

Characterisation and Quantification of Wind Farm Noise

By

Duc Phuc Nguyen
BEng (Civil), MEng (Acoustics)

*Thesis
Submitted to Flinders University
for the degree of*

Doctor of Philosophy
College of Science and Engineering
28 November 2022

To my Parents and Châm

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Duc Phuc Nguyen

BEng (Civil), MEng (Acoustics)

November 2022

Acknowledgements

My father likes to repeatedly tell me that I am the first in my entire extended family to go to university and now I have had the opportunity to pursue a doctoral degree. My mother is so supportive and always asks me how I am going at school and why it takes several years for me to finish “that homework” although I am now in my 30s. Being the first in my family to go to university has never been easy. Living and studying abroad for my first time has also been a real challenge, but it has also been joyful. I could not have imagined that my PhD journey was going to be so smooth, particularly when it started at the same time as the pandemic. This would not have been possible without the support of my principle supervisor, Dr Kristy Lee Hansen, who is the reason that this thesis exists. Kristy, thank you for your support, not only in academia, but also in life. Thank you for the long field trips during the final stages of your pregnancy. It was never easy to leave home at 6 am, hit two Kangaroos and tow the car back home for 200 km until 1 am the next day. Sometimes we drove for six hours just to fix microphone cables that had been chewed by sheep. All of your efforts and support were to make sure that I had comprehensive data sets for my PhD. My co-supervisors, Dr Branko Zajamsek and Professor Peter Catcheside, you both provided me with the perfect balance of guidance and freedom. Branko, I admire your effective problem solving that helped me to improve the technical quality of this thesis. Thank you also for reading countless drafts of papers and helping me to acquire a deep knowledge of acoustics. Peter, it was a great privilege to work closely with you. Conversations during the weekends or late nights in your office expanded my knowledge and encouraged me to feel that the skills that I learned during my PhD could be applied to other fields. Thanks to you, I became very interested in sleep medicine and respiratory physiology. Last but not least, thanks to

Professor Colin Hansen, I couldn't have asked for better support from you. I am lucky to have had the opportunity to work with you, a respected and leading acoustician. This thesis could not have been completed without your comprehensive and thoughtful comments on every piece of work in this thesis regardless of whether it was the weekend or Christmas holidays.

I would also like to thank other people on the Wind Farm project team: Bastien, Tessa, Claire, Felix and Gorica. Bastien, thank you for your help on most of the work in my thesis. I've really enjoyed collaborating with a very smart person like you. Claire, thank you for teaching me to setup a sleep study. I am still not as fast as you because you are the queen of this. I wish you all the best with your thesis. I also want to thank the people at the Adelaide Institute for Sleep Health (AISH): Danny, Amal, Alex, Carolin, Alison and Susanne. Thank you for your support, I really enjoyed working with you. Finally, I gratefully acknowledge the Flinders University Research Scholarship which supported my study.

Abstract

This thesis presents a study on the characterisation and quantification of wind farm noise (WFN) at long-range locations. The primary research goals in this thesis were to develop a machine learning model to detect amplitude modulation (AM) of WFN. This model was used to automate the AM detection process, and to quantify and characterise AM in long-term measured data sets. A further aim was to investigate the audibility of unique characteristics of WFN such as infrasound and amplitude modulated tones. The thesis concludes with an exploration of deep learning techniques, which were used to examine and automate the characterisation of WFN. Chapter 1 provides an overview of recent advances in WFN research, followed by a description of field work in Chapter 2 and then four major results chapters.

Chapter 3 presents an approach to detect and characterise AM in a comprehensive and year-long wind farm noise data set evaluated using human scoring. Benchmark AM characteristics were established towards further validation and calibration of results obtained using automated methods. Using these data, an advanced AM detection method was then developed, with predictive power close to the practical limit set by human scoring. However, given that noise impacts on humans remain of primary interest, human-based approaches should be considered as a benchmark method for characterising and detecting unique noise features most relevant to human WFN perception and impacts.

Chapter 4 quantifies and characterises AM over 1 year using acoustical and meteorological data measured at three locations near 3 wind farms. Substantial diurnal variation of outdoor AM prevalence was found, with nighttime prevalence approximately 2 to 5 times higher than daytime prevalence. On average, indoor AM occurred during the nighttime from 1.1 to

1.7 times less often than outdoor AM. However, the indoor AM depth was higher than that measured outdoors. An association between AM prevalence and sunset and sunrise was also observed. These data showed that AM occurs more often during downwind and crosswind compared to upwind conditions. These findings provide important insights into long term WFN characteristics needed to help better inform future WFN assessment guidelines.

Chapter 5 used a computational approach to assess the audibility of infrasound and amplitude modulated tones (AM tones) at long-range locations, which also considered the uncertainty associated with WFN measurements and human hearing variability. It was demonstrated that infrasound is rarely audible to residents with normal hearing who live at distances greater than 1 km from a wind farm, but that AM tones occurring at a low frequency are readily audible at distances up to 9 km. These results suggest that AM tones could be the main source of WFN complaints at long-range locations, and thus clearly warrant further attention towards ensuring that wind farms have minimal impacts on nearby residents located within 9 km of the nearest wind turbine.

Chapter 6 explored an approach for the characterisation and assessment of WFN. This was based on extraction of acoustic features from a pretrained deep learning model (referred to as deep acoustic features). Using data measured at a variety of locations, deep acoustic features were shown to contain meaningful information about noise characteristics. Deep acoustic features were also shown to reveal an improved spatial and temporal representation of WFN compared to traditional spectral analysis and statistical noise descriptors. These very promising novel findings provide a clear framework for improved WFN assessment in the future.

Taken together, this thesis work provides a major and important new contribution towards the understanding of some of the most prominent WFN features audible to humans. These new methods provide an important framework towards improved noise assessments and wind turbine designs better able to minimise impacts on surrounding communities. Ultimately, this approach, along with future improvements in wind farm planning, design, noise assessment and abatement strategies will all likely be needed to help ensure that wind energy is acceptable to surrounding communities.

Table of contents

List of figures	xiii
List of tables	xvi
Nomenclature	xvii
1 Introduction	1
1.1 Wind energy	1
1.2 Noise from wind farms	3
1.2.1 Amplitude modulation	3
1.2.2 Amplitude modulated tonal noise	7
1.2.3 Infrasound	7
1.2.4 Wind farm noise propagation	8
1.2.5 Impacts on humans	9
1.3 Machine learning in the context of environmental noise	12
1.4 Aims	13
1.5 Thesis outline	14
1.6 Research outputs	15
2 Field measurement	18
2.1 Study region	18
2.2 Wind farm characteristics	21
2.3 Measurement location and duration	23

2.4	Outdoor noise measurements	23
2.4.1	Measurement setup	23
2.4.2	Microphones	24
2.4.3	Calibration	26
2.4.4	Wind screens	27
2.4.5	Power requirements	28
2.5	Indoor noise measurements	29
2.5.1	Microphones	29
2.5.2	Microphone positions	30
2.5.3	Other setup details	30
2.6	Weather condition measurements	30
2.7	Wind farm operating data	31
3	Automated wind farm AM detection	32
3.1	Introduction	33
3.2	Methods	35
3.2.1	Overview of study region and data collection	35
3.2.2	Benchmark data set generation	36
3.2.3	Automated AM detectors	38
3.2.4	Random Forest classifier for AM detection	40
3.2.5	Audio feature extraction	42
3.2.6	Evaluation metrics	45
3.2.7	Benchmark AM characterisation	47
3.2.8	Data and statistical analysis	47
3.2.9	Data availability	47
3.3	Results	48
3.3.1	Benchmark data set	48
3.3.2	Benchmark AM characteristics	48
3.3.3	Diurnal and seasonal AM variation	52
3.3.4	Random forest-based AM detection	53

3.3.5	Performance of the automated detectors	56
3.3.6	Interpretable predictor	58
3.4	Discussion	59
3.5	Conclusions	62
4	Quantification and characterisation of AM	63
4.1	Introduction	64
4.2	Methods	65
4.2.1	Study region	65
4.2.2	Experimental design	66
4.2.3	Amplitude modulation detection	69
4.2.4	Wind direction categories	70
4.2.5	Data cleaning	71
4.2.6	Data and statistical analysis	72
4.2.7	Data availability	73
4.3	Results	75
4.3.1	Amplitude modulation characteristics	75
4.3.2	Outdoor-to-indoor variability	77
4.3.3	Diurnal and seasonal variability	79
4.3.4	Relationship between meteorological and power output conditions	79
4.4	Discussion	83
4.5	Conclusion	86
5	Audibility of wind farm noise	87
5.1	Introduction	88
5.2	Methods	90
5.2.1	Overview of study	90
5.2.2	Wind farm characteristics and experimental setup	90
5.2.3	Data cleaning	92
5.2.4	Frequency analysis	92

5.2.5	Hearing threshold variability	93
5.2.6	Probability of audibility	94
5.2.7	Quantification of noise measurement uncertainty	95
5.2.8	Regression analysis	96
5.2.9	Statistical analysis	97
5.2.10	Code availability	97
5.3	Results	98
5.3.1	Uncertainty in wind farm noise measurement	98
5.3.2	Infrasound characteristics	98
5.3.3	Low-frequency amplitude modulated tones	101
5.3.4	Hearing threshold variability	103
5.3.5	Audibility of wind farm infrasound	104
5.3.6	Audibility of amplitude modulated tones	106
5.4	Discussion	112
5.5	Conclusion	113
6	Beyond traditional wind farm noise characterisation	115
6.1	Introduction	116
6.2	Methods	117
6.2.1	Data sets	117
6.2.2	VGGish	118
6.2.3	Dimensionality reduction methods	120
6.2.4	Computational time	121
6.2.5	Correlation of pairwise distances	122
6.2.6	Performance of deep acoustic features	122
6.2.7	Statistical analysis	123
6.2.8	Data and code availability	124
6.3	Results	124
6.3.1	Deep acoustic features	124
6.3.2	Deep acoustic features reveal noise characteristics	125

6.3.3	Performance of dimensionality reduction methods	128
6.3.4	Deep acoustic features reveal spatial and temporal structures of wind farm noise	129
6.4	Discussion	132
6.5	Conclusion	133
7	Conclusions and Recommendations for Future Work	134
7.1	Overview of thesis aims	134
7.2	Summary findings	135
7.3	Conclusions	136
7.4	Future directions	137
	References	140

List of figures

1.1	Global distribution of wind farms	2
1.2	Wind farm noise characteristics	4
1.3	‘Swishing’ noise mechanism	6
1.4	Infrasound mechanism	8
1.5	Wind farm noise propagation	10
1.6	Relationship between annoyance and AM	11
1.7	Applications of machine learning	12
1.8	Thesis outline	15
2.1	Study region	19
2.2	Weather conditions	20
2.3	Wind farm power output	22
2.4	Outdoor noise measurements	25
2.5	Frequency response of microphone	26
2.6	Microphone performance	27
2.7	Secondary wind screen	28
2.8	Indoor noise measurements	29
3.1	Data selection process	37
3.2	MATLAB GUI for AM detection	38
3.3	Random forest classifier	41
3.4	Feature extraction	45
3.5	Characteristics of benchmark data sets	49

3.6	Benchmark AM characteristics	51
3.7	Variation of AM prevalence	54
3.8	Random Forest classifier	55
3.9	Performance of automated detectors	57
3.10	Interpretable classifier	59
4.1	Measurement locations and experimental set-up	67
4.2	Agreement between microphones	68
4.3	The relationship between wind farm power output capacity and local wind speed	69
4.4	Flow chart of AM detection method	70
4.5	Wind direction categories	72
4.6	Distributions of indoor SPL	73
4.7	Distributions of outdoor SPL	74
4.8	Characterisation of AM for outdoor data measured over one year	76
4.9	Outdoor and indoor AM prevalence and AM depth	78
4.10	Diurnal and seasonal variation of AM characteristics	80
4.11	AM prevalence for different meteorological and wind farm operating conditions	82
5.1	Study overview	91
5.2	Probability of audibility	95
5.3	Uncertainty of wind farm noise levels depending on measurement duration	99
5.4	Typical infrasound spectrum	100
5.5	AM tone at low frequency	102
5.6	Hearing threshold variability	104
5.7	Audibility of infrasound	105
5.8	AM tone at 80 Hz at wind farm A	107
5.9	AM tone of 25 Hz at wind farm B	109
5.10	Audibility of AM tones at wind farm C	111
6.1	VGGish and deep acoustic features	126

6.2	Deep acoustic features reveal unique noise characteristics	127
6.3	Dimensionality reduction methods	129
6.4	Spatial and temporal structure of WFN	131

List of tables

2.1	Characteristics of wind farms.	22
2.2	Measurement locations and duration	24
3.1	Value ranges of the hyperparameters used for random searching.	42
3.2	Feature descriptions	43
3.3	Evaluation metrics: Definitions and equations	46
3.4	Inter-scorer agreement	48
3.5	Intra-scorer agreement	50
3.6	Performance of the best classifier on the out-of-bag samples	53
3.7	Area under the precision-recall curves and optimal MCC of four methods.	56
4.1	Residential house dimensions and constructions at three measurement locations.	67
4.2	Percentage of detected AM in both 1.5 m and ground level data.	68
4.3	Performance of AM detectors	70
4.4	Data quality control and cleaning	71
4.5	Outdoor-to-indoor AM prevalence.	78
5.1	Wind farm infrasound levels over one-year long monitoring	101
5.2	AM tone characteristics	103
6.1	Software versions and parameters.	121
6.2	Hyperparameter tuning.	123
6.3	Optimised Hyperparameters.	123

Nomenclature

Acronyms / Abbreviations

AM Amplitude modulation

AUC Area under the receiver operating characteristic curve

BPF Blade-pass frequency

CNN Convolutional neural network

dBA A-weighted sound pressure level

dBC C-weighted sound pressure level

dB Decibel re $20\mu\text{Pa}$

dBG G-weighted sound pressure level

GW Gigawatts = 1 billion watts

IOA The UK Institute of Acoustics

PCA Principal component analysis

ROC Receiver operating characteristic curve

SPL Sound pressure level

tSNE t-distributed stochastic neighbor embedding

UMAP Uniform manifold approximation and projection

WFN Wind farm noise

WTN Wind turbine noise

XGBoost Extreme Gradient Boosting algorithm

Chapter 1

Introduction

In this chapter, I first provide an overview of the development of wind energy, followed by an introduction to wind farm noise (WFN)¹, its characteristics, mechanisms of generation and impacts. Recent advances in machine learning for acoustic applications are also introduced. Finally, with this context in place, I summarise the aims of the PhD thesis and provide an outline of the remainder of the thesis.

1.1 Wind energy

The global wind industry has experienced a remarkably rapid expansion in recent years (GWEC, 2019). This rapid expansion has been driven by rising concerns over global climate change and fossil fuel depletion, as well as energy security. The amount of global energy production has increased by an order of magnitude since the year 2000. At the time of writing this thesis, according to TheWindPower², the total global installed capacity is 1107.6 gigawatts (GW) with 22,385 wind farms around the world. The global wind energy growth rate is over 20% per year, with important variations from one country to another (GWEC, 2019). In the late 1990s, Denmark was leading the world with installed power less than 0.3 GW. The USA took this place in the 90's with over 3 GW. In the early

¹Throughout this thesis, I use WFN to refer to multiple wind turbines contributing to noise emissions at a residence. WFN is commonly used interchangeably with wind turbine noise (WTN) in the scientific literature.

²https://www.thewindpower.net/store_en.php

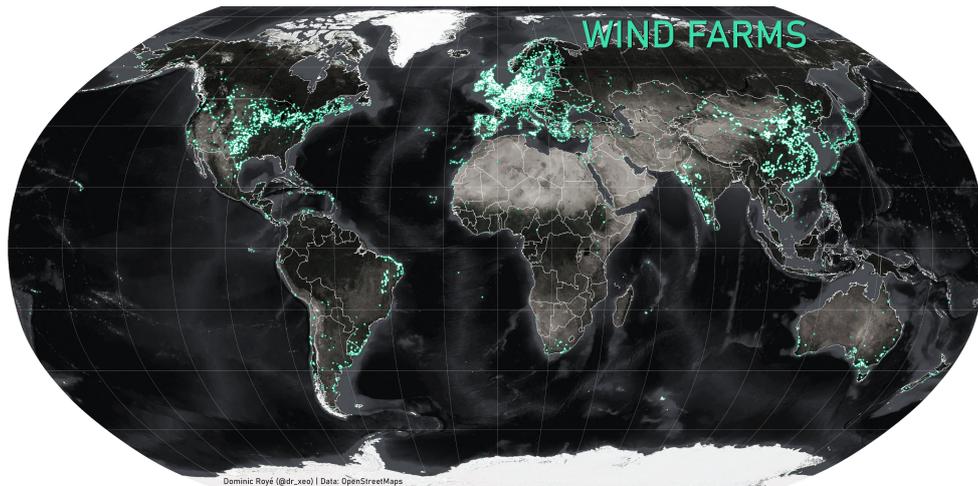


FIGURE 1.1. Global distribution of wind farms. This figure is reproduced with permission from Dr. Dominic Royé³.

2000's, Germany was leading the world with installed power of over 40 GW before the USA returned to this place in 2008. The total global installed capacity was just over 200 GW before 2010. The development of wind energy in China has been increasing exponentially since 2005 and leading the world since 2016. According to TheWindPower, the top three highest wind energy producing countries in 2022 are China (200 GW), the USA (163 GW) and the UK (73 GW). The global distribution of wind farms is shown in [Figure 1.1](#).

Wind turbine technology has been greatly improved to reduce installation cost, and increase energy capacity and reliability ([Dykes et al., 2019](#)). Wind turbines have become progressively larger to take advantage of higher wind speeds with greater ground clearances. Wind turbine blades have also been advanced through optimisation of blade profiles, limiting aerodynamic loads, minimising blade fouling and the use of composite materials with more favourable cost and performance characteristics ([Schubel and Crossley, 2012](#)). To maximise power output, modern wind turbines are pitch-regulated machines and variable speed, resulting in more optimal ratios between the blade-tip and wind speed. Three-bladed designs have also become dominant for modern wind turbines, as this configuration provides the current best compromise between aerodynamic efficiency, cost, rotational mass, structural integrity, inertial stability, relatively low tip-speed ratios and aesthetics ([Schubel and Crossley,](#)

³<https://dominicroye.github.io>

2012). While a number of early wind turbine designs integrated steel truss-type or concrete monopole (single support cylinder or partial-cone) towers into their designs, modern wind turbines consist of a steel, monopole structure with a reinforced concrete foundation.

1.2 Noise from wind farms

Despite achievements in the development of wind energy, wind farms continue to be associated with problems around social acceptance due to aesthetic, environmental and health impact concerns of nearby residents (Merlin et al., 2013). Noise generated from wind farms is a recurring source of complaints regarding annoyance and potential sleep disturbance from residents living near wind farms (Liebich et al., 2021a; Micic et al., 2018). The major noise source from a wind turbine is aerodynamic noise. Other noise sources such as mechanical and power transmission noise are generally weaker, especially at long-range locations. Aerodynamic noise sources generate unique characteristics of WFN such as infrasound, amplitude modulation and a low-frequency spectral dominance. A typical WFN spectrum and its characteristics are shown in Figure 1.2. The following sections provides a brief description of how wind turbines generate these unique noise characteristics.

1.2.1 Amplitude modulation

Amplitude modulation (AM) is defined as the periodic variation in amplitude of a noise or vibration signal. In relation to WFN, it is defined by the (UK) Institute of Acoustics (IOA) as (AMWG, 2015),

“...periodic fluctuations in the level of broadband noise from a wind turbine (or wind turbines), the frequency of the fluctuations being the blade-pass frequency of the turbine rotor, as observed outdoors at residential distances in free-field conditions.”

At distances several hundreds meters from a wind farm, AM is often perceived as a ‘swishing’ noise (Bakker et al., 2012; Doolan, 2013; Van den Berg, 2004). ‘Swishing’ noise is also referred to as ‘normal AM’, and it has been well characterised and quantified. Its mechanism has also been intensively investigated both on model scale rotors (Oerlemans et al.,

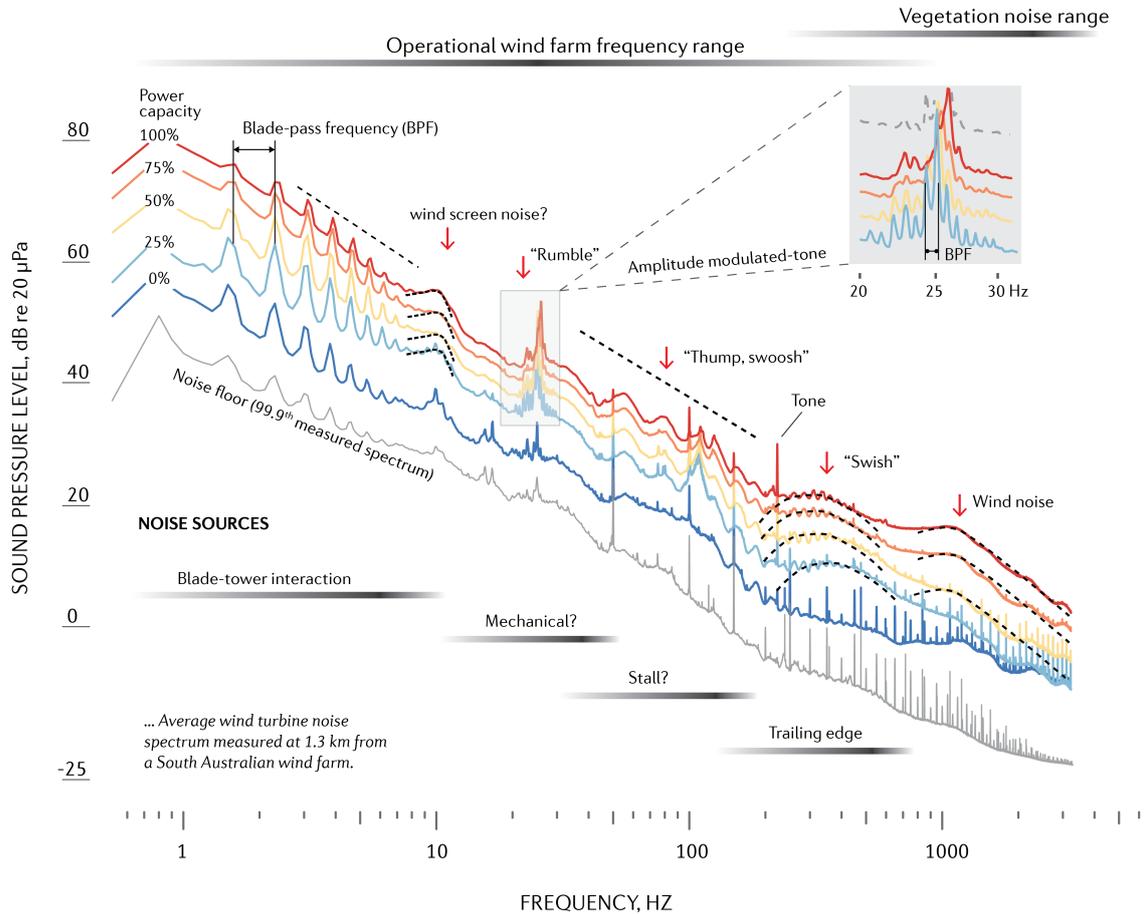


FIGURE 1.2. Wind farm noise characteristics. Wind farm Infrasound is observed as peaks on the spectrum below 10 Hz (see Section 1.2.3 for details). There are several amplitude modulated tones, which are characterised by a bell shape with side band spacing at the blade-pass frequency (BPF). The amplitude modulated tones are sometimes referred to as ‘rumbling’ noise that usually occurs at frequencies below 100 Hz. The modulated tones can be observed at distances up to 9 km (see Section 1.2.2). ‘Thumping’ noise or ‘enhanced AM’ occurs in the frequency range between 50 and 300 Hz. The noise is dominant at distances greater than 1 to 3 km (see Section 1.2.1). ‘Swishing’ noise or ‘normal AM’ occurs in the frequency range between 200 and over 1,000 Hz. This noise is dominant over several hundred meters from a wind farm (see Section 1.2.1).

2001) and full-scale wind turbines (Lee et al., 2012; Oerlemans et al., 2007). Experimental studies (Lee et al., 2012; Oerlemans et al., 2007; Ramachandran et al., 2014; Veggeberg and Zheng, 2008) using the beamforming technique have found that the outer part of the blades, creates most noise perceived on the ground (Figure 1.3a,b). The noise is mainly created during the downward movement of the blades, for which the noise is at least 10 dB higher than for the upward movement (Oerlemans et al., 2007). This strong asymmetry of the noise source produces ‘swishing’ noise (Oerlemans et al., 2007). The mechanism of this phenomenon is attributed to convective amplification and directivity of trailing-edge noise (Figure 1.3c). For a stationary receiver and moving source, convective amplification⁴ causes the sound pressure level (SPL) upstream of the source to be greater than that downstream at the same distance from the source. For an operating wind turbine, when the blades move and approach an observer on the ground, the noise level of the source will vary due to convective amplification. In addition to the convective amplification effect, the directivity of the trailing-edge noise-radiation contributes most to the asymmetry of the noise source (Oerlemans et al., 2007). The directivity remains constant relative to the blade. For high-frequency trailing edge noise directivity, the directivity pattern is ‘forward-looking’, which is the directivity index is higher at the downwind condition compared to the upwind condition. Since the blade is rotating, the relative position between the source and a stationary receiver varies periodically at the blade-pass frequency (BPF) (Figure 1.3c). This results in SPL variations at the receiver. A combination of convective amplification and directivity effects results in perception of ‘swishing’ noise (or ‘normal AM’) at receivers on the ground located within several hundred meters. This is most clearly perceived at crosswind locations (Figure 1.3d).

Compared to ‘swishing’ noise, there is no clear consensus in the literature to explain the mechanism of ‘thumping’ noise (or ‘enhanced AM’). Wind shear, a change in wind speed with height, is hypothesised to cause ‘thumping’ noise (Oerlemans, 2015). When the wind speed at the top of the rotor disc is much higher than the speed at the bottom of the disc, the angle of attack increases from the bottom to the top. Aerodynamic stall could occur at the tip of each turbine blade when it approaches the top of the rotor disc. This generates

⁴The phenomenon when a source moves and approaches an observer, its amplitude will rise.

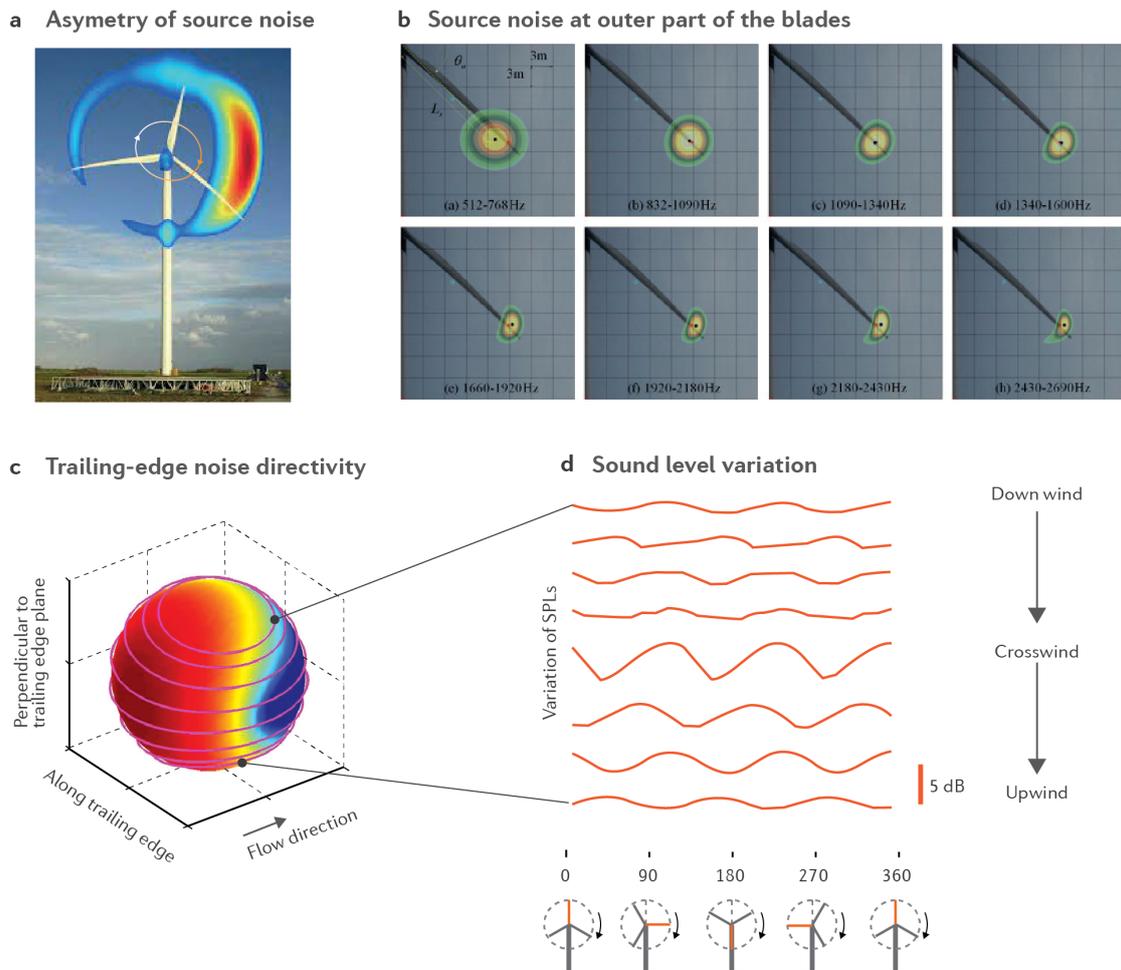


FIGURE 1.3. ‘Swishing’ noise mechanism. **a**, Noise is created during the downward movement of the blades. **b**, The primary noise source is at the outer part of the blades. The location of the blade sources moves outward for increasing frequency. **c**, Noise directivity of trailing edge plane and noise measured by ground-mounted microphones when the blades are rotating. **d**, ‘Swishing’ noise is clearly perceived during crosswind conditions due to combined directivity and convective amplification. Parts **a**, **c** and **d** are adapted with permission from [Oerlemans and Schepers \(2009\)](#), SAGE; Part **b** is adapted with permission from [Lee et al. \(2012\)](#), Elsevier.

separation-stall noise that is dominant at low-frequencies (Moreau et al., 2009). As the blade rotates, the periodic separation-stall noise, reported as ‘thumping’ noise, is created at the BPF. However, this explanation is yet to be comprehensively tested as separation-stall noise does not often occur with modern wind turbines with stall- or pitch-regulated controls (Lee et al., 2012). Previous studies have shown that ‘thumping’ noise occurs more often during the nighttime which usually corresponds with stronger wind shear compared to daytime conditions (Hansen et al., 2019a; Larsson and Öhlund, 2014; Van den Berg, 2005). However, it remains unclear whether ‘thumping’ noise is due to wind shear.

