noise is first propagated to 3 different distances (d $\in \{500, 1000, 1500\}$ m) for one propagation condition (see Section 3.1.1). The WT and the residual noise samples are then summed following the Signal-to-Residual-Ratio (SRR) defined as:

$$SRR = L_{A50,WT} - L_{A50,res.,on}, \quad (9)$$

where $L_{A,50,WT}$ is the A-statistical SPL of the WT component and $L_{A,50,res.,on}$ the A-statistical SPL of the residual component used for the ambient scene. The SRR is defined for {-9; -6; -3; 0; 3; 6; 9} dBA. When $SRR < 0$ dBA, the residual noise is predominant over the WT noise and inversely when $SRR > 0$ dBA. These SRR values correspond to 'exact' sound emergences $E_{ex.}$ of {0.5; 1.0; 1.8; 3; 4.8; 7.0; 9.5} dBA respectively. Finally, the simulated ambient scenes are the spectrograms V of NMF with dimensions F = 28 (i.e. number of third octave bands between 20 Hz and 10 kHz) and N = 600. The adjustments made to set the residual noise at a SPL corresponding to the SRR are also made on the last 10 minutes of the residual noise to assure its continuity. The Figure 5 displays for 2 SRR values the different parts of an ambient scene and Figure 7 summarizes the steps to build the ambient scenes.

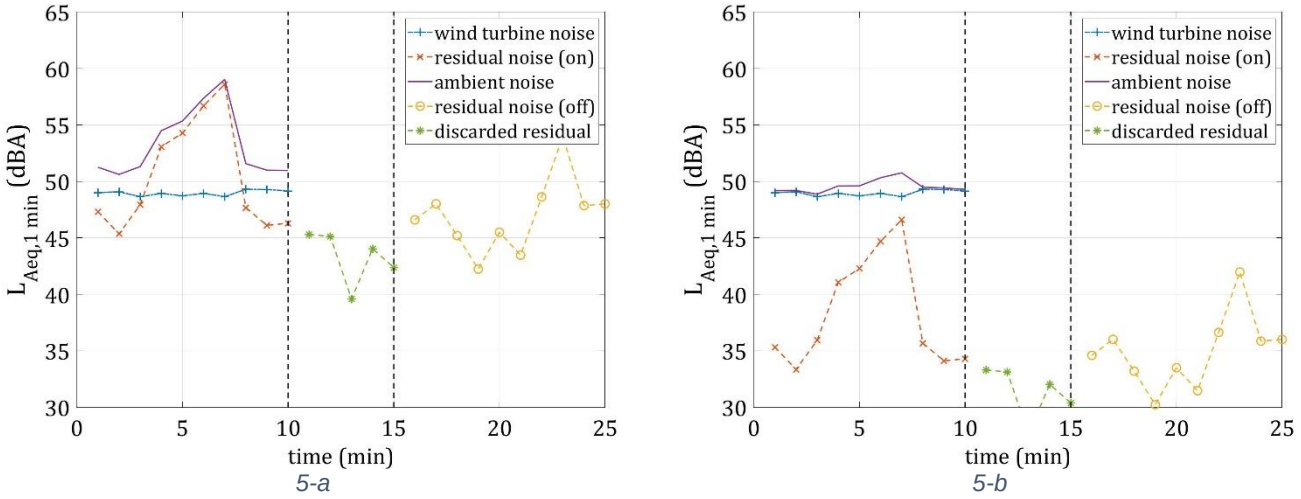


Figure 5: details on the building of the simulated ambient scene for SRR *= -6 dBA (5-a) and* SRR *= 6 dBA (5-b) (scene 6, WF1, d = 500 m).*

Finally, a pre-processing step is performed on the ambient scenes before considering it in NMF. This pre-process consists in the attenuation of the most energetic frames. One considers here that these frames are due to high emergences of residual noises. Thus, to help the approximation made by NMF and limit the influence of the residual noise, it might be beneficial to remove them. To do so, the temporal frames where the equivalent SPLs are superior to the $L_{A,10}$ are discarded and replaced by interpolated values to ensure coherent continuity in the spectrogram. Figure 6 displays an example. To know its impact, NMF is also performed without this operation.

