Geophysical Journal International

Geophys. J. Int. (2026) **245,** 1–22
Advance Access publication 2026 January 22
Research Paper

https://doi.org/10.1093/gji/ggag028

Machine learning for data-driven pattern recognition of seismic wind turbine emissions

Marie A. Gärtner [1,*] René Steinmann [2] Laura Gaßner [1] and Joachim R. R. Ritter [1]

[1]*Karlsruhe Institute of Technology (KIT), Geophysical Institute,* D-76187 *Karlsruhe, Germany. E-mail: marie.gaertner@kit.edu*
[2]*GFZ Helmholtz Centre for Geosciences,* D-14473 *Potsdam, Germany*

Accepted 2026 January 16. Received 2025 December 4; in original form 2025 August 7

SUMMARY

Seismic emissions from wind turbines (WTs) depend on the rotation of the WT blades and the wind direction-dependent movement of the WT. Mechanical coupling between the WT foundation and the subsurface generates complex seismic wavefields, making it challenging to manually separate the contributions of different signal sources, thus complicating data labelling. We address this challenge by applying unsupervised machine learning techniques that do not require labelled data. Our analysis focuses on seismic WT emissions recorded near Wind Farm Tegelberg in the eastern Swabian Alb, Southwest Germany. Specifically, we extract time-averaged wavelet features by temporal averaging the wavelet transformation of the continuous three-component seismic data and subsequently apply the clustering algorithm Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). The resulting clusters not only capture the variations in the WT rotation rate but also reveal a clear dependence on wind direction, associated with the radiation pattern of different surface waves. Our results demonstrate the potential of HDBSCAN to uncover meaningful, source-related patterns in continuous seismic records.

Key words: Machine learning; Wavelet transform; Computational seismology; Polarization analysis; Seismic noise; Surface waves and free oscillations.

1 INTRODUCTION

Wind turbines (WTs) are pivotal in the global transition to renewable energy. In Germany, a significant increase in both the density and capacity of installed WTs is anticipated in the coming years (section 4 paragraph 1d, EEG 2023 2024). However, WTs are large structures, and the mechanical coupling of their foundations with the subsurface results in the emission of ground motions (e.g. K. Stammler & L. Ceranna 2016; T. Zieger & J.R.R. Ritter 2018; S. Nagel *et al.* 2021). Seismic signals from WTs can be detected at a distance of more than ten kilometres and affect sensitive measurement instrumentation over a broad frequency range, primarily between 0.1 and 20 Hz (G. Saccorotti *et al.* 2011; K. Stammler & L. Ceranna 2016; S. Nagel *et al.* 2021; L. Gaßner *et al.* 2023; G. Diaferia *et al.* 2025). Although the resulting ground motion amplitudes are far below the human perception threshold of approximately 100 $\mu m\,s^{-1}$ (L. Gaßner & J. Ritter 2023b), local earthquakes and induced seismicity occur in a similar frequency range (J. Charléty *et al.* 2007; M. Hensch *et al.* 2019), which makes it more difficult to detect these seismic events (K. Stammler & L. Ceranna 2016; T. Neuffer & S. Kremers 2017; F. Limberger *et al.* 2023). Furthermore, seismic WT emissions can affect other sensitive scientific instrumentation, such as telescopes, and are, therefore, regarded as one potential disturbance factor in the site-selection process for the Einstein Telescope (G. Diaferia *et al.* 2025). In anticipation of the planned expansion of wind energy infrastructure, which will lead to more challenging conditions for detecting crucial seismic events, a more detailed understanding of the seismic WT emissions is vital.

WT emissions are associated with the eigenmodes of the tower-nacelle system and the rotation rate of the WTs (e.g. T. Neuffer *et al.* 2021; S. Nagel *et al.* 2021). The former corresponds to continuous, monochromatic signals independent of the WT's operational state. They can be detected even when the WT is not in operation, especially in close proximity to the WT (e.g. S. Nagel *et al.* 2021). Emissions related to the rotation of the WT blades are associated with the blade passing frequency (BPF)

*Now at: Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, *F-38000* Grenoble, France. E-mail: marie.gartner@univ-grenoble-alpes.fr

and its multiples. The BPF is given by BPF $= \frac{b \times \text{rpm}}{60}$, where b is the number of WT blades, typically three, and rpm is the rotation rate per minute. The recorded seismic signal is predominantly caused by the coupling of the WT foundation and the subsurface. The contribution from the acoustic-to-ground motion coupling is secondary in comparison (L. Gaßner & J. Ritter 2023b).

Understanding the complex patterns of seismic WT emissions requires advanced analysis. Variations in rotation rate, for instance, can be determined by analysing the frequency content of the recordings: higher rotation rates lead to an increase of the BPF and its multiples. In addition, the increase in rotation rate causes an increase in the amplitude of the WT eigenmodes. However, seismic WT emissions are not only related to the wind speed, which directly controls the rotation rate of the WT, but also to the wind direction, as shown by T. Neuffer *et al.* (2021) and L. Gaßner *et al.* (2023). To exploit the full energy of the wind, the WT's nacelle adjusts its orientation to face the wind. This reorientation may alter the generation and propagation of different types of surface waves, as proposed by T. Neuffer *et al.* (2021), although the underlying physical mechanism is not yet fully understood. In addition, seismic recordings in the vicinity of wind farms can be influenced, for example, by anthropogenic-related noise such as traffic and industry, or other meteorological conditions such as temperature changes (R. Steinmann *et al.* 2022a).

Continuous seismic monitoring produces vast amounts of data, and the manual separation of various signal sources is very time-consuming and not always feasible. This study aims to enhance the understanding of the signal characteristics of seismic WT emissions by utilizing machine learning (ML) techniques. ML, particularly unsupervised learning, is well-suited for this task, as no prior information is needed to identify groups of seismic signals with similar characteristics. Unsupervised learning has been applied in various settings to explore continuous seismograms such as tectonic settings (C.W. Johnson *et al.* 2020; L. Seydoux *et al.* 2020; P. Shi *et al.* 2021; R. Steinmann *et al.* 2022b) or volcanic settings (A. Köhler *et al.* 2010; A. Yates *et al.* 2023; Z. Zali *et al.* 2024; R. Steinmann *et al.* 2024; J. Rimpot *et al.* 2025). However, the application of unsupervised learning in the context of seismic WT emissions remains, to the best of our knowledge, unexplored.

The goal of this study is to investigate what insights can be gained from seismic WT signals using unsupervised ML methods. We aim to explore the capabilities and limitations of such an approach and assess its potential to reveal meaningful patterns in seismic WT emissions. Due to the noisy nature of continuous seismogram data, we apply Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) on time-averaged wavelet features, $\overline{W}$, extracting signal families with similar characteristics from the continuous seismic data.

For this study, we analyse a data set recorded during a field campaign in 2023 on the eastern Swabian Alb, southwestern Germany (Fig. 1), conducted as part of the Inter-Wind project (L. Gaßner *et al.* 2022).

First, we present the methodology and data set in Sections 2 and 3. The results and their interpretation are discussed in Section 4, followed by a summary and the conclusion in Section 5.

2 METHODS

In the following, we outline our workflow to group seismic signal families (Fig. 2). We first introduce HDBSCAN as our choice of a clustering algorithm (Section 2.1). We then explain how we extract time-averaged wavelet features, $\overline{W}$, from the continuous seismograms, which we use as input for HDBSCAN (Section 2.2).

2.1 Clustering continuous seismograms with HDBSCAN

Seismic time-series contain a wide variety of unidentified signals, making clustering a powerful tool for exploring this diversity without the need for labelled data. Given the many existing clustering algorithms, several ones have already been tested on seismic time-series. For instance, C.W. Johnson *et al.* (2020) used K-Means clustering to distinguish different types of ambient seismic noise in a dense array. Hierarchical clustering has been applied to continuous seismic recordings to separate local seismicity from anthropogenic sources (R. Steinmann *et al.* 2022b), and to cross-correlation data to identify different types of volcanic activity (A. Yates *et al.* 2023; P. Makus *et al.* 2023). In a recent example, J. Rimpot *et al.* (2025) employed density-based clustering, specifically Density-Based Spatial Clustering of Applications with Noise (DBSCAN), to reveal submarine volcanic activity in continuous seismic recordings.

The choice of a clustering algorithm should reflect the underlying structure of the data. Seismic signals are inherently noisy and typically exhibit substantial class imbalance, with rare transient events, such as earthquakes, embedded in long durations of ambient noise. These characteristics pose challenges for classical algorithms like K-Means, which assume spherical cluster shapes and balanced cluster sizes (A.K. Jain *et al.* 1999). Thus, a method is required that is both robust to noise and capable of identifying clusters with arbitrary shapes. DBSCAN addresses part of this problem by identifying clusters as dense regions while labelling sparse regions as noise. However, it is not designed to cluster data of varying density. To overcome this problem, we apply HDBSCAN, a hierarchical extension of DBSCAN, introduced by R.J.G.B. Campello *et al.* (2013) and implemented in Python by L. McInnes & J. Healy (2017). HDBSCAN constructs a cluster hierarchy and extracts the most stable structures (L. McInnes & J. Healy 2017), making it particularly well suited for seismic time-series in which signal families vary in both density and complexity. In this context, higher density relates to greater similarity among the analysed time windows. To our knowledge, HDBSCAN has not yet been applied directly to continuous seismic time-series, despite its strong conceptual fit for this type of data. However, it has been shown to be useful in separating earthquake swarms from background seismicity in event catalogues (D. Essing & P. Poli 2024).

To identify clusters within the data set and present their arrangement, HDBSCAN builds a dendrogram-like structure called a condensed tree (Fig. 2d), from which the most stable branches are extracted as clusters (L. McInnes & J. Healy 2017). A dendrogram is a tree-based diagram that represents the hierarchical relationships between clusters. At the bottom of the dendrogram, each point within a data set initially forms an individual cluster. These clusters are merged based on their similarity or dissimilarity, as quantified by a chosen distance metric (A.E. Ezugwu *et al.* 2022). In HDBSCAN, the condensed tree is constructed using a distance metric that incorporates local density information. The key principle is that dense regions form clusters, while low-density regions are classified as noise.

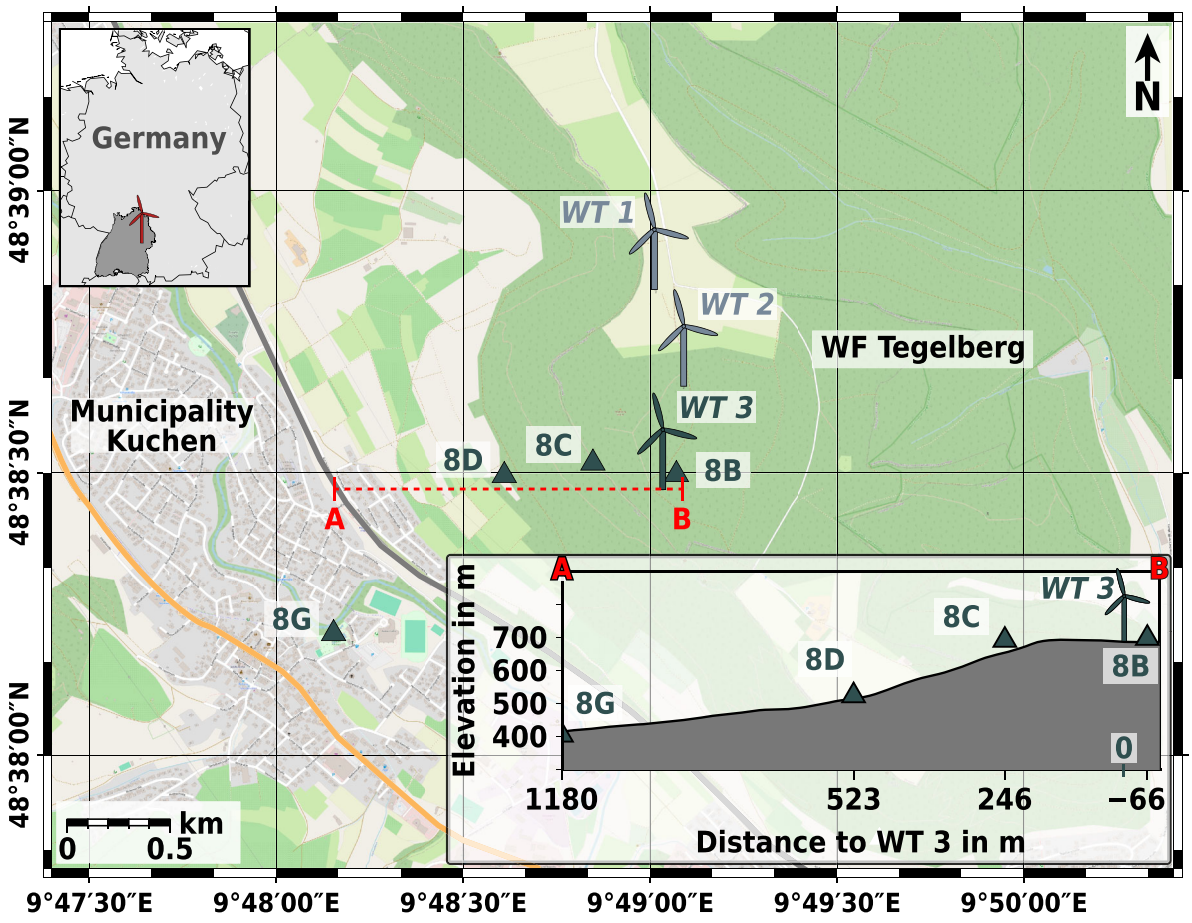

Figure 1. Map of wind farm (WF) Tegelberg, located east of the municipality of Kuchen on the escarpment of the eastern Swabian Alb in the federal state of Baden-Württemberg, Southwest Germany (see upper map inset, dark grey patch). The WF comprises three wind turbines (*WT 1 – 3*). The seismic stations used in this study are marked with triangles and labelled with their station codes (IW0∗∗). Distance and azimuth to *WT 3* are as follows: 8B: 66 m, 42°; 8C: 246 m, 291°; 8D: 523 m, 275°; 8G: 1180 m, 246°. The lower inset shows a cross-section along the dashed red line between A and B. The map background indicates land cover: dark green represents forest, white to light green denotes agricultural fields, and grey indicates developed areas. The light orange line shows a federal road and the darker grey line a major railway line.

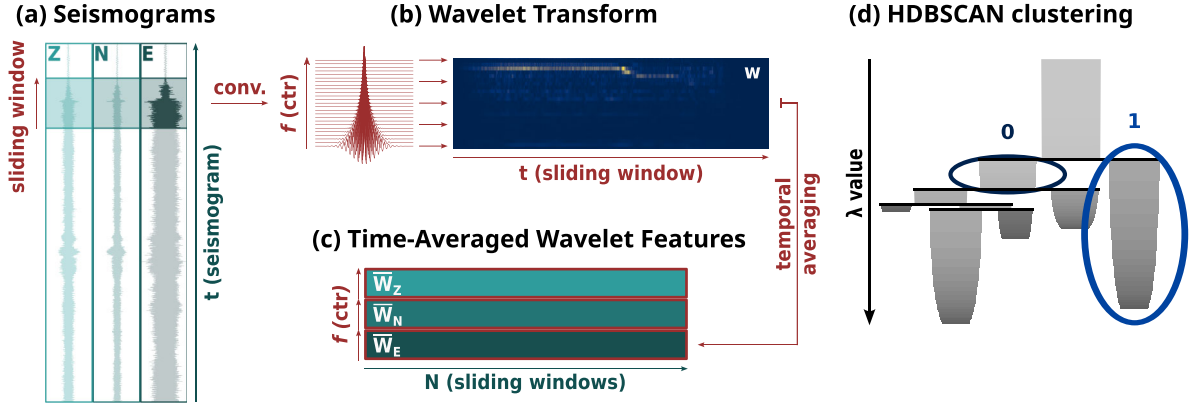

Figure 2. Application of the wavelet transformation to continuous three-component seismograms (a to c) and subsequent clustering (d). The data in each sliding 128 s long time window (a) are convolved with the filter bank consisting of 25 wavelets, each characterized by a specific centre frequency f_{crt} (b). After computing the modulus, the wavelet coefficients W are averaged over time, resulting in the time-averaged wavelet features $\overline{W}$ (c). These steps are conducted for all recording components and N sliding windows, and the resulting $\overline{W}$ are concatenated. Next, $\overline{W}$ are clustered with HDBSCAN (d). The clustering results are presented in a so-called condensed tree. λ, the inverse of the mutual reachability distance, describes the cluster density. The colour shading indicates the number of data points: light grey corresponds to a high accumulation of points, while dark grey indicates lower point density. The marked branches are the selected clusters. In this example, two clusters are extracted with the ID 0 and 1.