1.2.2 Amplitude modulated tonal noise

At long-range locations, it is common to observe amplitude modulated tones (AM tones), which are often described as ‘rumbling’ noise. AM tones have received relatively little attention in WFN studies to date (Hansen and Hansen, 2020). One likely reason is that this type of AM is not detected at all wind farms (Hansen et al., 2017). Also, ‘rumbling’ noise occurs at very low frequencies (< 100 Hz) and it is not always audible for humans with normal hearing. ‘Rumbling’ noise can travel over distances up to 9 km from wind farms (Hansen et al., 2019a). The mechanism(s) responsible for the AM tones is unclear, but it is hypothesised to have a mechanical origin and to be re-radiated from the blades and/or tower. The amplitude modulation of these tones is believed to occur due to periodic changes in loading on the blades (Hansen et al., 2017). Similar to the telecommunications field, the AM tone can be considered as the carrier signal, while the blade-tower interaction noise is the information signal. The result of this interaction is tonal noise with side-bands in the frequency domain. However, this hypothesis is yet to be investigated in detail due to the difficulty in collecting data from the wind turbine drive train.

1.2.3 Infrasond

Wind farm infrasond is primarily generated by the blades passing the tower (Doolan et al., 2012; Zajamsek et al., 2019). The wind flow passes around the tower, creating a region of

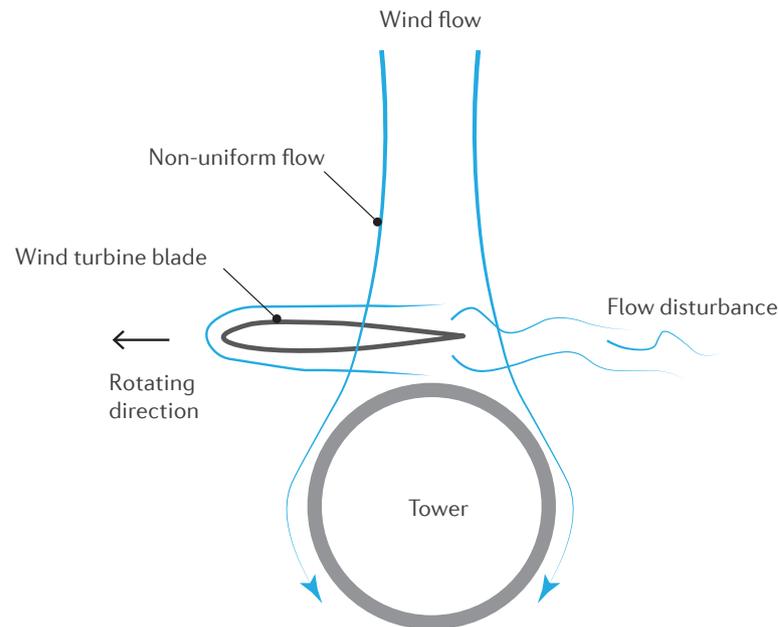


FIGURE 1.4. Infrasound mechanism. Note that for illustrative purposes, the relative size between the tower and blade is not shown to scale.

non-uniform flow upstream of the tower as shown in [Figure 1.4](#). The wind turbine blade angle of attack to the incoming flow changes due to this non-uniform flow region, resulting in a fluctuation in the lift force. This fluctuating load is the main mechanism of wind turbine infrasound generation. This phenomenon is more significant for the down wind turbine configuration ([Hubbard, 1990](#)), where the blade is located on the downwind side of the tower. However, most modern turbines use the upwind turbine configuration which generates lower levels of infrasound ([Hubbard and Shepherd, 1991](#)).

1.2.4 Wind farm noise propagation

The simplest way to understand how WFN travels to a residence is to use a ray tracing model ([Ostashev and Wilson, 2015](#)). In this model, the noise travels to the residence through ray paths, connecting the noise source and receiver, as shown in [Figure 1.5](#). Depending on the atmospheric conditions, the ray paths could be straight lines or curves. For example, if the wind speed and temperature profiles are constant, these ray paths follow straight lines because

the sound speed profile is also constant. However, it is more common that the wind speed profile is nonlinear (Figure 1.5a) and that the wind speed is higher at higher local altitudes above ground. Also, the temperature gradient is usually positive (higher at higher altitudes) during the nighttime, and negative during the daytime. Due to these atmospheric conditions, the sound speed is nonlinear i.e., negative or positive gradient relative to ground, resulting in ray paths that are curves rather than straight lines. Depending on whether the sound speed profile has a negative or positive gradient, the curves can bend upward or downward, respectively. Figure 1.5a shows an example of the ray paths bending downward. The ray paths can travel directly to the residence or reflect several times before reaching the residence. Due to the interference between these direct and reflected ray paths, the resulting sound field may exhibit local maxima and minima, as shown in Figure 1.5b. This example clearly shows that noise is not necessarily louder at closer distances to wind farms. Also, at specific locations, noise is significantly affected by atmospheric conditions and local topography. Therefore, long-term noise measurements are crucial to help assess the nature, prevalence and variability of the most prominent and potentially most problematic WFN characteristics.

1.2.5 Impacts on humans

Previous studies have shown that AM contributes to annoyance (Ioannidou et al., 2016; Lee et al., 2011; Schäffer et al., 2016) and possibly sleep disturbance (Bakker et al., 2012; Liebich et al., 2021a; Micic et al., 2018). Lee et al. (2011) investigated annoyance caused by AM using listening tests. The authors found that higher annoyance is associated with higher AM depth (Figure 1.6). Schäffer et al. (2016) found that WFN with AM is more annoying than WFN without AM. Also, increased annoyance to AM has been shown to be unrelated to its periodicity or randomness (Figure 1.6). Using a similar approach as Lee et al. (2011), Ioannidou et al. (2016) found that annoyance decreased when AM depth was reduced (Figure 1.6). Other studies (Von Hünenbein et al., 2013; Yokoyama et al., 2013) used synthesised AM stimuli to assess the impacts of AM. Von Hünenbein et al. (2013) showed that annoyance increased with AM depth, but this result was not statistically significant.

⁵<https://github.com/ducphucnguyen/FreeRay.jl.git>

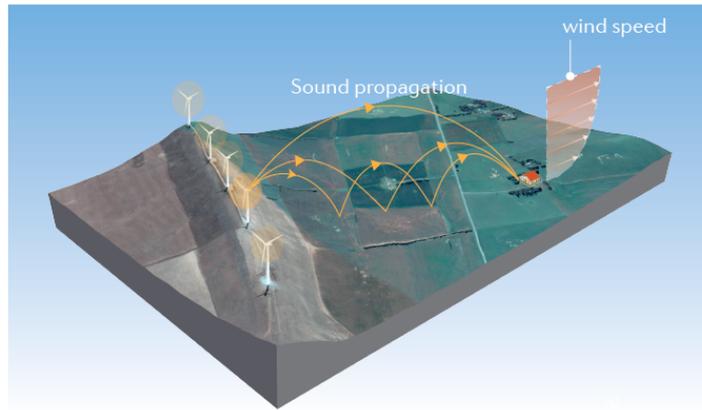
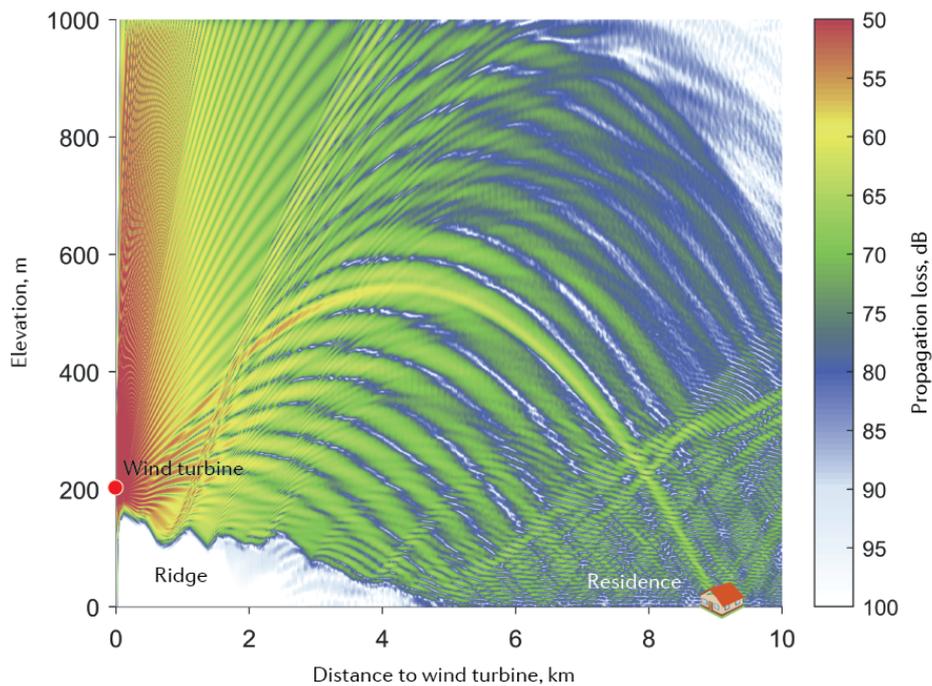
a Wind turbine noise travel to the residence**b** Propagation loss

FIGURE 1.5. Wind farm noise propagation. **a**, Wind farm noise travels to the residence through ray paths. **b**, An example of a sound field, in which a wind turbine is 80 m above the ridge. Sound propagation is modelled using FreeRay package⁵.

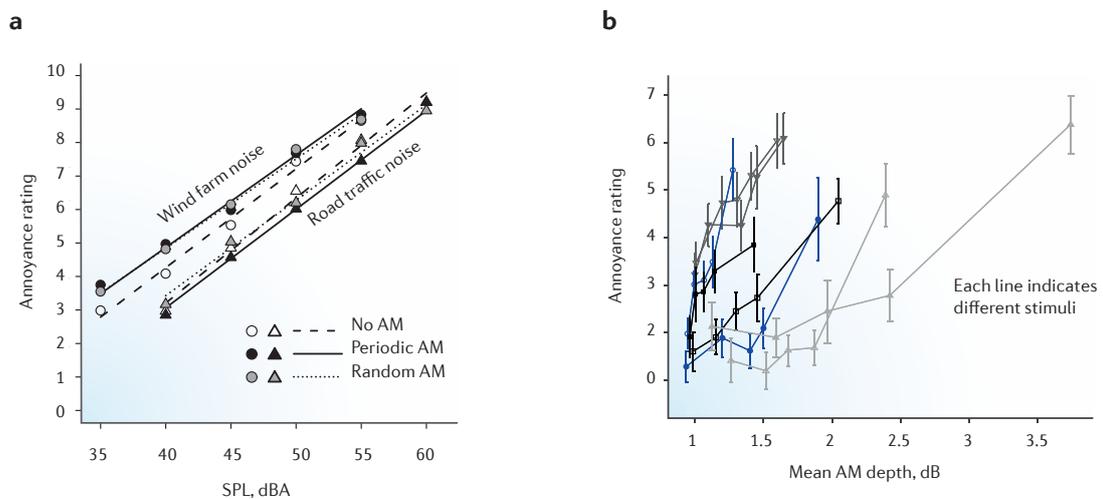


FIGURE 1.6. Relationship between annoyance and AM. **a**, Wind farm noise is consistently rated as more annoying than road traffic noise. Noise containing AM is also consistently rated as more annoying than noise without AM. **b**, A higher AM depth is associated with higher annoying ratings. The stimuli are measured at different distances and wind farm capacity. Part **a**, is adapted with permission from Schäffer et al. (2016), JASA; Parts **b** is adapted with permission from Ioannidou et al. (2016), JASA.

Yokoyama et al. (2013) found that noise fluctuations were noticeable when the AM depth exceeded 2 dB. In contrast to listening test measurements, Smith et al. (2020) assessed the impacts of WFN on sleep and found that AM of WFN may impact self-reported sleep while other measures of objective sleep did not differ significantly between nights with and without WFN. In a systematic review and laboratory study, Liebich et al. (2021a,b) found no consistent evidence to support that WFN increased sleep latency. However, inconsistent findings between studies (Liebich et al., 2021a) along with ongoing uncertainty regarding what WFN features and sound pressure levels may become problematic for sleep clearly warrant further investigation. Well-controlled experimental studies using realistic WFN features and sound pressure levels most relevant to real-world noise exposure settings are clearly needed. Studies clearly need both objective and subjective assessments using well established and validated outcome measures along with appropriate sample sizes towards more definitive conclusions regarding WFN impact on surrounding communities.

Transfer learning ([Brownlee, 2017](#)) is a modern research approach in deep learning. In this approach, a model already constructed and trained to solve a specific task is reused as the starting point to solve another similar task. For example, a model trained to classify noise types could be reused as a noise feature extractor. These extracted features can then be trained to detect AM or tonality in WFN. The benefit of this approach is that existing deep learning models already trained with massive data and substantial amounts of computer and time resources can be reused for additional applications. As a result, these deep learning models can provide significant advantages when applied to other relevant problems. Applying transfer learning to environmental noise has shown promising preliminary results. [Liu et al. \(2019\)](#) showed that transfer learning models can be incorporated into acoustic surveillance systems to detect chainsaw noise to help rainforest conservation. [Sethi et al. \(2020\)](#) further showed that acoustic features extracted from pre-trained models could be used to characterise noise. The authors also illustrated that the transfer learning approach was useful for real-time automated detection of irregular environmental behaviour including illegal logging and hunting. Using a pretrained deep model (DEEP-Hybrid DataCloud project), [Clark et al. \(2021\)](#) were able to characterise the details of diurnal and spatial community noise sources.

1.4 Aims

In this thesis, my aims were:

1. To develop and validate an advanced AM detection method based on a combination of traditional methods, machine learning-based methods and listening tests;
2. To quantify and characterise AM using year-long measurements of acoustical and meteorological data collected from different residential locations near several South Australian wind farms;
3. To investigate the audibility and characteristics of infrasound and modulated tonal noise based on a probabilistic approach and computer simulations;

4. To explore the transfer learning technique in machine learning for WFN quantification and characterisation.

1.5 Thesis outline

The overall structure of the thesis includes four major sections including *Introduction, Methods, Results & Conclusion* as shown in [Figure 1.8](#). The major Result section is divided into subsections, each with their own introduction, methods, results, discussion and conclusion. This reflects that each results Chapter was constructed as a stand-alone document for peer-review publication. The details are as follows:

- Chapter 2 presents how year-long data in the field were collected. This chapter provides details regarding the study region, wind farm characteristics, equipment and setup procedures.
- Chapter 3 presents a machine learning based method to detect AM. Benchmark characteristics of AM were established using a human-scored data set. The results presented in this chapter have been published as ([Nguyen et al., 2021b](#)).
- Chapter 4 used the classifier as developed in Chapter 3 to detect AM in long-term data sets. The relationship between AM characteristics and meteorological conditions was quantified and characterised. Differences between indoor and outdoor AM characteristics are presented and discussed. The results presented in this chapter have been published as ([Nguyen et al., 2021a](#)).
- Chapter 5 provides characteristics of infrasound and AM tones of WFN. In this chapter, a computer simulation was used to assess the audibility of key WFN characteristics at long-range locations. The results presented in this chapter have been published as ([Nguyen et al., 2022a](#)).
- Chapter 6 is an exploratory chapter. In this chapter, deep acoustic features extracted from a deep learning model were investigated to reveal whether they can be used to

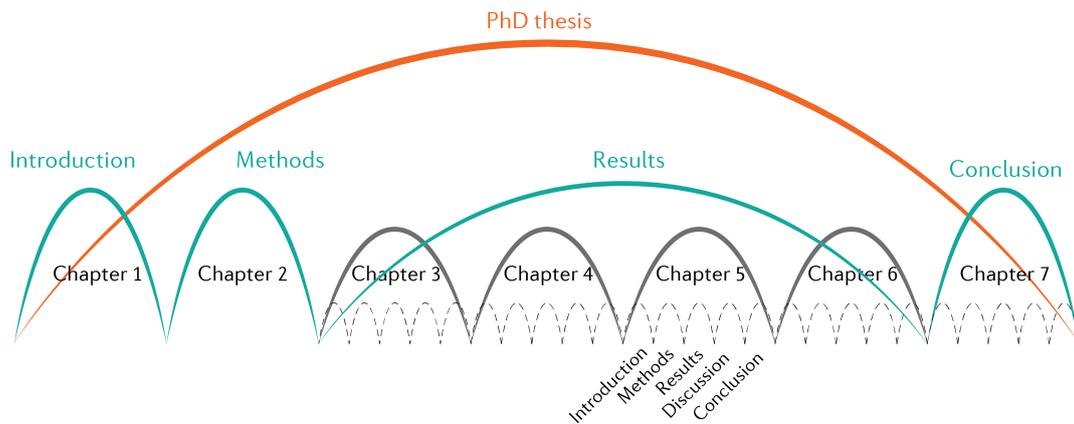


FIGURE 1.8. Thesis outline.

replace traditional acoustic features. The results presented in this chapter have been published as (Nguyen et al., 2022b).

- Finally, Chapter 7 summarise the main contributions of the work presented in this thesis and concludes the thesis as well as discusses future directions for research in this area.

1.6 Research outputs

I have published and submitted several papers as first author during my PhD candidature as follows:

1. **Nguyen, D. Phuc**, Hansen L. Kristy, Lechat Bastien, Catcheside Peter, Zajamsek Branko, and Hansen H. Colin. "Benchmark characterisation and automated detection of wind farm noise amplitude modulation." *Applied Acoustics* 183 (2021): 108286. <https://doi.org/10.1016/j.apacoust.2021.108286>
2. **Nguyen, D. Phuc**, Hansen L. Kristy, Catcheside Peter, and Hansen H. Colin, and Zajamsek Branko. "Long-term quantification and characterisation of wind farm noise amplitude modulation." *Measurement* (2021): 109678. <https://doi.org/10.1016/j.measurement.2021.109678>

3. **Nguyen, D. Phuc**, Hansen L. Kristy, Bastien Lechat, Zajamsek Branko, Hansen H. Colin and Catcheside Peter. "Going beyond traditional wind farm noise characterisation with deep acoustic features." *JASA Express Letters* (2022): 109678. <https://doi.org/10.1121/10.0010494>
4. **Nguyen, D. Phuc**, Hansen L. Kristy, Catcheside Peter, and Hansen H. Colin, and Zajamsek Branko. "Audibility of wind farm infrasound and modulated tonal noise observed at long-range locations." *Applied Acoustics* (2022) <https://doi.org/10.1016/j.apacoust.2022.109106>
5. **Nguyen, D. Phuc**, Hansen L. Kristy, Zajamsek Branko, Catcheside Peter and Hansen H. Colin. "Multi-input model uncertainty analysis for long-range wind farm noise predictions." *Applied Acoustics* (Dec/2021) (SUBMITTED REVISION)

I am a co-author on these following papers, to which I contributed more than 50% towards data analysis and collection:

6. Hansen L. Kristy, **Nguyen, D. Phuc**, Gorica Micic , Lechat Bastien, Catcheside Peter, Zajamsek Branko, and Hansen H. Colin. "Amplitude modulated wind farm noise relationship with annoyance: A year-long field study." *The Journal of the Acoustical Society of America* 150, no. 2 (2021): 1198-1208. <https://doi.org/10.1121/10.0005849>

During my candidature, I also collaborated with psychologists and physiologists to investigate the effects of WFN on sleep in following papers:

7. Tessa Liebich, Leon Lack, Gorica Micic, Kristy Hansen, Branko Zajamsek, Hannah Scott, Nicole Lovato, Claire Dunbar, Bastien Lechat, Felix Decup, **Nguyen, D. Phuc**, Peter Catcheside. "The effect of wind turbine noise on polysomnographically-measured and self-reported sleep latency in wind turbine noise naïve participants." *Journal of Sleep* (2021). <https://doi.org/10.1093/sleep/zsab283>
8. Dunbar, Claire, Peter Catcheside, Bastien Lechat, Kristy Hansen, Branko Zajamsek, Tessa Liebich, **Nguyen, D. Phuc**, et al. "EEG power spectral responses to wind farm

compared with road traffic noise during sleep: A laboratory study." *Journal of Sleep Research* (2021): e13517. <https://doi.org/10.1111/jsr.13517>

9. Sweetman, Alexander, Leon Lack, R. Doug McEvoy, Simon Smith, Danny J. Eckert, Amal Osman, Jayne C. Carberry, Douglas Wallace, **Nguyen, D. Phuc**, and Peter Catcheside. "Bi-directional relationships between co-morbid insomnia and sleep apnea." *Sleep Medicine Reviews* (2021): 101519. <https://doi.org/10.1016/j.smr.2021.101519>
10. Lechat, Bastien, Hannah Scott, Ganesh Naik, Kristy Hansen, **Nguyen, D. Phuc**, Andrew Vakulin, Peter Catcheside, and Danny J. Eckert. "New and Emerging Approaches to Better Define Sleep Disruption and Its Consequences." *Frontiers in Neuroscience* (2021): 1330. <https://doi.org/10.3389/fnins.2021.751730>

I also published a number of software packages that are publicly available on Github⁶ and gave several interviews that have been covered in the mainstream media internationally, featuring in ScienceDaily⁷, EcoVoice, RenewEconomy, CanberraOnlineNews, Techlive, Miragenews, MorningNews and ABC news.

⁶<https://github.com/ducphucnguyen>

⁷<https://www.sciencedaily.com/releases/2021/08/210818130533.htm>

Chapter 2

Field measurement

This chapter presents field work that I conducted from August 2018 to December 2019 under the supervision of my principal supervisor.

This study was approved by the Flinders University Social and Behavioural Research Ethics Committee (SBREC project number 7536). Residents living in the houses where measurements were conducted provided voluntary informed written consent and received a small reimbursement (AUD\$500) to compensate for the time and inconvenience associated with study participation.

2.1 Study region

Measurements were carried out in the mid-north region of South Australia as shown in [Figure 2.1](#), which has a Mediterranean climate with relatively mild winters and hot dry summers. Given the Southern latitude South Australia's seasons are at opposite times of the year compared to those in the northern hemisphere. Summer is from December to February; Autumn is from March to May; Winter is from June to August; and Spring is from September to November. This information is important because seasonal characteristics of WFN in this study were, in [Chapter 3](#), compared with previous studies mainly conducted in the northern hemisphere such as in Finland and Sweden.

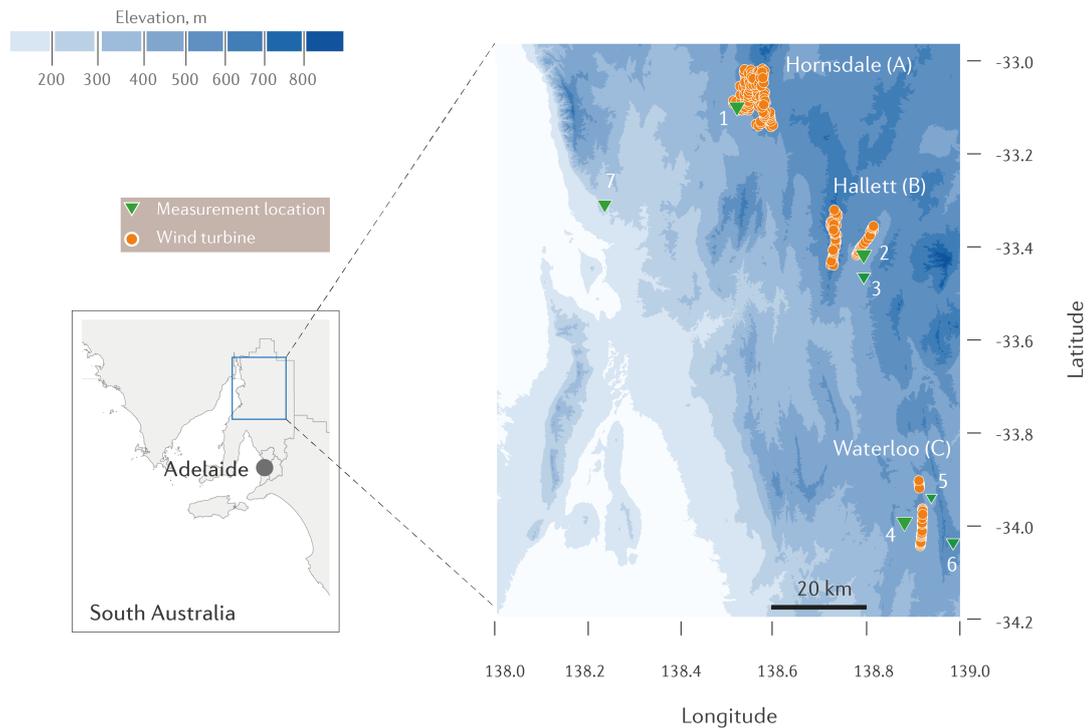


FIGURE 2.1. Study region. The study region is located in the north of South Australia, approximately 3 to 4 hours drive from Adelaide. The average elevation in this region is between 300 m and 600 m above the mean sea level. Elevation data are publicly available from Elvis - Elevation and Depth - Foundation Spatial Data¹.

During the time of measurements, the median difference in temperature between the daytime and nighttime was approximately 10°C, while the difference between summer and winter was greater than 15°C, as shown in Figure 2.2. The maximum temperature was approximately 50°C, while the minimum temperature was slightly below 0°C. The rainfall in South Australia was low, especially during the time of the measurements, during which all months (except August) were classified as dry months (rainfall < 60 mm per month). This condition is ideal for noise monitoring as it minimises the effects of wet conditions on the performance of microphones and windscreens.

¹<https://elevation.fsd.org.au/>

²<http://www.bom.gov.au/climate/data/index.shtml>.

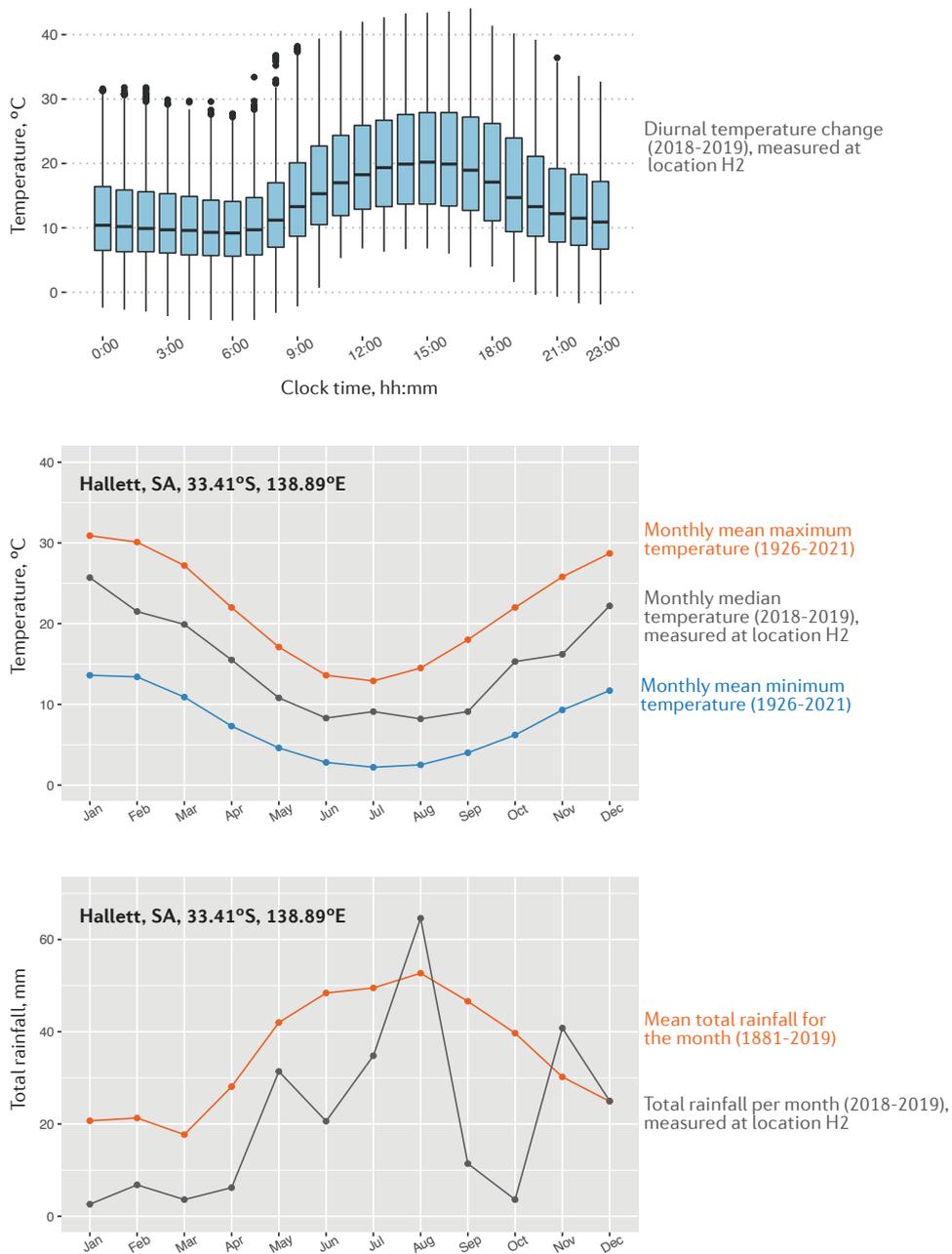


FIGURE 2.2. Weather conditions. Diurnal temperature change, monthly temperature and total rainfall. Long-term temperature and rainfall data were extracted from Bureau of Meteorology Australia².

2.2 Wind farm characteristics

Hornsedale wind farm (hereafter referred to as wind farm A), at the time of measurements, consisted of 99 wind turbines with a nominal capacity of 315 megawatts (MW), as shown in [Figure 2.1](#). Wind farm A consisted of Siemens 3.2-MW SWT-3.2-113 direct drive wind turbine generators. These turbines have a 95 m hub height and 55 m blade radius, making the tip height above ground 150 m. The annual power output during the time of the measurements was 125 MW, which was approximately 40% of the total output capacity, as shown in [Figure 2.3a](#). Measurement location 1 was approximately 1 km from the nearest wind turbine. Nearby wind turbines were at similar elevation levels to the measurement location.

At the time of measurements, Hallett wind farm (hereafter referred to as wind farm B) included four different smaller wind farms. This study only considered the characteristics of the two nearest wind farms (Hallett 1 and 5) to measurement locations 2 and 3, as shown in [Figure 2.1](#). Wind farm B had 1 Suzlon S97 and 69 Suzlon S88 wind turbines, each with nominal output power of 2.1 MW, for a total of approximately 148 MW ([Figure 2.3b](#)). The turbine rotor radius was 44 m, and the hub height 80 m, giving a blade tip height of 124 m. The annual power output capacity during the time of measurements was 36%. These wind turbines were positioned along the top of a ridge, approximately 70 m higher than measurement locations 2 and 3 ([Figure 2.1](#)).

Waterloo wind farm (hereafter referred to as wind farm C) consisted of 37 Vestas V90-3 MW turbines and 6 Vestas V117-3.45 MW turbines, with a nominal power output of 131 MW ([Figure 2.3](#)). The majority of wind turbines had a hub height of 80 m and a turbine blade radius of 44 m and all were located on the top of a ridge. The average height difference between the ridge and measurement locations 4, 5 and 6 was approximately 110 m, as shown in [Figure 2.1](#). Other details regarding the characteristics of wind farm A, B and C are also provided in [Table 2.1](#).

³<https://anero.id/energy/wind-energy/>

⁴https://en.wikipedia.org/wiki/Hallett_Wind_Farm

TABLE 2.1. Characteristics of wind farms.