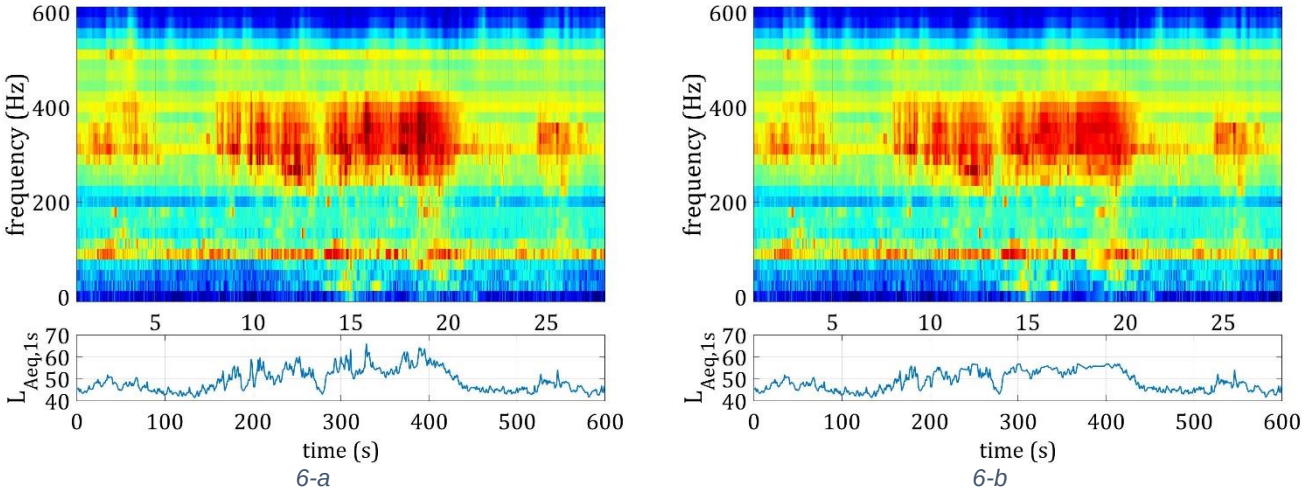


Figure 6: influence of the pre-processing step on the spectrogram and on the A-SPL. On the initial spectrogram (6-a), the residual noise provokes high emergences. On the pre-processed spectrogram (6-b), these emergences are reduced, for instance in the intervals {327; 334} s and {386; 396} s

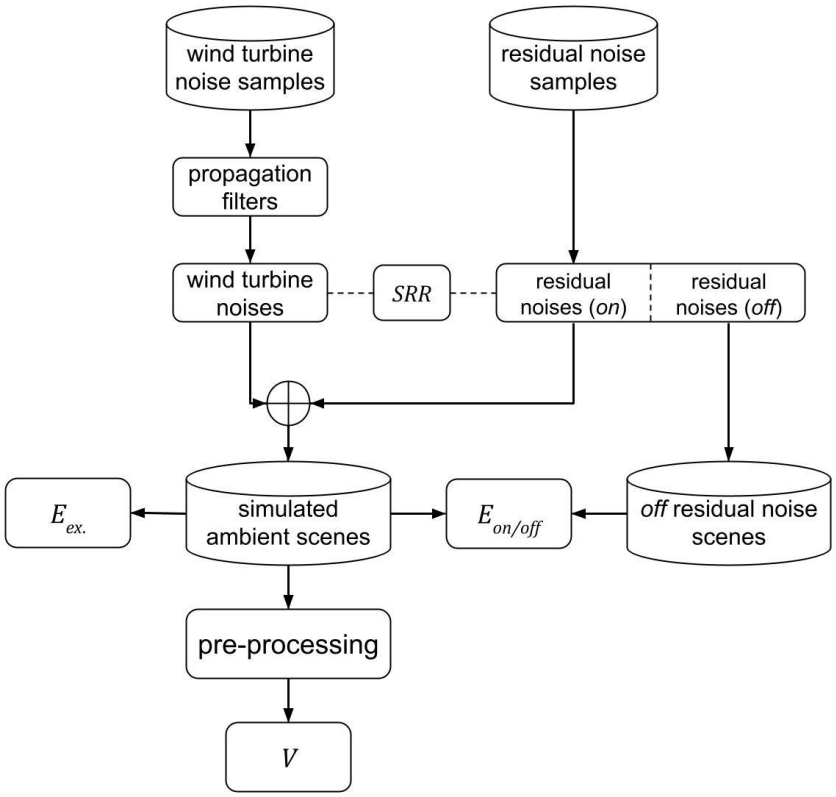


Figure 7: block diagram of the ambient scene corpus building.

3.2 Corpus for the dictionary W

To build a dictionary of WT spectra, measurements made during *on/off* cycle measurements on the same wind farm than in Section 3.1 are collected. To ensure to have spectra the less polluted by surrounding residual noise samples, only measurements made following the IEC standard protocol are considered as they are closed to WTs (see Figure 2). In addition, only the most emerging measurements are considered. They are then filtered by the estimated residual noise spectra measured during the *off* cycle as in Section 3.1.1, to ensure the most accurate WT spectra.