The λ-value on the ordinate of the condensed tree represents the inverse of the mutual reachability distance; higher λ indicates denser clusters (Fig. 2d). At the top of the tree, all samples belong to a single branch. As we move downwards, denser substructures are revealed and split into sub-groups. If a sub-branch contains fewer samples than the specified minimum cluster size (hyperparameter *min_cluster_size*), it is classified as noise and is removed from the tree. This step is indicated by the change in colour shading in Fig. 2(d). As a result, the upper part of the tree contains wider branches representing less dense, large-scale structures, while the lower part contains narrower branches corresponding to denser, finer-scale subclusters.

The stability of a branch is determined by the cumulative 'lifetime' of its samples, that is, by the range of λ-values over which each sample remains within the branch before either moving to a new branch or being labelled as noise. A branch is declared a cluster if its stability exceeds the combined stability of its primary descendant branches.

Here, we tune two key hyperparameters of HDBSCAN: *min_samples* and *min_cluster_size*. The *min_samples* parameter controls the number of samples used to estimate the local density. A smaller value tends to emphasize local structures in the data, whereas a larger value results in more conservative clustering (L. McInnes & J. Healy 2017). The *min_cluster_size* defines the minimum number of samples required for a group to be considered a valid cluster rather than classified as noise. The tuning of these hyperparameters is discussed in detail in Section 3.4.

2.2 Wavelet transformation for seismic data analysis

To identify clusters within continuous time-series data, we first extract time windows that can be grouped based on their similarity (Fig. 2a). However, seismic signals are sensitive to small deformations and translations, which pose challenges for direct clustering. In practice, this means that two time windows may contain signals of the same type but appear shifted or slightly distorted. Nevertheless, the chosen feature extraction approach should still recognize them as belonging to the same signal family. Therefore, it is essential to extract robust and informative features from the data that preserve information relevant for classification while reducing sensitivity to irrelevant variations.

Our feature extraction approach is inspired by the scattering network, a type of convolutional neural network originally applied to acoustic signal processing and image recognition by J. Bruna & S. Mallat (2013) and J. Andén & S. Mallat (2014). More recently, it has been successfully adapted to seismological data analysis (L. Seydoux *et al*. 2020; S. Barkaoui *et al*. 2021; Á.B. Rodríguez *et al*. 2022; R. Steinmann *et al*. 2022b, a, 2024; L. Moreau *et al*. 2023).

Extensive testing, presented in M.A. Gärtner (2024), has shown that a single-layer scattering network is sufficient to analyse the data used in this study. As only one layer is applied, the operation effectively reduces to a wavelet transform with subsequent temporal averaging. This transform is performed on continuous three-component seismograms using overlapping sliding windows (Fig. 2a). In the first step, a wavelet transform is applied separately to each component of the seismogram within each sliding time window (Fig. 2b). The signal $x(t)$ is convolved with a set of wavelets ψ_λ, called a filter bank, as follows:

$$W_\lambda(t) = |w_\lambda(t)| = |x(t) * \psi_\lambda(t)|, \tag{1}$$

where W_λ are the modulus of the wavelet coefficients w_λ, $*$ denotes the convolution operator, and ψ_λ are the wavelets with scaling factor λ given by

$$\lambda = 2^{\frac{k}{Q}}, \quad \text{where } k \in \{0, 1, \ldots, J \cdot Q - 1\}. \tag{2}$$

Here, λ is determined by the number of octaves $J \in \mathbb{Z}$ covered by the wavelets and the number of wavelets per octave $Q \in \mathbb{Z}$. Each wavelet has a centre frequency f_{ctr}. Following the proposal of J. Andén & S. Mallat (2014) the wavelet transform is conducted with a set of Morlet wavelets:

$$\psi(t) = \exp\left(-i2\pi f_{\text{ctr}} t\right) \exp\left(-\frac{t^2}{a^2}\right), \quad a = \frac{d}{f_{\text{ctr}}} = \frac{d}{\lambda f_{\text{Ny}}}, \tag{3}$$

with a describing the standard deviation of the Gaussian window, the centre frequency f_{ctr} and the bandwidth d.

Hence, to define the filter bank used in the wavelet transform, several parameters must be specified: the Nyquist frequency f_{Ny}, the number of octaves J, the number of wavelets per octave Q and the bandwidth of the wavelets. The latter is controlled by the so-called quality factor q. The result of the wavelet transform is a time–frequency representation of the signal, commonly referred to as a scalogram. Details on the parameter configuration are provided in Section 3.3.

After computing the wavelet transform, the resulting coefficients W_λ within each time window are averaged over time and concatenated to form a matrix, composed of the time-averaged wavelet features $\overline{W}$ (Fig. 2c). Depending on the data characteristics, other aggregation functions, such as median, maximum or minimum, might be better suited.

Since the input comprises three-component seismograms, the time-averaged wavelet features also consist of three recording components: the vertical component is transformed to $\overline{W}_Z$, the north component to $\overline{W}_N$ and the east component to $\overline{W}_E$ (Fig. 2c). The time-averaged wavelet features, $\overline{W}$, form the input for HDBSCAN.

3 DATA AND APPLICATION

We apply the wavelet transformation together with time averaging and subsequent cluster analysis to seismic data recorded at the wind farm (WF) Tegelberg on the eastern Swabian Alb in Southwest Germany (Fig. 1). There the Inter-Wind project was conducted as an interdisciplinary project with the aim of investigating both the acoustic and seismic emissions of WTs as well as the perception and annoyance of local residents. Furthermore, the project aimed to determine and mitigate the factors that influence the extent of perceived annoyance. During the Inter-Wind project, several measurement campaigns were carried out on the eastern Swabian Alb (Fig. 1; L. Gaßner *et al*. 2022, 2023; E. Blumendeller *et al*. 2023; L. Gaßner & J. Ritter 2023a, b; F.J. Müller *et al*. 2025). For this work, we focus on seismic data recorded during the IW08 campaign (campaign T3 in F.J. Müller *et al*. 2025). We analyse recordings from four recording stations: IW08B, IW08C, IW08D and IW08G, hereafter referred to as 8B, 8C, 8D and 8G. These recording stations were deployed between 2022 October 19 and 2023 February 21 near the southernmost WT of WF Tegelberg, marked as *WT 3* in Fig. 1. Recording station 8B was installed on top of Tegelberg mountain at an elevation of 699 m, in close proximity to *WT 3* (66 m distance). 8C was deployed in a distance of 246 m at 693 m of elevation. 8D was positioned on the escarpment of the Alb, a northwest-facing step

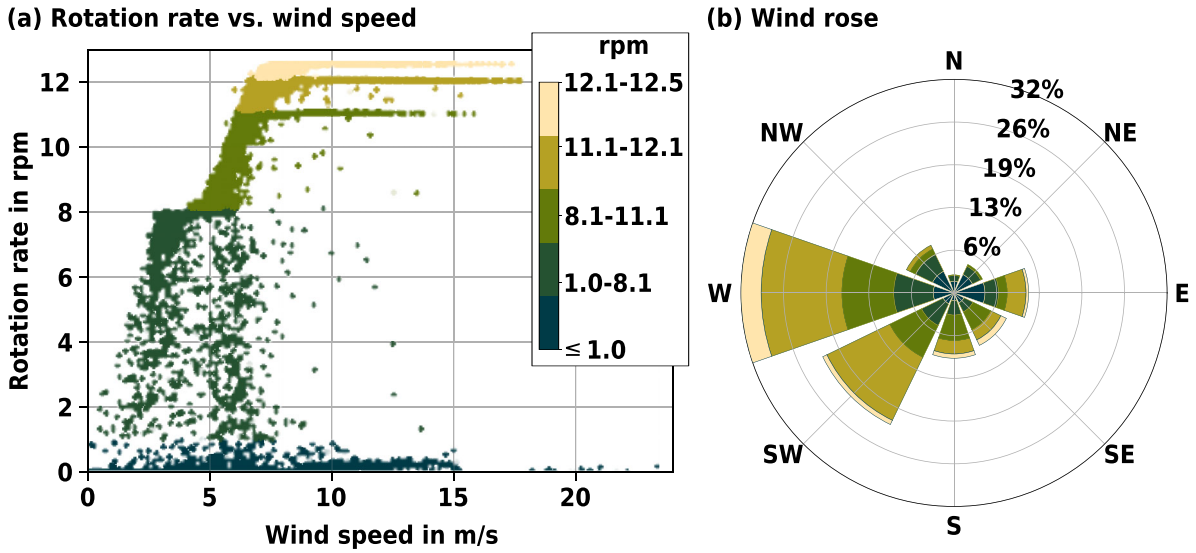

Figure 3. Operational data recorded at *WT 3* of wind farm Tegelberg from 2022 November 1 until 2023 February 20 with a 10-min sampling interval during measurement campaign IW08. (a) Rotation rate as a function of the wind speed for *WT 3* (wind farm Tegelberg). The cut-in wind speed is 3 m s^{-1} and the cut-off wind speed 18 m s^{-1}. The colours mark the different operational modes of *WT 3*, with brighter colours representing a higher rotation rate. Plateaus at 0, 7.9, 11, 12 and 12.5 rpm, correlate with the wind turbine operational stages: off, partial load, two noise reduced modes and full load. (b) Rotation rate and wind direction distribution in percentage of occurrence, using the same colour scale as in (a).

slope in a distance of 523 m to *WT 3* at 525 m of elevation. 8G was located in the valley below Tegelberg mountain, within the municipality of Kuchen, at an elevation of 407 m and a distance of 1180 m to the wind farm.

The sensors used were three-component Nanometrics Trillium Compact Posthole seismometers, with a corner period of 20 s and a sampling rate of 200 Hz was chosen to record signals up to approximately 90 Hz.

3.1 Operational data

To evaluate the seismic emissions from *WT 3*, we compare the seismic recordings with the operational data provided by the operators, sampled at 10 min intervals. In particular, we use the WT rotation rate, wind speed and nacelle position. The latter two were recorded at the hub height of 139 m. Since the nacelle of the WTs aligns with the wind direction, we refer to the nacelle position as wind direction throughout this work.

The WTs of WF Tegelberg are of the type General Electric (GE) 2.75–120 with a rated power of 2.78 MW. During the recording time, *WT 3* was operated in five different modes: off, partial load at 7.9 rpm, full load at 12.5 rpm and two noise-reduced modes at 11 and 12 rpm, respectively. Fig. 3(a) shows the rotation rate as a function of the wind speed, whereby the respective operational stages are colour-coded. The predominant wind direction, including the periods with the highest wind speeds, during the experiment was west, whereas the wind rarely came from north or northeast (Fig. 3b; F.J. Müller *et al.* 2025). The operational data used in this work are interpolated to match the sampling rate of the time-averaged wavelet features $\overline{W}$ (Section 3.3).

3.2 Seismic WT emissions

As described in the introduction, seismic recordings in the vicinity of WT type capture both the eigenmodes of the tower–nacelle system and signals associated with the BPF and its harmonics. These frequencies are extracted from consistently observed spectral peaks in the power spectral density (PSD) of the seismic recordings (L. Gaßner & J. Ritter 2023b). The eigenmodes of the analysed WT type are found at 1.2, 3.6, 6.0, 8.33 and 12.25 Hz. The eigenmodes and operational stages of the WTs of WF Tegelberg, *WT 1* to *3*, are summarized in Table 1. The eigenmodes are denoted by uppercase Roman numerals (Table 1a), while the frequencies associated with the BPF are labelled with lowercase Roman numerals (Table 1b). The 32nd multiple of the BPF is the first of the recorded data that appears distinctly in the spectrogram with a high amplitude (Figs 4 and 5). We therefore configured the filterbank of the applied wavelet transformation such

Table 1. Frequencies related to the seismic emissions of the General Electric (GE) 2.75–120 wind turbines (WTs) after L. Gaßner & J. Ritter (2023b). (a): Eigenmodes of the WTs tower–nacelle system. (b): Rotation rates in partial-mode (7.9 rpm) and full-load (≫12 rpm) with their corresponding blade passing frequency, BPF $= \frac{3\times\text{rpm}}{60}$ and its 32nd multiple. The IDs of the frequencies referenced in this study are also provided.

(a) WTs eigenmodes

Freq. in Hz	ID
1.20	I
3.60	II
6.00	III
8.33	IV
11.25	V

(b) WTs rotation rates and related frequencies

Rotation rate in rpm	BPF in Hz	32nd BPF in Hz	ID
7.9	0.395	12.64	i
11.0	0.550	17.60	ii
12.0	0.600	19.20	iii
12.5	0.625	20.00	iv

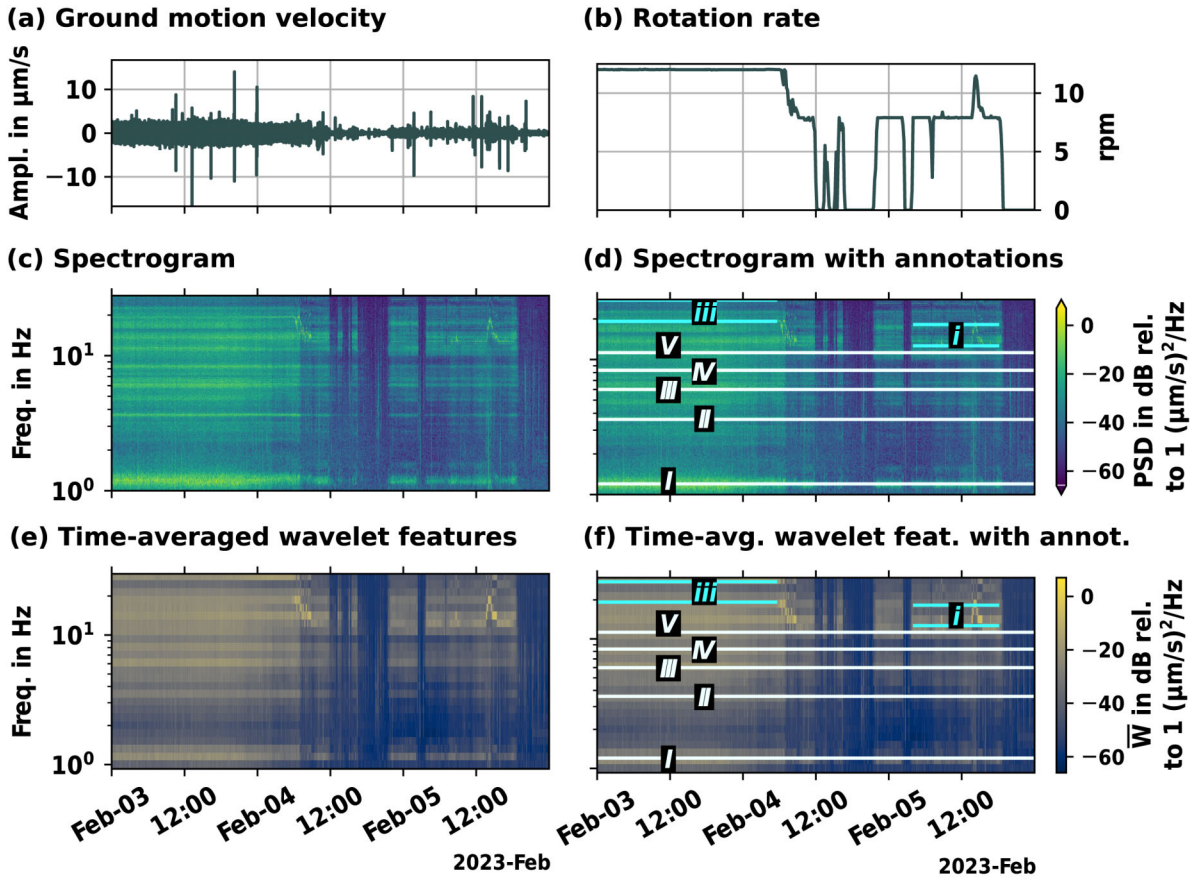

Figure 4. (a) Vertical ground motion velocity recorded at recording station 8B between 2023 February 3 and 5. (b) rotation rate of the southernmost wind turbine, *WT 3*, at WF Tegelberg. (c) Corresponding spectrogram with and (d) without annotations and time-averaged wavelet features $\overline{W}$, shown without (e) and with (f) annotations following the frequency nomenclature in Table 1. The white lines indicate the eigenmodes of the tower–nacelle system (capital roman number), while the cyan coloured lines represent the 32nd and 46th multiples of the blade pass frequency (small roman number). The change in frequency corresponds to the change in rotation rate from 12 to 7.9 rpm. A zoomed view of three selected excerpts is provided in the appendix (Fig. B1).

that the highest frequency wavelet encompasses the 32nd multiple across all possible rotation rates (Table 1b). In our analysis, we do not consider higher multiples of the BPF. Figs 4 and 5 show examples of the seismic WT emissions recorded during three days in February 2023. Figs 4(c) and (d) highlight the occurrence of the eigenmodes and BPF in the spectrogram. Comparing the spectrogram to the rotation rate of *WT 3* (Fig. 4b), we observe that the change in frequency above 10 Hz in the spectrogram, for example from 19.2 to 12.6 Hz (iii and i), coincides with the change in rotation rate from 12 to 7.9 rpm. In this example, the frequency changes of the 32nd and 46th BPF multiples are observable (Fig. 4d, highlighted in cyan). Alongside the frequencies associated with the rotation rate, continuous monochromatic signals corresponding to the eigenmodes of the tower–nacelle system are visible and marked with white lines (I to V). In this time window, the eigenfrequencies have their highest amplitude during the rotation rate 12 rpm and are barely visible when the WT is not in operation (Figs 4b and c).