Name	Hornsedale (Wind farm A)	Hallett (Wind farm B)	Waterloo (Wind farm C)
Nominal capacity (MW)	315	148	131
Turbine size (MW)	3.2	2.1	(3.0 & 3.3)
Turbine model	Siemens SWT-3.2-113	Suzlon S88 & S97	Vestas V90-3.0 & V117-3.45
Operational month and year*	Jul 2016	2012	Oct 2010
Wind farm latitude	-33.058	-33.367	-33.983
Wind farm longitude	138.544	138.728	138.900
Annual output (mean \pm s.d. MW)	125 \pm 97	54 \pm 44	45 \pm 38

* Hornsdale and Hallett wind farms were developed in several phases. Above operational dates are based on a fully operational status.)

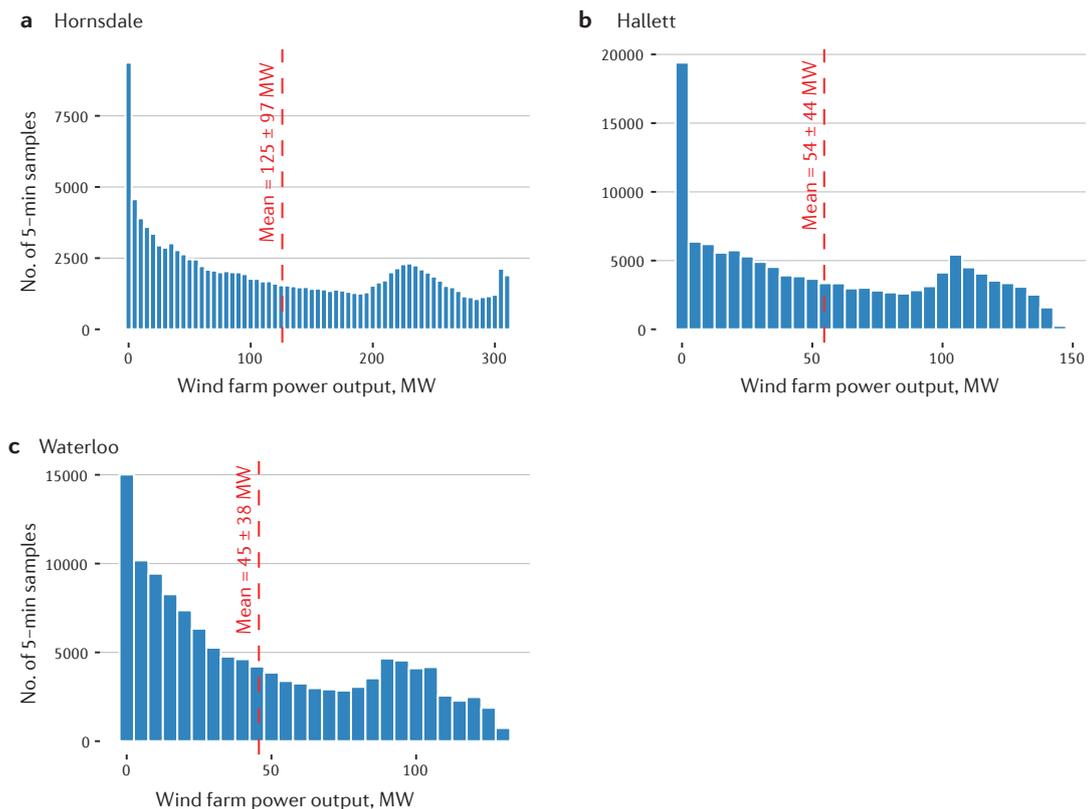


FIGURE 2.3. Wind farm power output. The data used for producing these figures were extracted from Australian Energy Market³ from June 2018 to September 2019. The data used for the Hallett wind farm plot is a summation of Hallett 1 and Hallett 5 wind farm power outputs (see wikipedia⁴ for wind farm details). The histogram bin-width is 5.0 MW. The values indicated on the red-dashed vertical lines indicate mean \pm one standard deviation.

2.3 Measurement location and duration

In this PhD project, I conducted measurements at 6 locations (location 1 - 4, 6, 7) as shown in [Figure 2.1](#), while the measurements at location 5 were conducted during a previous project. The main criteria for selecting these locations were residences located between 1 and 10 km from an operational wind farm. Resident were also required to consent for measurements to be carried out both indoors and outdoors of their residence for one year. The study was advertised via postal flyers that were dispatched to 585 residents living near surrounding wind farms. However, due to limited interest in the study and a limited number of measurement systems, only 6 measurements were conducted in this PhD project. Further details regarding measurement locations and data use in each chapter are summarised in [Table 2.2](#).

In this PhD project, I did not use all data sets described above for every study presented in Chapters 3 to 6. The most used data sets were measured at locations 1, 2 and 4 ([Table 2.2](#)). These data sets were large, and contained good quality outdoor and indoor data. These data sets were thus suitable for investigating the seasonal trends of WFN characteristics. Other data sets were smaller (i.e., locations 5 & 6), highly contaminated by other indoor noise sources (i.e., location 3) or did not contain WFN (i.e., location 7). These data sets were included for particular purposes. For example, the data set measured at location 7 was suitable for testing the false positive rate of the proposed AM detection algorithm as presented in Chapter 3. Therefore, the data selection in this PhD project was based on the suitability of the data sets to address specific research questions most relevant to each chapter. More details regarding data selection criteria are presented in each result chapter.

2.4 Outdoor noise measurements

2.4.1 Measurement setup

Noise was measured at 1.5 m above the ground and at ground level, as shown in [Figure 2.4](#). The microphone at a height of 1.5 m was expected to measure noise corresponding to the ear height of an average person. This microphone was mounted using a star-dropper to minimise

TABLE 2.2. Measurement locations and duration. The check marks indicate data sets included in each results chapter

Wind farm	Location	Distance* (km)	Duration (days)	Chapter 3	Chapter 4	Chapter 5	Chapter 6
A	1	1.0	342	✓	✓	✓	✓
B	2	1.3	391	✓	✓	✓	✓
	3	3.8	296			✓	
C	4	3.5	420		✓	✓	✓
	5	3.5	14	✓			
	6	8.8	70			✓	✓
	7	20	122	✓			

* The distance from the nearest wind turbine to the measurement location.

wind noise interference associated with the more conventional method of tripod mounting. The main disadvantage of measuring at this height is that noise can be contaminated by wind-induced noise, especially at low frequencies. To minimise wind-induced noise caused by wind interaction with the microphone, a microphone at ground level was also used, which was taped horizontally to the centre of a 1 m diameter aluminium plate of 3 mm thickness, as described in [IEC61400-11 \(2012\)](#). The main problem with this setup is that noise levels at ground level may not be representative of noise measurements at the other two measurement heights. This is due to interference between the direct and reflected sound waves, which is affected by the microphone height above the ground. A brief introduction to this phenomenon was provided in Chapter 1, Section 1.2.4. Throughout this PhD project, I predominantly analysed data measured at 1.5 m height. Data measured at ground level were only used cases where data at 1.5 m height were missing or unavailable. Further details regarding missing data and the need for ground versus 1.5 m height collected data are provided in each relevant result chapter.

2.4.2 Microphones

Low-frequency microphones were used for outdoor measurements. These microphones were GRAS 40AZ 1/2" Prepolarized Free-field Microphones, which can measure noise between 0.5 Hz and 20 kHz with an uncertainty of 2 dB. The frequency response of this microphone is shown in [Figure 2.5](#). This low frequency limit of the microphones is particularly important

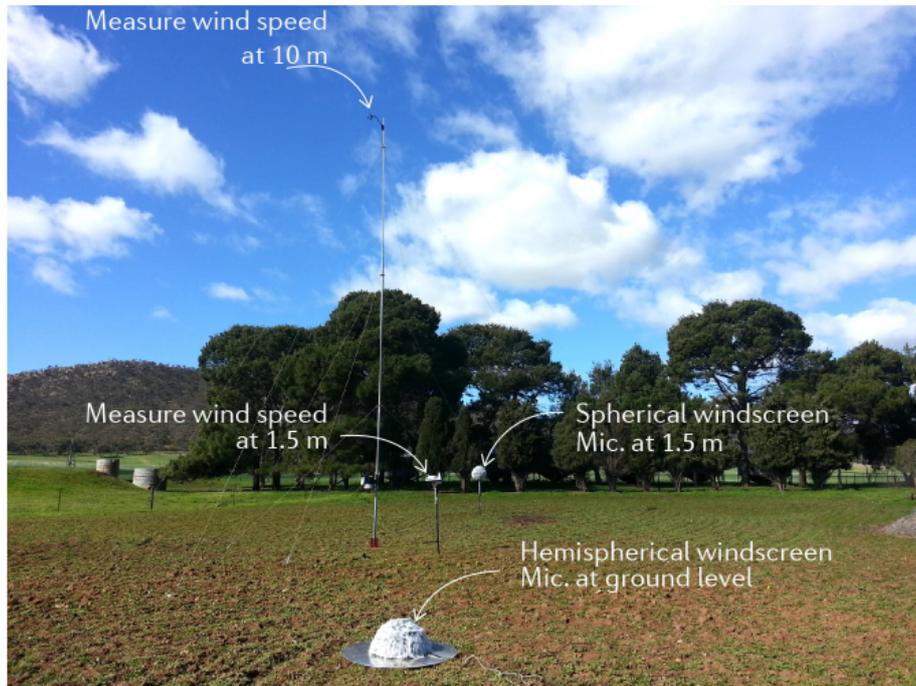


FIGURE 2.4. Outdoor noise measurements. This photo was taken by Kristy.

for capturing infrasonic frequency components of WFN and the fundamental blade-pass frequency, which is normally between 0.7 and 0.8 Hz. The upper frequency limit of these microphones is well above the frequency range of WFN, especially at long-range locations where WFN frequency components are typically below 1 kHz. The microphone is a pre-polarised type that is less sensitive to dust and moisture, which is particularly advantageous for outdoor measurements. The dynamic range of the microphone, which reflects the upper and lower limit of the SPL that the microphone can measure, is between 14 and 148 dB (the reference sound pressure in air used throughout this thesis is $20 \mu\text{Pa}$). The temperature-range of these microphones is between -40 and 120°C , well suited to the measurement region where outdoor temperatures are typically between 0 and 50°C (see Section 2.1 for details).

Time series data from the microphones were recorded using B&K LAN-XI 24-bit data acquisition hardware, which was operated using Pulse software. This system had a flat frequency response between 0 Hz and 51 kHz and a dynamic range between 0 and 160 dB.

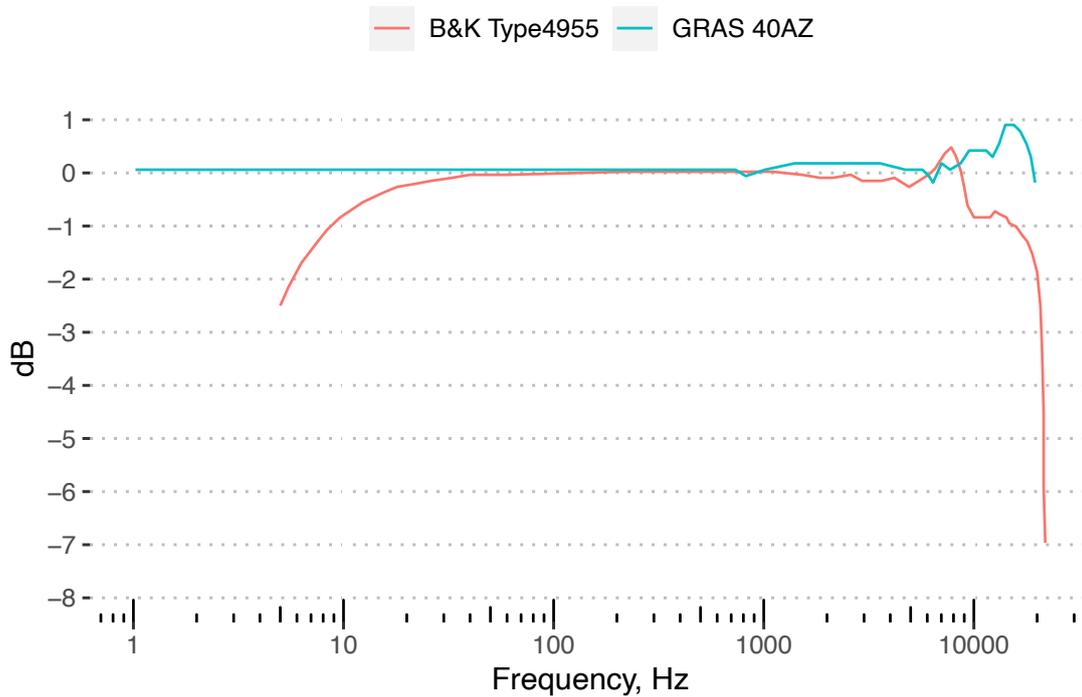


FIGURE 2.5. Frequency response of microphone. Microphone GRAS 40 AZ is a low-frequency microphone that was used for all outdoor measurements. Microphone B&K Type 4955 is a low-noise microphone that was used for indoor measurements (see [section 2.5](#) for details).

Acoustic data were acquired with a sampling frequency of 8,192 Hz at all measurement locations.

2.4.3 Calibration

The microphones used in this PhD project were purchased when the study commenced, and thus accredited calibration was arranged by the manufacturer. To further verify microphone performance, I compared data measured simultaneously using four microphones in a low noise setting. The microphones were placed in a bedroom where two low-noise and two low-frequency microphones were positioned as close to one another as possible to capture room background noise. The purpose of this experiment was to test if the performance of the same type of microphones was consistent. The results showed that microphones of the same type produced very consistent measurements, as shown in [Figure 2.6](#). As expected, the largest difference between two types of microphones was in the frequency range below 10

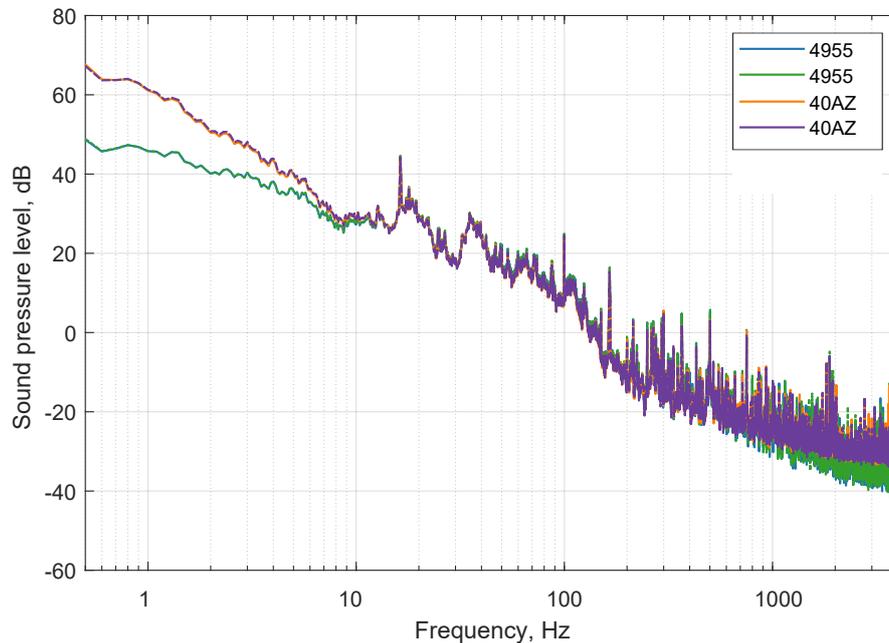


FIGURE 2.6. Microphone performance. Microphone B&K Type 4955 is a low-noise microphone, while microphone GRAS 40 AZ is a low-frequency microphone. The performance of the same type of the microphones was consistent.

Hz. In the frequency range between 10 Hz and 1000 Hz, both types of microphones showed good agreement.

In the field, microphones were calibrated at the beginning and end of each measurement period. At locations 1 and 4, cables were chewed by sheep during the measurements, and thus microphones were calibrated after fixing the cables. The calibrator was a SVANTEK SV 36 Class 1 calibrator, which produces a calibration tone of 94 dB at 1000 Hz. All systems were confirmed to be working properly with no significant drift. Changes in the required adjustment gain were within acceptable limits, equating to a difference of less than 0.5 dB for all microphones.

2.4.4 Wind screens

To minimise the effect of wind-induced noise, I used a standard wind screen consisting of a 90-mm diameter solid sphere of open-cell foam that fitted snugly over the microphone, as recommended in ANSI/ASA [ANSI/ASA \(1993\)](#). However, this small wind screen is only



FIGURE 2.7. Secondary wind screen.

effective for overall A-weighted SPL measurements in wind speeds up to 5 m/s. To further reduce wind-induced noise, I used a secondary wind screen, which has a diameter of 450 mm and was located symmetrically over the smaller 90-mm primary wind screen. The secondary wind screen was handmade and consisted of a thin-wire steel frame of outer diameter 450 mm (supplied by Bunnings⁵), covered by a 16 mm layer of acoustic foam with a porosity of 4 to 8 pores per 10 mm (supplied by Unifilter⁶), and a layer of SoundMaster acoustic fur (Figure 2.7).

2.4.5 Power requirements

In all measurements, recording equipment was powered by a 12 V battery connected to a charger, which automatically charged the battery when the voltage was below 10.5 V. Compared to plugging the equipment directly into the grid power system at the residence, using a battery ensures that the instrumentation is provided with a constant source of power. This is important when measuring noise in rural areas as it can reduce the number of field trips

⁵https://www.bunnings.com.au/gardman-45cm-black-georgian-hanging-basket_p2891375.

⁶<https://www.uniflow.com.au/>.



FIGURE 2.8. Indoor noise measurements.

required to troubleshoot issues associated with an unreliable power supply. This approach also minimises 50 Hz noise associated with mains power.

2.5 Indoor noise measurements

The indoor measurements were similar to the outdoor measurements. The only differences were as follows:

2.5.1 Microphones

Noise levels measured indoors are consistently lower than outdoors so low-noise microphones were used for indoor measurements. The microphones used were B&K Type 4955, which have a dynamic range between 6.2 and 110 dB re 20 μ Pa. The frequency response of these microphones is between 5 Hz and 20 kHz with an uncertainty of 3 dB as shown in

Figure 2.5. The temperature range is between -20 and 80°C. Although these microphones cannot accurately capture the frequency content lower than 5 Hz, this was considered acceptable for indoor measurements because it was a reasonable compromise to allow measurements of the noise down to 6 dB. This was important because all measurements in this PhD project were conducted in rural areas where noise levels are typically low.

2.5.2 Microphone positions

I measured the noise at two positions in a bedroom as shown in **Figure 2.8**. Most bedrooms in the study were unoccupied rooms (except at location 3, where noise was measured in a living room). There was always one microphone in the room corner, which is an anti-node for all room resonant acoustic modes and is therefore the location with the highest associated SPL at all frequencies. The corner was chosen to make sure that it is most out of the way of residents and furthest to indoor noise sources. The other microphone was at a height of 1.5 m above the floor and at least 0.5 m away from walls. These positions are specified in the Danish guidelines (DPA, 1997).

2.5.3 Other setup details

The advantage of the indoor measurements was that the wind-induced noise was minimal. Thus, I only used a 90-mm primary wind screen for these indoor microphones, which was mounted on a tripod instead of a star-dropper as used outdoors, as shown in **Figure 2.8**. Also, a star-dropper is not practical for indoor setup as it could hammer into the floor.

2.6 Weather condition measurements

The local wind speed, wind direction and rainfall were measured concurrently at heights of 1.5 and 10 m using a Davis Vantage Vue and a Davis Vantage Pro weather station (cup anemometer), respectively (**Figure 2.4**). These cup anemometers can measure wind speed and direction with an accuracy of 0.4 m/s and 22.5°, respectively. These cup anemometers

were easy to maintain in the field study, and the coarse resolution of the wind direction was considered adequate to determine if the receiver was in a downwind, crosswind or upwind direction from the wind farm. The anemometer at a height of 1.5 m was placed on a star-dropper, while the anemometer at a height of 10 m was attached to a low cost TV antenna mast adapted to suit the weather station and to facilitate set-up.

2.7 Wind farm operating data

I was not able to obtain rich data related to wind farm operating conditions. The only publicly available data were the total wind farm power output for each 5-minute interval accessed via Australian Energy Market⁷.

⁷<https://anero.id/energy/wind-energy/>

Chapter 3

Benchmark characterisation and automated detection of wind farm noise amplitude modulation

This chapter presents work that I published in [Nguyen et al. \(2021b\)](#). My coauthors were primarily involved in an advisory role so I led and undertook the majority of the work.

Contributions: My contributions to this chapter included:

- I generated a benchmark data set which was scored by human experts to identify the presence or absence of AM.
- I developed and validated an advanced AM detection method with a predictive power close to its practical limit set by human scoring.
- I established benchmark AM characteristics, based on the benchmark data set. The benchmark characteristics were compared with results from previous studies.
- I made all source codes and the benchmark data set publicly available. This is important for reproducibility and future development of this research area.

3.1 Introduction

Amplitude modulation of WFN is a unique feature known to contribute to annoyance (Ioannidou et al., 2016; Lee et al., 2011; Schäffer et al., 2016) and possibly sleep disturbance (Bakker et al., 2012; Liebich et al., 2021a; Micic et al., 2018). AM in the context of WFN is defined as a periodic variation in SPL at the blade-pass frequency (Bass et al., 2016; Hansen et al., 2017), typically between 0.4 and 2 Hz, and is typically most prominent during the evening and night-time when environmental conditions tend to be more favourable for AM (Conrady et al., 2020; Hansen et al., 2019a; Larsson and Öhlund, 2012). AM is a highly variable phenomenon, depending on meteorological conditions (Conrady et al., 2020; Larsson and Öhlund, 2014; Paulraj and Välisuo, 2017), distance from the wind farm and wind farm operating conditions (Hansen et al., 2019a), making AM challenging to detect. Subsequently, characterising AM also becomes a challenging task because it depends on the performance of AM detectors.

Despite the difficulty in detecting AM, this noise phenomenon is commonly detected using simple engineering methods (Hansen et al., 2017) using specific noise features (single predictor). For example, one of the first frequency domain-based methods, as proposed by Lundmark (2011), detected and quantified AM using the AM spectrum of the time variation of instantaneous SPLs. To detect AM in field measurements of wind farm noise, this method was extended by specifying additional criteria such as a valid spectral peak frequency range of 0.6 – 1.0 Hz, and critical values of the maximum spectral peak of 0.4 dB (Larsson and Öhlund, 2014) or 0.6 dB (Conrady et al., 2020). Time domain-based methods typically detect AM using SPL variations, where AM is classified as the difference between the 5th and 95th percentile of SPL greater than 2 dB (Fukushima et al., 2013) or as a peak-to-trough difference of 3 dB or 5-6 dB (Bass, 2011; Cooper and Evans, 2013). Recently, the IOA has developed a hybrid method (Bass et al., 2016), which is a combination of time and frequency domain methods. This method uses the prominence ratio, a ratio of peak and masking level, as a predictor of AM occurrence. The main advantage of these engineering methods is the ease of their implementation and computational speed, which makes them suitable for automated

analysis of large data sets (Conrady et al., 2020; Hansen et al., 2019a; Larsson and Öhlund, 2014). However, evaluation of the performance of these methods is currently limited to false positive rates alone, or to small data sets (Bass, 2011; Bass et al., 2016; Larsson and Öhlund, 2014) or is lacking altogether (Fukushima et al., 2013; Nordtest, 2002).

Detection and quantification of AM using automated detectors has been adopted in many previous studies (Conrady et al., 2020; Hansen et al., 2019a; Larsson and Öhlund, 2012; Larsson and Öhlund, 2014; Paulraj and Välisuo, 2017). This approach is practical and efficient as the analysis of AM is usually implemented on large data sets. In fact, using automated detectors, several unique AM features can be identified and possible associations between weather conditions Larsson and Öhlund (2014), wind farm operation conditions, distances to wind farm (Hansen et al., 2019a), and the diurnal and seasonal variation of AM (Conrady et al., 2020; Paulraj and Välisuo, 2017) can be identified. However, the above AM characteristics and associated variables have been identified based on the assumption that the performance of currently available AM detectors is reasonable.

Human detectors are usually considered as a benchmark (or gold-standard) method for classification tasks which require unique skills to detect target features (Warby et al., 2014). Although this approach is likely impractical to use for detecting AM in year-long data sets, it has some merits. A small subset can be extracted from a large data set using statistical sampling methods (Hastie et al., 2009). If AM samples in this subset are identified by skilled scorers, this information can be used to detect and quantify AM in the large data set. Additionally, this human-scored subset is useful for developing advanced AM detectors such as machine learning methods. In fact, machine learning methods are emerging in many acoustical applications (Bianco et al., 2019) such as noise predictions (Iannace et al., 2019; Valente, 2013), sound propagation (Hart et al., 2016) and source noise classification (Paulraj and Välisuo, 2020; Välisuo, 2017). These methods allow for the combination of multiple, otherwise isolated noise features into one robust classifier. This overcomes one of the major issues associated with traditional AM detection methods, which is the reliance on a single noise feature, which poorly accounts for the highly variable and multifaceted phenomenon of AM (Hansen et al., 2017).

The aims of this study were twofold: (1) to establish benchmark characteristics of AM based on the results of expert human detectors, and (2) to develop an advanced AM detection method based on the benchmark data set. To create the benchmark data set, 6,000 10-sec audio files were randomly extracted from a database including one year measurements at two residences located near different wind farms. AM samples in this subset were then identified by a single scorer using a listening experiment under controlled conditions. Subsequently, the benchmark AM characteristics were established and compared with previous published findings. Finally, using above benchmark data set, an advanced AM detection method was developed which is based on the random forest classification algorithm (Breiman, 2001). Three widely-used AM detection methods (Bass et al., 2016; Fukushima et al., 2013; Larsson and Öhlund, 2014) were also evaluated. In particular, I demonstrate a promising method to reliably establish AM characteristics. Also, the advanced method described in this paper, which is based on a state-of-the-art algorithm, outperformed current methods and is effective for exploration of large wind farm noise data sets.

3.2 Methods

3.2.1 Overview of study region and data collection

The acoustical data sets used in this study were measured at four residences (i.e., location 1, 2, 5 and 7 as shown in Figure 2.1 Chapter 2). To make this chapter is easier to follow, the data measured at location 1, 2, 5 and 7 were renamed as H1 to H4, respectively. The residences located approximately 1.0 km (H1), 1.3 km (H2), 3.5 km (H3) and 30 km (H4) from the nearest wind turbine of South Australian wind farms. These distances are relevant to wind farms in Australia, where residences usually located greater than 1 km from wind farms. Residence H4 was unoccupied and located far away from wind farms, and thus it was assumed that AM did not exist at this location. Noise data were measured for one year at locations H1 and H2 and two weeks and four months at locations H3 and H4, respectively. The H3 data set also contained approximately three days of measurements of background noise when the wind farm was not operating.

The data measured at H1 and H2 were used for establishing benchmark AM characteristics as well as training and validating the AM detection algorithm. The data measured at H3 and H4 were used for false positive rate validation of the proposed AM detection method and previously published methods. The characteristics of wind farms at the time of measurements are shown in [Table 2.1](#) in Chapter 2.

A typical measurement setup included a microphone that was positioned at 1.5 m above ground level (except H1 where ground level microphone was used) and protected using a spherical secondary windshield with a diameter of 450 mm (see [Hansen et al. \(2014b\)](#) for details). The microphone was typically positioned at least 10 m away from the residence and surrounding vegetation to minimize façade reflections and wind-induced vegetation noise. Further details of the experimental setup are described in Section [2.4](#) Chapter 2.

3.2.2 Benchmark data set generation

One benchmark data set contained 6,000 10-second audio files of WFN and the second one of equal size contained no WFN (i.e., environmental background noise only). The first data set was used for establishing benchmark AM characteristics and developing the AM detection method, while the latter data set was specifically constructed for testing false positive detection. These data sets were selected randomly from recorded data using the resampling without replacement technique. The data selection process is shown in [Figure 3.1](#).

The WFN benchmark data set was primarily scored by a single scorer using a validated rating experiment procedure based on detection theory ([Macmillan and Creelman, 2004](#)). The scorer was an acoustician experienced with wind farm noise AM through both field measurements and listening tests. The scorer was also familiar with AM characteristics in the time and frequency domains. Acoustician scorers familiar with the acoustic features of AM were selected to avoid potential confounding and bias by other acoustic and non-acoustic features unrelated to AM through the use of non-acoustician scorers. Intra-scorer variability was validated in which the scorer re-scored a sub-set of the data (100 samples) in a blinded manner. To further evaluate inter-scorer agreement, another skilled scorer also rated a sub-sample of 100 randomly chosen audio samples. These scorers listened

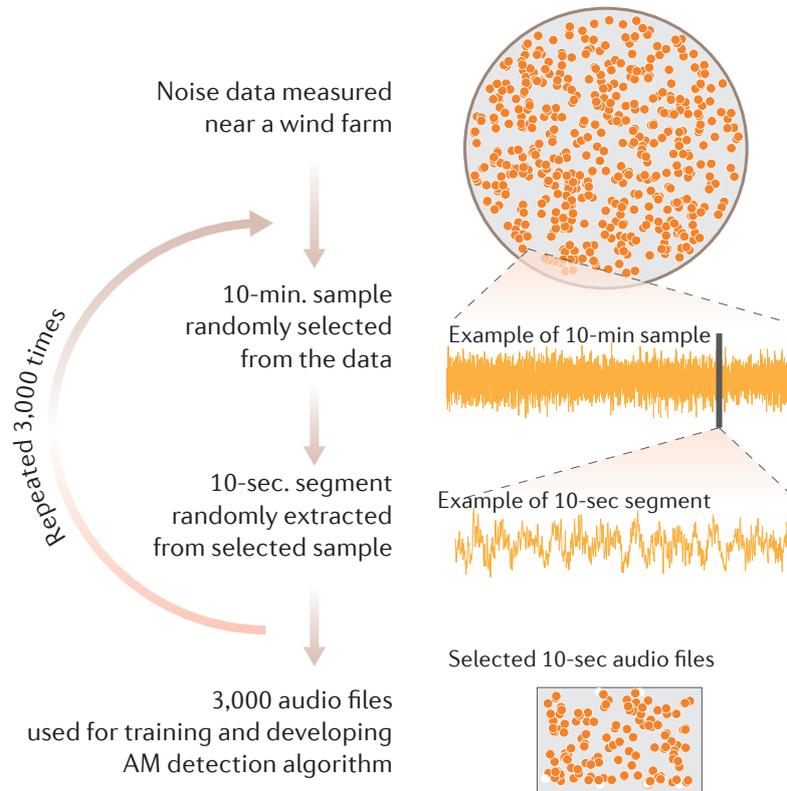


FIGURE 3.1. Data selection process. For illustration purposes, each orange dot inside the top circle represents a 10-minute sample in a measured data set. To extract 10-second audio files, I randomly selected 3,000 10-minute samples from each data set using the resampling without replacement technique (i.e., each 10-minute sample has only one chance to be selected in the data set). There were 6,000 audio files used for training and developing the algorithm (i.e., 3,000 audio files per site). From each selected sample, a 10-second duration segment was randomly selected then extracted. The segments were then converted to audio files (.wav). Exclusion criteria were not specified for the data extraction (i.e., raining, dogs barking and farming machinery noise etc).

to the audio files and scored the presence versus absence of AM. AM presence was rated based on confidence level which varied from high confidence of AM absence (rating '1'), to uncertainty between AM presence/absence (rating '3'), to high confidence of AM presence (rating '5'). For this particular AM identification task, the modulated frequency and duration of AM presence were not identified by the scorer. A MATLAB GUI was designed for the experiment as shown in [Figure 3.2](#). To maximise the performance of detection task, the scorers were allowed to adjust headphone volume level and to listen the audio multiple times before rating. Therefore, AM samples, regardless of their audibility, were detected by the scorer. The visual characteristics of AM were also presented to the scorers, as shown in

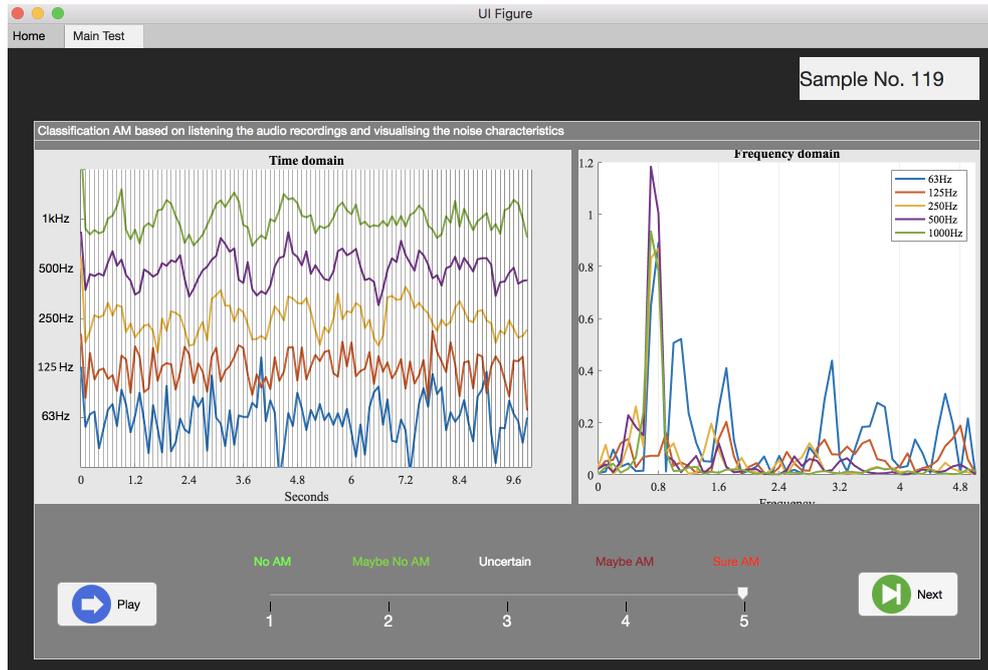


FIGURE 3.2. MATLAB GUI for rating the presence of amplitude modulation audio files

Figure 3.2. This additional information was expected to further improve the scorer’s AM detection performance. The rating experiment was performed in a bedroom at the Adelaide Institute for Sleep Health. The noise reproduction system consisted of Bose Quite Comfort II headphones and a RME Babyface Pro sound card. The background noise in the headphone cavity was approximately 22 dBA during the experiment.