In all, 14200 third octave band spectra (ranking from 20 to 10 kHz) are extracted. This large number of spectra cannot be directly in the multiplicative update algorithms (Eqs. 4) as it will highly increase the computation cost. Furthermore, these spectra present redundant information. Consequently, a K-mean algorithm reduces these 14200 spectra to $K \in \{5, 10, 20\}$.

As the ambient scenes simulate the WT noise contribution at 3 distances from the front line of a wind farm, an attenuation 'transfer function' filters is also applied on the initial dictionary at the considered distances. Instead of using PE, one chooses a simpler approach to facilitate the implementation of the proposed method in an industrial process. The Weyl's Van der Pool equation is then considered and solved by the sound rays model (Salomons, 2012). The sound emission models the WT noise by 3 moving monopoles (hub height = 80 m, speed rotation = 20 rpm). A downward meteorological condition is chosen with a linear vertical profile of the effective celerity: $c_{eff}(z) = c_0 + b_{lin}z$ with c_0 the sound celerity (m/s) with an air temperature of 10°C, b_{lin} = 0.1 m^{-1} and z the height (m). The ground impedance is defined by the Miki's model (Miki, 1990) with an air flow resistivity representative of countryside (σ = 500 kN.s.m^{-4}). The corresponding attenuation filters are displayed on Figure 8. They are generated for the 3 distances with a reference microphone corresponding to the one on the ground located at 150 m from the front line of the wind farm (see Figure 2). Finally, each filtered spectra of the dictionary is normalized such as $\|W_k\| = 1$.

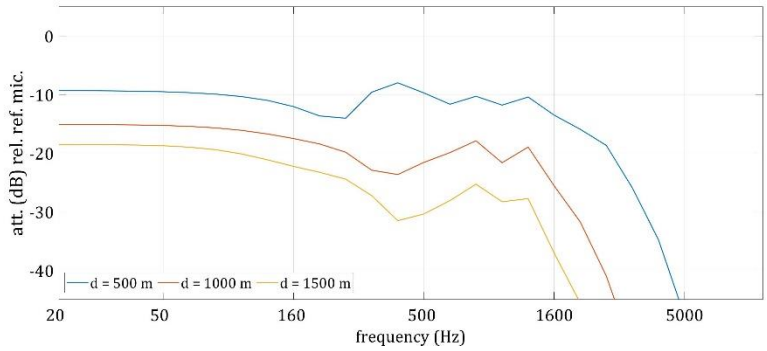


Figure 8: attenuation filters for the dictionary with a reference receiver located at 150 m from the front line of the wind farm and on the floor.

4. Computation and discussions

The objective of the experiment is to find an optimal setting of NMF (β, K, J, α_{sm}, number of iteration, estimation by $[WH]$ or $[YZ]$, pre-processing step) on 30 simulated ambient scenes with the WT noise propagated at different distances ($d \in \{500, 1000, 1500\}$ m), for one downward propagation condition and for two locations of residual noises (WF1 and WF2). For each ambient scene, one gets the 'exact' sound emergence $E_{ex.}$, thanks to the simulation process, the estimated emergence by the simulated *on*/*off* cycle scenes $E_{on/off}$ and the estimated one by NMF $\tilde{E}_{WH/YZ.}$.

To compare these indicators, the Mean Absolute Error MAE is considered. It is expressed such as:

$$MAE_{WH/YZ} = \frac{\sum_{m=1}^{M} \left| \tilde{E}_{WH/YZ,m} - E_{ex.,m} \right|}{M} \quad (10)$$

This metric is also computed with $E_{on/off}$ instead of $\tilde{E}$. M is the number of considered scenes. It can be computed for the total number of scenes (M= 180, see Section 4.1.1) and for each distance and residual noise (see Section 4.1.2).

4.1.1. Global results on all the corpus of simulated ambient scenes

The MAE error is computed for each association of setting (384 associations in all) on all the corpus (M = 180 scenes). Instead of summarized all of these, only the ones with the lowest errors are displayed in Table 1 according to the dimension J, the use of the pre-processing step (*no* or $L_{A,10}$) *and* the choice of the component in $\tilde{V}$ used to estimate the sound emergence ($[WH]$ or $[YZ]$). Then, the association of settings reaching the lowest error is the optimal SS NMF formulation the most efficient that is considered.

Table 1: lower errors of SS NMF among all the tested association of settings according to the dimension J, the choice of the component in $\tilde{V}$ estimation and the pre-processing step. In bold letters, the best results.