At recording stations 8C and 8D, we observe the same pattern but with lower amplitudes (Figs 5a and b). At recording station 8G, however, we cannot observe the 32nd and 46th multiple of the BPF and only barely the eigenmodes of the WT (Fig. 5c), as it has a distance of more than 1 km to the WF and is set up in the municipality Kuchen (Fig. 1). At this side, the anthropogenic noise of the municipality is dominant.

The WT-related ground motion recorded at station sites 8B and 8C is predominantly emitted by *WT 3* (Fig. A1). *WT 1* and *WT 2* have only a minor influence at these recording stations. Recording station 8D, on the other hand, receives stronger emissions from *WT 1* and *WT 2*. The amplitudes, observed when only *WT 3* is running, are comparable to those when *WT 1* and *WT 2* are in operation. However, in the latter case, the WT-related frequency peaks appear more distinct. For simplicity, we treat *WT 3* as the single dominant source, acknowledging that wave-interference effects generated by multiple WTs may arise, as demonstrated by, for example, F. Limberger *et al.* (2022) and L. Gaßner *et al.* (2023).

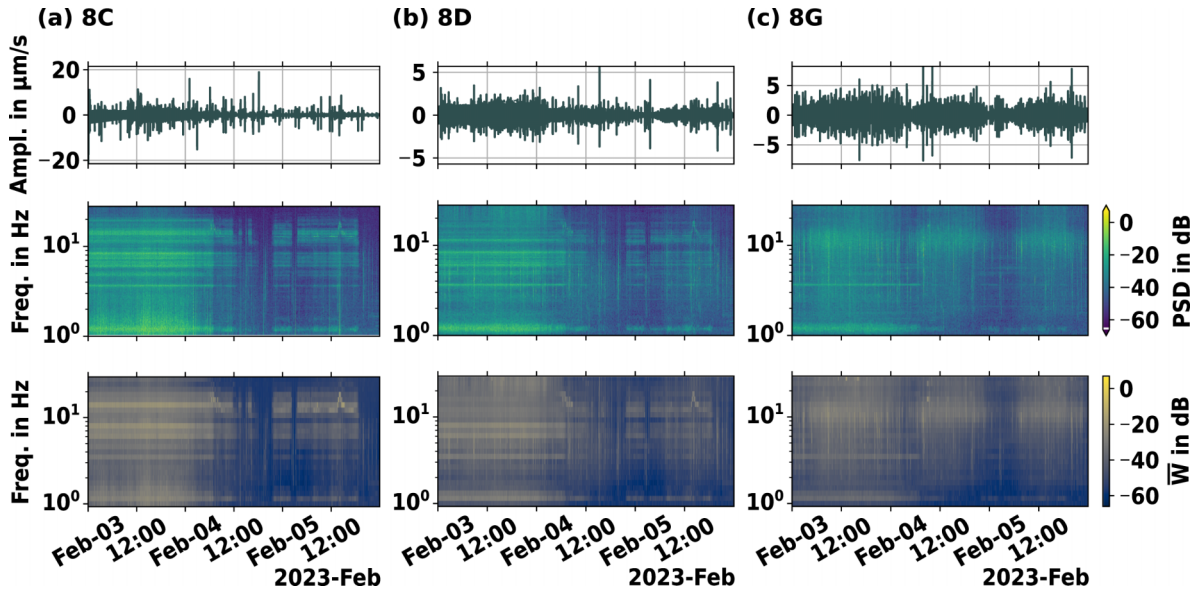

Figure 5. Example of wind turbine emissions presented in Fig. 4 recorded on the vertical component at recording stations (a) 8C, (b) 8D and (c) 8G. They are arranged from left to right in order of increasing distance from *WT 3*, located 246, 523 and 1180 m away, respectively. The rows show, from top to bottom, the ground motion velocity, the spectrogram and the time-averaged wavelet features $\overline{W}$. The amplitudes of the power spectral density (PSD) and $\overline{W}$ are given in dB relative to 1 $(\mu m\,s^{-1})^2 Hz^{-1}$..

3.3 Design of the wavelet transform

First, the seismic recordings are pre-processed by subtracting their average, removing their linear trend and converting the data to ground motion velocity. In addition, we downsample the data to 64 Hz. With this, the highest frequency captured by the wavelet transform is 32 Hz and the highest centre frequency f_{ctr} of the wavelets is 27.86 Hz. Hence, we capture the frequencies of the 32nd multiple of the BPF of all operational stages (Table 1b). The seismograms are segmented into 128 s long time windows with an overlap of 50 per cent between consecutive windows to capture changes in rotation rate (Fig. 2a). Furthermore, we taper each sliding window with a Tukey window to avoid edge effects.

Since the lowest detectable eigenmode of this data set is 1.2 Hz, we set the number of octaves J to five. Therefore, the centre frequencies f_{ctr} of the filter banks range between 1.00 and 27.86 Hz. Using a frequency resolution of five wavelets per octave and a quality factor of four, we ensure that each frequency peak corresponding to the WT eigenmodes is captured by a different wavelet. Hence, the filter bank contains 25 wavelets (Fig. 2b). Each wavelet is normalized in time by its $L1$-norm, to preserve the relative amplitudes of the signal (R. Morel *et al.* 2023).

To capture both the influence of monochromatic signals associated with the WT eigenmodes and the varying frequency content linked to the BPF multiples, we apply the temporal average as an aggregation function to collapse the transformed time windows along the time axis. The resulting time-averaged wavelet features $\overline{W}$ have a sampling rate of 64 s and comprise 3 x 25 dimensions (Fig. 2c). With the chosen setup, no additional pre-processing steps, such as dimensionality reduction, are required prior to clustering, and we effectively avoid the curse of dimensionality (R.E. Bellman 1961), while still retaining relevant frequency information associated with seismic WT emissions.

In Figs 4 and 5, we compare the spectrograms with the corresponding $\overline{W}$ for the selected time window. The eigenmodes I, II and IV are clearly distinguishable, whereas III and V appear to be blurred with the surrounding frequency content. The 32nd multiple of the BPF is observable, shifting from 19.2 to 12.64 Hz, while the 46th is only faintly visible, decreasing from 27.6 to 18.17 Hz. Similar to the spectrograms, $\overline{W}$ capture the reduction in amplitude with increasing distance from the WT (Fig. 5). This effect is especially pronounced at higher frequencies, which are more strongly attenuated, as can be seen in Fig. 5(c), where the frequency shift, due to the change in rotation rate, is not visible. A more detailed comparison of the time-averaged wavelet coefficients $\overline{W}$ and the PSD of three individual time windows excerpted from Fig. 4 is provided in Appendix B.

3.4 HDBSCAN hyperparameter tuning

We will now turn to the tuning of the hyperparameters of HDBSCAN. Two key hyperparameters, *min_samples* and *min_cluster_size*, are optimized using the time-averaged wavelet features $\overline{W}$ from station 8B. This recording site is the closest to *WT 3*, and we expect the clusters to be dominantly influenced by WT operational stages and meteorological conditions.

As described in Section 3.2, the WT operated in five distinct modes: off, partial load, two noise-reduced modes and full load (Table 1). Since the frequency content of both noise-reduced modes is captured by the same wavelet, they are expected to produce a high-amplitude time-averaged wavelet feature at the same frequency scale. As a result, we anticipate identifying at least four distinct clusters related to WT operational modes.

To ensure that the clusters are meaningful, we require that at least four clusters contain more samples than the median number of samples across all clusters (Fig. C1). When multiple hyperparameter combinations meet this criterion, we select the combination that maximizes the number of samples in the smallest cluster, to avoid overly small or fragmented clusters. Fig. C1

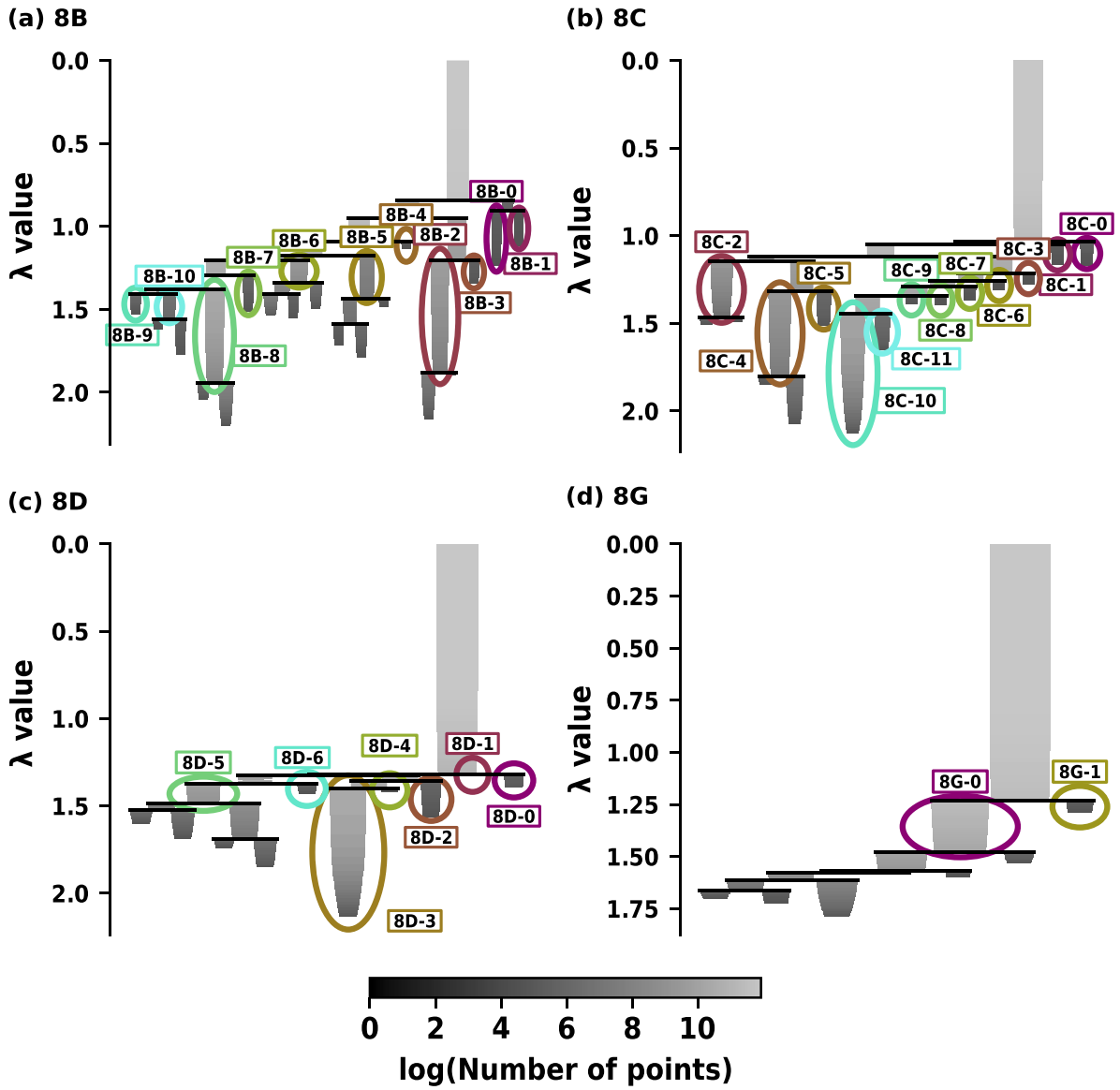

Figure 6. Condensed tree with highlighted cluster IDs calculated from the time-averaged wavelet features $\overline{W}$ of recording stations (a) 8B, (b) 8C, (c) 8D and (d) 8G. The coloured ovals indicate the selected cluster IDs and the respective colours are consistently used in other figures throughout this work. HDBSCAN hyperparameters were tuned based on $\overline{W}$ from recording station 8B. The hyperparameter *min_cluster_size* was set to 100 and *min_samples* to 50 (Fig. C1).

presents the resulting cluster solutions obtained during the hyperparameter testing for recording station 8B. Based on this tuning procedure, we set *min_samples* to 50 and *min_cluster_size* to 100.

4 EVALUATION OF THE CLUSTERING RESULTS

The time-averaged wavelet features $\overline{W}$ of each recording station is individually clustered using HDBSCAN (Fig. 2d). This approach allows us to investigate not only the influence of the WT operational stages, meteorological conditions and other anthropogenic activities, but also the impact of source–receiver distance on the clustering results.

4.1 Overview of the clustering results

Using the selected hyperparameters, the clustering analysis yields 11 clusters for station 8B (labelled as *8B-ID*), 12 clusters for station 8C (*8C-ID*), seven clusters for station 8D (*8D-ID*) and two clusters for station 8G (*8G-ID*), indicated with coloured ovals in Fig. 6.

Each cluster in the condensed trees of stations 8B, 8C and 8D exhibits at most two hierarchical levels of subclusters, indicating relatively simple cluster structures. In contrast, cluster *8G-0* reveals extensive branching, suggesting a more complex internal data structure that likely reflects higher variability in signal characteristics, possibly due to the proximity of the recording site to an urban environment (Fig. 1). Table D1 summarizes the population of each cluster.

To investigate the correlation between the clustering results and WT operational stages, as well as the meteorological conditions, we analyse the distribution of each cluster with respect to rotation rate, wind speed, wind direction, day of the week and hour of the day (Figs D1 to D4). Clusters from all stations, except 8G, show correlations with the operational stages of *WT 3*. Given that the rotation rate is linked to the wind speed, this relationship is also reflected in the cluster distributions. Additionally, several clusters exhibit dependencies on wind direction, which we will discuss in more detail below.

Most clusters from stations 8B, 8C and 8D, along with cluster *8G-0*, exhibit no clear time-dependent patterns (Figs D1 to D4). In contrast, cluster *8G-1* is exclusively populated by time windows recorded between Monday and Friday, from 06:00 to 15:00 UTC, strongly indicating an anthropogenic origin linked to weekday working hours (Fig. D4).

4.2 Detailed analysis of selected clusters

A detailed discussion of all clusters is beyond the scope of this paper; therefore, we focus on those associated with the higher noise-reduced operational mode, corresponding to a rotation rate of 12 rpm, which was the primary operational stage during the recording period. Specifically, we highlight clusters *8B-5*, *8B-8*, *8C-7*, *8C-10*, *8C-11*, *8D-0* and *8D-3* from Fig. 6.