3.2.3 Automated AM detectors

The proposed AM detection method was compared against three previously published AM detection methods. The first method, labelled a1 (Bass et al., 2016), uses a ‘hybrid’ approach involving analysis in both the time- and frequency-domains. The other two methods labelled a2 (Larsson and Öhlund, 2014) and a3 (Fukushima et al., 2013) are implemented in the frequency- and time-domains, respectively. To make these methods consistent, all methods were implemented using audio samples with a 10 second period and a fast time weighting of 100 ms.

Method a1 band-pass filters the signal over the expected AM frequency range, calculates the fast-time weighted SPL time series, detrends the data, then transforms the detrended SPL time series data to the frequency-domain. AM is then detected where the prominence ratio (PR), the ratio between the spectral peak in the blade-pass frequency range and the noise floor, is greater than four (Bass et al., 2016). The pseudo code is provided in algorithm 1.

Algorithm 1: Hybrid AM detection algorithm (Bass et al., 2016)

Data: 10-second audio files
Result: Prominence ratio, present or absent AM
for *band-pass filter between ([50-200 Hz],[100-400 Hz],[100-800 Hz])* **do**
 Apply band-pass filter to the input signals;
 Apply the A-weighting filter to the obtained signal;
 Calculate the fast (100 ms) time-weighted SPL, $L_{Aeq}(fast)$;
 Detrend the SPL signal using a third order polynomial curve fit, resulting in ΔLA ;
 Transform the detrended SPL (ΔLA) to the frequency domain;
 Find spectral peaks located in the range of 0.4 to 1.0 Hz;
 $A = \max$ (peaks) in the range [0.4 to 1 Hz];
 $B = \text{mean}$ (4 spectral lines) which are closest to the max peak location (except the two peaks closet to the max peak);
 Calculate the prominence ratio, $PR = A/B$;
 if $PR \geq 4$ **then**
 | AM present, and the AM depth is calculated ;
 else
 | Report AM absent;
 | Go to the end of the loop;
 end
end

Method a2 is implemented by first applying a low-pass filter at 1 kHz, calculating the fast-time weighted SPL and then transforming this time series into the frequency-domain. The AM_{factor} , the maximum spectrum amplitude between 0.6 Hz and 1 Hz, is then used to obtain the threshold for AM detection. The suggested threshold is 0.4 (Larsson and Öhlund, 2014) (see algorithm 2).

Method a3 is implemented by applying a low-pass filter at 1 kHz and then detrending the fast-time weighted SPL. After quantifying the variation of detrended SPL via calculating the difference between statistical noise levels L_{95} and L_5 , this value, referred to as DAM , is used as a threshold for detecting AM. The suggested threshold varies from 2 dB to 6dB (Bass,

Algorithm 2: Frequency-domain AM detection algorithm (Larsson and Öhlund, 2014)

Data: 10-second audio files
Result: AM factor, present or absent AM
 Apply low-pass filter at 1kHz;
 Apply the A-weighting filter;
 Calculate the fast (100ms) time-weighted SPLs, $L_{Aeq}(fast)$;
 Detrend the SPL signal using a third-order polynomial curve fit;
 Transform the detrended SPL to the frequency domain using an FFT;
 Calculate the amplitude modulation spectrum;
 $AMS = \sqrt{2} * |FFT(detrendedLA(fast))|/N$;
 Calculate the AM factor ;
 $AMfactor = \max(AMS(f))$ with $[0.6 \quad 1Hz]$;
if $AMfactor \geq 0.4$ **then**
 | AM present ;
else
 | AM absent;
end

2011; Cooper and Evans, 2013; Fukushima et al., 2013) (see algorithm 3). Also, the source code for method a1, as provided by Coles et al. (2017) was re-implemented using MATLAB in this study.

Algorithm 3: Time-domain AM detection algorithm (Fukushima et al., 2013)

Data: 10-second audio files
Result: DAM
 Apply low-pass filter at 1kHz;
 Apply the A-weighting filter;
 Calculate slow (1s) and fast (100ms) time-weighted SPLs, $L_{Aeq}(slow)$ and $L_{Aeq}(fast)$;
 Detrend the SPL signal: $\Delta LA = L_{Aeq}(fast) - L_{Aeq}(slow)$;
 Calculate the 95th and 5th percentiles of ΔLA ;
 $DAM = \Delta LA5 - \Delta LA95$;

3.2.4 Random Forest classifier for AM detection

A random forest classifier (Breiman, 2001) consists of decision trees, which represent possible outcome maps for a series of related choices. Decision trees are easy to use and generally

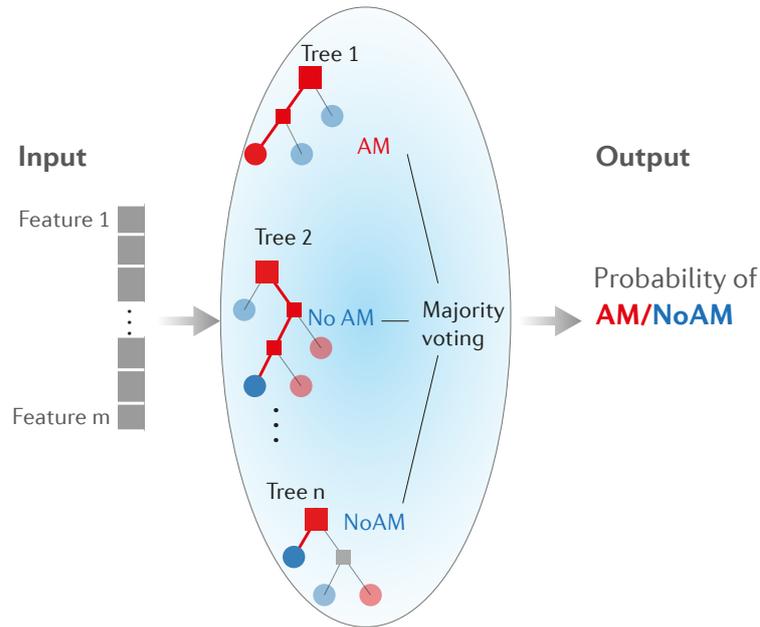


FIGURE 3.3. Random forest classifier

work very well with the data used to create them, but are more problematic for predictive learning models requiring more flexibility for accurate classification of new data (Hastie et al., 2009). To overcome these decision tree problems, the random forest classifier uses bootstrap sampling and random variable selection to build multiple trees (Biau and Scornet, 2016), which are then combined into a random forest classifier as shown in Figure 3.3. To classify an input sample (i.e., AM or no AM), the relevant audio features are plugged into every predictor (tree) in the classifier. Then each predictor classifies the sample as ‘AM’ or ‘no AM’. Finally, a majority voting approach is used to decide if the input audio can be classified as containing ‘AM’ or ‘no AM’. This achieves a probabilistic classifier, where the ratio between the number of trees voting ‘AM’ out of the total tree population represents the probability of AM being present.

Optimisation of hyperparameters, that is parameters which are set before the learning begins, was done using a random searching technique (Bergstra and Bengio, 2012). The following set of hyperparameters were adjusted: number of trees, number of features considered for splitting at each leaf node, maximum number of decision splits, and the minimum

TABLE 3.1. Value ranges of the hyperparameters used for random searching.

Hyperparameter	Range
Num tree	{2, 4, 8, ...1024}
Max num feature	{1, 2, 3, ...31}
Max num split	{2, 4, 8, ...4096}
Max leaf size	{2, 4, 8, ...1024}

number of data points allowed in a leaf node. The random searching technique utilises a range of realistic hyperparameter values, as shown in [Table 3.1](#).

3.2.5 Audio feature extraction

WFN spectra are dominated by lower-frequencies, particularly at distances greater than 1 km from a wind farm ([Hansen et al., 2017](#)). Also, WFN can contain both tonal AM ([Hansen et al., 2019a](#)) and/or broadband AM. Furthermore, AM can occur at frequencies ranging from 30 Hz to more than 1 kHz, and the peak-to-trough magnitude can vary between each successive oscillation period ([Larsson and Öhlund, 2014](#)). To help capture the highly variable and evolving nature of WFN, which likely influences AM characteristics and consequently detection performance, a comprehensive range of 31 noise features were used in this study as shown in [Table 3.2](#). The noise features were divided into four categories, including frequency domain features, overall noise features, time domain features and features extracted from the other automated AM detection methods described in [Section 3.2.3](#).

TABLE 3.2. Feature descriptions

No.	Type	Feature (description)	Ref.
1-13	Frequency domain features	<i>spectralCentroid</i>	Alías et al. (2016); Sharma et al. (2020)
		<i>spectralCrest</i>	
		<i>spectralDecrease</i>	
		<i>spectralEntropy</i>	
		<i>spectralFlatness</i>	
		<i>SpectralFlux</i>	
		<i>spectralKurtosis</i>	
		<i>spectralRollofPoint</i>	
		<i>spectralSkewness</i>	
		<i>spectralSlope</i>	
		<i>spectralSpread</i>	
		<i>pitch</i>	
		<i>harmonicratio</i>	
14-17	Overall noise features	<i>LA</i> (LAeq)	Bies et al. (2017)
		<i>ratioLGLA</i> (LGeq/LAeq)	
		<i>ratioLCLA</i> (LCeq/LAeq)	
		<i>diffLCLA</i> (LCeq-LAeq)	
18-27	Time domain features	<i>peakLoc</i> (Peak location)	proposed
		<i>peakVal</i> (Peak value)	
		<i>posSlope</i> (Mean positive slope)	
		<i>negSlope</i> (Mean negative slope)	
		<i>peakloc_unweightedSPL</i>	
		<i>L1000</i> (Var. octave-band SPL at 1000 Hz)	
		<i>L500</i> (Var. octave-band SPL at 500 Hz)	
<i>L250</i> (Var. octave-band SPL at 250 Hz)			
<i>L125</i> (Var. octave-band SPL at 125 Hz)			
<i>L63</i> (Var. octave-band SPL at 63 Hz)			
28-31	Automated methods	<i>LA</i> (LAeq)	Bass et al. (2016);
		<i>PR</i> (Prominence ratio)	Fukushima
		<i>Fo</i> (Fundamental frequency)	et al. (2013);
		<i>AMfactor</i>	Larsson
		<i>DAM</i>	and Öhlund (2014)

The frequency domain feature categories (feature 1 to feature 13) have been explained in detail in previous reviews (Alías et al., 2016; Sharma et al., 2020) and the pseudo code for extracting these features can be also found in (Sharma et al., 2020). Figure 3.4a shows the process to extract these audio features. A hamming window of 125 ms (50% overlap) is applied to the input signals which are then transformed to the frequency domain using an FFT. The signals are then filtered using bark scale critical bands and the spectral shape features are calculated for each hamming window. The outcome of the process is a matrix (no. of features \times no. of windows). The mean values of the rows in this matrix were calculated, resulting in a single value for each feature. The overall noise feature category (Feature 14 to feature 17) such as A, C and G-weighted SPLs were also extracted as shown in Figure 3.4b. The selected features were $LGeq/LAeq$, $LCeq/LAeq$, and $LCeq - LAeq$, as these measures are expected to be indicative of WFN presence and spectral balance (Hansen et al., 2017; Kelley, 1987; Tokita et al., 1984). The $LAeq$ was selected as it has been used as a metric for analyzing AM in previous studies (Energy, 2014; Larsson and Öhlund, 2014). The time-domain feature category (Feature 18 to feature 27) was extracted as shown in Figure 3.4c. The fast-time weighted SPL (125 ms overlapping 100 ms) was calculated, similar to the method for calculating the prominence of impulsive sounds outlined in Nordtest (2002). The derived SPL (40 Hz sampling frequency) was further smoothed using a moving average window of 5 samples. To estimate AM fundamental frequency of the smoothed SPL, the first derivative of the smoothed SPL was calculated and then transformed to the frequency domain. The highest peak (Feature 19) and its corresponding frequency (Feature 18) of the derivative in the frequency domain were obtained. Also, the average ramp-up and ramp-down of SPL were estimated by calculating mean values of positive and negative values of the derivative signals (Feature 20 and 21, respectively). Using the derivative signals are advantageous because the fluctuation frequency of the derivative signal is similar to the smoothed SPL, while its amplitude is less variant compared to the smoothed SPL. As a result, the blade-pass frequency peaks were clearer in the frequency domain. Feature 22 was calculated in a similar way to feature 18, except using the unweighted SPL. Features 23-27 are variations (calculated as L5-L95) of the octave-band unweighted SPL for octave-band centre frequencies between

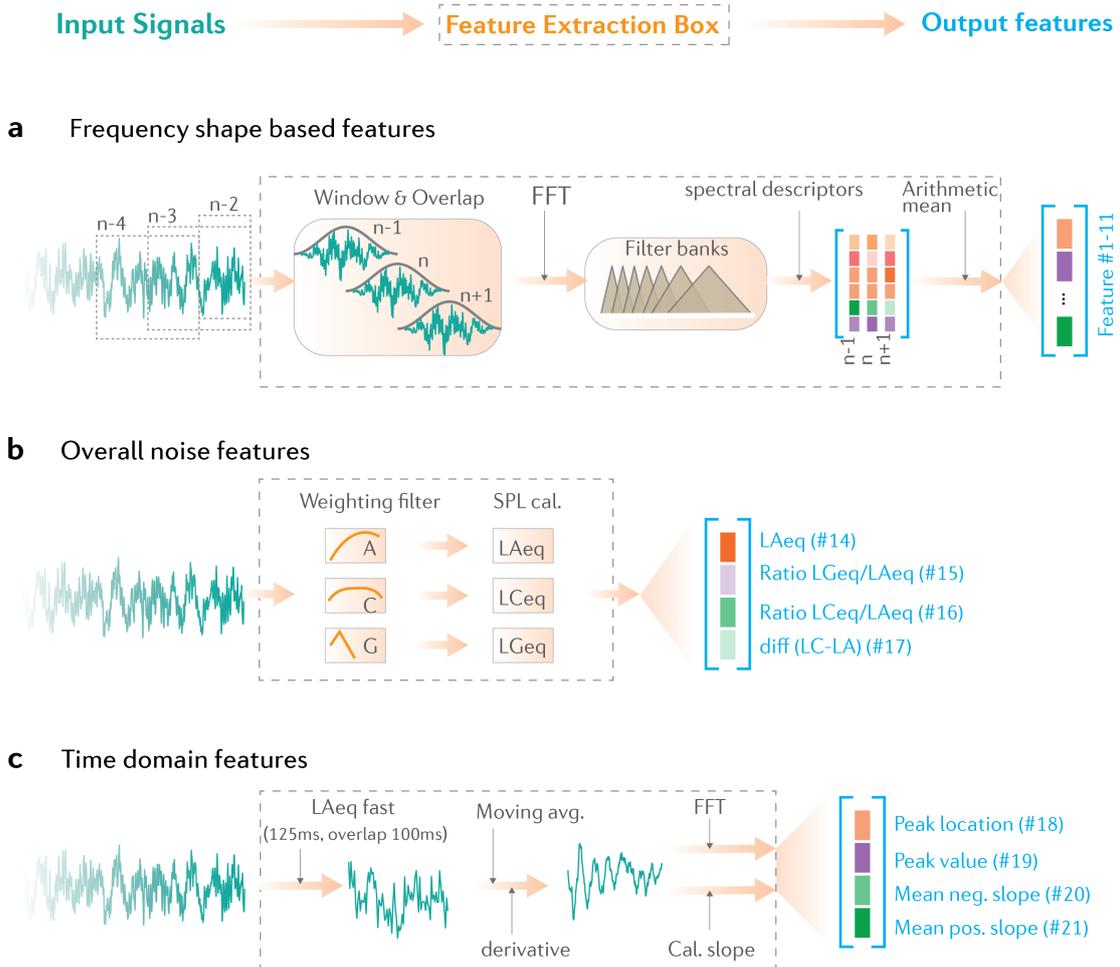


FIGURE 3.4. Feature extraction. **a**, The audio features are grouped into three main groups including features extracted from the shape of the noise spectrum. **b**, Common environmental noise indicators; **c**, Characteristics in the time domain.

63 Hz to 1000 Hz. The automated methods (a1, a2 and a3) were also used as noise features (Feature 28 to 31).

3.2.6 Evaluation metrics

The performance of the automated AM detection methods was evaluated using both a precision-recall curve (PR) and the Matthews correlation coefficient (*MCC*), which are well suited to imbalanced data sets (Lever et al., 2016). To construct the PR curve, pairs (*precision*, *recall*) were calculated from the counts of true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*) as follows

TABLE 3.3. Evaluation metrics: Definitions and equations

Definitions:	
True positive (TP):	Correct detection of AM
False positive (FP):	Incorrect detection of AM
True negative (TN):	Correct detection of no AM
False negative (FN):	Incorrect detection of no AM
Fundamental equations:	
F1-score:	$F1score = 2 \frac{precision \times recall}{precision + recall}$
Specificity:	$Specificity = \frac{TN}{TN + FP}$
Accuracy:	$Accuracy = \frac{TP + TN}{N}$ where $N = TP + TN + FP + FN$
Cohen's kappa:	$k = \frac{TP + TN + FP + FN - Pr(e)}{1 - Pr(e)}$ where
	$Pr(e) = \frac{TP + FN}{N} \frac{TP + FP}{N} + (1 - \frac{TP + FN}{N})(1 - \frac{TP + FP}{N})$

$$recall = \frac{TP}{TP + FN}; \quad precision = \frac{TP}{TP + FP} \quad (3.1)$$

The aggregate metric of the *MCC* is a more informative and faithful score of overall classification performance compared to common metrics such as the accuracy or *F1*-score (Chicco and Jurman, 2020). The *MCC* ranges from -1 (classification is always wrong) to 0 (classification is no better than random guess) to 1 (classification is always correct), and it is calculated as follows

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (3.2)$$

The use of a single metric, and even an aggregate metric like *MCC*, can be misleading without careful inspection of the underlying results. Thus, in this study, additional metrics including Cohen's kappa, accuracy, area under ROC curve, etc., (Lever et al., 2016), were also calculated as secondary results (Table 3.3).

3.2.7 Benchmark AM characterisation

The diurnal and seasonal variation of AM prevalence were compared against previously published AM characteristics obtained using WebPlotDigitizer¹ (Rohatgi, 2017). Specifically, diurnal variation of AM prevalence was extracted from Figures 7 and 8 (Larsson and Öhlund, 2012), Figure 12e (Hansen et al., 2019a) and a mean value of the data in Figures 4a, c and e (Conrady et al., 2020). The seasonal variation data were extracted from Figure 3 ($AM_{0,4}$) (Conrady et al., 2020) and Table 1 (Paulraj and Välisuo, 2017).

3.2.8 Data and statistical analysis

Audio signal analyses were implemented in MATLAB, in which the noise feature extraction was implemented using the Audio Toolbox. The random forest model was implemented using the Statistics and Machine learning Toolbox. Statistical analysis and visualisation were implemented in R programming language². The statistical significance threshold used was $\alpha = 0.05$. All data are reported as mean [95% confidence interval], unless otherwise indicated. The 95% CI range of performance metrics was estimated using a bootstrapping method with 2,000 simulations. Pearson correlation coefficients were used to examine the strength of linear relationships between features and AM quantification metrics.

3.2.9 Data availability

The MATLAB code used to extract features and build the random forest-based AM detection method can be found in the GitHub open repository³. The audio and scored AM files are publicly available at Nguyen (2021a).

¹<https://automeris.io/WebPlotDigitizer/>

²<https://www.r-project.org>

³https://github.com/ducphucnguyen/WFN_AM_Detection

3.3 Results

3.3.1 Benchmark data set

The benchmark data set of 6,000 10-second audio files was unbalanced with around 40% of audio samples containing AM (Figure 3.5a). The AM confidence rating was transformed into a binary score (AM vs. no AM) using a confidence rating threshold of three. Samples with ratings greater than three were classified as AM, and all other samples were classified as no AM. Both positive and negative skewness was observed from the rating distribution, indicating high confidence in scorer rating. The *MCC*, Cohen's kappa coefficient (κ) and *F1*-score for inter-scorer agreement were (0.65 [0.49, 0.80], 0.64 [0.48, 0.8] and 0.77 [0.66, 0.87], indicating a high degree of agreement (Warby et al., 2014) (see Table 3.4 for other metrics). Also, intra-scorer agreement was higher than inter-scorer agreement (*MCC* = 0.71 [0.56, 0.85], κ = 0.7 [0.56, 0.85], *F1*-score = 0.82 [0.71, 0.91]; see Table 3.5 for other metrics). Distributions of scored audio files over months, hours and wind farm power output relative to capacity were also nearly uniform, consistent with ecological validity (Figure 3.5b).

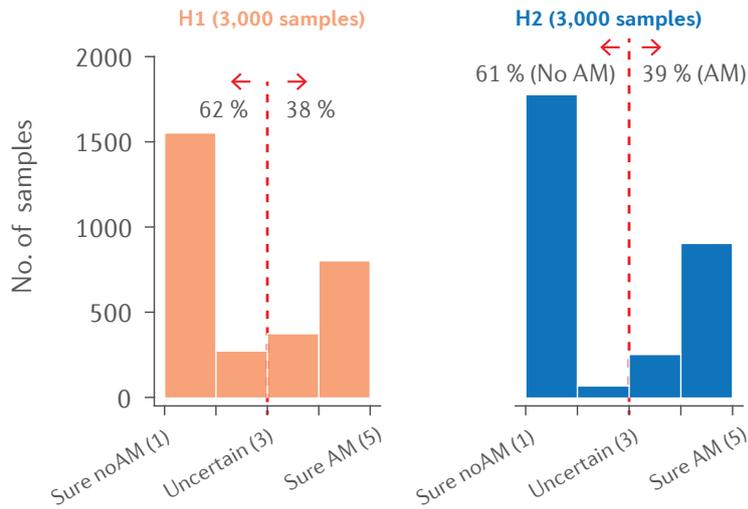
TABLE 3.4. Inter-scorer agreement

Metric	Mean [95%CI]
Recall	0.79 [0.66 0.92]
FPR	0.14 [0.06 0.23]
FNR	0.21 [0.008 0.34]
Specificity	0.86 [0.77 0.94]
Precision	0.75 [0.6 0.89]
FDR	0.25 [0.12 0.4]
FOR	0.11 [0.03 0.19]
NPV	0.89 [0.81 0.96]
Accuracy	0.84 [0.76 0.91]
<i>F1</i> -score	0.77 [0.66 0.87]
<i>MCC</i>	0.65 [0.49 0.8]
Cohen's kappa	0.64 [0.48 0.8]

3.3.2 Benchmark AM characteristics

At the residential locations investigated, which were approximately 1 km from the nearest wind turbine, Less than 10% of the AM samples in the benchmark data set had an associated

a Rating response



b Distribution of benchmark data sets

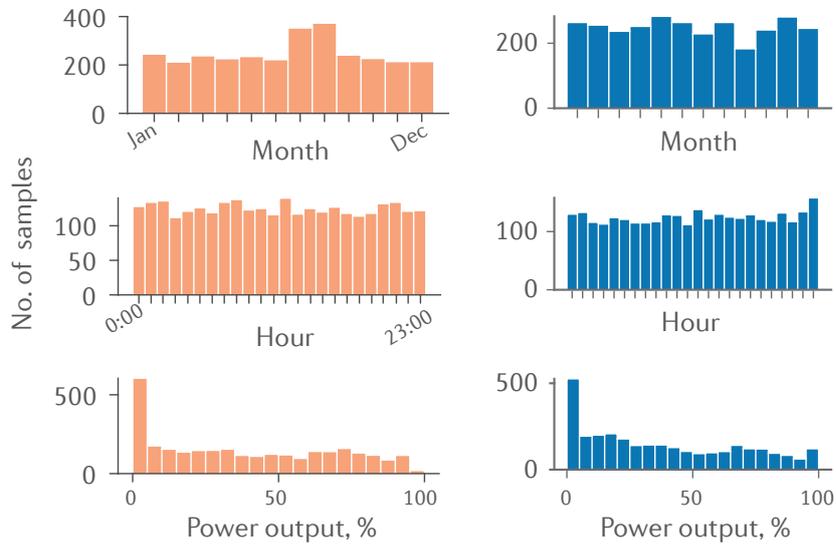


FIGURE 3.5. Characteristics of benchmark data sets. a, scorer rating distribution with corresponding binary classification. **b**, distributions of audio files per month, hour and wind farm power percentage output relative to capacity.

TABLE 3.5. Intra-scorer agreement

Metric	Mean [95%CI]
Recall	0.86 [0.74 0.97]
FPR	0.14 [0.05 0.23]
FNR	0.13 [0.03 0.26]
Specificity	0.86 [0.77 0.95]
Precision	0.78 [0.65 0.91]
FDR	0.24 [0.09 0.45]
FOR	0.08 [0.02 0.16]
NPV	0.92 [0.84 0.98]
Accuracy	0.86 [0.79 0.93]
F1-score	0.82 [0.71 0.91]
MCC	0.71 [0.56 0.85]
Cohen's kappa	0.70 [0.56 0.85]

A-weighted SPL between 30 and 50 dBA ([Figure 3.6a](#)). This supports the feasibility of using a threshold of 30 dBA to trigger AM analysis ([Larsson and Öhlund, 2014](#)), at least for data recorded at similar distances from the wind farm. This could thus reduce false positive rates and/or exclude samples with low SPLs which are likely to be less relevant for assessing community annoyance. I noted that these results can be considered as an upper bound of AM prevalence as both audible and inaudible AM samples were quantified. The audible AM is more relevant to human response to the noise such as annoyance response. The prevalence of audible AM can be determined using the approach proposed by [Hansen et al. \(2019a\)](#) by considering the normal hearing threshold curve.

There are three common metrics (i.e., *AMdepth*, *AMfactor* and *DAM*) to quantify the strength of SPL variations (see Methods and [algorithm 1](#) to [algorithm 3](#) for calculation details). The magnitude of AM hereafter is referred as the AM depth, despite differences in AM depth values obtained using each metric. The distributions of AM depth as quantified by the three metrics are shown in [Figure 3.6b-d](#). More than 50% of the AM samples had an AM depth greater than 2 dBA using the *AMdepth* and *DAM* metrics, which is the fluctuation sensation threshold ([Bowdler and Leventhall, 2011](#)). All three above metrics evaluated AM depth using A-weighted overall SPLs, resulting in an underestimation of the AM depth at low frequencies in comparison to 1/3-octave band-pass filtered data. Although A-weighted SPL aims to replicate AM audibility, this metric is poorly correlated with annoyance and

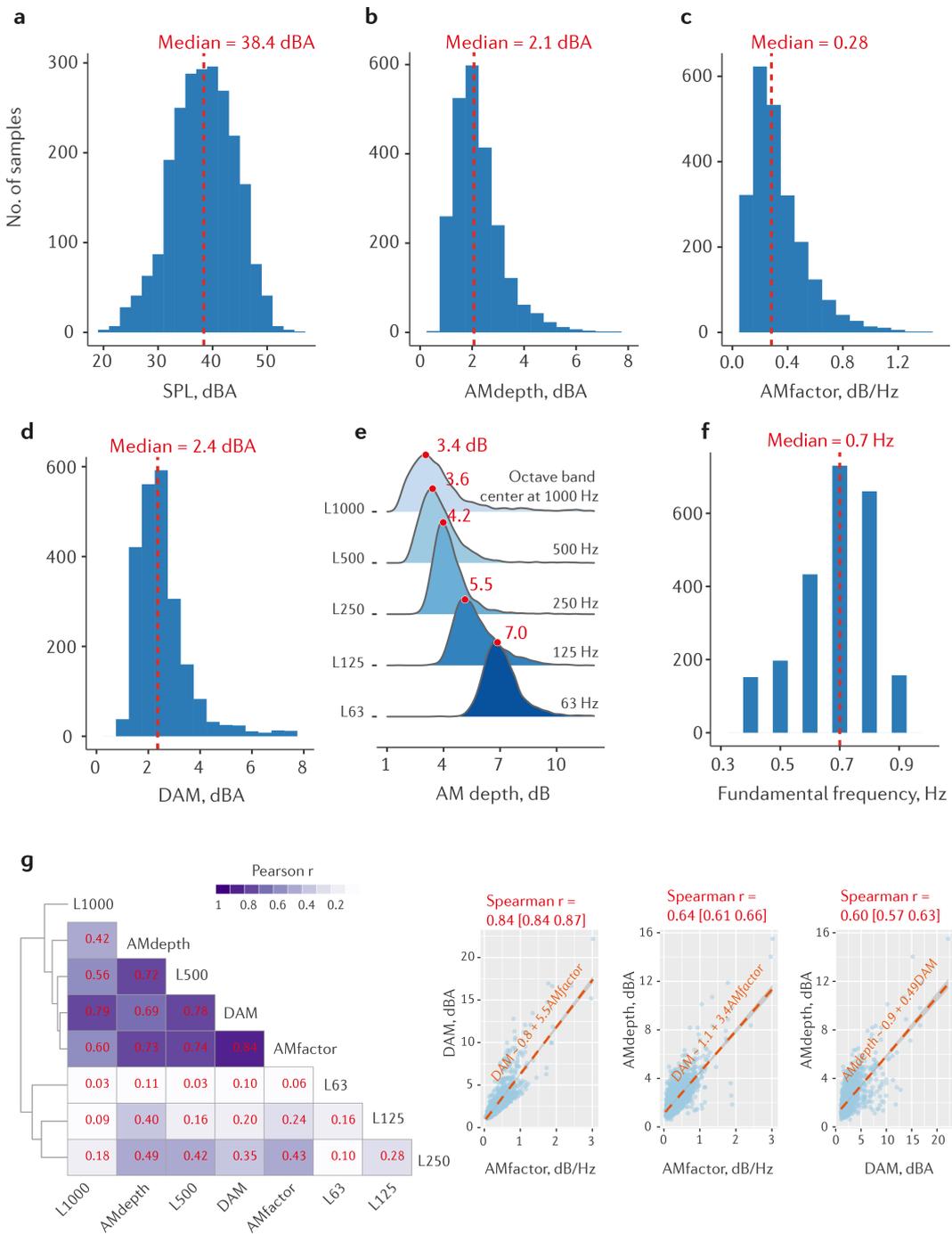


FIGURE 3.6. Characteristics of 2,329 AM samples in the benchmark data set. a, A-weighted SPL. **b-d**, AM depth as quantified using three common metrics. **e**, Distribution of AM depth in each octave band. **f**, Fundamental frequency of AM. **g**, Correlation between AM depth metrics.

perceived loudness (A. Kjellberg, 1984; Bruel, 2001). Furthermore, AM at low-frequencies would be barely discernible in the overall A-weighted SPL due to the high penalties applied at low frequencies. Thus, to better capture AM at low frequencies, data should be analysed in 1/3-octave or octave bands, as suggested by (McCabe, 2011). Future research should focus on suitably designed listening tests to support this approach. The distributions of AM depth as quantified in each octave band from 63 Hz to 1000 Hz are shown in Figure 3.6e, where it can be seen that the AM depth increased for low-frequency bands. The modulation frequency as estimated using algorithm 1 was dominant between 0.6 Hz and 0.8 Hz, accounting for approximately 80% of AM samples. This frequency corresponds to the expected blade-pass frequency when the wind turbines are operating at their nominal speed of 14 to 16 rpm.