Method	K	J	α_{sm}	Iteration	Pre-processing step	Component in $\tilde{V}$	MAE (dBA)
Baseline	-	-	-	-	-	-	2.7 ($\pm$ 2.8)
SS NMF	5	2	0.5	50	*no*	[WH]	2.1 ($\pm$ 1.8)
	5	2	0.5	50	$L_{A,10}$	[WH]	**1.9 ($\pm$ 1.7)**
	5	5	0.5	50	*no*	[WH]	2.1 ($\pm$ 1.8)
	5	5	0.5	50	$L_{A,10}$	[WH]	**1.9 ($\pm$ 1.7)**
	5	2	1	5	*no*	[YZ]	6.2 ($\pm$ 5.8)
	5	2	0.5	1000	$L_{A,10}$	[YZ]	3.1 ($\pm$ 2.0)

5	5	1	1000	*no*	[YZ]	4.0 (± 4.5)
5	5	0.5	1000	$L_{A,10}$	[YZ]	2.5 (± 1.8)

The lowest error is reached when emergences are estimated from $[WH]$, with the pre-processing step and for J ∈ $\{2; 5\}$. More generally, the lowest MAE errors are obtained when the emergence is estimated from $[WH]$ instead of $[YZ]$. This choice happens to be the most influential as the NMF settings are for the most part similar (i.e. the dimensions K and J, the number of iteration, the smoothness weight α_{sm} stay the same).
The addition of the pre-processing step improves the results, even if its impact is more reduced when emergences are estimated from $[WH]$. This result can be understandable as it limits on the spectrogram V the presence of the residual noise, which can naturally help the approximation of the WT component by NMF.
Furthermore, the increase of basis in Y only improves the results when $[YZ]$ is considered. The approximate WT component is then constant for J = 2 or J = 5. This behavior reveals that the estimation of the $[WH]$ component is quite robust and the increase of the dimension J does not result in more confusion. For the next, we consider J = 5 as it is also where the estimations of the sound emergence with $[YZ]$ are the best.
Finally, the proposed method shows lower errors with reduced standard deviations than the baseline method on this corpus, which reveals that the estimation by SS NMF is more accurate. The baseline method, based on *on/off* measurement process, makes the hypothesis that residual noise is constant over time, which might not be the case. This is particularly the case for the samples that come from the WF2 wind farm. Applied directly to the ambient scenes, this versatility does not affect SS NMF, in addition to not having to shut down the wind farm.

4.1.2. Behaviors of the optimal SS NMF detailed

From this main results, the behavior of the optimal SS NMF (K = 5, J = 5, α_{sm} = 0.5, 50 iterations, with the pre-processing step and with the emergence deduced from $[WH]$ component) is now according to the distance and the location of the residual noises in Figure 9.

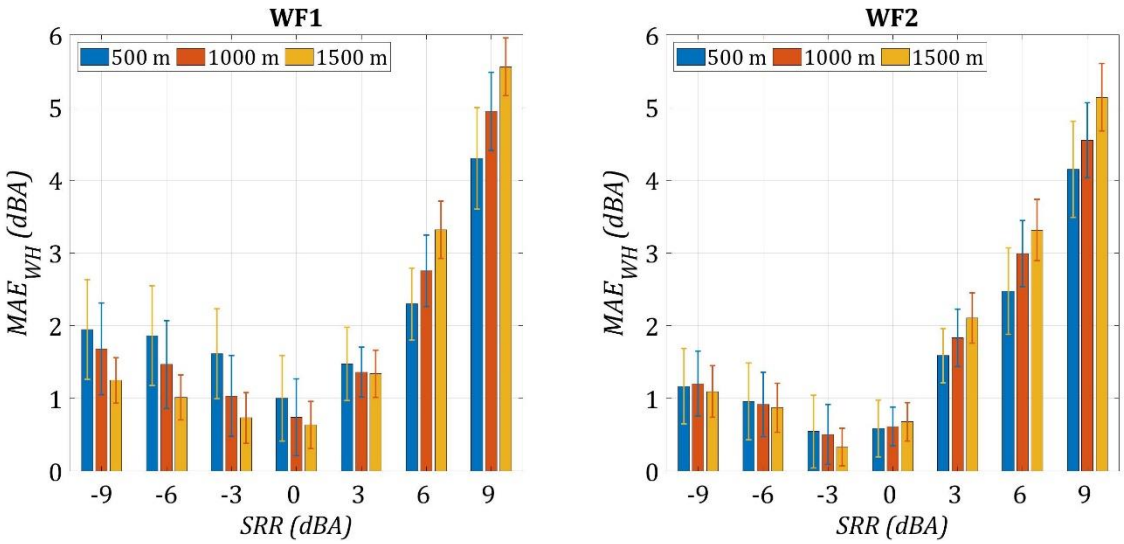
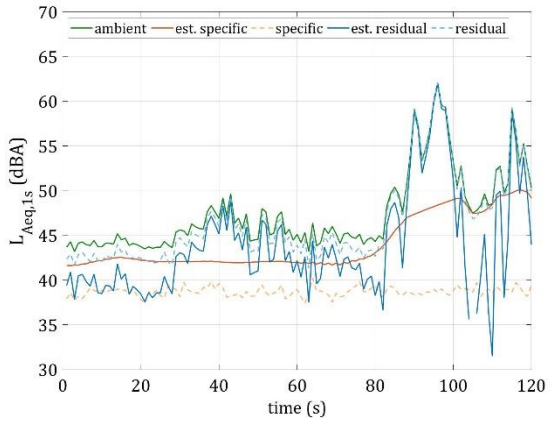