While these clusters share similar distributions of rotation rate and wind speed, they differ in their associated wind directions (Figs 7a to c). For stations 8B and 8D (Figs 7a and c), one cluster corresponds to periods when the wind predominantly blew from the east and southeast (clusters *8B-5* and *8D-0*), whereas the other is associated with wind from the west and southwest (clusters *8B-8* and *8D-3*). For station 8C (Fig. 7b), three clusters correspond to the higher noise-reduced mode: clusters *8C-7* and *8C-11* are linked to winds from the east and southeast, while cluster *8C-10* is associated with wind from the southwest and west. In agreement with the dominant western wind direction during the recording time of the stations (Fig. 3b), the clusters associated with the wind from the southwest and west are approximately between ten and 100 times higher populated compared to the other clusters shown in Figs 7(a) to (c).

To assess whether these clusters are activated simultaneously, we plot their cumulative sums in Fig. 8. Due to the classification of certain time windows as noise, some time windows may not be present in all clusters at every accumulation point (Table D1). Clusters corresponding to wind from the west and southwest are filled nearly simultaneously (Fig. 8a). Fig. 8(b) shows those clusters activated during winds from the southeast or east. In particular, cluster *8B-5* exhibits a higher overall accumulation, yet the activation times of the other clusters correlate closely with those of *8B-5*. *8C-7* and *8D-0* are only populated during the weekend between 19:00 and 2:00 UTC, and 20:00 and 7:00 UTC, respectively (Figs D1 and D3). Fig. 8(b) shows that these two clusters were only active during a single weekend.

In general, these results indicate that clusters associated with the same wind directions at different recording sites tend to be activated concurrently.

4.2.1 *Effect of distance to WT on the clustering results*

Several studies analysed the amplitude decay of seismic WT emissions with distance (e.g. N. Lerbs *et al.* 2020; F. Limberger *et al.* 2021; L. Gaßner & J. Ritter 2023b; L. Gaßner *et al.* 2023). This effect is clearly visible in both, the spectrogram and $\overline{W}$, shown in Figs 4 and 5. In this section, we investigate how the WT–station distance influences the clustering results and, consequently, how amplitude decay affects the clustering performance. To visualize this effect, we plot the mean amplitude of $\overline{W}$ associated with each cluster and component (Fig. 9). The highlighted WT eigenmodes (I to V, Table 1) and the frequency peak corresponding to the 32nd multiple of the BPF, labelled iii (Table 1), decay in amplitude across the subplots. These peaks are most prominent in the clusters of stations *8B* and *8C* (Figs 9a and b). In contrast, station *8D* exhibits clear peaks only at the eigenmodes, while peak iii is significantly reduced or absent (Fig. 9c). At station *8G*, only eigenmode I can be identified, while the other WT-related frequency peaks are no longer visible (Fig. 9d). These observations are consistent with the attenuation of the signals visible in Figs 4 and 5. Furthermore, the presence of peak iii in Fig. 9(d) cannot be reliably attributed to WT emissions due to the influence of attenuation and increased anthropogenic noise from the surrounding municipality of Kuchen, where a federal road and a major railway line cross the study area (Fig. 1).

Additionally, going back to Fig. 7(d), the histograms reveal no clear correlation between the operational stages and the clustering results. While cluster *8G-0* contains time windows from all operational modes, cluster *8G-1* lacks the higher noise-reduced mode and full load. Furthermore, *8G-0* has a significantly larger population compared to *8G-1*. This also might explain the large variance observed in amplitudes of $\overline{W}$ of cluster *8G-0* (Fig. 9d).

It should be noted that the hyperparameter tuning was performed using the time-averaged wavelet features $\overline{W}$ of recording station 8B. Using the 8G time-averaged wavelet features $\overline{W}$ to tune the hyperparameters may have revealed clusters related to WT emission. Especially, as we can identify the WT eigenmodes in Fig. 5(c). Notably, cluster *8G-0* has the overall largest number of subclusters (Fig. 6) and a detailed analysis could potentially uncover WT-related patterns. However, such an investigation is beyond the scope of this study, and station *8G* is, therefore, excluded from further analysis.

Overall, clustering performs well for the stations in proximity to the wind farm, namely 8B, 8C and 8D (Figs 7a to c and D1 to D3). At greater distances, signals from other sources, like anthropogenic noise, dominate the ground-motion recordings, and a clear separation of WT-related signals becomes difficult (Figs 7d, D4, and 9d). It may be worthwhile to investigate whether WT-related clusters can still be identified within the subclusters in future work.

4.2.2 *Separation of clusters with identical rotation rates under varying wind directions*

We analysed the 3-D particle ground motion velocity during selected time windows in which all investigated clusters are active, to examine the separation of clusters with the same rotation rate but different wind directions (Fig. 7). We choose one window under west wind conditions (Figs 10a to c) and another under east wind conditions (Figs 10d to f). Extending earlier work from, for example, G. Saccorotti *et al.* (2011), N. Lerbs *et al.* (2020),

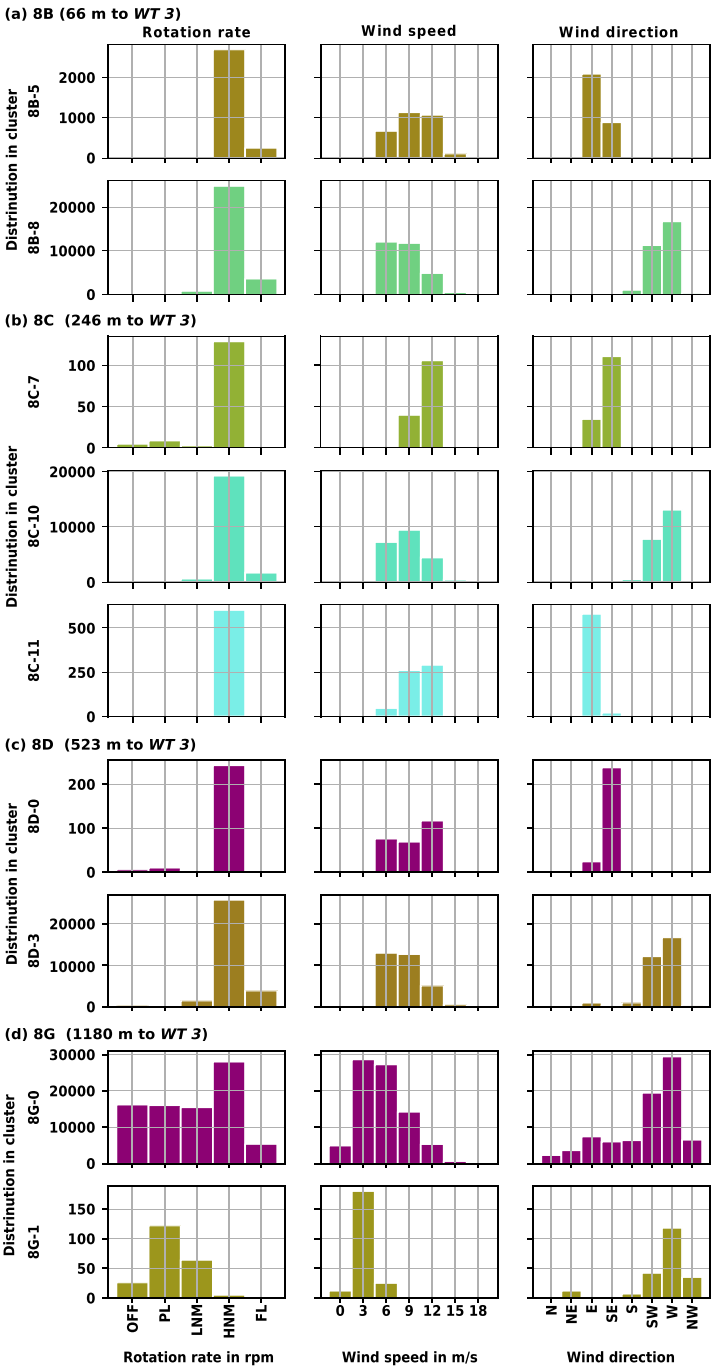

Figure 7. Selected clusters analysed in relation to the rotation rate of *WT 3* (OFF: ≤ 1 rpm, PL (partial load): $1.0 - 8.1$ rpm, LNM (lower noise reduced mode): 8.1–11.1 rpm, HNM (higher noise reduced mode): 11.1–12.1 rpm, FL (full load): 12.1–12.5 rpm, wind speed and wind direction. (a) to (c) show only the clusters that correlate with the higher noise reduced mode (HNM), while (d) shows all clusters of station 8G. A comprehensive overview of all clusters, including their distributions by day of the week and hour of the day, is provided in the appendix (Figs D1 to D4).

T. Neuffer *et al.* (2021) and G. Diaferia *et al.* (2025), we analyse the particle motion at the 32nd multiple of the BPF, the frequency peak with the highest amplitude at stations 8B and 8C. We bandpass filter the signal to extract the frequency content at 19.2±0.25 Hz. Afterwards, we rotate the vertical (Z), north (N) and east (E) components of the seismograms to Z, radial (R) and transverse (T) components, using *WT 3* as the source location. Other windows with similar wind conditions exhibit consistent particle motion patterns.

Note that station 8B is located to the northeast, 8C to the west-northwest and 8D to the west of *WT 3* (Fig. 1). Consequently, station 8D is positioned nearly in line with the analysed wind directions and the WT, while station 8C is only approximately aligned. 8B and 8C are oriented nearly perpendicular to each other. Station 8B is situated on flat terrain at the same elevation as *WT 3*, station 8C lies at the transition from the plain to the slope of the Alp Cuesta and station 8D is installed on the slope itself (see inset in Fig. 1).

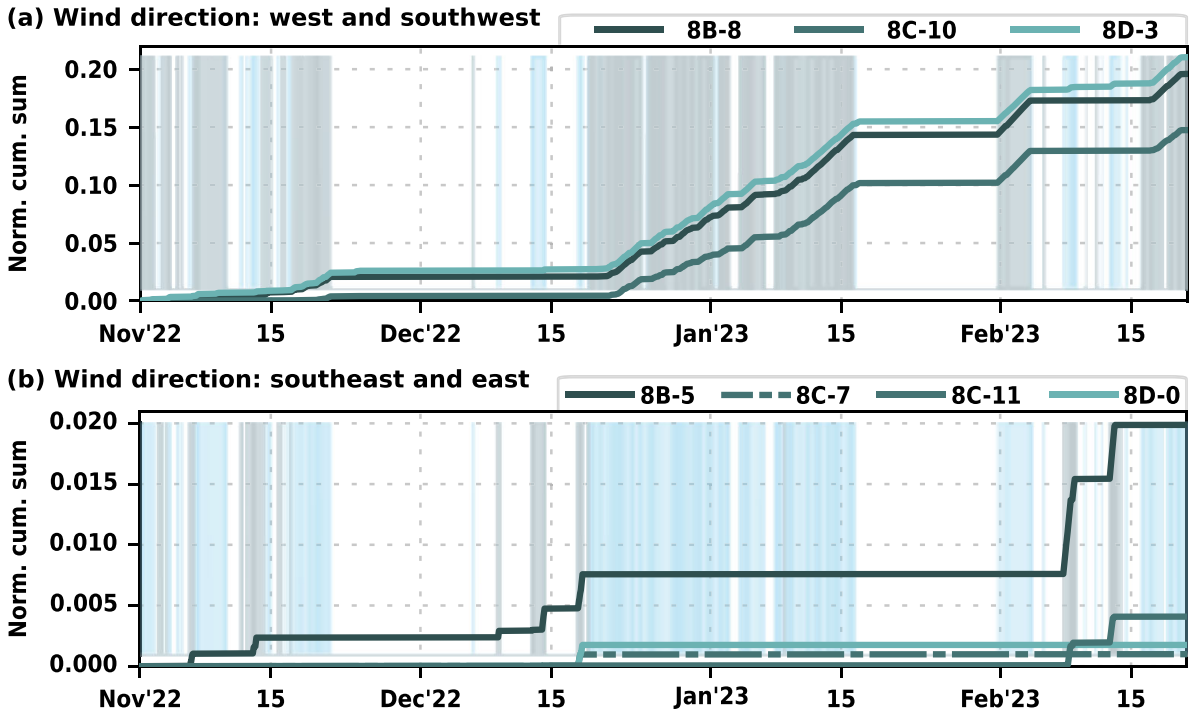

Figure 8. Normalized cumulative sum of clusters correlating with the higher noise reduced mode (grey and light blue areas) and wind from (a) west and southwest, and (b) southeast and east. The periods during which *WT 3* was running in higher noise reduction mode are highlighted in grey and in light blue. Grey indicates the times with the corresponding wind direction, which is displayed in the respective subplots.

Under west wind conditions, we observe minor contributions from the vertical components (Figs 10a to c). At station 8C, the amplitude of the vertical ground motion is approximately half that of the horizontal components (Fig. 10b). The transverse component at station 8B is dominant (Fig. 10a), while at stations 8C and 8D, the radial and transverse components contribute almost equally, and are correlated at 8C (Figs 10b and c).

During periods of east wind, the amplitude of the vertical component is slightly greater (Figs 10d to f). The vertical component of station 8B has an amplitude approximately two times lower than the horizontal components. In contrast to the west wind conditions, the radial and transverse components at station 8B are anticorrelated (Fig. 10d), whereas at station 8C, the transverse component is dominant. At station 8D, all three components have nearly equal amplitudes. However, the amplitude of the particle motion at station 8D is approximately four times larger during west wind conditions, while the amplitudes at the other two stations are comparable under both east and west wind conditions.

Our analysis of the particle motion indicates that the recorded wave field is a superposition of Love and Rayleigh waves. We observe a dominant influence of Love waves at a BAZ of 222° (8B) with a westerly wind (Fig. 10a). At a BAZ of 111° (8C) and a BAZ of 105° (8D), the influences of Love and Rayleigh waves are indistinguishable (Figs 10b and c). Furthermore, the radial and transverse components correlate at station 8C. An approximately equal contribution of Love and Rayleigh waves at these stations could explain the comparatively small vertical amplitude. For the same wind farm but a different measurement campaign, L. Gaßner & J. Ritter (2023b) found that at a station located between *WT 1* and *WT 2* (south of *WT 1*), a dependence of the vertical ground motion amplitude on both, the wind direction and the rotation rate. During partial and full-load operation, the vertical ground motion was largest under westerly winds. They measured maximum amplitudes around 1 μms^{-1}, whereas the vertical amplitudes we observe are significantly smaller. This difference may be due to the different station azimuth, which would support our explanation for the low vertical ground motion. Another explanation could be wavefield interferences, as the station analysed by L. Gaßner & J. Ritter (2023b) is located between two WTs.

During east wind conditions, the dominant influence of the Love wave excitation can be observed at recording station 8C (Fig. 10e), while in the recordings of station 8B, Love and Rayleigh waves are not distinguishable (Fig. 10d). Under these conditions, the motions of radial and transverse components are anticorrelated.