The AM depth is one of the most important characteristics of AM, as its magnitude is directly related to the levels of annoyance. Thus, to further characterise AM depth, the Spearman correlation coefficients (Spearman's r) between pair metrics are shown in Figure 3.6g. Two clusters were observed from pair as shown in the dendrogram. The first cluster included the three above metrics with AM depth quantified for mid- to high-frequency bands (i.e., 500 and 1000 Hz). The second cluster included the metrics used to quantify AM depth for low-frequency bands (i.e., 63, 125 and 250 Hz). Additionally, a linear relationship between three common metrics is shown in Figure 3.6g on the left. A strong correlation between these metrics was observed, especially between (*AMfactor*, *DAM*) pair, followed by (*AMfactor*, *AMdepth*) and (*DAM*, *AMdepth*) pair, respectively.

3.3.3 Diurnal and seasonal AM variation

AM appeared to be more prevalent during the evening and night (Figure 3.7a). Previous studies (Conrady et al., 2020; Hansen et al., 2019a; Larsson and Öhlund, 2014) showed that AM occurs approximately 20% to 40% of the nighttime (defined on the basis of 22:00 to 6:00) and around 20% during the daytime. Amplitude modulation was detected using method a1 (for the study of (Hansen et al., 2019a)) and method a2 (for the studies of (Conrady et al., 2020; Larsson and Öhlund, 2012; Paulraj and Välisuo, 2017)). From the comparison of daytime and nighttime, it appears that although the automated detection methods can

TABLE 3.6. Performance of the best classifier on the out-of-bag samples

Metric	Out-of-bag validation value (mean [95%CI])
Recall	0.81 [0.8 0.82]
FPR	0.17 [0.17 0.18]
FNR	0.19 [0.18 0.20]
Specificity	0.83 [0.82 0.83]
Precision	0.7 [0.69 0.72]
FDR	0.3 [0.28 0.31]
FOR	0.11 [0.1 0.11]
NPV	0.89 [0.89 0.90]
Accuracy	0.82 [0.81 0.83]
F1-score	0.75 [0.74 0.76]
MCC	0.62 [0.60 0.63]
Cohen's kappa	0.61 [0.60 0.63]
AUC	0.88 [0.88 0.89]
AUPRC	0.85 [0.84 0.86]

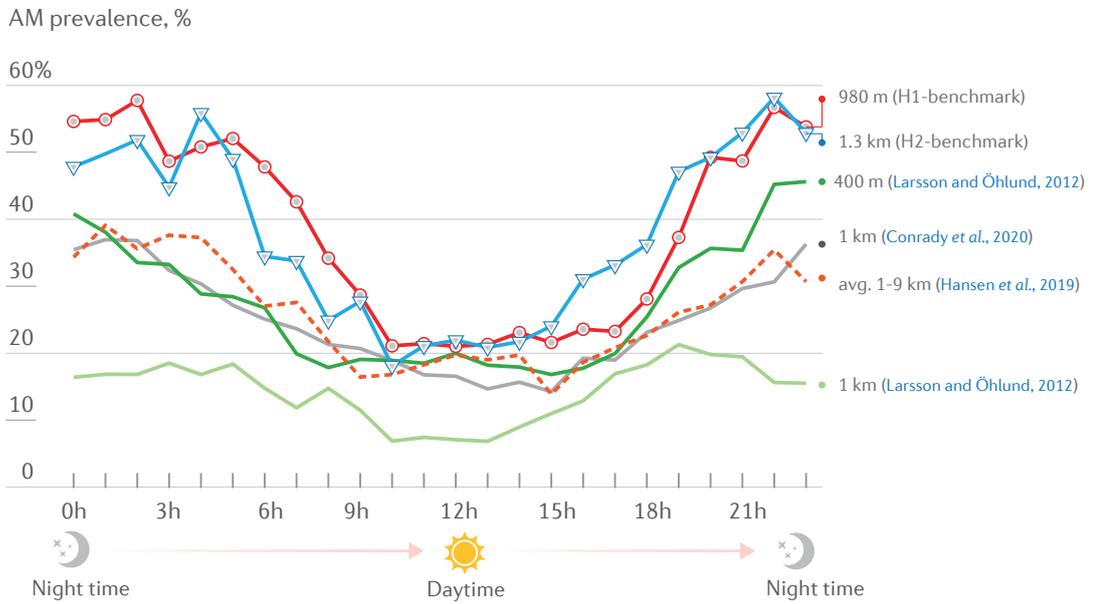
capture a general pattern of diurnal variation, AM prevalence was lower compared with the benchmark data set, especially during the nighttime. Note that AM prevalence was also substantially affected by the difference in meteorological conditions, distance to wind farms, geographical conditions and wind farm layout. On the other hand, seasonal variation is likely to have a negligible effect on AM prevalence, as shown in [Figure 3.7b](#).

3.3.4 Random forest-based AM detection

Hyperparameters were estimated using the out-of-bag samples, which comprised approximately 37% of the total samples not used for training the classifier. The hyperparameters were chosen after 500 iterations by maximising the area under the precision-recall curve (*AUPRC*), ([Breiman, 1996](#)) ([Figure 3.8a](#)). The optimal hyperparameter settings were: 1,024 trees, a maximum of 16 features, a maximum of 2,048 splits and a minimum of 4 samples in the leaf nodes. The precision-recall curve in [Figure 3.8b](#) shows the optimal random forest classifier based on these hyperparameters with *AUPRC* = 0.85 [0.84, 0.86] (see [Table 3.6](#) for other metrics).

Some selected features may not be useful for AM prediction given a cluster of highly correlated variables in the dendrogram (showing the hierarchical relationship between features)

a Diurnal variation



b Seasonal variation

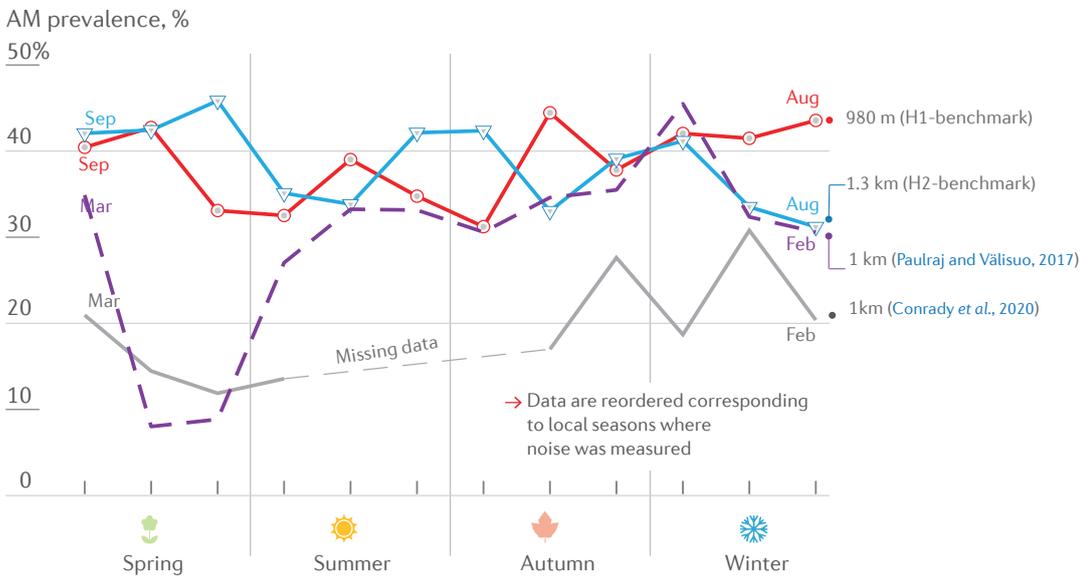


FIGURE 3.7. Variation of AM prevalence. **a**, diurnal variation. **b**, Seasonal variation. Previously published data are from Australia Hansen et al. (2019a), Sweden Conrady et al. (2020); Larsson and Öhlund (2012) and Finland Paulraj and Välisuo (2017).

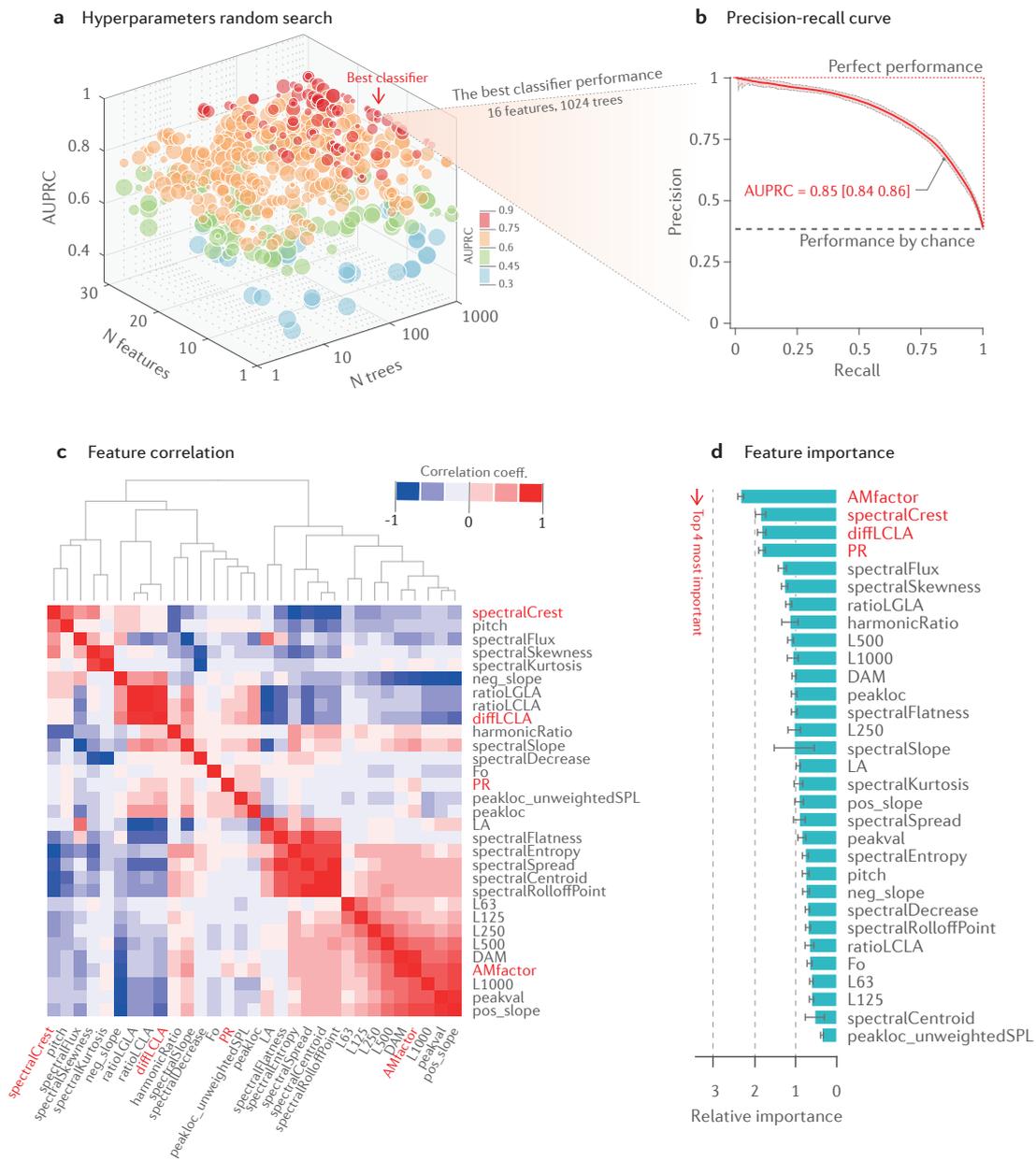


FIGURE 3.8. Random Forest classifier. **a**, hyperparameter tuning using a randomized search technique. The size of the circles represents the maximum splits. Minimum leaf node samples are not shown. **b**, the precision-recall curve of the best random forest classifier. The shaded area indicates 95% CI. **c**, Pearson correlation coefficient (Pearson's r) map with dendrogram for illustrating clusters. **d**, feature importance in descending order from top to bottom. Error bars indicate 95% CI.

and high Pearson correlation coefficient in [Figure 3.8c](#). The four most importance features for predicting AM are *AMfactor*, *SpectralCrest*, *diffLCLA* and *PR* ([Figure 3.8d](#)).

TABLE 3.7. Area under the precision-recall curves and optimal MCC of four methods.

Method	<i>AUPRC</i>	Max <i>MCC</i>
Random forest	0.85 [0.84 0.86]	0.62
a1	0.55 [0.52 0.58]	0.29
a2	0.47 [0.45 0.49]	0.32
a3	0.43 [0.40 0.44]	0.28

3.3.5 Performance of the automated detectors

The performance of the random forest-based AM detection method was compared to three automated detectors (a1-a3) on precision-recall plots (Figure 3.9a). The test set for detectors a1-a3 was all samples in the benchmark data set while the out-of-bag samples were used as the test set for the random forest detector. The random forest-based method outperformed the other methods (ANOVA P -value < 0.001), with an *AUPRC* of 0.85. The performance of a1-a3 was poor with the mean *AUPRC* ranging from 0.43 to 0.55 (Table 3.7). The performance of a1 was better than a2 and a3 (all $P < 0.001$), and a2 performed better than a3 ($P < 0.001$).

The performance of AM detection algorithms has previously been described in terms of the false positive rate (*FPR*) (Bass et al., 2016; Larsson and Öhlund, 2014), and thus this metric was also examined (Figure 3.9b). As the random forest classifier is based on probabilistic values, a threshold of 0.5 was used for binary classification of AM. Thus, if more than 50% of trees in the classifier voted for ‘AM’, the sample was classified as an AM sample, otherwise ‘no AM’ was declared. The cut-of values for method a1-a3 were 4, 0.2 and 2, respectively (See Methods section). The false positive rate of the random forest classifier was low (1.6%) compared to methods a1-a3 (50%, 19% and 62%, respectively). The false positive rate of methods a1 and a3 was not reported in the original descriptions of these methods (Bass et al., 2016; Fukushima et al., 2013), but was reported to be 2.6% for method a2 (Larsson and Öhlund, 2014), and thus substantially lower than in my data set analysed in this study.

To evaluate if the performance of all detectors could be improved using different threshold values, thresholds for each method were varied systematically to find the highest *MCC* values as shown in Figure 3.9c. The optimal threshold for the random forest classifier was 0.44 (44% of trees voted “AM”). The optimal threshold for method a1 was PR=6.7, which

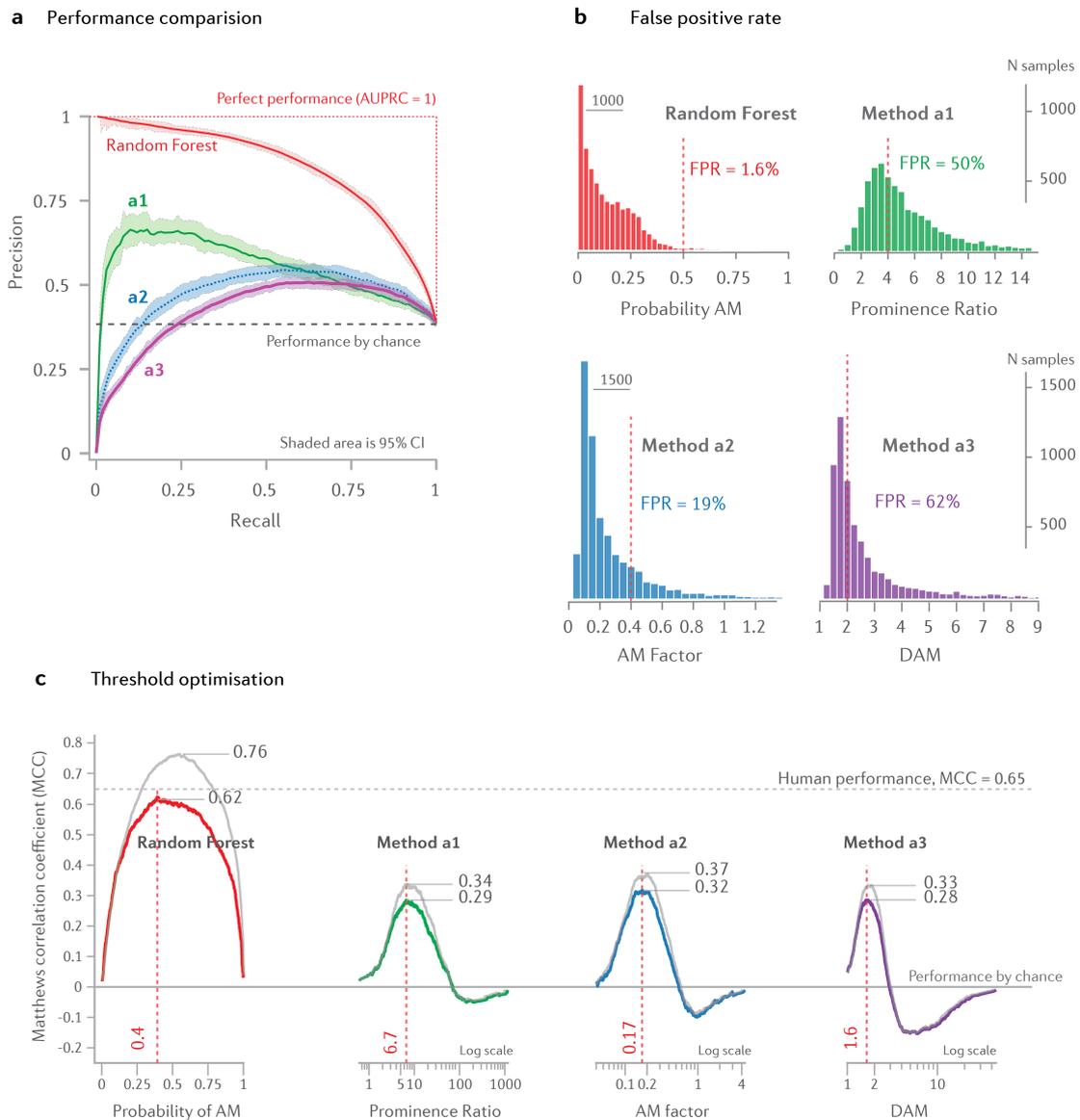


FIGURE 3.9. Performance of automated detectors. **a**, performance using the benchmark data set, where the values associated with each curve are mean [95% confidence interval]. The shaded area is the 95% CI. **b**, false positive rate of each detection method estimated from the no wind farm noise data set. The dashed lines indicate the AM classification threshold. **c**, optimal AM detection threshold according to MCC, where negative values indicate performance worse than by chance. The grey lines are the performance of the proposed algorithm when only samples corresponding with certain responses of the scorer (i.e., samples rated a score of 5 during the manual AM identification process) are used.

is higher than the original reported value of $PR = 4$ in (Bass et al., 2016) and the value obtained using a Receiver Operating Characteristic curve ($PR=3$) in (Hansen et al., 2019a). In contrast, the optimal thresholds for method a2 and a3 were lower than original suggested values (Fukushima et al., 2013; Larsson and Öhlund, 2014). For comparison, the MCC between two scorers was calculated and considered as the ceiling value for the AM detection task ($MCC = 0.65$), supporting that the performance of the random forest classifier was remarkably close to human performance, as evaluated using inter-scorer agreement ($MCC = 0.62$ vs. 0.65). I further investigated if the performance of automated methods could be improved when using only samples corresponding with certain responses of the scorer (i.e., sure AM with responses > 4.5 vs sure no AM with responses < 1.5). The performance of all automated methods increased, especially the Random Forest based method, which showed an approximately 22% increase in performance ($MCC = 0.76$, Figure 3.9c). This was expected as clearer AM or no AM events were likely detected with higher confidence.

3.3.6 Interpretable predictor

The random forest classifier with 31 features and 1,024 trees outperformed traditional detection methods and showed performance comparable with human classifiers. However, random forest classifiers work much like a black box, which is difficult to interpret. The classifier also requires skilled human and computer resources to implement. Given the findings of the importance of *AMfactor*, *diffLCLA*, *SpectralCrest* and *PR* features, this study thus aimed to build a simplified classifier, which can be used as a simpler and more portable classifier for AM detection. This simplified classifier was a single decision tree built from four features, as shown in Figure 3.10. The performance of the single decision tree showed $AUCPR = 0.68$ [0.64, 0.71] (accuracy = 82%), which is lower than the random forest classifier, yet still higher than methods a1-a3. These results further illustrate that a simple combination of several features outperforms traditional single feature detection methods.

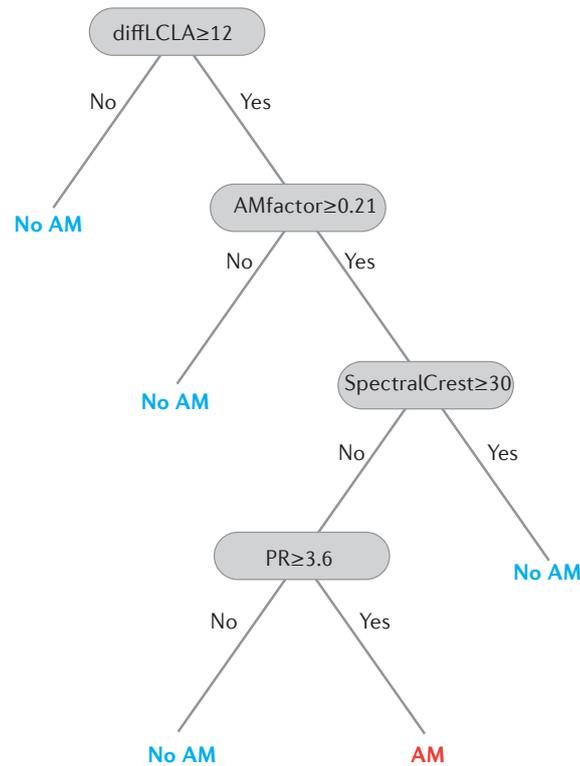


FIGURE 3.10. Performance of automated detectors. A simplified single tree classifier utilising the four most important features for identified by the random forest classifier AM detection. The order of the decision tree is optimised by the algorithm to obtain the highest accuracy.

3.4 Discussion

In summary, I presented a new and promising approach to characterise AM in a large data set using an expert human scoring method. The resulting estimates of benchmark AM characteristics such as AM depth, frequency, and diurnal and seasonal variations are important for validation and calibration of the results using automated methods. I further showed that it is possible to develop an advanced AM detection method with a predictive power close to its practical limit set by human scoring. This approach shows major promise as an effective automated tool which could be used for detecting WFN AM presence in large data sets, such as for research or to support wind farm noise regulations.

Although AM identification by humans was a benchmark approach to establish high quality scored data, it is striking to find that an advanced machine learning algorithm performed close to the human limit. In fact, AM is a challenging signal to detect, as its

characteristics vary depending on meteorological conditions. As a result, the spectral content and time varying features are not constant. Despite these changes, the human auditory system can still recognize the presence of wind farm AM. Thus, my presented algorithm sought to incorporate the most important acoustical features predictive of human scored AM. The selected features cover the whole range of the most dominant WFN characteristics, including noise level variation (or AM), tonality and low-frequency content. Two of the features incorporate noise level variations (*AMfactor* and *PR*); the difference between *LCeq* and *LAeq* is an indicator of low-frequency noise presence; and the spectral crest provides a simple measure of tonality. My findings support the idea that human perception of AM is more complex than assumed by previous AM detection methods that are based on noise level variations alone. Hence, it is not surprising that the method presented here achieved substantial improvements in performance compared to previous methods.

Very high false positive rates were found for methods a1-a3, which is inconsistent with previous reports in (Bass et al., 2016; Larsson and Öhlund, 2014). However, it is worth noting that method a1 was originally designed and evaluated on 10-minute samples, as opposed to the 10-second samples used in my work, and method a1 classifies AM if more than 50% of 10-second blocks within 10 minutes contain AM. By introducing the above criterion, the false positive rate may be substantially reduced, as reported in (Bass et al., 2016). However, 10-second long samples appear to have higher validity, as typical AM events usually last around 10-15 seconds (Larsson and Öhlund, 2014). With regards to the false positive rate for method a2, an arbitrary 30 dBA L_{Aeq} cut-off was imposed in the original evaluation, which was not used in my study, and likely helps to explain the large discrepancy between the originally reported 2.6% (Larsson and Öhlund, 2014) and the 19% false positive rate in my study. If the 30 dBA cut-off is applied to my data before method a2 is used to detect AM, the false positive rate is reduced from 19% to 9%. This number is expected to further reduce if data were measured in a quiet area, where many samples would have associated noise levels less than 30 dBA. Therefore, my findings further support that false positive rate metrics are problematic for evaluating detection performance (Warby et al., 2014), as this only represents one parameter in a confusion matrix.

A limitation of the present study is the under-representation of noise data measured greater than 1 km. As a result, the benchmark AM characteristics are not relevant at other distances. The proposed classifier also may not work well for detecting AM measured several kilometers from the nearest wind turbine, where AM may have different characteristics (Hansen et al., 2019a). The classifier could not be tested on data sets measured outside of South Australia, where weather conditions and topography near wind farms will inevitably vary. Using a single scorer to classify the AM is not ideal as the perception of the noise varies between individuals (Maffei et al., 2015). Although human scoring is a subjective process, it is still considered the most common and reliable method to established human-labelled scored data sets (Gemmeke et al., 2017). I used a single scorer to identify the presence of AM to minimise inter-scorer variability effects which are typically higher than intra-scorer variability. Nevertheless, it remains unclear how generalizable these findings may be to AM more broadly, for which inter-scorer differences as well as noise source and climatic effects could be important. As suggested by Wendt et al. (2015), two or more scorers and a consensus scoring approach may be preferable to a single scorer to help ensure broader generalisability. Future studies should examine if residents living near wind farms identify AM similarly to acoustician and algorithm scored AM, and how strongly AM identification ratings are related to annoyance ratings. Nevertheless, a single scorer is more practical and avoids the potential effects of poor inter-scorer agreement. Also, good inter-scorer agreement was found in a smaller subset of the data, supporting this approach.

Although detector a1 clearly warrants improvements in order to increase accuracy, the source code (Coles et al., 2017) is readily available, making it easy to understand the methodology and to implement the method. Although the other methods were reproduced as closely as possible, my codes may be different from the original codes. This is a similar problem previously identified for the reproduction of the tonality assessment code in Søndergaard et al. (2019). Thus, depositing source code to open source repositories, together with relevant data sets would greatly advance the development of practical and robust amplitude modulation detection methods.

3.5 Conclusions

In conclusion, this study demonstrated that human scoring is a feasible and promising approach to identify AM. This approach is invaluable for detecting unique characteristics of wind farm noise in cases where the performance of automated detectors is low or not validated. The advanced AM detector based on the random forest approach demonstrated high performance, and substantially outperformed traditional AM detection methods to achieve a classification performance close to that of humans. It was also shown that a simplified classifier based on a single decision tree using the four main features identified through the random forest approach also achieved good classification performance. This approach is readily interpretable and easy to implement without the need for extensive computer resources. I hope that, in the future, further insight into the prevalence of AM and associated meteorological conditions, and impacts on humans will help to explain underlying noise generation mechanisms relevant to human perception. Ultimately, this will improve the design of wind turbines such that they are less disturbing and hence, more acceptable to surrounding communities.

Chapter 4

Long-term quantification and characterisation of wind farm noise amplitude modulation

This chapter presents work that I published as first author in [Nguyen et al. \(2021a\)](#). My coauthors were primarily involved in an advisory role so this is predominantly my work.

Contributions: My contributions in this chapter include:

- I showed comprehensive AM characteristics including day/night prevalence, overall depth, octave band depth, duration, modulated frequency and relationships with SPL.
- I quantified the outdoor-to-indoor reduction of AM prevalence and depth. I also showed relationships between these characteristics and the overall noise reduction levels.
- I found that AM occurred most frequently at nighttime compared to the daytime, with a substantial increase in AM prevalence after sunset.
- I illustrated that AM prevalence is high during both crosswind and downwind directions depending on the distance between measurement locations and the noise source wind farm.

4.1 Introduction

Wind energy is one of the fastest-growing renewable energy sectors in the world (GWEC, 2019), reaching approximately 870 GW in 2021 ((WindPower, 2021). Despite the benefits of wind energy, some concerns remain regarding social (Kitzing et al., 2020; Krohn and Damborg, 1999; Wolsink, 2007), ecological (Schuster et al., 2015; Thaker et al., 2018) and environmental impacts (Vautard et al., 2014; Zhou et al., 2012). The noise generated by wind turbines is a recurring source of complaints regarding annoyance and potential sleep disturbance from residents living near wind farms (Liebich et al., 2021a; Micic et al., 2018). WFN contains unusual spectral and time-varying features that may exacerbate annoyance (Perkins et al., 2016) and increase loudness (Jurado et al., 2019), including a low-frequency dominated spectrum (Ingieiewicz et al., 2014; Zajamsek et al., 2016), tonality (Liu et al., 2012) and AM, which is a periodic variation of the noise level primarily related to blade rotational effects (Hansen et al., 2017). Wind farm AM is commonly described as ‘swish swoosh’ or ‘rumble’ and is of particular research interest due to its propensity to contribute to annoyance (Ioannidou et al., 2016; Lee et al., 2011; Schäffer et al., 2016) and possible sleep disturbance (Smith et al., 2020). However, its characteristics such as depth (or degree), duration, consistency and occurrence time could vary between wind farms (Zagubień, 2018).