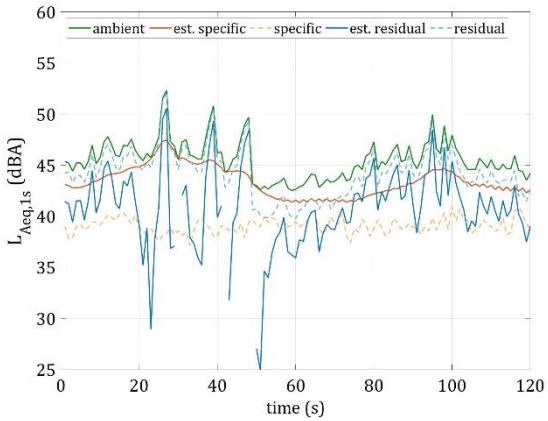
Figure 9: errors of the optimal SS NMF details according to the distance d, the location of the residual noises and the SRR.

First, one notices the strong dependence of the results to the SRR values. The evolution of the errors according to the residual noise location or the distance follows the same pattern. The MAE_{WH} tends to decrease when SRR increases and is inferior to 0 dBA. Then, it increases strongly with the SRR as the WT become more and more predominant over the residual noise.
To apprehend these results, the 1 second equivalent SPL are displayed in Figure 10 for SRR = -6 dBA and for SRR = 6 dBA in Figure 11.

In these representative examples, when SRR is negative, the estimated WT noise is overestimated, which results in the underestimation of the estimated residual noise and so the overestimation of the sound emergence.
Conversely, when SRR = 6 dBA, the estimated WT noise is more equivalent to its 'exact' value. Consequently, the strong errors observed in Figure 9 are not due to a bad estimation of the WT component, but to a higher sensitivity of the emergence indicator in these cases when SSR is negative. For instance, an underestimation estimation of the WT SPL of 1 dBA for SRR = 9 dBA results of an error on the residual noise SPL of 4 dBA. This behavior happens specifically when the sound emergence is high (above 5 dBA) and can still be detected.

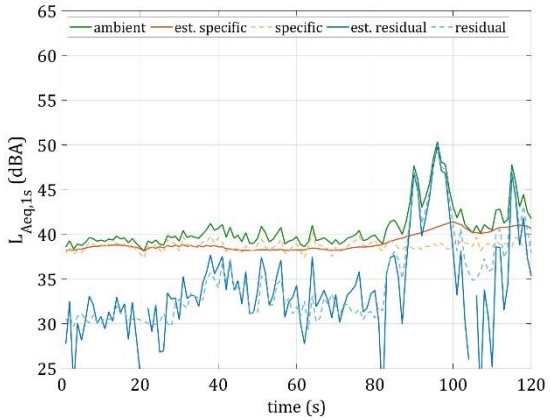


10-a

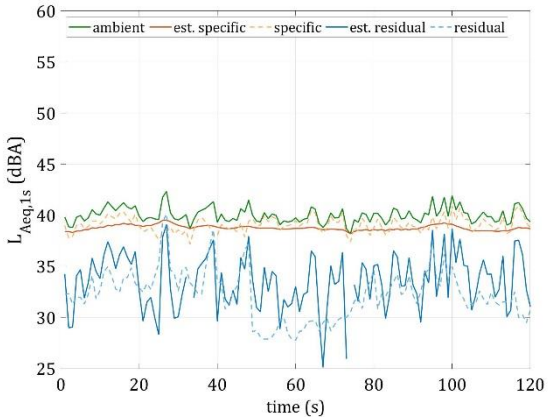


10-b

Figure 10: a 120 seconds extract of the 1 second equivalent SPL for SRR *= -6 dBA with residual noise from WF1 (scene 4, d = 500 m) (10-**Erreur ! Source du renvoi introuvable.**) and WF2 (scene 4, d = 500 m) (10-**Erreur ! Source du renvoi introuvable.**).*



11-a



11-b

Figure 11: a 120 seconds extract of the 1 second equivalent SPL for SRR = +6 dBA with residual noise from WF1 (scene 4, d = 500 m) (11-a) and WF2 (scene 2, d = 500 m) (11-b).