At station 8D, at a larger distance from *WT 3*, the low-amplitude WT emissions are superimposed by signals from additional sources, such as the nearby forest with shaking trees, that influence the particle motion induced by the WT. Furthermore, the signals recorded at station 8D are likely to be affected by wavefield interference related to the emissions of *WT 1* and *WT 2* (Fig A1). This makes it difficult to clearly determine how the recorded wavefield correlates with the emissions of *WT 3*. In this case, the separation of the clusters is likely driven by differences in signal amplitude. During west wind conditions, the observed stronger ground motion amplitudes could, for instance, be related to effects due to the topography of the Alb escarpment (Fig. 1; Section 3; F. Limberger *et al*. 2022; L. Gaßner *et al*. 2023) and the adjacent forest, which is exposed to west winds. In contrast, wind from the east blows across the more or less flat plain forming the top of the Alb. That flat topography, may explain the comparatively low amplitudes of the ground

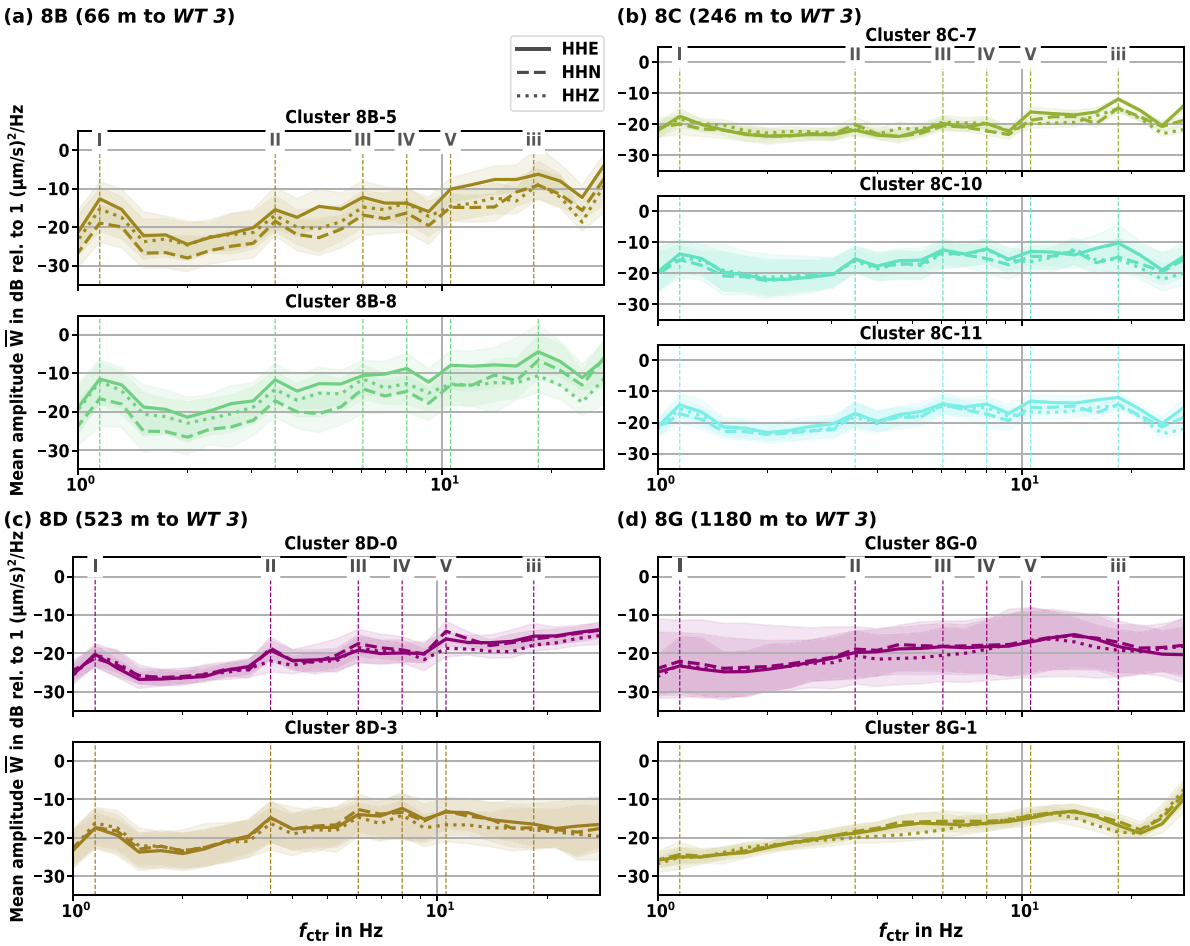

Figure 9. Mean amplitude of the time-averaged wavelet features $\overline{W}$ of each recording component (E, N, Z) within the individual clusters. The coloured zones mark the range between minimum and maximum amplitude. (a) to (c) show only the clusters that correlate with the higher noise reduced mode, while (d) shows all clusters of station 8G. The frequencies corresponding to the eigenmodes of the wind turbine and 32nd multiple of the blade pass frequency are highlighted by dashed lines (see Table 1).

or particle motions for winds from east (Fig. 10f; F.F. Lott *et al.* 2017).

Both, T. Neuffer *et al.* (2021) and L. Gaßner *et al.* (2023), analysed the direction-dependent radiation of seismic WT emissions, focusing on the WT eigenmodes. They suggest that the emission of surface waves depends on the frequency, either Love waves are radiated in the downwind direction (parallel to the wind direction), and Rayleigh waves in the cross-wind direction (perpendicular to the wind direction), or vice versa. In our study, recording stations 8C and 8D are approximately in line with the main wind direction. They are located upwind, when the wind is blowing from the west, or downwind, when the wind is blowing from the east (Fig. 1). No station was located in the crosswind direction in this experiment. Rather than analysing the eigenmodes of the WT, we examine the particle motion of the ground motion at the 32nd multiple of the BPF at 19.2 Hz. For this mode, the emitted waves are similar to the radiation pattern observed by T. Neuffer *et al.* (2021), and the particle motion suggests that Love waves are excited in the downwind direction. In the upwind direction, the particle motion recorded at 8C and 8D is composed of both, Love and Rayleigh waves. The observed wind-dependent variation in particle motion likely primarily contributes to the separation of clusters that are observed during the same rotation rate of the WT.

The analysis of the well-recorded 32nd multiple of the BPF could possibly be used to analyse the state-of-health (SOH) of the WT blades and generator. W. Hu *et al.* (2018) demonstrates the potential of seismic measurements to monitor the SOH of WTs. They investigate the acceleration of the tower at the frequency of a tower bending mode of 3.5 Hz, as well as the drivetrain vibrations between 2.5 and 12.5 Hz. Our data set is too short to assess the capability of our method for SOH monitoring, but future studies with longer time-series may reveal its potential in this context.

4.3 Capabilities and limitations of the method

The time-averaged wavelet features, $\overline{W}$, capture all relevant frequency peaks associated with seismic WT emissions for the analysed type of WT. They are sparse, stable and easy to tune, making them well-suited for clustering tasks. In contrast, PSDs often contain strong fluctuations and irregularities, which complicate the use of clustering algorithms without further pre-processing, such as dimension reduction. Owing to their sparsity and frequency selectivity, $\overline{W}$ can be easily adapted to the seismic emission of other WT types, making them a promising feature representation, for example, for SOH monitoring.

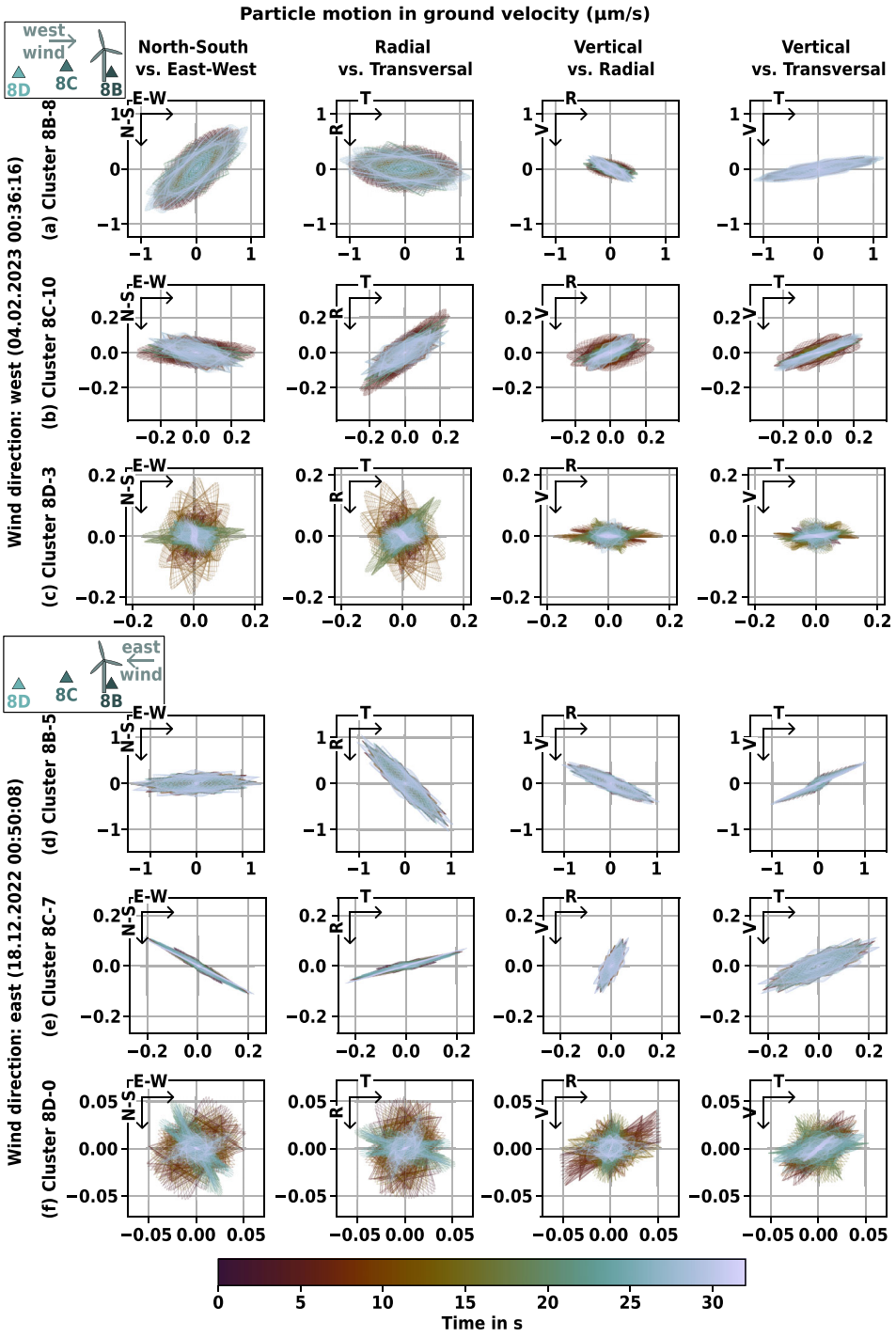

Figure 10. Particle motion of ground motion velocity for two time windows corresponding to clusters associated with the higher noise-reduced mode, under west-wind conditions (a to c) and east-wind conditions (d to f), respectively. The duration of the displayed time windows corresponds to half the length of the sliding window used as input for the wavelet transform, reflecting the 50 per cent overlap. The signal has been bandpass filtered to 19.2 ± 0.25 Hz. Station 8B is located to the northeast [222° backazimuth (BAZ)], 8C to the west-northwest (111° BAZ) and 8D to the west of *WT 3* (95° BAZ).

With $\overline{W}$ as input for HDBSCAN, the clustering algorithm effectively separates clusters that correlate with rotation rate and wind direction. We associated the latter with the differences in particle motion related to the complex behaviour of the WT foundation. We choose HDBSCAN because it does not require a pre-defined number of clusters. This proves to be advantageous, as we observed not only clusters correlating with the WTs' rotation rate, which could have allowed for a pre-defined number of clusters, but also multiple clusters associated with the same rotation rate. These clusters correlate with the wind direction. The hierarchical nature of HDBSCAN allows clusters of different sizes to be identified, as is the case in our study (Table D1).

However, clustering $\overline{W}$ with HDBSCAN is very sensitive to the choice of the hyperparameters (Fig. C1) and tuning them requires operational data to ensure that the number of 'dominant' clusters matches or exceeds the number of different operational stages of the WT (Appendix C). Consequently, knowledge of WT operational stages is necessary. Without reference data, it is difficult to extract the rotation rate, as the corresponding frequency peaks are not easily automatically identifiable (e.g. L. Gaßner & J. Ritter 2023b).

For verification and discussion of clustering results, a reference data set containing operational data is thus beneficial. However, it may be sufficient to use only a short time window of operational data as long as it contains all possible rotation rates and wind directions. This assumption should be further evaluated and validated to better constrain the multiparameter problem. Additionally, it could be interesting to analyse $\overline{W}$ in a supervised manner, using identifying stable seismic WT emissions properties.

Another limitation of the approach is the relatively high number of time windows that the clustering algorithm labels as noise (Table D1), which may be due to the still relatively high dimensionality of $\overline{W}$. As a density-based clustering algorithm, HDBSCAN can struggle in higher dimensional spaces, where sparse regions become more common. More generally, a high-dimensional feature space can cause difficulties for any clustering algorithm because of the curse of dimensionality (R.E. Bellman 1961). We tested prior dimensional reduction in M.A. Gärtner (2024), but direct clustering of $\overline{W}$ yielded the most promising results. As we use short input time windows (128 s), the analysed three months of recordings provide us a data set of approximately 150 000 time windows for each recording station. Hence, even with the high number of time windows declared as noise, we can still ensure a statistically robust data set.

Nevertheless, even with these limitations, our method represents a promising and novel approach for analysing seismic WT emissions and is expected to benefit from further validation at additional wind farms.

5 SUMMARY AND CONCLUSION

In this study, we apply a data-driven approach to investigate seismic emissions from WTs. We analyse three-component seismograms from four broad-band stations. The workflow involves applying a wavelet transform to sliding time windows, followed by the calculation of the temporal average, and subsequently clustering the resulting time-averaged wavelet features $\overline{W}$ using HDBSCAN. To our knowledge, HDBSCAN has previously been applied only once to continuous seismic time-series, namely by C. Chai *et al.* (2025), who clustered Fourier spectra after reducing their dimensionality with UMAP. In contrast, our approach enables the direct use of $\overline{W}$ as input to HDBSCAN, eliminating the need for an additional dimensionality reduction step and providing a representation that is inherently well-suited for clustering. Although our approach may still result in a comparatively large number of time windows being labelled as noise, which is likely a consequence of the remaining dimensionality of $\overline{W}$, it nevertheless offers a streamlined workflow and delivers clusters that are closely related to the underlying signal content, as we were able to confirm through particle motion analysis.

Our findings demonstrate that the resulting clusters are sensitive not only to the rotation rate of the WT but also to the wind direction. The influence of the WT operational stage is reflected in $\overline{W}$ of each cluster. Furthermore, we show how wind direction influences the emitted wavefield of the WT, enabling the separation of clusters even when the rotation rate is similar. Our interpretation suggests that the different clusters associated with the different wind directions can be attributed to varying radiation patterns generated by the coupling of the WT foundation and the surrounding subsurface. In particular, we analysed the radiation pattern associated with the 32nd multiple of the BPF. During east wind conditions, we observe the radiation of Love waves in the downwind direction, and a superposition of Love and Rayleigh waves at a BAZ of 222°. Under west wind conditions, the pattern is reversed.

These findings illustrate that our method is sensitive to the radiation pattern emitted by WTs and can thus successfully extract meaningful patterns related to the WT operation and the prevailing wind direction. Further insight might be gained by applying this approach to seismic data recorded at other wind farms and under varying wind directions. It might be worth testing our approach with long time-series to evaluate its suitability for SOH monitoring. Moreover, the application of HDBSCAN has proven to be useful in extracting the relevant signal families from seismic data with varying occurrence rates. Thus, we encourage applying HDBSCAN as a well-adapted tool to explore continuous, noisy seismograms with strong class imbalances.

ACKNOWLEDGMENTS

We thank the editor and reviewers for their constructive feedback, which improved the manuscript.

We thank the local authorities of the municipality Kuchen for their support, as well as the Stadtwerke Schwäbisch Hall and the KWA Contracting AG for providing access to *WT 3* and the WT operating data. Felix Bögelspacher, Leon Merkel and Rune Helk helped with the installation and maintenance of the ground motion sensors.

We are grateful to Yvonne Fröhlich for providing the wind turbine symbol and for her support in creating the map shown in Fig. 1.

We thank Dario Eickhoff and Mathieu Herbelot for proofreading the manuscript and for their valuable feedback and suggestions.

The Inter-Wind project, including the data acquisition, was funded by the German Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag (grants 03EE2023A-D, 2019/12/13).

ISTerre is part of Labex OSUG@2020 (ANR10 LABX56).