Previous long-term WFN measurements found wind farm AM to be associated with wind direction (Conrady et al., 2020; Larsson and Öhlund, 2014; Paulraj and Välisuo, 2017), sound speed gradient, solar elevation angle, turbulence intensity (Larsson and Öhlund, 2014), and diurnal meteorological variations (Conrady et al., 2020; Hansen et al., 2019a). The majority of these studies were carried out in cold climates where ground cover with snow during winter months and other climactic effects are clearly different from warmer climates without snow. Snow covered ground has a very high sound absorption coefficient, even at very low frequencies, and thus attenuates noise much more effectively than other ground surface types (Bies et al., 2017; Hansen et al., 2017; Ostashev and Wilson, 2015). Previous long-term studies (Conrady et al., 2020; Larsson and Öhlund, 2014) recorded only low time and frequency resolutions of acoustic data such as 1/3-octave bands or fast time-weighted

SPLs which limited analyses to conventional AM detection methods (Bass et al., 2016; Larsson and Öhlund, 2014) unable to reliably detect AM. Long-term quantification of AM has been predominantly carried out at distances of 1 km or less from wind farms, where WFN is dominated by mid to high frequencies (> 200 Hz). At larger wind farm setback distances, much more typical for Australia, AM is dominated by lower frequencies (< 200 Hz) (Hansen et al., 2019a). However, to date, low-frequency AM has not been systematically studied over a long period of time. Furthermore, although indoor WFN noise character is much more relevant to human perception, annoyance and sleep disturbance than outdoor levels, long-term characterisation and quantification of indoor AM has not been attempted to date, especially at long-range distances to wind farm.

The purpose of this study was to quantify and characterise AM, and to examine relationships between AM, meteorological conditions and wind farm operational data over one year. To detect AM, I used a previously developed AM detection method based on machine learning (Nguyen et al., 2021b). This allowed for accurate and reliable detection of AM in three long-term data acoustic sets measured near South Australian wind farms at locations up to 3.5 km away from the nearest wind turbine.

4.2 Methods

In this section, I provide key information regarding measurement locations and wind farm characteristics. Other details such as equipment and measurement setup were provided in Chapter 2.

4.2.1 Study region

Measurements were conducted in the mid-north region of South Australia (Figure 4.1a and Figure 2.1), which has a Mediterranean climate with relatively mild winters and hot dry summers (Figure 2.2). Noise was measured both outdoors and indoors at three residential houses (i.e., location 1, 2, and 4 as shown in Figure 2.1 in Chapter 2). In this chapter, these locations were renamed as H1, H2 and H3 for convenience, respectively. These measurement

locations were located between approximately 1 and 3.5 km from the nearest wind turbines of three wind farms. The selected wind farms included one with nearby turbines positioned at similar elevation levels as the residence (Wind farm A, [Figure 4.1a](#)), and two with all turbines positioned along the top of ridges (Wind farms B and C, [Figure 4.1a](#)). The average height difference between each ridge and residence was 70 m for Wind farm 1 and 110 m for Wind farm 3. The chosen wind farms layouts, turbine types and total power capacity are presented in [Figure 4.1a](#) and [Table 2.1](#).

4.2.2 Experimental design

WFN noise was measured for more than one year, from May 2018 to August 2019. The outdoor noise measurement systems consisted of two low-frequency microphones located at ground level and 1.5 m above ground ([Figure 4.1b](#)). Indoor noise was measured at the top and bottom wall corners ([Figure 4.1c](#)). The room dimensions and constructions were showed in [Table 4.1](#). Indoor noise levels in rural area are normally low ([Hansen et al., 2014c](#)), and thus, low-noise microphones were used for indoor measurements. At location H1, the outdoor data measured at a height of 1.5 m were not available, and thus data measured at ground level were used. Although the noise measured using a microphone mounted at ground level and 1.5 m is not exactly the same ([Hansen et al., 2014b](#)), the use of 1.5 m data was still reliable, particularly for AM quantification. The agreement between AM detection results in both data sets was high (accuracy = 0.82, F1-score = 0.78, AUC = 0.9) ([Figure 4.2](#)). The AM prevalence quantified using data measured at ground level may be higher compared with 1.5 m height ([Table 4.2](#)) most likely due to lower wind-induced noise at ground level.

Local wind speed and direction were measured concurrently at 1.5 and 10 m. Although a general relationship between local wind speed and wind farm power output was observed ([Figure 4.3](#)), this relationship is highly uncertain. The wind speed and direction accuracy of these weather stations is 0.4 m/s and 22.5 degrees, respectively. The coarse resolution of wind direction was considered adequate to determine if the receiver was in a downwind, crosswind or upwind direction from the wind farm.

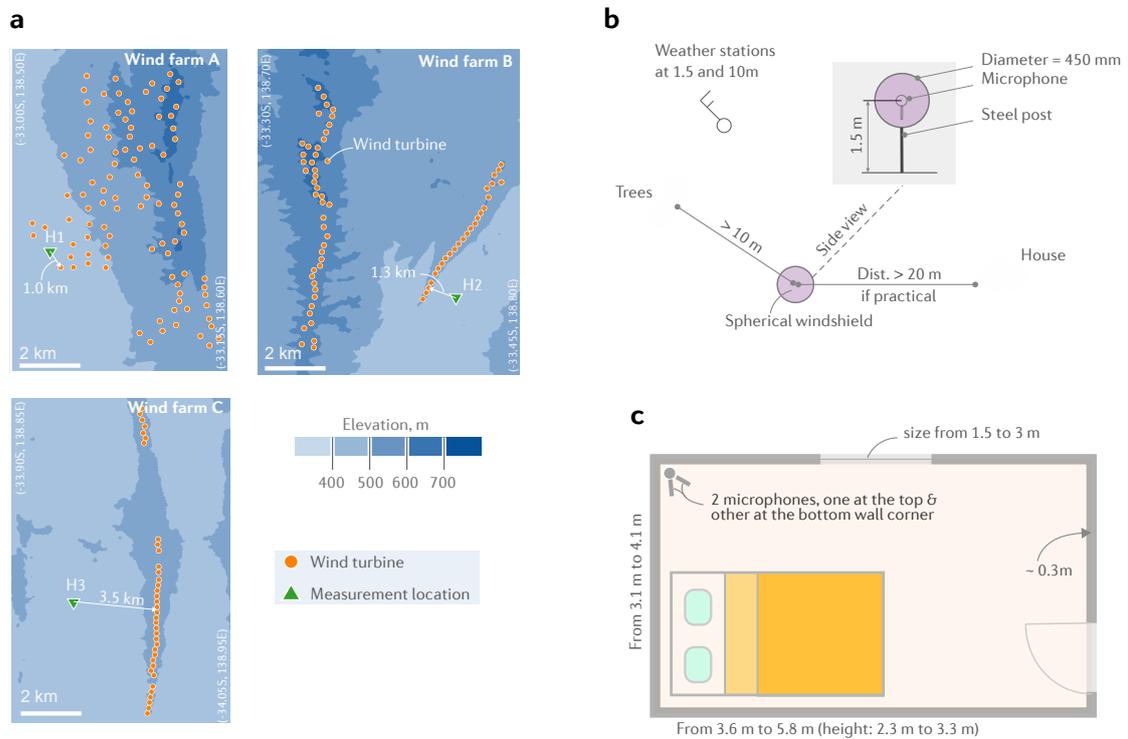


FIGURE 4.1. Performance of automated detectors. (a) Wind farm layouts and measurement locations. (b) and (c) Typical outdoor and indoor microphone position set-up.

TABLE 4.1. Residential house dimensions and constructions at three measurement locations.

Location	Room dimension*	Construction
H1	3.7 × 5.8 × 2.3	Timber walls; single glazed windows; pink batts ceiling insulation; concrete floor.
H2	3.1 × 3.6 × 3.3	Red brick walls (300 mm); single glazed windows (installed double glazed windows during the measurement); pink batts ceiling insulation; wooden floor.
H3	4.1 × 4.3 × 3.2	Thick stone/cement brick walls (350 mm); small-medium single-pane wood-framed sash design windows; corrugated sheet steel roof; plaster panel ceiling.

* Dimension is length × width × height in m.

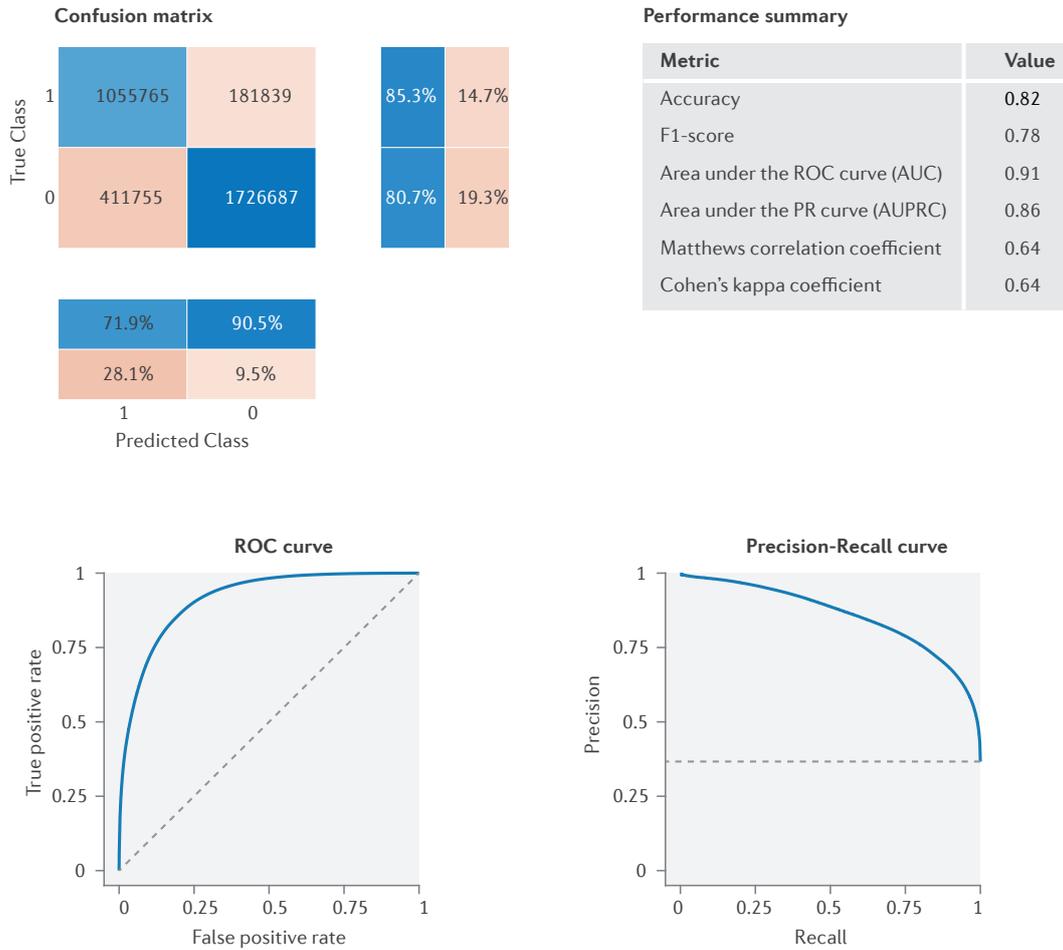


FIGURE 4.2. Agreement between microphones at 1.5 m height and at ground level in terms of detecting AM. The confusion matrix: true class is AM samples identified using data measured at 1.5 m and the predicted class is AM samples identified using data measured at the ground level.

In addition to acoustic and meteorological data, wind farm power output capacities and digital elevation data were obtained from online data repositories (see Section 4.2.7).

TABLE 4.2. Percentage of detected AM in both 1.5 m and ground level data.

Measure position	Class	Count (10-sec sample)	Percent
At 1.5 m	No AM	2,138,442	63.34
	AM	1,237,604	36.66
At ground	No AM	1,908,526	56.53
	AM	1,467,520	43.47

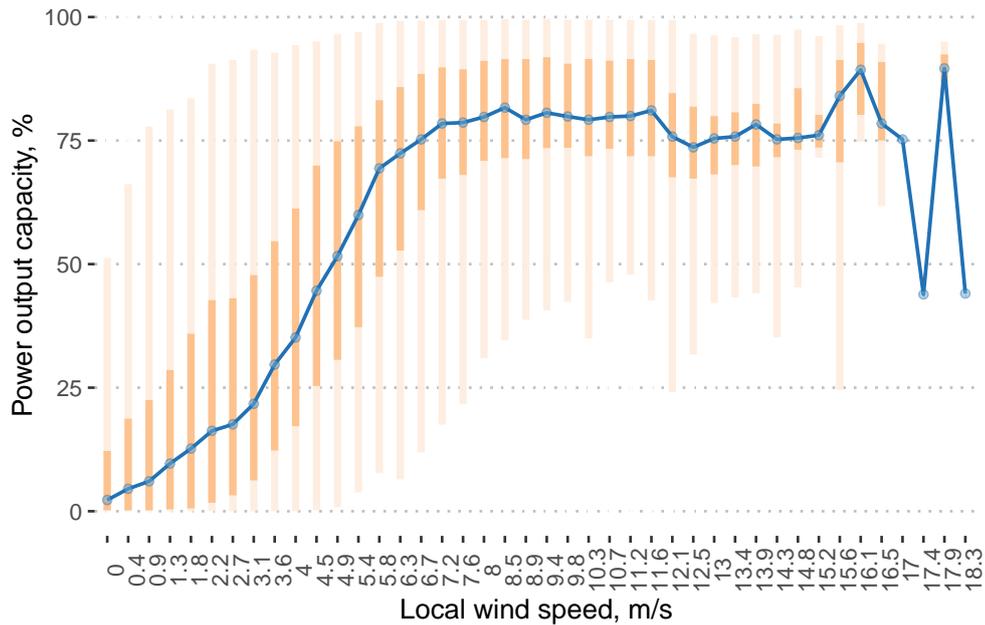


FIGURE 4.3. The relationship between wind farm power output capacity and local wind speed. The data were measured at 10 m at H2. The blue line indicates the median while the darker orange bars indicate [25 75] and the lighter [2.5 97.5] quantile ranges.

4.2.3 Amplitude modulation detection

My machine learning-based random forest method was used for detecting AM and was validated using a benchmark human-scored data set (Nguyen et al., 2021b). In brief, for validation purposes, an acoustic engineer listened to 6,000 10-sec audio samples randomly extracted from measured data and manually classified them as either ‘containing AM’ versus ‘no AM’.

Due to the imbalance between noise samples containing AM and no AM in the data sets of the current study ($\approx 20\%$ AM versus $\approx 80\%$ no AM samples), a Random Undersampling Boosting (RUSBoost) classifier (Seiffert et al., 2009), a simpler and faster alternative to SMOTEBoost (Chawla et al., 2003), was used to improve the previous developed method. Also, to maximise the performance of AM detection, a separate classifier was used for each data set (three classifiers for outdoor and three for indoor data sets). All classifiers implement the RUSBoost algorithm which was trained using all features listed in Table 3.2. The training data sets included the 2 data sets described in Chapter 3 and 4 and data extracted from

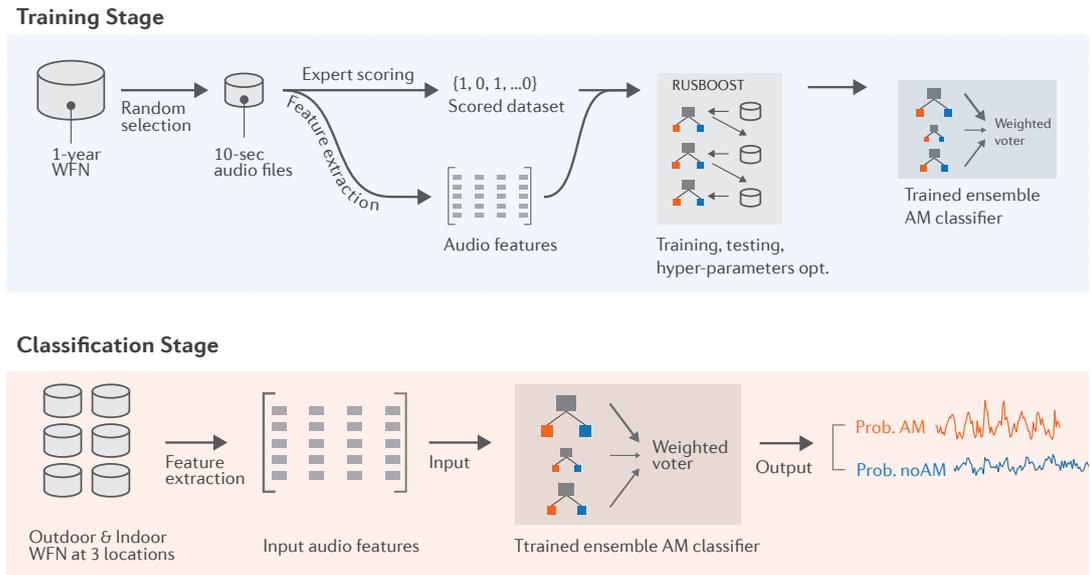


FIGURE 4.4. Flow chart of AM detection method.

TABLE 4.3. Performance of AM detectors.

Detector	Applied to data	No. of sam- ples	%AM %noAM	: F-1 score	MCC
1	Outdoor WF1	3,000	38.4 : 61.6	0.77	0.62
2	Outdoor WF2	5,000	37.9 : 62.1	0.78	0.65
3	Outdoor WF3	1,000	13.5 : 86.5	0.71	0.75
4	Indoor WF1	1,000	13.4 : 86.6	0.64	0.59
5	Indoor WF2	1,000	21.6 : 78.4	0.71	0.63
6	Indoor WF3	1,000	8.5 : 91.5	0.80	0.78

outdoor and indoor measurements at WF2 and WF3 (Table 4.3). A schematic overview of the AM detection method is shown in (Figure 4.4). This machine learning approach showed high performance with $F-1$ scores from 0.64 to 0.8 and Matthew correlation coefficients from 0.59 to 0.78 (Table 4.3) and higher accuracy than previous methods (Bass et al., 2016; Larsson and Öhlund, 2014; Yokoyama et al., 2013) (see Nguyen et al. (2021b) for comparison details).

4.2.4 Wind direction categories

To determine downwind, upwind or crosswind directions, firstly the centroid of all turbines within 5 km from the house was determined (red dot point) (Figure 4.5). In general, the direction from the centroid to the house (a1 line) ± 45 deg was considered as the downwind

TABLE 4.4. Data quality control and cleaning.

Location	WF1 (Horns- dale)	WF2 (Hal- lett)	WF3 (Water- loo)
Total number of samples	49,815	56,267	51,179
Number excluded outdoor data	0	0	9
% excluded indoor data	7	0	9

direction (red angle). The reverse direction was considered as the upwind direction (green angle). To determine crosswind directions, a line a_2 which is perpendicular to a_1 was determined. The direction surrounding this line ± 45 deg was considered as the crosswind direction (yellow angle). A simpler method considers the direction from the nearest wind turbine to the house as the downwind direction. However, in the case where there are several nearby turbines, this method is not suitable, as shown in the WF1 layout below and our method is generally better suited to most cases. We considered turbines within 5 km from a wind farm only because we assumed that turbines at distances greater than 5 km contribute less to the noise at the residence as shown in (Hansen et al., 2019a). The centroid of the turbines (points) $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$ is calculated as follow:

$$\left(\frac{x_1 + x_2 + x_3 + \dots + x_n}{n}, \frac{y_1 + y_2 + y_3 + \dots + y_n}{n} \right) \quad (4.1)$$

4.2.5 Data cleaning

The present study analysed the outdoor WFN noise data measured at 1.5 m above ground level (except at wind farm 1 where noise was measured at ground level) and indoor data measured in a top room corner. To ensure data quality of outdoor and indoor WFN measurements, a plot of the L_{Aeq} of all data against time was constructed, and extraneous noise events were detected visually and manually excluded if noise contamination was confirmed through listening to the file. The most common contaminated sources included farming machines, heavy rain and thunderstorms. The indoor contaminated sources consisted mainly of human activity and equipment (i.e., fan during summer nights). Less than 10% of the total measured samples were excluded (Table 4.4, Figure 4.6 and Figure 4.7).

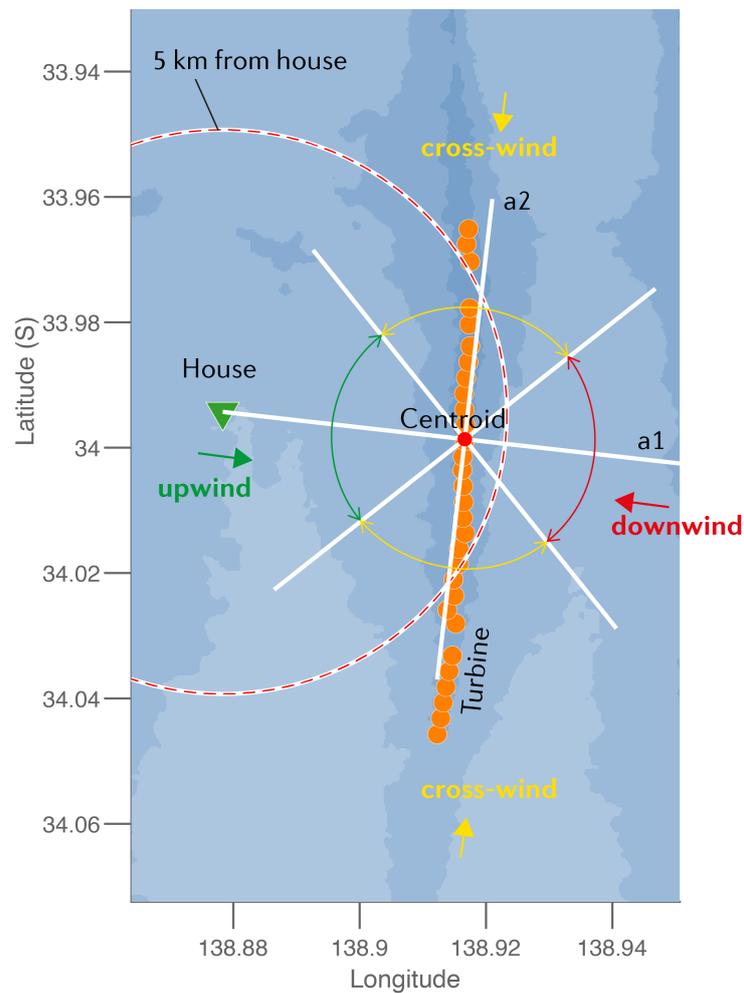


FIGURE 4.5. Wind direction categories at the Waterloo wind farm (WF3).

4.2.6 Data and statistical analysis

All signal processing and data analysis were implemented in MATLAB¹, while statistical analysis (two-tailed t -test and linear regression as appropriate) were implemented in R² version 4.0.0. The statistical significance threshold was set at $P < 0.05$.

¹<https://www.mathworks.com>.

²<https://www.r-project.org>.

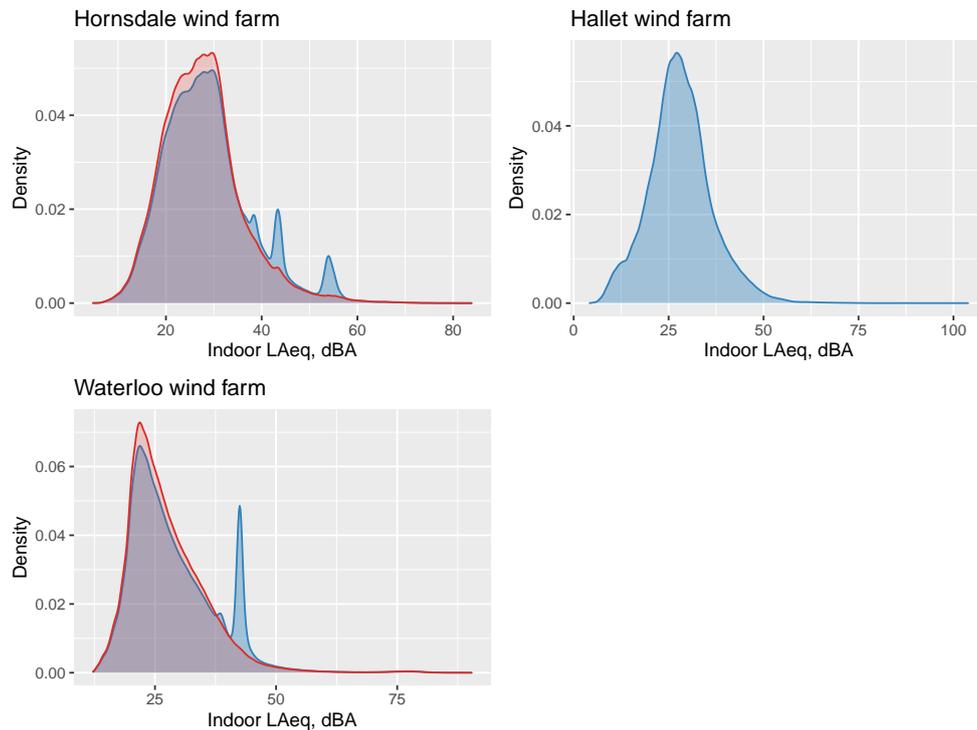


FIGURE 4.6. Distributions of indoor SPL. The original distribution is shown using blue shading and the distribution after removing extraneous noise is shown using red shading.

4.2.7 Data availability

Data to support the findings of this study are based on publicly available data including wind farm power output capacity data for each 5-minute interval accessed via Australian Energy Market³, Digital Elevation Data (DEM) extracted from Geoscience Australia⁴ and AM detection algorithms are available at my Github repository⁵.

³<https://anero.id/energy/wind-energy/>

⁴<http://www.ga.gov.au/scientific-topics/national-location-information/digital-elevation-data>

⁵<https://github.com/ducphucnguyen/Quantification-AM>

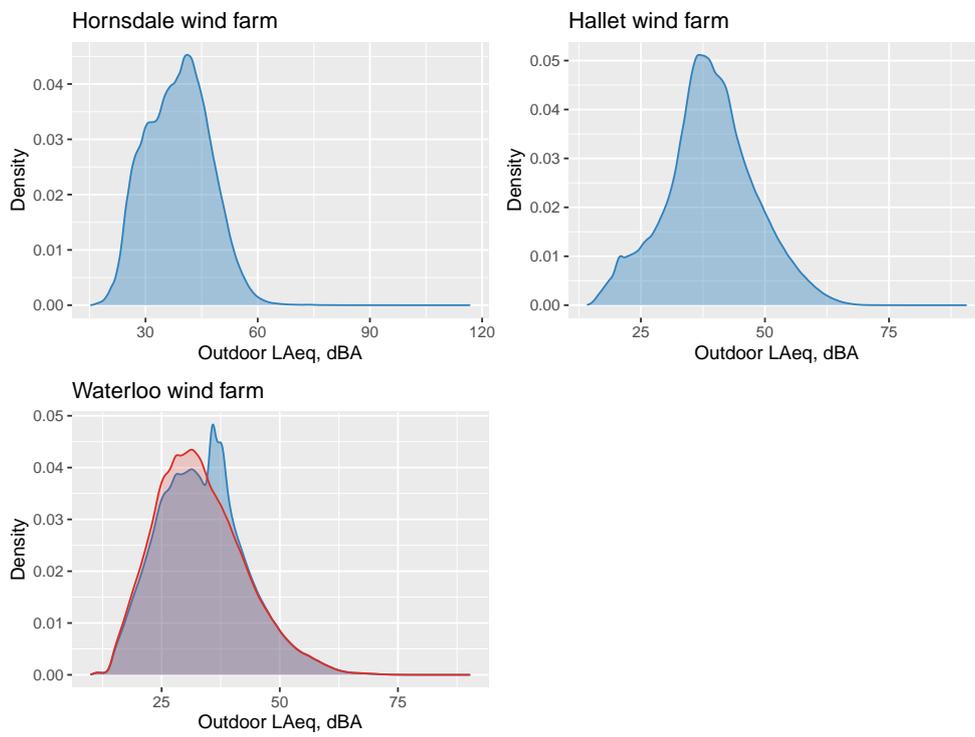


FIGURE 4.7. Distributions of outdoor SPL. The original distribution is shown using blue shading and the distribution after removing extraneous noise is shown using red shading.

4.3 Results

4.3.1 Amplitude modulation characteristics

AM occurred more often during the nighttime compared to the daytime (Figure 4.8a, two-sample t -test, all P -values < 0.001). At locations H1 and H2 which were within 1.3 km of the nearest wind turbine, AM occurred on average for more than 50% and 25% of the nighttime and daytime, respectively. Similar trends were also observed at location H3, but with a lower prevalence of around 25% AM during the nighttime and only 3% during the daytime, where the nighttime value is comparable to previous observations for similar distances (Hansen et al., 2019a). The AM depth, which is a measure of the peak-to-trough variation in the overall SPL, varied between measurement locations (Figure 4.8b). The AM depth was calculated as the difference between statistical noise levels L_{5th} and L_{95th} of the fast time-weighted and frequency A-weighted SPL. This metric is reported in this study because it is commonly used in laboratory listening experiments assessing annoyance potential of AM (Von Hünenbein et al., 2013; Yokoyama et al., 2013). The AM depth of WFN was inversely related to frequency at H1, as shown in Figure 4.8c.

AM was an intermittent phenomenon as fewer than 20% of consecutive 10-second AM events spanned more than one hour, as shown in Figure 4.8d. Furthermore, Figure 4.8d shows that a larger number of consecutive 10-second AM events were observed at closer locations to the wind farm, equivalent to AM lasting between 1 and 3 hours. The modulation frequency was consistently between 0.5 and 1 Hz (Figure 4.8e), which is as expected for modern wind turbines which rotate at a speed of 10 to 20 revolutions per minute (Hansen et al., 2017). However, a large number of AM events did not show a clear periodic variation (modulation frequencies between 0 and 0.5 Hz). Reasons for this may be (1) a false positive detection of AM or (2) an intermittent, rather than periodic variation in SPL. AM events were dominant at particular SPLs (median values of 39.4, 36.4 and 29.1 dBA for H1-H3, respectively) (Figure 4.8f).