If the SRR is the main influence parameter, the location and the distance have also a significant impact (but reduced). According to their evolution on Figure 9, it might be some interaction phenomena between these parameters. At 500 m for SRR< 0 dBA, for WF1 residual samples, the errors are higher than at 1500 m. They decrease with WF2 residual noise samples. When, SRR becomes positive, these patterns reverse. At d = 500 m, residual samples from WF1 are less noisy, which results in more similar spectral shapes to the WT noise and then generates confusion. With the distance, the attenuation filters altered the spectra, which results in a better fit of the estimated and the 'exact' WT noise spectra. This observation can be seen in Figure 12 where the mean spectra of each component are displayed. Finally, when SRR increases, the error becomes higher with the distance. This might be due to the difference in the propagation

filter applied on the dictionary W (see Section 3.2) emphasized with the distance and the high value of the SRR where the sound emergence is more sensitive.

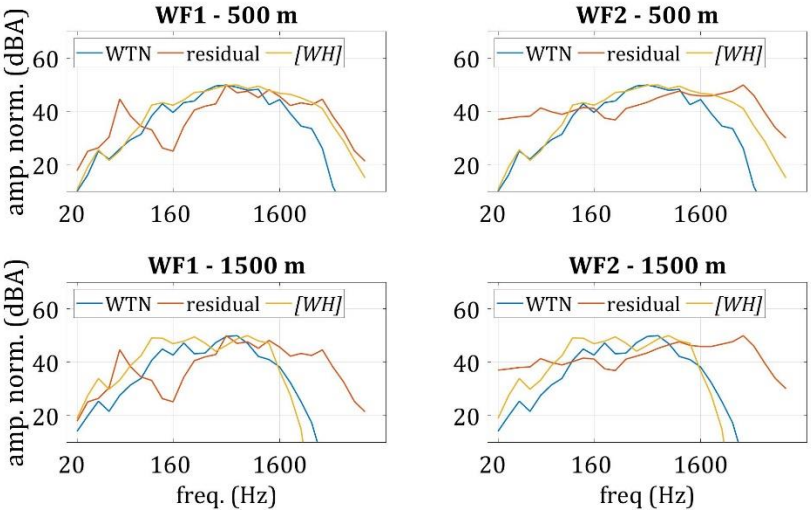


Figure 12: influence of the distance and the location of the wind farms of the residual noise samples on the mean spectra of wind turbine noise, residual noise and estimated WTN ($[WH]$) (scene 2, SRR = -6 dBA).

4.1.3. Comparison of the estimated residual noises

In this last part, the residual noise deduced from the estimated WT noise and the residual noise directly computed with $[YZ]$ are compared. To do so, the best setting with $[WH]$ and $[YZ]$ (see Table 1) are selected and their errors according to SRR are displayed in Figure 13.

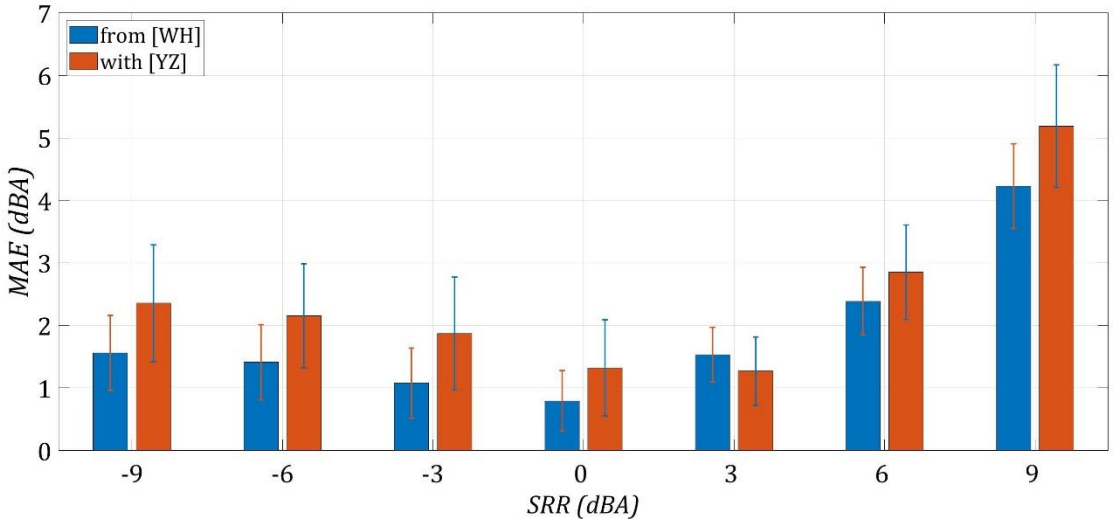


Figure 13: MAE errors according the SRR of the optimal error of SS NMF and the one that reached the lowest MAE error with the residual noise estimated with $[YZ]$ (see Table 1).