AUTHOR CONTRIBUTIONS

Marie A. Gärtner (Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing—original draft), René Steinmann (Conceptualization, Methodology, Software, Writing—review and editing), Laura Gaßner (Conceptualization, Data curation, Supervision, Writing—review and editing), and Joachim R. R. Ritter (Conceptualization, Funding acquisition, Project administration, Supervision, Writing—review and editing)

DATA AVAILABILITY

The seismological data of the project Inter-Wind is published under the network code 4C available at http://ws.gpi.kit.edu/fdsnws/. Part of the dataset of the Inter-Wind project is published by J. Ritter & L. Gassner 2022; however, the stations used in this study are not individually assigned a DOI. Operational data of the wind turbines is the property of the wind turbine owner, and a clearance needs to be asked for individually. The data was processed in Python. We pre-processed the data with ObsPy (M. Beyreuther *et al*. 2010; T. Megies *et al*. 2011; L. Krischer *et al*. 2015), computed the time-averaged wavelet feature with the python package ScatSeisNet (L.S. Seydoux *et al*. 2025) and performed the clustering with HDBSCAN (L. McInnes *et al*. 2017). The python scripts we used for the analysis are available at http://github.com/MarieAGaertner/SeisWTanalysis. The workflow and parameter testing are described in more details in M.A. Gärtner (2024). The figures were mainly visualized using Matplotlib (J.D. Hunter 2007) with the Scientific colour maps (F. Crameri 2023). The map in Fig. 1 was generated with PyGMT (D. Tian *et al*. 2025), using a background map from OpenStreetMap contributors (2017).

REFERENCES

Andén, J. & Mallat, S., 2014. Deep scattering spectrum, *IEEE Trans. Signal Proc.,* **62**(16), 4114–4128.

Barkaoui, S. *et al.*, 2021. Anatomy of Continuous Mars SEIS and Pressure Data from Unsupervised Learning, *Bull. seism. Soc. Am.,* **111**(6), 2964–2981.

Bellman, R.E., 1961. *Adaptive Control Processes,* Princeton Univ. Press.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. ObsPy: a python toolbox for seismology, *Seism. Res. Lett.,* **81**(3), 530–533.

Blumendeller, E., Gaßner, L., Müller, F.J., Pohl, J., Hübner, G., Ritter, J. & Cheng, P.W., 2023. Quantification of amplitude modulation of wind turbine emissions from acoustic and ground motion recordings, *Acta Acust.,* **7,** 55. doi:10.1051/aacus/2023047.

Bruna, J. & Mallat, S., 2013. Invariant scattering convolution networks, *IEEE Trans. Pattern Anal. Mach. Intell.,* **35**(8), 1872–1886.

Campello, R.J.G.B., Moulavi, D. & Sander, J., 2013. Density-based clustering based on Hierarchical density estimates, in Advances in Knowledge Discovery and Data Mining, Vol. 7819, pp. 160–172, eds Hutchison, D. *et al*, Springer Berlin Heidelberg, Berlin, Heidelberg. .

Chai, C., Marcillo, O., Maceira, M., Park, J., Arrowsmith, S., Thomas, J.O. & Cunningham, J., 2025. Exploring continuous seismic data at an industry facility Using Unsupervised Machine Learning, *The Seismic Record,* **5**(1), 64–72.

Charléty, J., Cuenot, N., Dorbath, L., Dorbath, C., Haessler, H. & Frogneux, M., 2007. Large earthquakes during hydraulic stimulations at the geothermal site of Soultz-sous-Forêts, *Int. J. Rock Mech. Mining Sci.,* **44**(8), 1091–1105.

Crameri, F., 2023. *Scientific Colour Maps*[Software], Zenodo

Diaferia, G. *et al.*, 2025. Seismic noise characterization for the BuddusÃ²–Ala dei Sardi wind park (Sardinia, Italy) and its impact on the Einstein Telescope candidate site, *Solid Earth,* **16**(6), 441–456.

EEG 2023, 2024. Gesetz für den Ausbau Erneuerbarer Energien. Available at: https://www.gesetze-im-internet.de/eeg_2014/BJNR106610014.html

Essing, D. & Poli, P., 2024. Unraveling earthquake clusters composing the 2014 Alto Tiberina earthquake swarm via unsupervised learning, *J. geophys. Res.: Solid Earth,* **129**(1), e2022JB026237.

Ezugwu, A.E., Ikotun, A.M., Oyelade, O.O., Abualigah, L., Agushaka, J.O., Eke, C.I. & Akinyelu, A.A., 2022. A comprehensive survey of clustering algorithms: state-of-the-art machine learning applications, taxonomy, challenges, and future research prospects, *Eng. Appl. Artif. Intell.,* **110,** 104743. doi:10.1016/j.engappai.2022.104743.

Gärtner, M.A., 2024. Data-driven pattern recognition of seismic wind turbine emissions with machine learning, Master's thesis, Karlsruhe Institute of Technology (KIT), Geophysical Institute (GPI).

Gaßner, L. & Ritter, J., 2023a. Wind turbine emissions: Interdisciplinary analysis and mitigation approaches – Project Inter-Wind. Description of data sets "Inter-Wind" and "Inter-Wind (recorder log Files)", Tech. rep., Karlsruhe Institute of Technology (KIT), Geophysical Institute (GPI).

Gaßner, L. & Ritter, J., 2023b. Ground motion emissions due to wind turbines: observations, acoustic coupling, and attenuation relationships, *Solid Earth,* **14**(7), 785–803.

Gaßner, L. *et al.*, 2022. Joint analysis of resident complaints, meteorological, acoustic, and ground motion data to establish a robust annoyance evaluation of wind turbine emissions, *Renew. Energ.,* **188,** 1072–1093.

Gaßner, L., Gärtner, M.A. & Ritter, J., 2023. Simulation of ground motion emissions from wind turbines in low mountain ranges: implications for amplitude decay prediction, *J. Seismol.,* **27,** 933–952.

Hensch, M., Dahm, T., Ritter, J., Heimann, S., Schmidt, B., Stange, S. & Lehmann, K., 2019. Deep low-frequency earthquakes reveal ongoing magmatic recharge beneath Laacher See Volcano (Eifel, Germany), *Geophys. J. Int.,* **216**(3), 2025–2036.

Hu, W., Barthelmie, R.J., Letson, F. & Pryor, S.C., 2018. A new seismic-based monitoring approach for wind turbines, *Wind Energy,* **22**(4), 473–486.

Hunter, J.D., 2007. Matplotlib: a 2D graphics environment, *Comput. Sci. Eng.,* **9**(3), 90–95.

Jain, A.K., Murty, M.N. & Flynn, P.J., 1999. Data clustering: a review, *ACM Comput. Surv. (CSUR),* **31**(3), 264–323.

Johnson, C.W., Ben-Zion, Y., Meng, H. & Vernon, F., 2020. Identifying different classes of seismic noise signals using unsupervised learning, *Geophys. Res. Lett.,* **47**(15), e2020GL088353. doi:10.1029/2020GL088353.

Köhler, A., Ohrnberger, M. & Scherbaum, F., 2010. Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps, *Geophys. J. Int.,* **182**(3), 1619–1630.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. Obspy: a bridge for seismology into the scientific python ecosystem, *Comput. Sci. Disc.,* **8**(1), 014003. doi:10.1088/1749-4699/8/1/014003.

Lerbs, N., Zieger, T., Ritter, J. & Korn, M., 2020. Wind turbine induced seismic signals: the large-scale SMARTIE1 experiment and a concept to define protection radii for recording stations, *Near Surf. Geophys.,* **18**(5), 467–482.

Limberger, F., Lindenfeld, M., Deckert, H. & Rümpker, G., 2021. Seismic radiation from wind turbines: observations and analytical modeling of frequency-dependent amplitude decays, *Solid Earth,* **12**(8), 1851–1864.

Limberger, F., Rümpker, G., Lindenfeld, M. & Deckert, H., 2022. Development of a numerical modelling method to predict the seismic signals generated by wind farms, *Sci. Rep.,* **12**(1), 15516. doi:10.1038/s41598-022-19799-w.

Limberger, F., Rümpker, G., Lindenfeld, M. & Deckert, H., 2023. The impact of seismic noise produced by wind turbines on seismic borehole measurements, *Solid Earth,* **14**(8), 859–869.

Lott, F.F., Ritter, J.R.R., Al-Qaryouti, M. & Corsmeier, U., 2017. On the analysis of wind-induced noise in seismological recordings: approaches to present wind-induced noise as a function of wind speed and wind direction, *Pure appl. Geophys.,* **174**(3), 1453–1470.

Makus, P., Sens-Schönfelder, C., Illien, L., Walter, T.R., Yates, A. & Tilmann, F., 2023. Deciphering the whisper of volcanoes: monitoring velocity changes at Kamchatka's Klyuchevskoy group with fluctuating noise fields, *J. geophys. Res.: Solid Earth,* **128**(4), e2022JB025738.

McInnes, L. & Healy, J., 2017. Accelerated Hierarchical density based clustering, in *2017 IEEE International Conference on Data Mining Workshops (ICDMW),* pp. 33–42, IEEE, New Orleans, LA.

McInnes, L., Healy, J. & Astels, S., 2017. HDBSCAN: Hierarchical density based clustering, *J. Open Source Softw.,* **2**(11), 205. doi:10.21105/joss.00205.

Megies, T., Beyreuther, M., Barsch, R., Krischer, L. & Wassermann, J., 2011. ObsPy–what can it do for data centers and observatories?, *Ann. Geophys.,* **54**(1), 47–-58

Moreau, L., Seydoux, L., Weiss, J. & Campillo, M., 2023. Analysis of microseismicity in sea ice with deep learning and Bayesian inference: application to high-resolution thickness monitoring, *The Cryosphere,* **17**(3), 1327–1341.

Morel, R., Rochette, G., Leonarduzzi, R., Bouchaud, J.-P. & Mallat, S., 2023. Scale dependencies and self-similar models with wavelet scattering spectra, *Appl. Comput. Harmonic Anal.,* **75,** 101724. doi:10.1016/j.acha.2024.101724.

Müller, F.J., Blumendeller, E., Gaßner, L., Cheng, P.W., Ritter, J., Pohl, J. & Hübner, G., 2025. Wind turbine noise annoyance–an interdisciplinary three-year field study, *Environ. Int.,* **202**(10961), 4. doi:10.1016/j.envint.2025.109614.

Nagel, S., Zieger, T., Luhmann, B., Knödel, P., Ritter, J. & Ummenhofer, T., 2021. Ground motions induced by wind turbines, *Civil Eng. Design,* **3**(3), 73–86.

Neuffer, T. & Kremers, S., 2017. How wind turbines affect the performance of seismic monitoring stations and networks, *Geophys. J. Int.,* **211**(3), 1319–1327.

Neuffer, T., Kremers, S., Meckbach, P. & Mistler, M., 2021. Characterization of the seismic wave field radiated by a wind turbine, *J. Seismol.,* **25**(3), 825–844.

OpenStreetMap contributors, 2017. Planet dump retrieved from https://planet.osm.org, https://www.openstreetmap.org.

Rimpot, J. *et al.*, 2025. Self-supervised learning of seismological data reveals new eruptive sequences at the Mayotte submarine volcano, *Geophys. J. Int.,* **240**(1), 1–12.

Ritter, J. & Gassner, L. 2022. Inter-Wind. GFZ Data Services. Dataset/Seismic Network. doi:10.14470/9P982225.

Rodríguez, Á.B., Balestriero, R., De Angelis, S., Benitez, M.C., Zuccarello, L., Baraniuk, R., Ibanez, J.M. & De Hoop, M.V., 2022. Recurrent scattering network detects metastable behavior in Polyphonic Seismo-Volcanic Signals for Volcano Eruption Forecasting, *IEEE Trans. Geosci. Remote Sens.,* **60,** 1–23.

Saccorotti, G., Piccinini, D., Cauchie, L. & Fiori, I., 2011. Seismic noise by wind farms: a case study from the virgo gravitational wave observatory, Italy, *Bull. seism. Soc. Am.,* **101**(2), 568–578.

Seydoux, L., Balestriero, R., Poli, P., de Hoop, M., Campillo, M. & Baraniuk, R., 2020. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning, *Nat. Commun.,* **11**(1), 3972. doi:10.1038/s41467-020-17841-x.

Seydoux, L.S., Steinmann, R., Gärtner, M., Tong, F., Esfahani, R., Mouaoued, S., Shapiro, N. & Campillo, M., 2025. *Scatseisnet, a Scattering Network for Seismic Data Analysis*[Computational notebook], Zenodo

Shi, P., Seydoux, L. & Poli, P., 2021. Unsupervised learning of seismic wavefield features: clustering continuous array seismic data during the 2009 L'Aquila earthquake, *J. geophys. Res.: Solid Earth,* **126**(1), e2020JB020506.

Stammler, K. & Ceranna, L., 2016. Influence of wind turbines on seismic records of the Gräfenberg Array, *Seism. Res. Lett.,* **87**(5), 1075–1081.

Steinmann, R., Seydoux, L. & Campillo, M., 2022a. AI-based unmixing of medium and source signatures from seismograms: Ground freezing patterns, *Geophys. Res. Lett.,* **49**(15), e2022GL098854. doi:10.1029/2022gl098854.

Steinmann, R., Seydoux, L., Beaucé, É. & Campillo, M., 2022b. Hierarchical exploration of continuous seismograms with unsupervised learning, *J. geophys. Res.: Solid Earth,* **127**(1), e2021JB022455. 10.1029/2021JB022455.

Steinmann, R., Seydoux, L., Journeau, C., Shapiro, N.M. & Campillo, M., 2024. Machine learning analysis of seismograms reveals a continuous plumbing system evolution beneath the Klyuchevskoy volcano in Kamchatka, Russia, *J. geophys. Res.: Solid Earth,* **129**(3), e2023JB027167.

Tian, D. *et al.*, 2025. *PyGMT: A Python interface for the Generic Mapping Tools*[Software], Zenodo

Yates, A., Caudron, C., Lesage, P., Mordret, A., Lecocq, T. & Soubestre, J., 2023. Assessing similarity in continuous seismic cross-correlation functions using hierarchical clustering: application to Ruapehu and Piton de la Fournaise volcanoes, *Geophys. J. Int.,* **233**(1), 472–489.

Zali, Z., Mousavi, S.M., Ohrnberger, M., Eibl, E.P. & Cotton, F., 2024. Tremor clustering reveals pre-eruptive signals and evolution of the 2021 Geldingadalir eruption of the Fagradalsfjall Fires, Iceland, *Commun. Earth Environ.,* **5**(1), 1. doi:10.1038/s43247-023-01166-w.

Zieger, T. & Ritter, J.R.R., 2018. Influence of wind turbines on seismic stations in the Upper Rhine Graben, SW Germany, *J. Seismol.,* **22**(1), 105–122.

APPENDIX A: INFLUENCE OF WIND TURBINES *WT 1* AND *WT 2* COMPARED TO *WT 3*

Fig. A1 shows the contributions of *WT 1* and *WT 2* compared to those of *WT 3* on the seismic recordings (Fig. 1). In this section, we set the focus on the 32nd multiple of the blade passing frequency (BPF), corresponding to 20 Hz in Figs A1(a) and (c), and 19.2 Hz in Fig. A1(b). During periods when all three WTs operate at the same rotation rate (Fig. A1a), pronounced spectral peaks at 20 Hz are observed across all three stations. The amplitude of this peak is highest in the recordings of station 8B, while the recordings of stations 8C and 8D lead to nearly identical, yet slightly lower, amplitudes.

During times when only *WT 3* is in operation (Fig. A1b), the PSD amplitudes at 19.2 Hz at recording stations 8B and 8C are comparable to those observed in the time window in Fig. A1(a). However, the amplitude at station 8D is notably lower.

When *WT 1* and *WT 2* are operating without *WT 3* (Fig. A1c), the spectral peak of 20 Hz is reduced in the recordings of stations 8B and 8C. In contrast, this spectral peak becomes more pronounced at station 8D compared to the time window in Fig. A1(b); nevertheless, the absolute amplitude of the peak at station 8D is the same in both Figs A1(b) and (c).

These observations suggest that *WT 1* and *WT 2* have only a minor influence on the recordings of stations 8B and 8C. In comparison, the recordings of station 8D appear to be more sensitive to the operation of *WT 1* and *WT 2* than to *WT 3*, at least under conditions when *WT 1* and *WT 2* are operated at the same rotation rate.