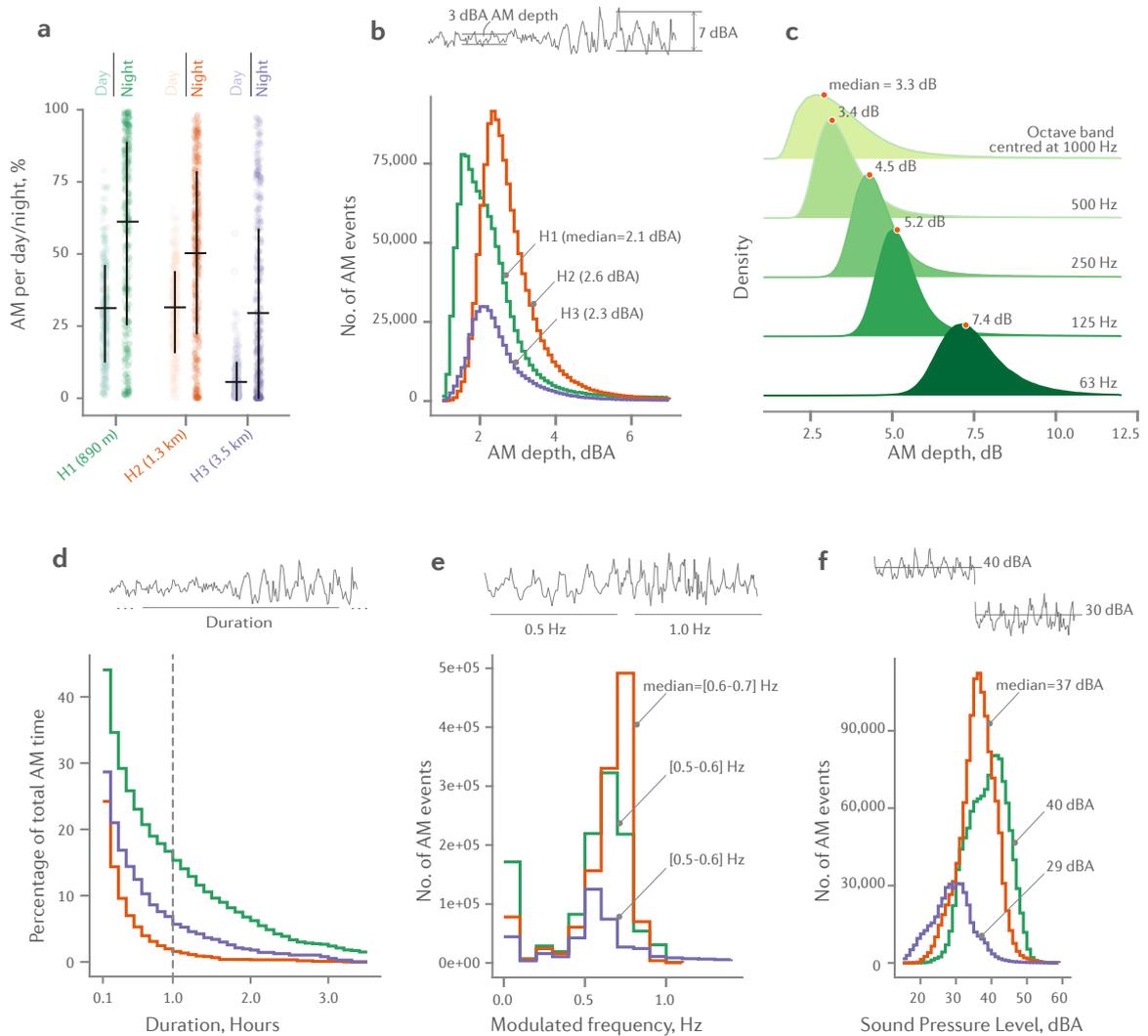


FIGURE 4.8. Characterisation of AM for outdoor data measured over one year. (a) The percentage of AM during the daytime (from 7:00 to 22:00) and nighttime (22:00 to 7:00). Horizontal and vertical error bars represent the mean and s.d. (b) AM depth calculated as the difference between statistical noise levels L5 and L95. (c) AM depth quantified in each octave band for measurements at H1. (d) AM duration measured as the continuous occurrence of AM events in consecutive and uninterrupted 10 s blocks. (e) AM frequency. (f) A-weighted sound pressure level associated with AM events. Examples of the time domain representation of the unweighted sound signals are provided above each histogram (b,d,e & f).

4.3.2 Outdoor-to-indoor variability

More AM events were detected outdoors compared to indoors, with the exception of location H3 during the daytime, as shown in [Figure 4.9a](#). On average, outdoor AM prevalence was approximately 1.5 times higher than that measured indoors (see [Figure 4.9a](#) and [Table 4.5](#)). The outdoor-to-indoor AM prevalence reduction at H1 and H2 was similar, ranging between 1.5 and 2.2. In contrast, the difference between outdoor and indoor AM prevalence was smaller for data measured at H3 during the night-time (reduction = 1.1), and indoor AM occurred more often than outdoor AM during the daytime (reduction = 0.4). The AM depth measured indoors was higher than that measured outdoors (see [Figure 4.9b](#), two sample *t*-test, all $P < 0.001$). Lower indoor background noise (or masking noise), as shown in [Figure 4.9d](#), most likely explains the higher AM depth measured indoors. It is also possible that the false positive detection of the algorithm classified local indoor noise as AM.

To examine if differences between outdoor and indoor AM prevalence could be attributed to house insulation, the distribution of simultaneously occurring outdoor and indoor noise levels are presented in [Figure 4.9c](#). A greater A-weighted SPL reduction was observed for H1 and H2, compared to H3. This may explain some of the differences between the relative outdoor and indoor AM prevalence for H3 ([Figure 4.9a](#)). The outdoor-to-indoor SPL reduction at H3 was poor for outdoor SPL < 40 dBA. It should be noted that the outdoor-to-indoor noise reduction as characterised using overall noise levels depends not only on building materials, but also the characteristics of the noise spectrum and indoor background noise. The lowest level of indoor background noise measured inside H3 was higher than those in H1 and H2 ([Figure 4.9d](#)). This may affect the relationship between indoor and outdoor noise levels as shown in [Figure 4.9c](#). Although it is not very accurate to characterise the outdoor-to-indoor reduction using overall noise levels ([Hansen et al., 2015](#); [Thorsson et al., 2018](#)), this is a simple approach and the results are easy to interpret.

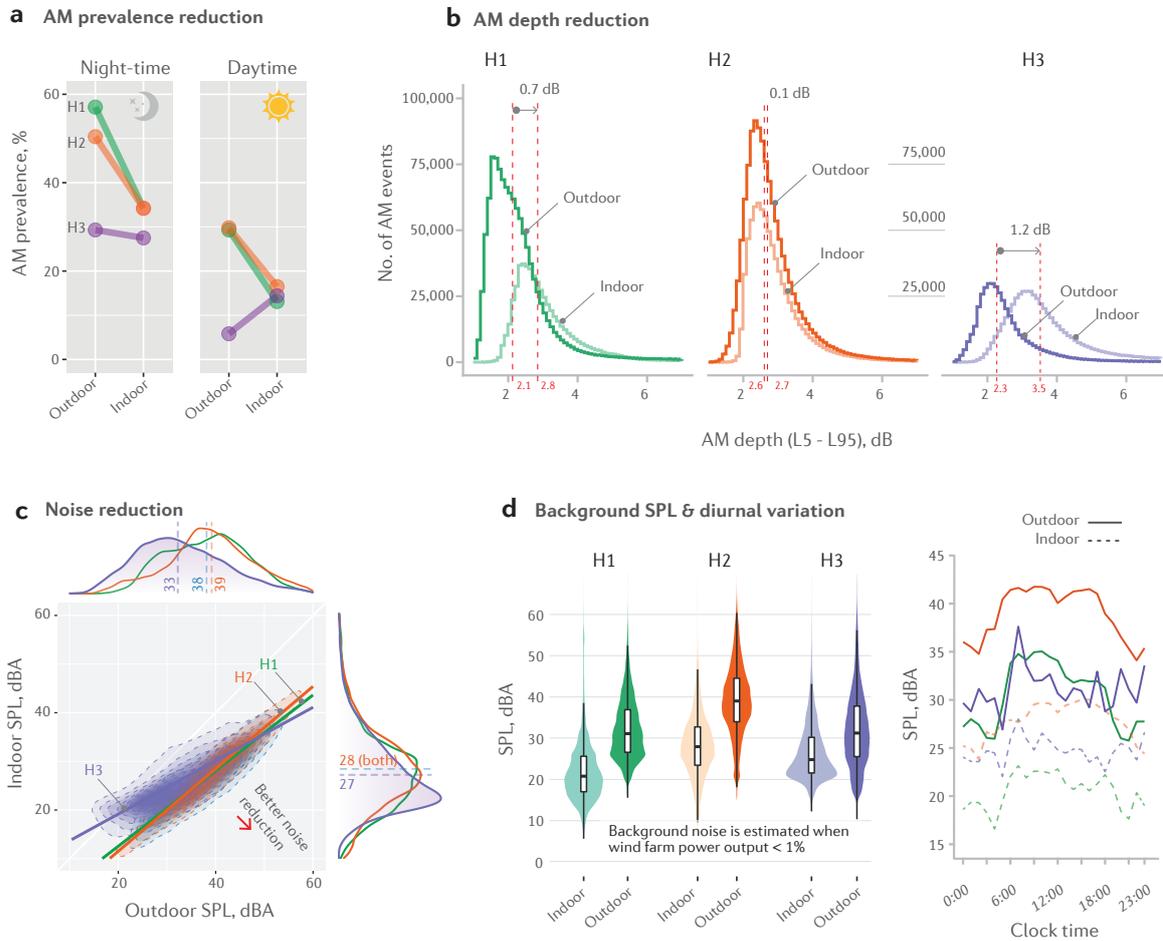


FIGURE 4.9. Outdoor and indoor AM prevalence and AM depth. (a) AM prevalence percentage difference during the nighttime and daytime. (b) AM depth distributions measured indoors and outdoors at three locations with median values indicated by vertical red dashed lines. AM depth is estimated using the DAM method which specifies use of the A-weighted SPL. (c) Comparison between indoor and outdoor SPL. Outdoor SPL distributions with median values are shown on the top and side, respectively. (d) Indoor and outdoor background noise as calculated corresponding with the wind farm power output capacity < 1%.

TABLE 4.5. Outdoor-to-indoor AM prevalence.

		Outdoor	Indoor	Outdoor/indoor
H1	Night	57.1	34.2	1.7
	Day	29.3	13.1	2.2
H2	Night	50.4	34.2	1.5
	Day	29.8	16.5	1.8
H3	Night	29.3	27.5	1.1
	Day	5.8	14.4	0.4
	Mean			1.5

4.3.3 Diurnal and seasonal variability

AM occurred most frequently at nighttime between 22:00 pm - 4:00 am, whereas the lowest AM prevalence was observed at midday around 12:00 pm (Figure 4.10a). Similar distributions of AM prevalence were observed at H1 and H3. For these locations, the highest and lowest AM prevalence were approximately 60% and 20%, observed at 0:00 am and 12:00 pm, respectively. For location H3, less than 5% of AM prevalence was observed during the daytime, but this number increased to more than 30% during the nighttime. The background noise during the nighttime was also found to be lower compared to the daytime as shown in Figure 4.9d. This was anticipated as the noise associated with human activities was expected to be lower at nighttime. Additionally, higher AM prevalence observed during the night-time could be partly attributed to lower background noise levels at nighttime compared to daytime.

The mean AM prevalence for each season is shown in Figure 4.10b. The mean AM prevalence was not notably different between months in my data. However, when AM prevalence was averaged over an hour, as shown in Figure 4.10c, clear monthly and hourly variations of AM were evident. At all measurement locations, during the winter and spring months, AM prevalence significantly increased after 16:00, which corresponds to the timing of sunset during these seasons. AM prevalence significantly increased after 20:00 in summer and autumn months, which also corresponds to the timing of sunset during these seasons. This pattern clearly corresponded to sunrise and sunset times (dashed line, Figure 4.10c) and is consistent with Larsson and Öhlund (2014) findings, where the authors observed a strong association between AM prevalence and solar elevation angle.

4.3.4 Relationship between meteorological and power output conditions

An increase in the local wind speed did not always correspond to higher AM prevalence (Figure 4.11a). The relationship between wind speed and AM prevalence at all three locations followed a similar pattern. Specifically, AM prevalence increased as the wind speed increased in specific regions (i.e., [0 4) at H1, [0 4) at H2 and [0 2) at H3), but dropped rapidly at higher wind speeds. AM was rarely detected when the local wind speed was greater than

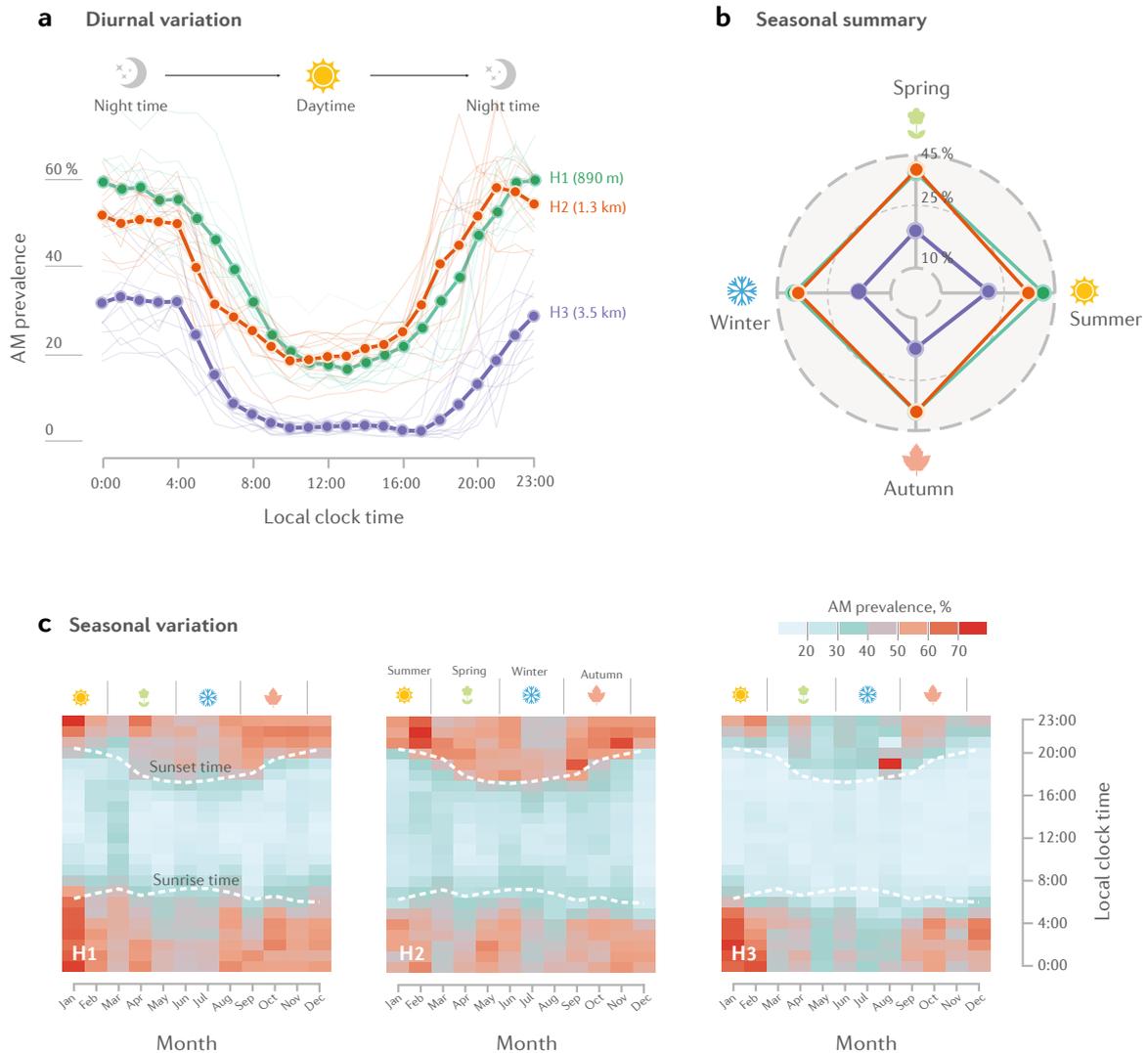


FIGURE 4.10. Diurnal and seasonal variation of AM characteristics. (a) Diurnal variation of AM prevalence. Thicker lines are the average trend over the year for three locations. Light lines indicate the trend for each month. (b) Seasonal summary of AM prevalence, calculated as the mean AM over each season (i.e., Summer from Dec-Feb, Autumn from Mar-May, Winter from Jun-Aug and Spring from Sep-Nov). (c) Relationship between diurnal and seasonal variation of AM prevalence. Dashed lines indicate sunset and sunrise time. Outdoors data sets were used for this analysis.

approximately 7 m/s for H1 and 10 m/s for H2 and H3. Note that these wind speed data were measured locally at each residence, rather than at the wind turbine nacelle. The wind speed data were measured at 10 m above ground level for H2 and H3 and at 1.5 m above ground level for H1, where data was not available at 10 m for the latter location.

AM prevalence measured at H2 and H3 increased with an increase of the calculated wind gradient (see [Figure 4.11b](#) for the relevant equation) for lower values of the wind gradient (between 0 and 1 s^{-1}). However, for higher values of wind gradient, the prevalence of AM was reduced. AM prevalence increased with humidity ([Figure 4.11c](#), linear regression, $R^2 = 0.4$, $P = 0.001$) as expected given that higher humidity is more favourable for sound propagation ([Hansen et al., 2017](#)).

At all measurement locations, the AM prevalence in the upwind direction was less than for other wind directions ([Figure 4.11d](#), all $P < 0.001$) (see [Figure 4.5](#) for definition of wind direction category). At location H1, a mean AM prevalence of 60% was detected for downwind data, as opposed to 30% for upwind data. A similar trend was also observed for data measured at location H3, with around 40% of AM events occurring during downwind conditions and less than 5% during upwind conditions. At location H2, more AM events were detected for crosswind directions with a mean AM prevalence of around 50% compared with 40% downwind and 25% in the upwind direction.

The maximum wind farm power output did not correspond with the highest AM prevalence ([Figure 4.11e](#)). At locations H2 and H3, AM occurred more often when the wind farm operated at around 50% of its capacity. At location H1, AM occurred more often at low power output (between 10% and 60%). These findings are consistent with previous studies ([Hansen et al., 2019a](#)). Compared to H1 and H3, the relationship between AM prevalence and wind farm power output capacity at H2 was skewed to lower wind farm power output ([Figure 4.11e](#)). The reasons for this observation are unclear. Possible contributions include the difference in rated turbine power output, wind farm layout, terrain characteristics or false positive detection of the algorithm. The relative importance of these factors is unclear, and thus more data and modelling approaches are needed to understand this relationship.

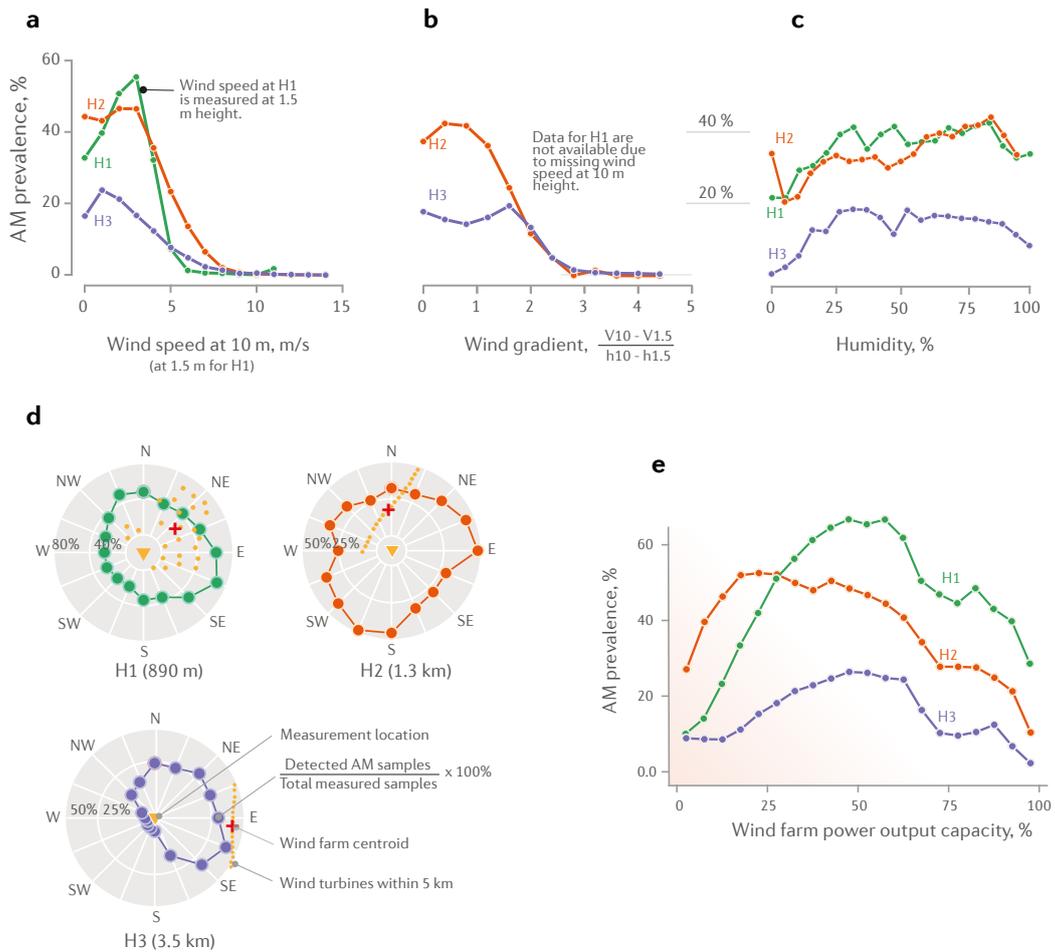


FIGURE 4.11. AM prevalence for different meteorological and wind farm operating conditions. (a,b and c) The relationship between AM prevalence factors predicted to influence sound propagation (i.e., wind speed, wind gradient and humidity). (d) The dominance of AM prevalence for particular wind directions. The yellow dot points inside the grey circles indicate wind turbines within 5 km. The grey circles are within a 5 km radius of the measurement locations (yellow triangle points). (e) The relationship between wind farm percentage power output capacity and AM prevalence.

4.4 Discussion

This paper presented long-term AM characteristics of WFN through analysis of acoustical and meteorological data measured at three South Australian wind farms. I showed comprehensive information regarding the prevalence and diurnal distribution of AM at three locations with different wind farm layouts, wind turbine types, housing constructions and wind farm separation distances. The resulting estimates of AM depth, duration, frequency and associated sound pressure levels are important for both laboratory human trials and physical modelling of WFN (Barlas et al., 2017; Makarewicz and Gołębiewski, 2019), as these characteristics can be used to design relevant noise stimuli and also used to validate wind farm noise models most relevant to real-world noise exposure conditions in the field. To the best of my knowledge, this is the first study to characterise and quantify AM using comprehensive data measured several kilometres from a wind farm, and to detect wind farm AM using a comprehensively validated, machine learning-based algorithm.

AM was found to occur most often during the nighttime, consistent with previous studies (Conrady et al., 2020; Hansen et al., 2019a). This is expected because nighttime provides favourable weather conditions for sound propagation and perhaps also better conditions for AM generation (stable atmospheric conditions, high humidity, strong temperature inversion, high wind shear) (Stull, 2012). During these conditions, sound waves are refracted towards the ground surface in the case of downwind and crosswind conditions (although wind shear does not contribute in the latter case) (Ostashev and Wilson, 2015). This also most likely explains the high AM prevalence observed in the downwind and crosswind directions.

I found that AM is an intermittent characteristic of wind farm noise as most AM events lasted for several minutes only. This is comparable to another study (Larsson and Öhlund, 2014) where the authors observed that typical main AM events lasted around 15 seconds and were followed by weaker AM. There are several factors that could lead to AM intermittency, particularly at large distances from a wind farm. These factors include varying source noise level, varying influences on noise propagation and varying background noise at the receiver location. Source noise levels vary with time as they depend on wind speed, atmospheric conditions and wake effects. Noise propagation is influenced by constantly changing atmo-

spheric conditions. Masking noise levels at measurement locations are expected to vary with time due to changes in the local wind speed and direction and varying levels of extraneous noise. A combination of these factors could contribute towards shorter AM events observed.

A large difference was found between outdoor and indoor AM. At long-range, spectral imbalance of wind farm noise arises due to the higher atmospheric and ground absorption at mid to high frequencies (Ostashev and Wilson, 2015). In fact, Hansen et al. (2019a) found that AM usually occurs at very low frequencies (i.e., around 50 Hz) at several kilometres from a wind farm. In addition, low-frequency noise is poorly attenuated by building structures, resulting in lower outdoor-to-indoor noise level reduction at low frequencies (Hansen et al., 2015). These results could explain the relatively small outdoor-to-indoor reduction in AM prevalence that was observed at H3 at nighttime. The increase in AM events measured indoors during the daytime at H3 may have been a result of high outdoor ambient noise that masked the outdoor AM but not the indoor AM. These findings suggest that the outdoor-to-indoor noise reduction also impacts AM prevalence. Also, a greater AM depth is associated with higher annoyance (Lee et al., 2011; Schäffer et al., 2016; Yokoyama et al., 2013), and thus AM may be more annoying when people are indoors with low ambient background noise, which is exaggerated during the nighttime. These observations are particularly relevant for cases where AM is only measured outdoors.

I investigated the seasonal variations of AM prevalence measured outdoors, but found no significant differences in outdoor AM prevalence between seasons. This contrasts with the study conducted by Conrady et al. (2020) where the authors reported more frequent AM during the Winter compared to spring, but with more limited data from a much colder climate in Sweden. Interestingly, I found a remarkably strong temporal relationship between sunset and sunrise times and the beginning and end of AM, most likely indicative of temperature inversion effects (Stull, 2012).

While directivity of broadband trailing edge noise causes swishing noise which is prominent close to the turbine (within 1-2 rotor diameters) (Oerlemans, 2015), the wind gradient has been hypothesised to cause AM perceived at larger distances (hundreds of metres to several km) from the wind farm. This is due to an increased difference in aerodynamic

loading between the upper and lower parts of the wind turbine blade trajectory (Bowdler, 2008; Oerlemans, 2015; Van den Berg, 2005). This change in loading could then affect blade aerodynamic noise production such as trailing edge, leading edge and loading noise sources which would show greatest variation between the lowest and highest parts of the blade trajectories, thereby resulting in AM. It is thus expected that higher wind gradients provide more favourable conditions for AM generation and that AM prevalence increases with increasing wind gradient. However, my data did not support this as AM prevalence was reduced with increasing local wind gradient, a finding comparable with other studies (Conrady et al., 2020; Cooper and Evans, 2013). However, it is also worth noting that the wind gradient measured in my study was based on local wind speeds between 1.5 and 10 m above ground which is likely not representative of wind gradients at higher altitudes more relevant for AM generation at the noise source. Furthermore, wind gradients over a ridge are significantly modified by wind speed-up effects (Ngo and Letchford, 2009). Therefore, the association between wind shear and AM is still unclear and more suitable wind speed data closer to the noise source are needed to confirm the wind gradient hypothesis.

Although I measured comprehensive acoustical data, a limitation of my study is a lack of comprehensive meteorological data measured at hub-height most relevant to the noise source. Thus, relationships between AM and meteorological conditions remain unclear and further studies are needed to more comprehensively assess these relationships. This limitation calls for better data sharing practices between wind farm operators and researchers (Kusiak, 2016) to allow for more in depth analysis of relationships between wind farm noise and meteorological conditions. Future studies are needed to answer questions if continuous or discrete of AM duration is more annoying for residents living nearby wind farms. Also, how interaction between the AM prevalence and AM depth could increase the perception of human on wind farm noise AM.

4.5 Conclusion

In summary, this study characterised and quantified wind farm noise AM for a large data set measured over one year at three relatively long-range distances from three wind farms in South Australia. At nighttime, AM prevalence was lower indoors than outdoors, but there was an increase in AM depth in the indoor data. My findings also showed a dependence of AM prevalence with respect to time (i.e., diurnal and monthly variations). I further found that AM occurred more often during downwind and crosswind directions, compared to upwind conditions. The measured data can also be used for validating wind farm propagation models, particularly those that attempt to model AM. Ultimately, improved wind farm noise assessment guidelines and more accurate noise prediction models will make wind energy more acceptable to surrounding communities.

Chapter 5

Audibility of wind farm infrasound and modulated tonal noise observed at long-range locations

This chapter presents work that I published as first author in (Nguyen et al., 2022a). My coauthors were primarily involved in an advisory role, so this is predominantly my work.

Contributions: My main contributions to this chapter include:

- I used a novel computational approach to assess the audibility of infrasound and AM tones at long-range locations, which considered the uncertainty associated with WFN measurements and human hearing variability.
- I characterised and quantified infrasound and amplitude modulated (AM) tones at long-range locations and revealed relationships with weather conditions.
- I showed that infrasound is highly unlikely to be audible to residents with normal hearing living at distances greater than 1 km from a wind farm.
- I found that AM tones occurring at a low frequency are highly perceivable at distances up to 9 km.

5.1 Introduction

Wind energy has grown rapidly in Australia, from less than 1% of Australia's total electricity generation in 2005 to 9% in 2020 (Department of Industry and Resources, 2021), mirroring global trends (WindPower, 2021). Although wind energy is one of the fastest growing sources of renewable energy in Australia, increasing annually by 14%, significant problems remain with social acceptance due to aesthetic, environmental and health concerns (Merlin et al., 2013). Noise generated from wind farms is also a recurring source of complaints regarding annoyance and sleep disturbance from residents living near wind farms (Liebich et al., 2021a; Micic et al., 2018). Laboratory studies clearly demonstrate that WFN characteristics such as low-frequency spectral dominance, tonality and AM contribute to annoyance (Ioannidou et al., 2016; Lee et al., 2011; Schäffer et al., 2016; Smith et al., 2020), supporting that low frequency WFN characteristics are the most prominent and problematic features.

Infrasound, which comprises acoustic signals below 20 Hz generated by blade-tower interaction (Van den Berg and van den Berg, 2006; Zajamsek et al., 2019), is a major source of controversy due to uncertainties regarding whether or not infrasound has negative impacts on people. Much of this debate reflects problems with the terminology, which implies a clear-cut 20 Hz threshold of audibility. Previous studies (Baumgart et al., 2021; Ingieiewicz et al., 2014; Jakobsen, 2005; Turnbull et al., 2012; Zagubień and Wolniewicz, 2019) have found that infrasound levels measured near wind farms (between 100 and 500 m from the nearest wind turbine) are well below the average audible infrasound threshold of a person with normal hearing. For example, Van den Berg (2005) showed that infrasound levels measured close to wind turbines are 15-20 dB below the average infrasound perception threshold (95 dBG Jakobsen (2005)). Comparing 1/3-octave band levels with the normal hearing threshold curve for infrasound, Baumgart et al. (2021) revealed that measured infrasound was lower than this curve by at least 20 dB. At some wind farms, tones are modulated at the blade-pass frequency to create AM tones, which span the infrasonic and low-frequency ranges (Hansen et al., 2017). These AM tones are often perceived as 'rumbling' and have been found to exceed the OHC threshold at distances up to 4 km (Zajamsek et al., 2016).