For both settings, the errors follow a similar respond according to the SRR, meaning that the estimation of the residual component is similar too. To illustrate this observation, Figure 14 summarizes 120 seconds of the residual components of one scene.

The behavior of the two residual components are similar in this example. The one deduced from $[WH]$ does not present the effects of the pre-processing step (between 46 s and 66 s, see Figure 14) as it is deduced from the ambient noise and from $[WH]$. Conversely, $[YZ]$ is learned on processed spectrograms. Finally, both approaches reveal to be very similar and to be mostly influenced by the approximation of the WT component by $[WH]$, which provokes the overestimation of the residual component (between 74 s and 83 s for instance, see Figure 14).

By using $[YZ]$, the main limit is the reduced number of spectra to approximate this sound source. The results in Table 1 reveal that the increase of the dimension J is beneficial (MAE_{YZ} = 3.1

($\pm$ 2.0) dBA with J = 2 and MAE_{YZ} = 2.5 ($\pm$ 1.8) dBA with J = 5). However, increasing this dimension also increases the risk that the cost function (Eq. 3) might be minimized mainly with this component at the expanse of $[WH]$, which is moreover subject to a temporal constraint. Consequently, the dimension J must stay low in order to not include in $[YZ]$ some WT noise component.
Furthermore, the reduced number of spectra in Y might be limited in some cases to model a component that includes many kinds of sound sources and might be insufficient to model it properly. From $[WH]$, the WT noise is more constant over time, if the dimension K is also low, it is still sufficient to approximate this component. Consequently, deduced directly from the ambient signal, the sound diversity of the residual component is better considered even if the estimated WT component presents some mistakes. This way limits the error and then improves the estimation.

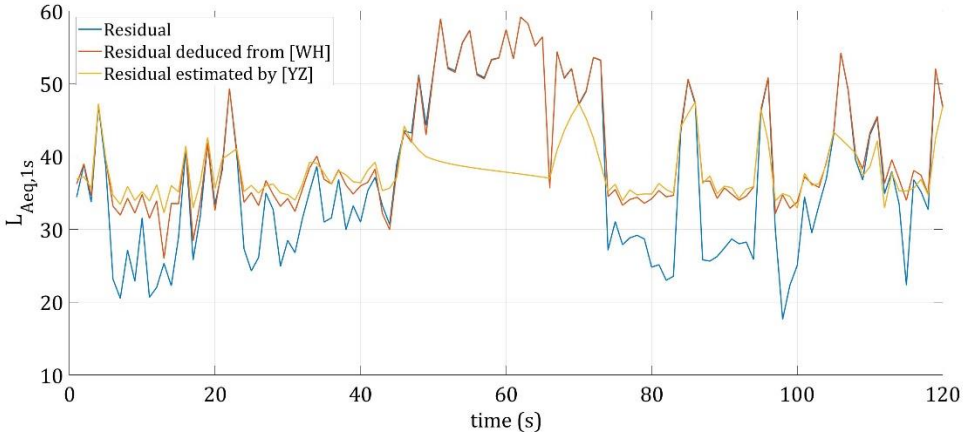


Figure 14: extract of 120 seconds of the exact residual components and obtained with SS NMF from $[WH]$ and from $[YZ]$ (scene 10, WF2, d = 500 m).