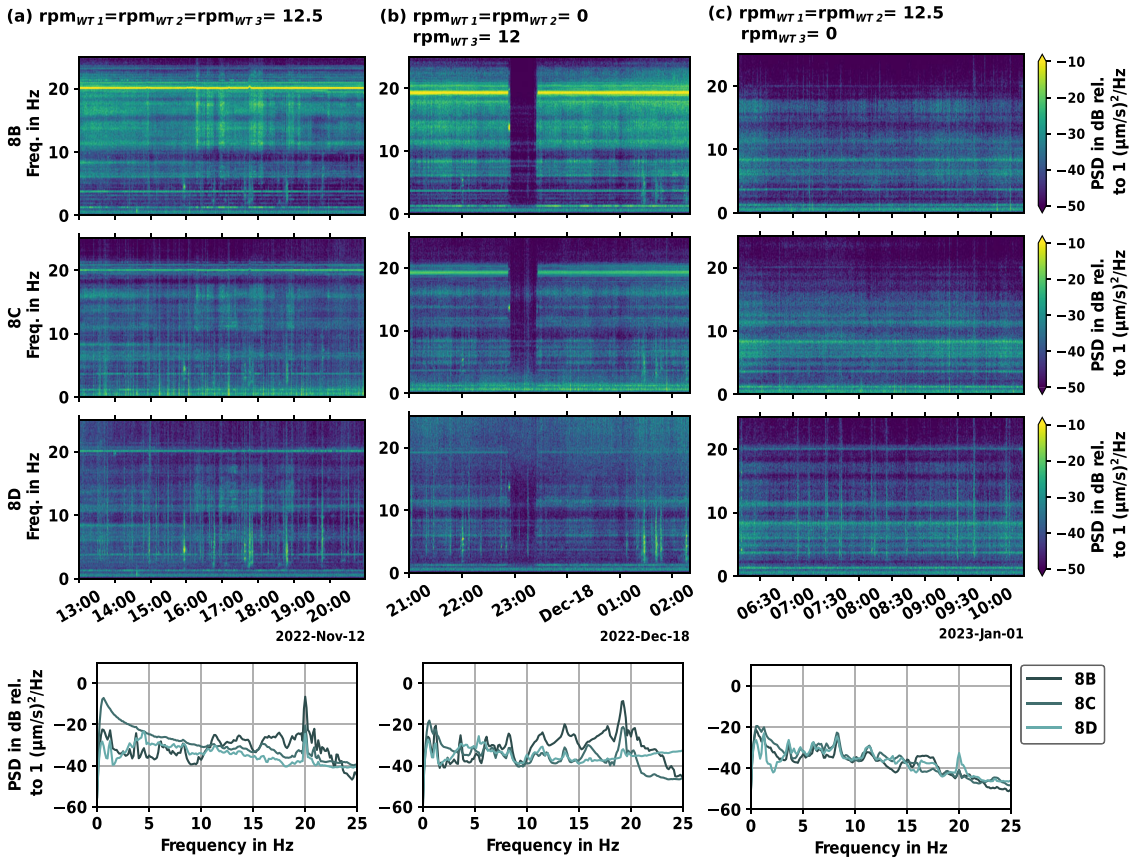

Figure A1. Comparison of the impact of the emissions of *WT 1* and *WT 2* on the recorded seismic data compared to those of *WT 3* (Fig. 1). The columns (a–c) show three time windows during different operational states of the wind farm: (a) all three WTs are operating at full load (12 rpm), (b) only *WT 3* is running, (c) *WT 1* and *WT 2* are operating at full load while *WT 3* is not in operation. The first three rows display the spectrograms of the vertical component recorded at stations 8B, 8C and 8D, respectively. The fourth row shows the corresponding power spectral densities (PSD) for each time window.

APPENDIX B: COMPARISON OF THE TIME-AVERAGED WAVELET TRANSFORMATION WITH THE PSD

The spectral peaks associated with the WT eigenmodes and the multiples of the BPF in Figs 4 and 5 appear much sharper in the spectrograms than in the time-averaged wavelet features $\overline{W}$. This is partly due to the shorter 10 s time window over which each spectrogram is computed, but primarily because the spectrograms have a substantially higher frequency resolution. Each spectrogram contains 321 frequency bins per seismometer component, compared with only 25 bins of $\overline{W}$.

In Fig. B1, we present the vertical ground motion of three 128 s time windows (left panels), excerpted from Fig. 4, and matching the input window length used in the wavelet transform. The corresponding power spectral densities (PSDs; blue) are compared with $\overline{W}$ (black) in the right panels.

Apart from subtracting the mean, removing the linear trend, converting the data to ground-motion velocity and downsampling to 64 Hz, the signals shown in the left panels are not further processed. Ocean microseisms generate the dominant low-frequency fluctuations.

$\overline{W}$ are considerably sparser than the PSDs. Nevertheless, $\overline{W}$ still follows the overall spectral shape without the high-frequency fluctuations that would complicate clustering. Unlike the Fourier transform, whose frequency bins are linearly spaced, the wavelet filter bank is logarithmically spaced. This enables us to tune the filter bank of the wavelet transformation to capture each of the WT eigenmodes and the 32nd multiples of the BPF of all operational stages of *WT 3*, with the exception of the two noise-reduced modes, with an individual wavelet.

The time-averaged wavelet features $\overline{W}$ are therefore advantageous for clustering: they exhibit a naturally clusterable dimensionality and do not retain the rapid spectral fluctuations present in the PSD.

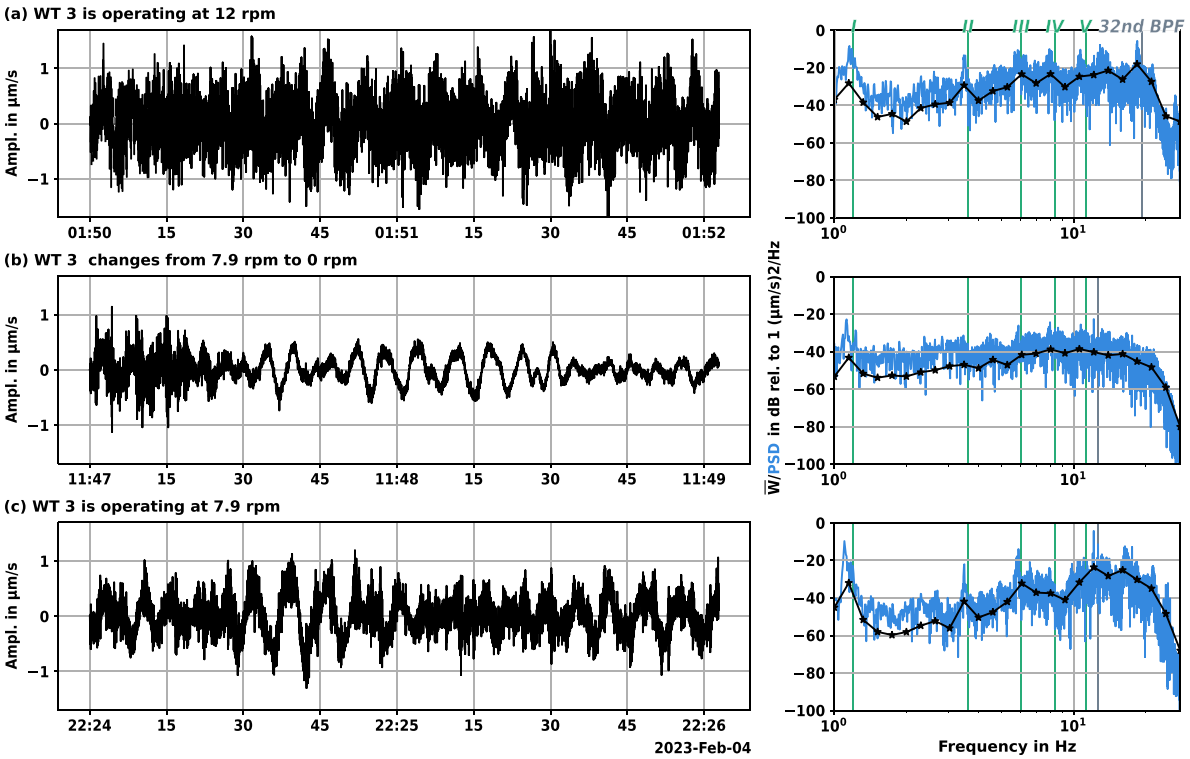

Figure B1. Zoom-in on three 128 s time windows from Fig. 4 in which *WT 3* operated in (a) full load (12 rpm), (b) transition from partial load (7.9 rpm) to shut-down and (c) partial load (7.9 rpm). The left panels show the vertical ground motion, and the right panels compare the time-averaged wavelet features $\overline{W}$ with the power spectral densities (PSDs). Vertical lines mark the WT eigenmodes (green) and the 32nd BPF multiples (grey) at 12.64 and 19.20 Hz.

APPENDIX C: HDBSCAN HYPERPARAMETER TUNING

Fig. C1 shows our results from testing the HDBSCAN hyperparameters *min_samples* and *min_cluster_size*. The hyperparameter *min_samples* defines the number of samples that is used to estimate the local density; the larger the number, the more conservative the clustering result. The hyperparameter *min_cluster_size* controls the minimum number of samples a cluster can have before it will be declared as noise. We utilize the time-averaged wavelet features $\overline{W}$ of recording station 8B as input for the parameter testing. The histograms show the number of samples populating each cluster of the tested HDBSCAN parameters. The number in the upper left corner of each panel describes the number of samples populating the smallest cluster. We expect at least four dominant clusters (Section 2.1). We define a cluster as dominant when its population is larger than the median cluster population marked by the red line. The panels of all parameter sets that fulfill this condition are framed by a dashed red line. Since multiple parameter sets satisfy this condition, we chose the parameter set that leads to the largest population of the smallest cluster, to avoid overly small or fragmented clusters. Based on this tuning procedure, we set *min_samples* to 50 and *min_cluster_size* to 100 (panel framed by a solid red line).

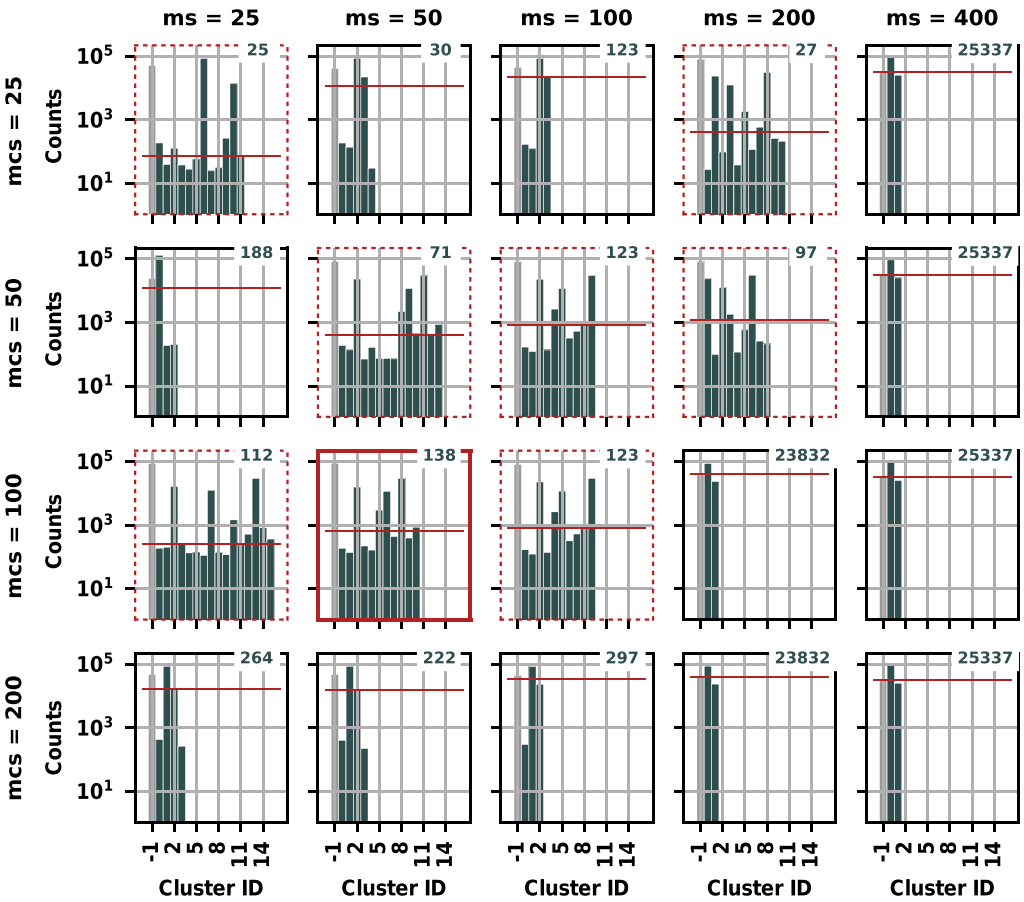

Figure C1. Results of the HDBSCAN hyperparameter testing. The histograms show the number of samples populating each cluster of the tested HDBSCAN parameters. The clusters with ID -1 include all samples, which are declared as noise during the clustering, and the respective bars are coloured in grey. The rows show the variation of the HDBSCAN hyperparameter *min_cluster_size* (mcs) and the columns of *min_samples* (ms). The horizontal line marks the median value of each cluster population. The number in the upper right corner of each panel describes the number of samples populating the smallest cluster.

APPENDIX D: OVERVIEW OF THE HDBSCAN CLUSTERING DISTRIBUTIONS

In Table D1 we summarize the global population, how many time windows were clustered and how many were declared as noise, and the population of each cluster. In summary, more than 50 per cent of the time windows are declared as noise for stations 8B, 8C and 8D. From the clustered time windows, each data set has two to three heavily populated clusters (> 18.7 per cent): for station 8B these are clusters 8B-2, 8B-6 and 8B-8, for station 8C clusters 8C-2, 8C-4 and 8C-10, for station 8D clusters 8D-3 and 8D-5. Recording station 8G has only two clusters of which cluster 8G-0 is the significantly higher populated one. The population of all other clusters lies between 0.21 per cent and 1.39 per cent with one cluster having a population of 4.77 per cent.

Figs D1 to D4 show the clustering distribution in relation to the rotation rate of *WT 3*, the wind speed and wind direction, and days of the week, as well as hours of the day in UTC.

Fig. D1 presents the HDBSCAN clustering distribution of recording station 8B. Clusters *8B-0* to *8B-3* are associated with periods during which *WT 3* was not in operation. Among them, clusters *8B-0* and *8B-3* correspond to similar wind speeds but differ with respect to wind direction and the weekdays during which they are active. In contrast, clusters *8B-1* and *8B-2* are distinguished by wind speed: *8B-1* is predominantly active during high wind speeds exceeding 9 m s^{-1}, whereas *8B-2* is active when wind speeds are below 9 m s^{-1}. Notably, *8B-2* is the largest cluster within this group. Cluster *8B-6* is the only one that correlates with periods in which *WT 3* operated under partial load conditions, while cluster *8B-4* is the sole cluster associated with full-load operation. The lower noise-reduced operational mode is captured by clusters *8B-7*, *8B-9* and *8B-10*. Cluster *8B-7* is characterized by a wind direction from south and southeast, whereas clusters *8B-9* and *8B-10* are correlated with a dominant westerly wind direction. They are differentiated by their associated wind speed ranges. Clusters *8B-5* and *8B-8* are discussed in detail in Section 4.

Fig. D2 presents the HDBSCAN clustering distribution of recording station 8C. Among the twelve identified clusters, Clusters *8C-0*, *8C-1* and *8C-3* to *8C-5* are active during periods when *WT 3* was not in operation. Clusters *8C-0*, *8C-1* and *8C-5* are populated during similar wind speed conditions but differ in the dominant wind directions. Clusters *8C-3* and *8C-4* are distinguished by high versus low wind speed regimes. Similar to the clustering results of recording station 8B (Fig. D1), only a single cluster, *8C-2*, correlates with times when the WT operated in partial load. In contrast, no individual cluster is associated with full load. Clusters *8C-6*, *8C-8* and *8C-9* are associated with the lower noise-reduced operational mode. Cluster *8C-6* is active

Table D1. Global and cluster population of the analysed data sets. The global population is separated in time windows declared as noise and clustered time windows. The cluster population lists the population in percentage of each cluster relative to all clustered time windows in the data set.

Global population:							
Noise windows:							
8B	58.42 per cent	8C	70.65 per cent	8D	66.79 per cent	8G	45.70 per cent
Clustered windows:							
8B	41.58 per cent	8C	29.35 per cent	8D	33.21 per cent	8G	54.30 per cent

Cluster population (per cent of clustered windows):							
Station 8B:		Station 8C:		Station 8D:		Station 8G:	
8B-0	0.30 per cent	8C-0	0.56 per cent	8D-0	0.53 per cent	8G-0	99.73 per cent
8B-1	0.22 per cent	8C-1	0.40 per cent	8D-1	0.21 per cent	8G-1	0.27 per cent
8B-2	25.56 per cent	8C-2	21.55 per cent	8D-2	0.71 per cent	–	–
8B-3	0.36 per cent	8C-3	0.27 per cent	8D-3	63.28 per cent	–	–
8B-4	0.27 per cent	8C-4	23.42 per cent	8D-4	0.33 per cent	–	–
8B-5	4.77 per cent	8C-5	1.01 per cent	8D-5	34.47 per cent	–	–
8B-6	18.72 per cent	8C-6	0.30 per cent	8D-6	0.47 per cent	–	–
8B-7	0.70 per cent	8C-7	0.34 per cent	–	–	–	–
8B-8	47.08 per cent	8C-8	0.30 per cent	–	–	–	–
8B-9	0.64 per cent	8C-9	0.29 per cent	–	–	–	–
8B-10	1.38 per cent	8C-10	50.17 per cent	–	–	–	–
		8C-11	1.39 per cent	–	–	–	–

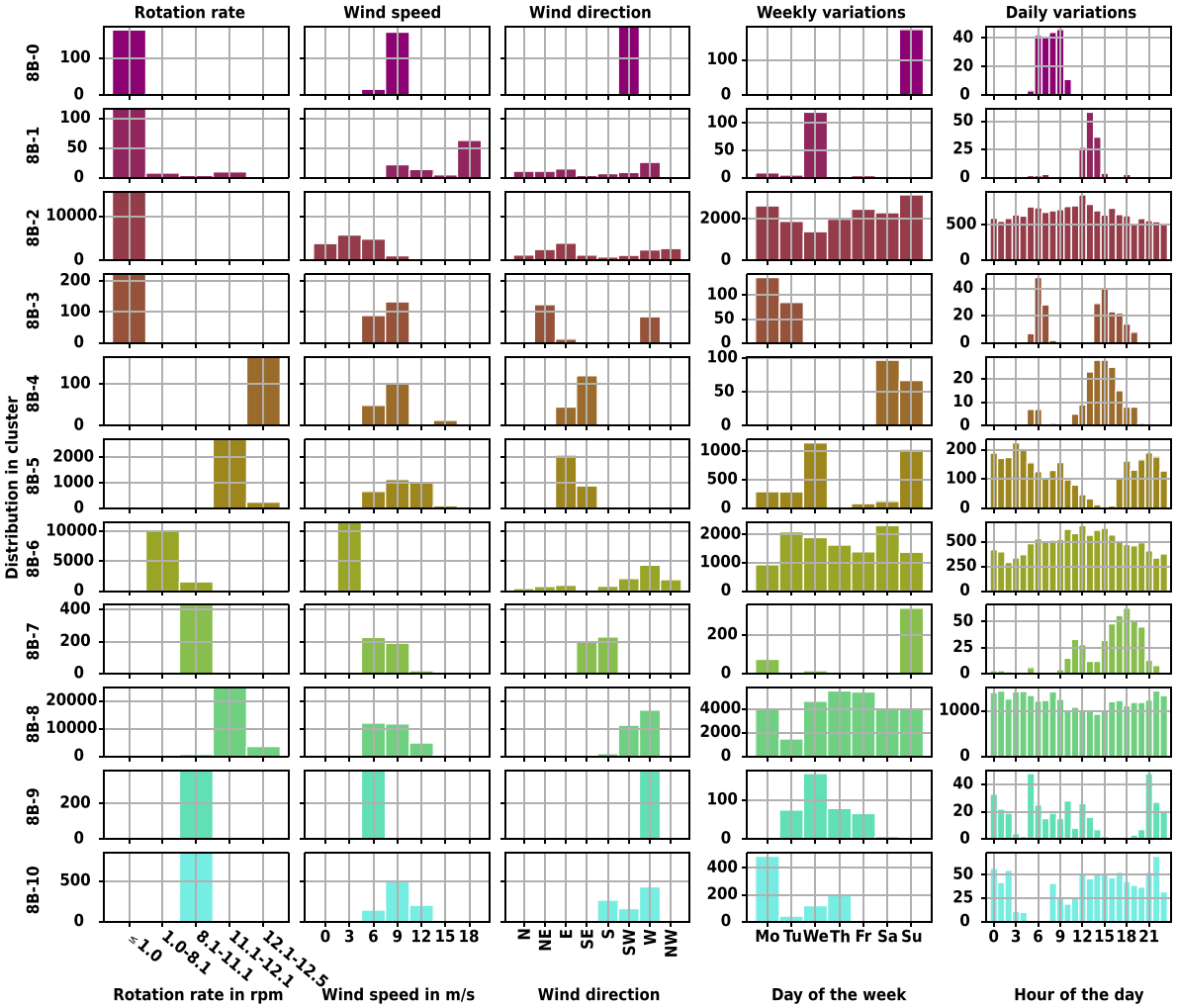

Figure D1. Clustering distribution of the feature matrix of recording station 8B using the HDBSCAN hyperparameters *min_sample* set to 50 and *min_cluster_size* to 100. The distribution of the clusters is analysed with regard to the rotation rate of *WT 3*, the wind speed and wind direction, and days of the week, as well as hours of the day in UTC. The displayed rotation rate is grouped into the five operational stages of *WT 3* (Fig. 3a). Each row displays the distribution of a single cluster.

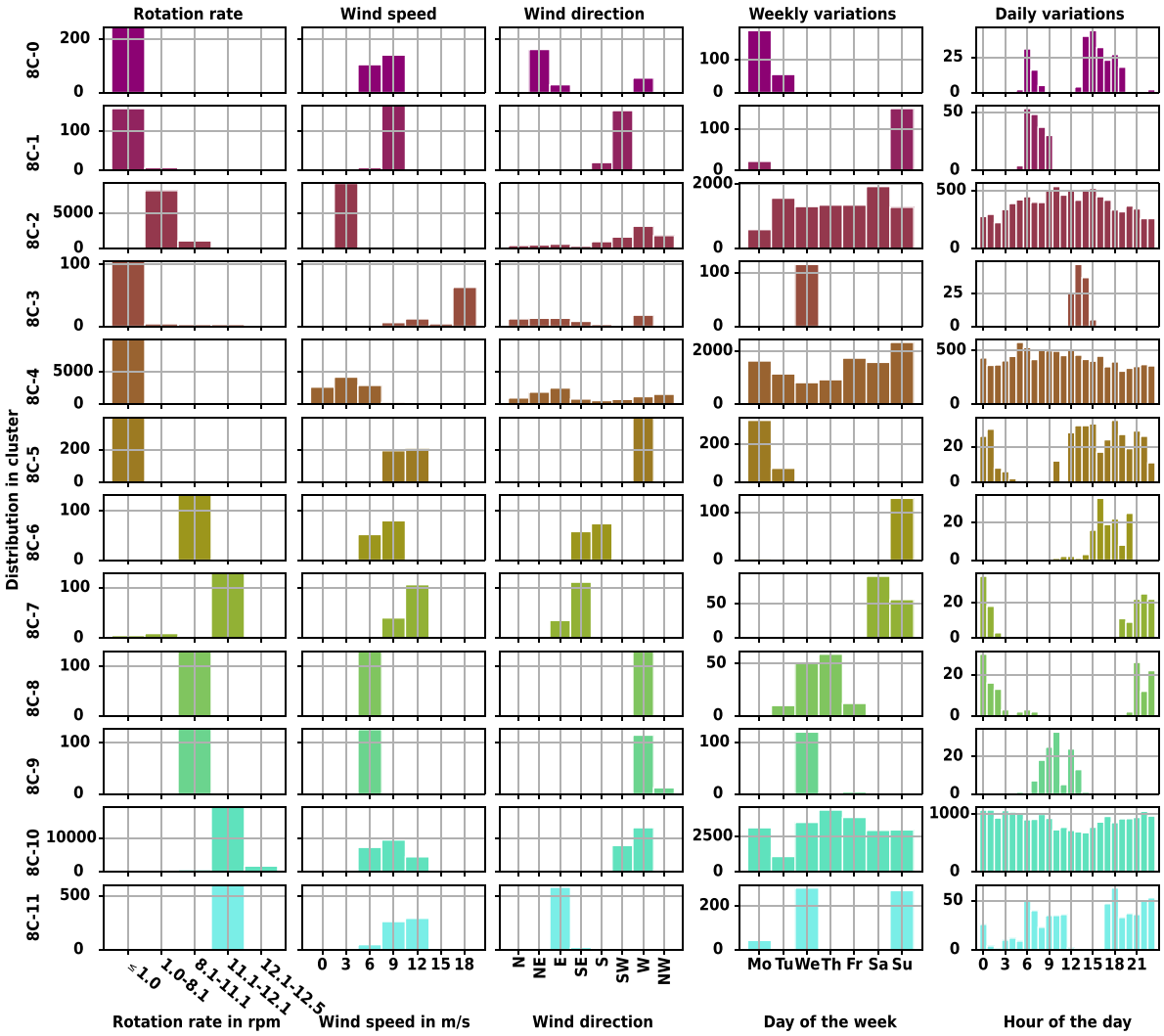

Figure D2. Clustering distribution of the feature matrix of recording station 8C using the HDBSCAN hyperparameters *min_sample* set to 50 and *min_cluster_size* to 100. The distribution of the clusters is analysed with regard to the rotation rate of *WT 3*, the wind speed and wind direction, and days of the week, as well as hours of the day in UTC. The displayed rotation rate is grouped into the five operational stages of *WT 3* (Fig. 3a). Each row displays the distribution of a single cluster.

during times with south and southeastern winds, while clusters *8C-8* and *8C-9* are populated during periods in which the wind blows from the west. Cluster *8C-8* is primarily populated during nighttime, and Cluster *8C-9* during daytime. Clusters *8C-7*, *8C-10* and *8C-11*, associated with the higher noise-reduced mode, are discussed in detail in Section 4.

Fig. D3 presents the HDBSCAN clustering distribution of recording station 8D. The feature matrix of station 8D reveals seven clusters, none of which are associated with a dominant rotation rate corresponding to partial or full load operation. Clusters *8D-1*, *8D-5* and *8D-6* are active during periods when *WT 3* is not in operation. Among these, Cluster *8D-6* is distinguished by its occurrence under comparatively higher wind speeds. Cluster *8D-1* is inactive during the early hours of the day but otherwise does not differ significantly from cluster *8D-5*. Clusters *8D-2* and *8D-4* are associated with the lower noise-reduced operational mode and can be distinguished based on the prevailing wind direction during which they are active. Clusters *8D-0* and *8D-3* are populated during times in which *WT 3* is operated in the higher noise-reduced mode and are discussed in detail in Section 4.

Fig. D4 presents the HDBSCAN clustering distribution of recording station 8G. Cluster *8G-0* is the largest cluster of this data set. Its population is distributed almost evenly across all analysed variables, with no clear pattern discernible. In contrast, cluster *8G-1* is active predominantly on weekdays during working hours (local time: UTC+1). It is further characterized by wind speeds below 9 ms^{-1} and rotation rates below 11.1 rpm.

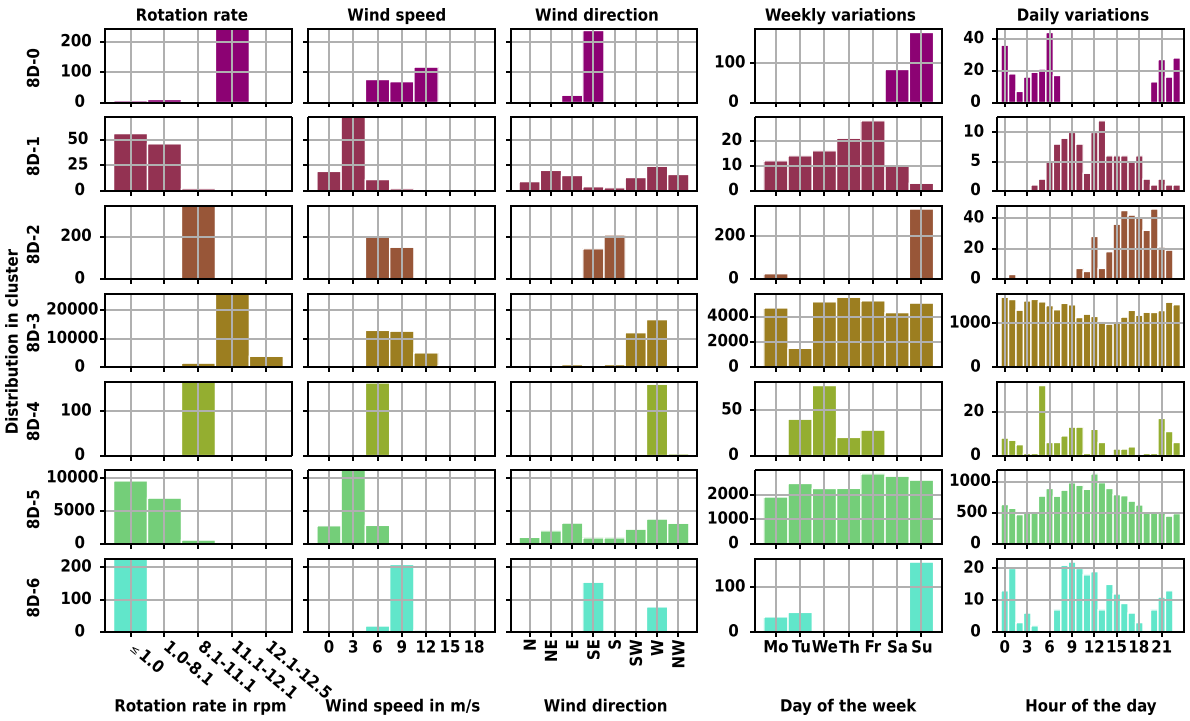

Figure D3. Clustering distribution of the feature matrix of recording station 8D using the HDBSCAN hyperparameters *min_sample* set to 50 and *min_cluster_size* to 100. The distribution of the clusters is analysed with regard to the rotation rate of *WT 3*, the wind speed and wind direction, and days of the week, as well as hours of the day in UTC. The displayed rotation rate is grouped into the five operational stages of *WT 3* (Fig. 3a). Each row displays the distribution of a single cluster.

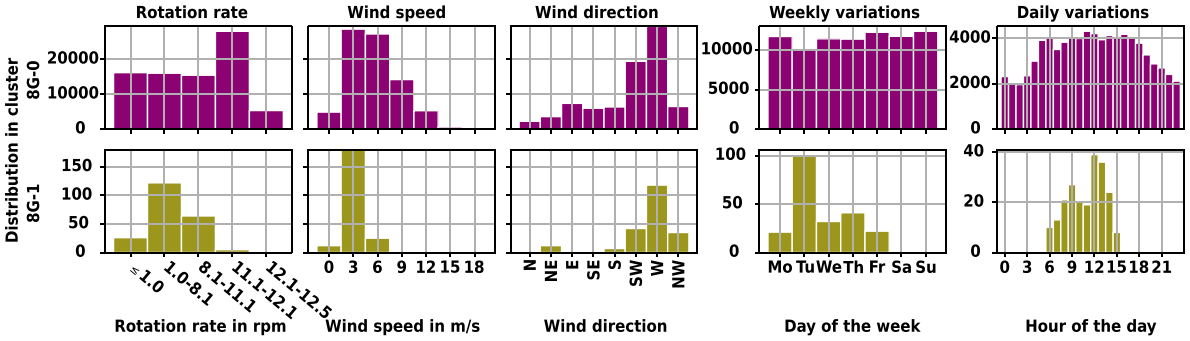

Figure D4. Clustering distribution of the feature matrix of recording station 8G using the HDBSCAN hyperparameters *min_sample* set to 50 and *min_cluster_size* to 100. The distribution of the clusters is analysed with regard to the rotation rate of *WT 3*, the wind speed and wind direction, and days of the week, as well as hours of the day in UTC. The displayed rotation rate is grouped into the five operational stages of *WT 3* (Fig. 3a). Each row displays the distribution of a single cluster.