5. Conclusions

Semi-Supervised Non-negative Matrix Factorization (SS NMF) has been considered to estimate the sound emergence of wind turbine noise. To estimate the optimal settings, simulated ambient scenes have been built. This process makes it possible to compare the estimated emergences with their 'exact' values, which is not possible with *in situ* measurements without stopping the machines. The semi-supervised learning consists in adding in the dictionary a free part that can be adapted on each ambient scene. In addition, a regulation term is added through a temporal continuity constraint applied only on H (i.e. the temporal activations of the wind turbine spectra). This constraint makes it possible to have more continuous shapes of the activations and then more similar to the wind turbine noise behavior.
The common formulation of SS NMF has been adapted to correspond to the form of the input data through an energetic sum instead of a linear sum. This change required generating new update algorithms.
After computation, the settings of the optimal SS NMF have been found (K = 5, J = 5, α_{sm} = 0.5, 50 iterations, with the pre-processing step and with the emergence deduced from $[WH]$ component), which result in a global MAE error of less than 2 dBA. The simulated ambient scenes enables the study of this result according to the SRR, the distance and the location of the residual noise. The SRR reveals to be the more influential setting as it impacts directly the sound emergence. For negative SRR, SS NMF performs appropriately, but the estimation errors increase significantly as the wind turbine noise becomes more predominant. This behavior is more due to a high sensitivity of the emergence indicator than to a poor estimation of wind turbine noise. The different locations of residual noise in the scene have also an impact on the results

according to its similarity of its audio spectra with the wind turbine noise: the more similar they are, the higher the risk of confusion.
Finally, this proposed method, despite its limitations, still generates estimation errors lower than the ones based on *on/off* cycle measurements, which is a great improvement for wind farm operators. Now, these results and observations make it possible to consider this method using real *in situ* measurements and allow a better apprehension of the estimations according to the dynamic of the ambient noise or the distance between the wind farm and the measurement point.

References

Cotté, B. (2019). Extended source models for wind turbine noise propagation. *The Journal of the Acoustical Society of America*, *145*(3), 1363-1371. https://doi.org/10.1121/1.5093307

Dikmen, O., & Mesaros, A. (2013). Sound event detection using non-negative dictionaries learned from annotated overlapping events. *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 1-4. https://doi.org/10.1109/WASPAA.2013.6701861

Févotte, C., & Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. *Neural Computation*, *23*(9), 2421-2456.

Févotte, C., Vincent, E., & Ozerov, A. (2018). Single-Channel Audio Source Separation with NMF : Divergences, Constraints and Algorithms. In S. Makino (Éd.), *Audio Source Separation* (p. 1-24). Springer International Publishing. https://doi.org/10.1007/978-3-319-73031-8_1

Gloaguen, J.-R., Ecotière, D., Gauvreau, B., Finez, A., Petit, A., & Lebourdat, C. (2020). *Automatic estimation of the sound emergence of wind turbines using Non-negative Matrix Factorization : A preliminary study*.

Gloaguen, J.-R., Ecotière, D., Gauvreau, B., Finez, A., Petit, A., & Lebourdat, C. (2021). Automatic estimation of the sound emergence of wind turbine noise with Non-negative Matrix Factorization. *In review*.

Heittola, T., Mesaros, A., Virtanen, T., & Eronen, A. (2011). Sound event detection in multisource environments using source separation. *in Workshop on Machine Listening in Multisource Environments, CHiME2011*.

Kayser, B., Cotté, B., Ecotière, D., & Gauvreau, B. (2020). Environmental parameters sensitivity analysis for the modeling of wind turbine noise in downwind conditions. *The Journal of the Acoustical Society of America*, *148*(6), 3623-3632. https://doi.org/10.1121/10.0002872

Kayser, B., Ecotière, D., Bourdat, C., & Gauvreau, B. (2018, juin 12). *A new experimental database for wind turbine noise propagation in an outdoor inhomogeneous medium*. 17th International Symposium on Long Range Sound Propagation, Lyon, France.

Kitamura, D., Saruwatari, H., Yagi, K., Shikano, K., Takahashi, Y., & Kondo, K. (2014). Music Signal Separation Based on Supervised Nonnegative Matrix Factorization with Orthogonality and Maximum-Divergence Penalties. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *E97.A*(5), 1113-1118. https://doi.org/10.1587/transfun.E97.A.1113

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788-791. https://doi.org/10.1038/44565

Lee, D. D., & Seung, H. S. (2000). Algorithms for Non-negative Matrix Factorization. *In NIPS*, 556-562.

Ludeña-Choez, J., Quispe-Soncco, R., & Gallardo-Antolín, A. (2017). Bird sound spectrogram decomposition through Non-Negative Matrix Factorization for the acoustic classification of bird species. *PLOS ONE*, *12*(6), e0179403. https://doi.org/10.1371/journal.pone.0179403

Miki, Y. (1990). Acoustical properties of porous materials-Modifications of Delany-Bazley models-. *Journal of the Acoustical Society of Japan (E)*, *11*(1), 19-24. https://doi.org/10.1250/ast.11.19

Rogers, J. (2020). Optimal strategies for wind turbine environmental curtailment. *Wind Energy*, *23*(5), 1331-1350. https://doi.org/10.1002/we.2489

Salomons, E. M. (2012). *Computational Atmospheric Acoustics*. Springer Science & Business Media.