However, the variability of the hearing threshold has not been considered when assessing the audibility of WFN characteristics such as infrasound and AM tones. Previous studies (Baumgart et al., 2021; Turnbull et al., 2012; Van den Berg, 2005; Zajamsek et al., 2016) were mainly based on the median hearing threshold in ISO 226:2003 (ISO226:2003, 2003) or the average infrasound perception threshold. However, Møller and Pedersen (2004) found that individual hearing thresholds could be lower than the median hearing threshold by up to 20 dB. Uncertainty is not only associated with hearing acuity, but is also a feature of WFN measurements themselves. Due to variable and complex effects of weather and wind farm operating conditions, the measurement of WFN is highly uncertain. For example, the difference in the sound pressure level (SPL) at some particular frequencies could be over 20 dB during wind farm operational versus non-operational conditions (Zajamsek et al., 2016). The difference could also be over 10 dB during nighttime versus daytime operational conditions (Nguyen et al., 2021a). However, measurement variability is rarely considered when assessing WFN. Most previous studies have used the median SPL (Baumgart et al., 2021; Turnbull et al., 2012) or the maximum SPL (Zajamsek et al., 2016), potentially biasing towards under- or over-estimation of WFN impacts. The deterministic approach for assessing WFN, which uses the median values of SPL and the normal hearing threshold, could systematically underestimate the impacts of wind farm noise on nearby communities. It is more appropriate to assess the audibility of the noise based on a probabilistic approach. Thus, consideration of uncertainty associated with both WFN measurements and variability in hearing acuity is needed to better understand potential WFN impacts on humans.

The purpose of this study was to investigate the audibility of WFN characteristics such as infrasound and amplitude modulated (AM) tones at long-range locations. I first quantified the uncertainty associated with the measurement of WFN using long-term data sets measured at locations between 1 and 9 km from the nearest wind turbine. I leveraged these large data sets (over 4-years of data in total) and previous data describing the variability of the hearing threshold to estimate the probability of audibility of infrasound and AM tones. This approach demonstrates that AM tones are highly perceivable at least up to 9 km, especially during high wind farm power output and downwind conditions.

5.2 Methods

The following section briefly summarises key information regarding measurement locations, wind farm characteristics and experimental setup. All other information regarding field measurements is provided in detail in Chapter 2.

5.2.1 Overview of study

I measured WFN at five locations (1-4 and 6) close to three wind farms (A-C) in South Australia as shown in [Figure 5.1a](#). The distances from the measurement locations to the nearest wind turbine of wind farms ranged from approximately 1 to 9 km (see [Table 2.2](#) for details). All measurements were carried out over one year, except at location 6, where data were collected over three months. These long-term data sets were analysed using spectral analysis techniques such as power spectral density (PSD) or 1/3-octave band analysis to characterise key features unique to WFN ([Figure 5.1d, e](#)). These potentially most annoying characteristics of WFN at long-range locations (i.e., infrasound and AM tones) were then investigated to quantify their audibility ([Figure 5.1f](#)).

5.2.2 Wind farm characteristics and experimental setup

Wind farm A was located in an area with flat terrain, while wind farms B and C were located on ridges. The ridges were approximately 70 to 110 m higher than the measurement locations. [Figure 5.1b](#) shows wind farm B, where several nearby wind turbines were positioned along the ridge. At all wind farms, the wind turbines were a three-bladed design with a rated power from 2.1 to 3.3 MW (see [Table 2.1](#) in Chapter 2 for further details).

The typical setup included two microphones, which were positioned outdoors at a height of 1.5 m and at ground level ([Figure 5.1c](#)). Both microphones were equipped with primary and secondary windshields with diameters of 90 mm and 450 mm, respectively. Acoustic data were acquired using a Bruel and Kajer LAN-XI Type 3050 data acquisition system with a sampling rate of 8,192 Hz and a G.R.A.S type 40 AZ microphone with a 26CG preamplifier, which has a noise floor of 16 dBA and a flat frequency response down to 0.5 Hz. Wind speed

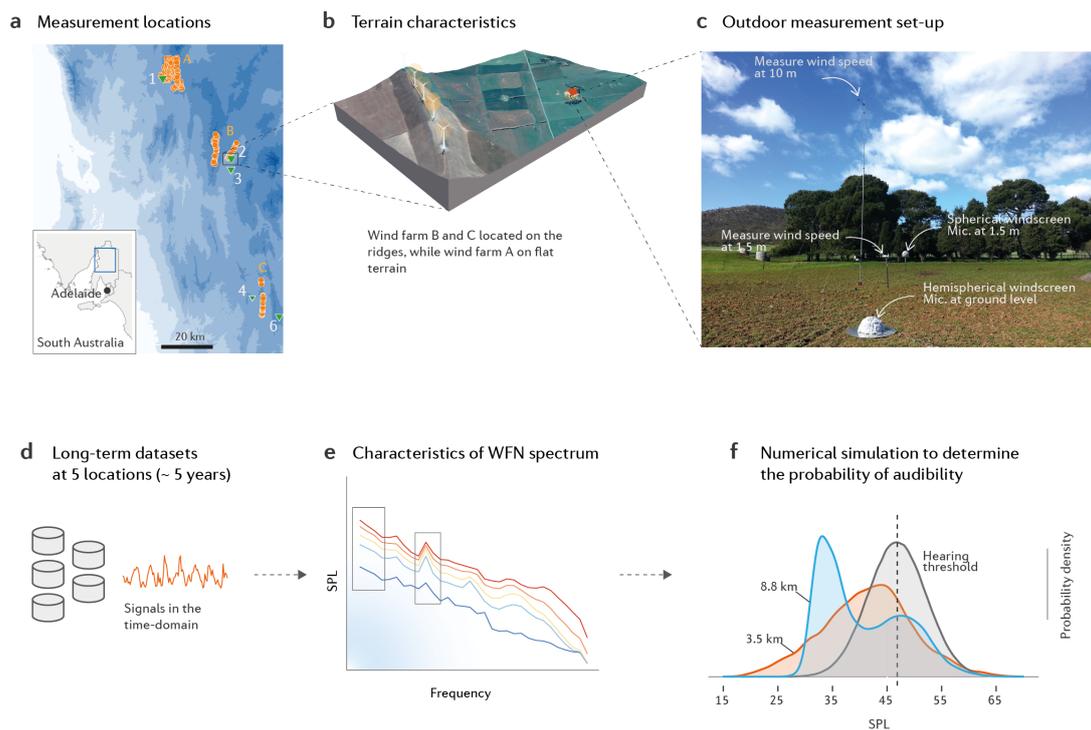


FIGURE 5.1. Study overview. **a** Measurement locations located in South Australia. Orange circles represent wind turbines, while green triangles represent measurement locations 1-5. Three wind farms are indicated using the capital letters A, B and C. **b** 3D view of location 2 near wind farm B. The map is generated using Blender software (<https://www.blender.org>) and Google Earth data. **c** Typical setup of the measurements. **d-e** An overview of the process used to analyse data. Acoustical data were transformed to the frequency domain **e** to identify unique noise features. **f** The noise features were compared with the 1/3-octave band hearing threshold to assess their audibility.

and direction was measured at two levels (i.e., 1.5 and 10 m) using Davis Vantage Vue and Davis Vantage Pro weather stations. The accuracy of the corresponding wind speed and direction measurements is 0.4 m/s and 22.5 degrees, respectively.

5.2.3 Data cleaning

I removed all data measured during high wind speed conditions since data were contaminated by wind-induced microphone noise as well as background noise from surrounding trees. All data corresponding with 10-minute wind speed average greater than 5 m/s at 1.5 m height were removed from the analysis. Also, no criteria were provided for gusts and thus I would have to make an arbitrary choice of which threshold to use for data exclusion in this study. This cut-off level is recommended by WFN South Australian EPA guidelines ([SA-EPA, 2021](#)) for a microphone with a typical 90-mm wind shield. Although, I used a 450-mm secondary wind screen in all setups, and anticipated that the relevant cut-off wind speed may be higher, I conservatively removed all data points corresponding with wind speeds greater than 5 m/s. Furthermore, data corresponding with periods of rainfall were completely discarded. To detect other extraneous noise events, I also plotted the L_{Aeq} of all samples against time. Extraneous noise events were detected visually and manually excluded from the data sets (see [section 4.2](#) for other details regarding this process). The most common data contamination sources included farming machines and thunderstorms. In total, the percentage of data excluded in this study was less than 13%.

5.2.4 Frequency analysis

I used the PSD to visualise WFN in the frequency domain. The PSD was implemented using *pwelch*, a built-in function in MATLAB with the following setup: 10-minute sample length, sampling frequency $F_s = 8,192$ Hz, Hanning window of 81,920 discrete Fourier transform points ($10 \times F_s$) and 50% overlap between segments. This setup gave a 0.1 Hz frequency resolution, allowing for clear visualisation of wind farm infrasound peaks in the frequency domain. The SPLs of the PSD were calculated using the reference sound pressure

$p_{ref} = 20\mu\text{Pa}$. The SPL, L_p in decibels is related to the PSD, S , in pascals by the following equation:

$$L_p = 10 \log_{10} \left(\frac{S}{p_{ref}^2} \right) \quad (5.1)$$

The 1/3-octave band SPLs were also calculated for comparison with the hearing threshold established at 1/3-octave band frequencies. To calculate the 1/3-octave band SPL centered at f_c , I first calculated the lower edge frequency $f_l = f_c/C$ and upper edge frequency $f_u = f_c \times C$ of the 1/3-octave band with $C = 10^{0.05}$. The PSD data points, L_{pi} between the frequency range f_l and f_u were then summed to estimate the SPL using the following equation (Hansen et al., 2017):

$$L_{po} = 10 \log_{10} \sum_{f \in [f_l, f_u]} (B_{enbw} \times \Delta f) \times 10^{L_{pi}/10}, \quad (5.2)$$

in which B_{enbw} is the equivalent noise bandwidth of the window function (i.e., $B_{enbw} \approx 1.5$ for a Hanning window of length 81,920), $\Delta f = 0.1$ Hz is the frequency resolution.

5.2.5 Hearing threshold variability

Hearing thresholds vary between participants. This variability was quantified in terms of the standard deviation, which varied between studies (Møller and Pedersen, 2004). A comprehensive summary of the hearing threshold standard deviation between subjects is provided in Figure 10 (Møller and Pedersen, 2004). I used a validated WebPlotDigitizer tool¹ to manually extract data points from published figures for which data are not publicly available. These data were based on 21 studies conducted between 1933 and 2001. To find the relationship between the standard deviation, SD and frequency, f , I fitted regression models of the SD on f with orders up to the 3rd degree using the following equation:

$$SD = \beta_0 + \beta_1 f + \beta_2 f^2 + \beta_3 f^3 + \varepsilon, \quad (5.3)$$

¹<https://automeris.io/WebPlotDigitizer/>

where ε is an unobserved random error. I determined the optimal maximum order of the terms of f in Equation 5.3 by increasing the order from 1 to higher order, after which higher order models did not statistically significantly improve model fits. In this study, I assumed that the hearing threshold of humans has a normal distribution (Kurakata and Mizunami, 2008). The mean value of the hearing threshold was based on the ISO 226:2003 (ISO226:2003, 2003) hearing threshold curve (for frequencies > 20 Hz) and Møller and Pedersen (2004) (for frequencies < 20 Hz).

5.2.6 Probability of audibility

The audibility of WFN is usually assessed by comparing the median or mean values of measured SPL and the median hearing threshold in 1/3-octave bands as specified in the ISO 226:2003 standard, resulting in a binary outcome (i.e., audible or not audible). Although this approach is simple, it may not accurately estimate the effects of WFN on humans. For example, the measured SPL as shown in Figure 5.2 was assessed as not audible because its median value was well below the median value of the hearing threshold. However, due to the variability of the SPL and hearing threshold, WFN should be audible to those within the larger overlapped area on the distributions (Figure 5.2).

The probabilistic approach used to estimate the probability of WFN audibility of 1/3-octave band tones f is as follows: First, I randomly selected N samples in the long-term data set, $S(f)$, and N samples in the distribution of hearing ability, $H(f)$. I then calculated the difference between these functions, $g(H, S, f) = S(f) - H(f)$. The occurrence of $g(S, H, f) > 0$ was counted, which is the number of audible events, n_a . The probability of audibility, p_a , was estimated as $p_a(f) = \frac{n_a}{N}$. The value of N was determined using a convergence analysis (see subsection 5.3.5 for details). The process is illustrated in Figure 5.2. The benefit of the probabilistic approach is that the probability could be interpreted as the percentage of the population expected to hear the noise at the measurement locations. The proposed method does not take into account the perception of multiple tones within a critical band and the effects of masking noise, as outlined in ISO 1996-2, resulting in possible underestimation or overestimation of the probability of audibility, depending on the level of masking noise and

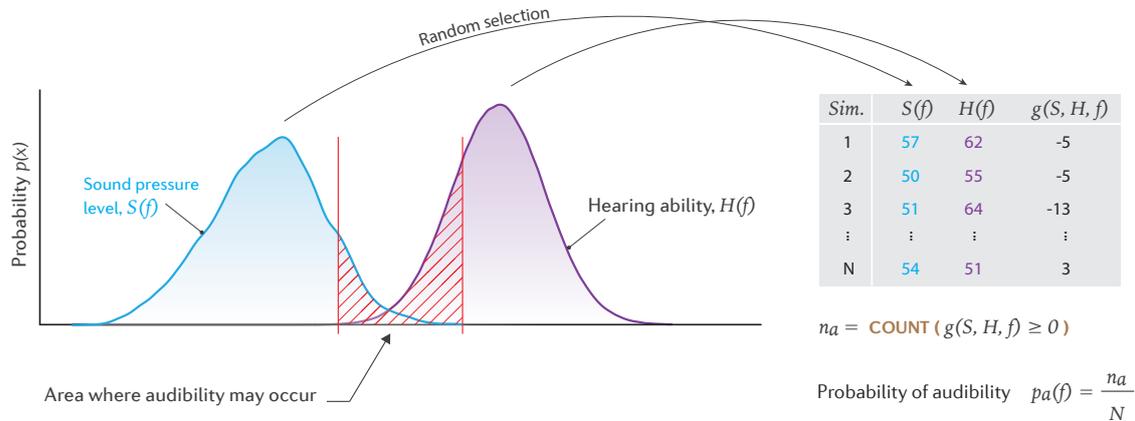


FIGURE 5.2. Probability of audibility. The measured SPL varied due to weather and wind farm operating conditions. The hearing threshold also varied between individuals. The hatched area indicates the region where audibility may occur. The table shows the process used to estimate the probability of audibility.

the number of tones within a critical band. However, the proposed method is simple, efficient and considers the human hearing threshold. It is thus the most feasible approach at present to quantify the audibility of tonal AM for a large data set, such as the one that was the focus of this thesis.

In this study, I assumed that if the noise level in each 1/3-octave band was above the hearing threshold, the noise should be audible, which is similar to the assumption made in (Keith et al., 2019). This assumption implies that the local background noise is much lower than wind farm noise. This is reasonable as my measurement locations were in rural areas where the background noise is very low (Hansen et al., 2014a). Also, the audibility of a modulated tone was calculated with the assumption that there is only one tone in a particular critical band. This assumption could underestimate the probability of audibility as the energy of multiple tones within a critical band is not considered. However, this assumption was necessary to make the analysis of the large data set in this study tractable.

5.2.7 Quantification of noise measurement uncertainty

To estimate the uncertainty in measuring WFN (i.e., 1/3-octave band SPLs) associated with the measurement duration, I ran a computer simulation as shown in algorithm 4. I estimated

the uncertainty corresponding with the length of the measurement from 1 to 90 days. I used a maximum of 90 days because the uncertainty was expected to converge with this measurement length. This analysis was used to determine a reasonable measurement length to ensure that the uncertainty of the measurement would be lower than the auditory perception threshold (see Section 5.3.1 for details). The SPL depends on multiple factors such as diurnal and seasonal weather conditions, daytime activity, wind farm characteristics (i.e., number of turbines, power output and wind farm layout). These complex sources of uncertainty cannot be described by the simple relationship with the length of measurement. Therefore, the Monte Carlo simulation proposed in this thesis could be a suitable approach to establish the dependency of SPL on measurement duration.

Algorithm 4: Uncertainty in measuring wind farm noise

Data: 1-year long data set

Result: Uncertainty in 1/3-octave band SPL

$Uncertainty \leftarrow$ matrix $[90 \times 26]$

for $Nday \leftarrow 1$ to 90 **do**

Randomly select 100 sub-data sets, df_i , with $Nday$ sequential days of data;

for $freq \leftarrow 3.15$ to $1,000$ Hz **do**

Calculate the median values of SPL for the 1/3-octave band centered at $freq$ in df_i , resulting in a vector $M = [M_1, M_2, \dots, M_{100}]$;

Calculate the 2.5th and 97.5th percentiles ($L_{2.5}, L_{97.5}$) of the vector M ;

The uncertainty in the 1/3-octave band SPL centered at $freq$ during $Nday$ is:

$Uncertainty[Nday, freq] = L_{2.5} - L_{97.5}$;

end

end

5.2.8 Regression analysis

I used Spearman's rank correlation analysis to investigate if the probability of audibility increased with wind farm power output. I implemented this analysis in R², a programming language for statistical computing and graphics, using the *cor.test* function and two-sided tests.

²<https://www.r-project.org>

To investigate if wind direction could affect the probability of audibility, I used regression models that included the wind farm power output, *power*, (including higher order terms up to the fifth) and wind direction, *dir* (four categories North, East, South and West) as follows:

$$Probability = \beta_0 + \beta_1 power + \dots + \beta_2 power^5 + dir + \varepsilon, \quad (5.4)$$

where ε is an unobserved random error. I determined the optimal order terms of *power* in Equation 5.4 by increasing the order from 1 to 5 until the model fit showed no statistically significant improvement. This model can be used to estimate the contribution of wind direction to the probability of audibility, after taking into account the contribution of power output.

I used a similar approach to above to see if the probability varies with months and day versus night. I fitted a regression model of the probability based on month (12 categories: Jan-Dec) and day-night (2 categories: day, night). All regression models were implemented using R base package *lm* version 3.6.2.

5.2.9 Statistical analysis

The SPL measured over the year was summarised using the median and 90 to 10 percentile range presented below as median [10 percentile 90 percentile]. The 90 and 10 percentile of the hearing threshold was converted from the standard deviation (SD) using 1.84 SD. Unless specified, I used this form throughout this study. The significance threshold used was $\alpha = 0.05$. All statistical analysis and visualisation were performed using R version 3.0.1.

5.2.10 Code availability

All regression models were implemented using R base package *lm* version 3.6.2. The source code used to generate the main figures is published at my Github repository ³.

³<https://github.com/ducphucnguyen/WFNSpectrum>

5.3 Results

5.3.1 Uncertainty in wind farm noise measurement

The uncertainty in the measured 1/3-octave band SPLs exponentially reduced with measurement duration as shown in [Figure 5.3](#). Although the uncertainty was different for different 1/3-octave bands, a reducing trend was clearly observed ([Figure 5.3](#)). The median uncertainty over 1/3-octave frequency bands was above 15 dB if the noise was measured over one day. The uncertainty reduced to 5 dB after 14 days of measurement. This means that if two sets of measurements were conducted over 14 sequential days at different times of the year, the median SPL calculated from these data sets could vary up to 5 dB. This variability was expected as WFN depends on meteorological and wind farm operating conditions ([Hansen et al., 2017](#)), and thus measurements taken over a short period of time would not capture longer-term variability of WFN. Although the two-week data collection period is required by WFN guidelines ([SA-EPA, 2021](#)), long-term measurement is important to reduce uncertainty. To determine a reasonable measurement length, I used the auditory perception threshold. Specifically, [Miller \(1947\)](#) found that the ability to discriminate changes in SPL of a broadband noise was typically 0.4 to above 3 dB, depending on the SPL of the noise stimulus. Assuming a threshold of 3 dB, the length of the measurements would need to be at least 50 days. In our measurements, the shortest duration was 70 days while the longest duration was over 400 days, supporting that the data set spanned a sufficient duration to reliably capture long-term WFN variability relevant to human perceptible differences.

5.3.2 Infrasound characteristics

Typical spectrum of infrasound

Wind farm infrasound in the frequency domain was characterised by distinct peaks, ranging from 0.5 to 10 Hz as shown in [Figure 5.4](#). The first peak was the fundamental frequency, which is directly related to the blade-pass frequency. In our measurements, the fundamental frequency was between 0.7 and 0.8 Hz. This fundamental frequency range is common for

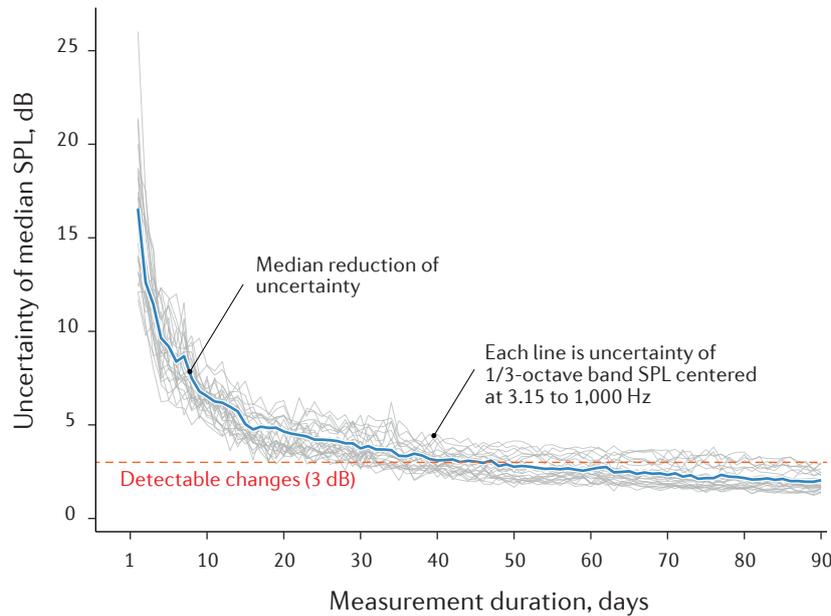


FIGURE 5.3. Uncertainty of wind farm noise levels depending on measurement duration. Each grey line indicates the uncertainty in different 1/3-octave bands over a range of 3.15 - 1,000 Hz. The blue line is the median uncertainty over the 1/3-octave bands.

a modern three bladed wind turbine, which has a nominal rotational speed between 15 and 18 rpm. The fundamental frequency increased with power output (Figure 5.4) which was expected, as wind turbines rotate faster when producing a higher power output. A typical infrasound spectrum also included several harmonic frequencies, which were observed up to approximately 8 Hz.

I also observed an AM tone at 10 Hz, which is unlikely to be caused by a blade-tower interaction mechanism, given that it is higher than the 8th harmonic tone at 7 Hz. Higher harmonics of the blade tower interaction noise are expected to decay linearly. Furthermore, the tone at 10 Hz could be generated by an amplitude modulated tonal noise source. The spectral broadening seems to be stronger at higher power outputs (that is at higher wind speeds) where the power of the tone is completely distributed into the broadband field while at lower power outputs some of the energy still remains in the tone which is hence visible. The AM feature of WFN is not uncommon and it can also be observed in the infrasonic and low-frequency ranges.

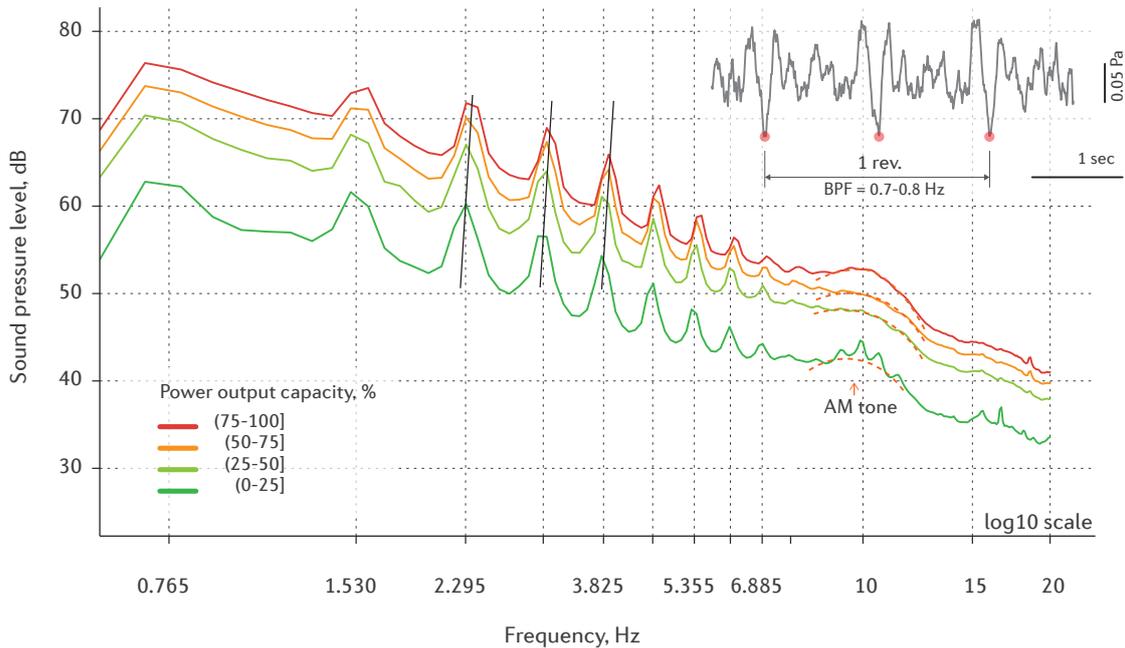


FIGURE 5.4. Typical infrasound spectrum. Typical infrasound spectrum measured at H2 (1.3 km from the nearest wind turbine) at different power output capacities. An example of the acoustic signal is shown in the top-right of the figure. An impulsive signal is clearly observed at the rate of the tower blade-pass frequency (BPF).

Long-term monitoring of infrasound levels

Overall, the infrasound level measured at long-range distances (> 1 km) did not exceed 75 dBG (Table 5.1), which is comparable to previous studies (Jakobsen, 2005; Turnbull et al., 2012). This level was 10 dBG lower compared to the average infrasound hearing threshold of 85 dBG, indicating that infrasound was unlikely to be audible to people with normal hearing living at long-range locations. The infrasound perception level is between 85 and 100 dBG (Broner, 2010), in this study, I used consistently 85 dBG as the average infrasound hearing threshold. The infrasound levels in 1/3-octave bands were also well below the infrasound hearing threshold (see subsection 5.3.4 for hearing threshold calculation details).

The slope of the infrasound spectrum decreased at a rate of 3 to 6 dB per octave band. For example, the slope was 3 dB at locations 1-2, while the slope increased to 6 dB at locations 3-5. A steeper slope was observed at locations further from wind farms. This is consistent

TABLE 5.1. Wind farm infrasound levels over one-year long monitoring

Wind farm	Loc. (dist.)	3.15 Hz	4 Hz	5 Hz	6.3 Hz	8 Hz	10 Hz	L_{Geq}
A	H1(1.0)	54 [41 65]	52 [40 63]	51 [40 62]	49 [38 59]	48 [38 58]	48 [38 57]	58 [50 67]
B	H2(1.3)	56 [43 65]	54 [41 62]	53 [39 60]	50 [38 58]	50 [38 57]	50 [39 58]	61 [50 69]
	H3(3.8)	53 [42 66]	50 [40 63]	48 [38 60]	44 [35 57]	43 [34 55]	42 [33 52]	54 [47 66]
C	H4(3.5)	57 [42 73]	55 [41 71]	53 [40 68]	50 [38 65]	47 [37 62]	45 [35 58]	58 [51 72]
	H6(8.8)	48 [36 64]	45 [34 61]	42 [32 57]	39 [30 53]	38 [29 50]	37 [28 47]	54 [43 66]
Hearing threshold		117 [110 124]	114 [107 120]	110 [103 117]	106 [99 113]	103 [96 109]	98 [91 105]	85

with significant attenuation of higher frequency signal content by the atmosphere, resulting in noise with low-frequency dominance at greater distances from the noise source.

5.3.3 Low-frequency amplitude modulated tones

One of the most prominent characteristics unique to WFN at long-range locations was AM tones occurring at low-frequency. The mechanism(s) responsible for these AM tones is unclear, but it is hypothesised to have a mechanical origin and to be re-radiated from the blades and/or tower. The amplitude modulation of these tones is believed to occur due to periodic changes in loading on the blades (Hansen et al., 2017). Despite the unclear mechanism, I observed this characteristic of WFN at all measurement locations, as summarised in Table 5.2.

An example of this phenomenon is shown in Figure 5.5a, where AM tones were observed at multiple frequencies between 20 and 63 Hz. This phenomenon manifests in the frequency domain as a broadband tone with a bell shape with multiple peaks spaced at the blade-pass frequency (i.e., 0.8 Hz for this wind farm) (Figure 5.5b,c). In the spectrum, the level of the AM tones was clearly above the levels observed in the surrounding frequency region at frequencies higher than 40 Hz (Figure 5.5a). The AM tones were potentially perceivable to some individuals as their levels were within three standard deviations of the hearing threshold (Figure 5.5a).

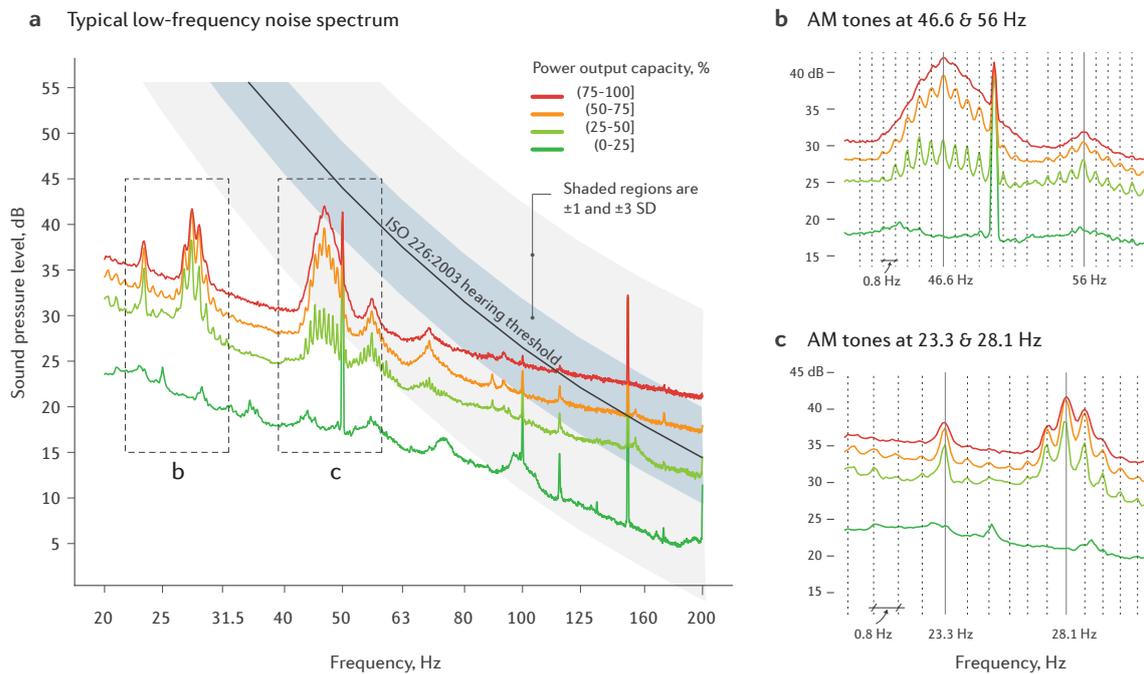


FIGURE 5.5. AM tones at low frequency measured at H4 (3.5 km from the nearest wind turbine). (a) A typical spectrum containing AM tones at low frequencies. The hearing threshold curve and its variability are shown for comparison. Electrical noise occurring at 50 Hz and its harmonic frequencies is observed in the spectrum, but the power spectrum lines at these frequencies are removed from all analyses in this study. Figures (b) and (c) are zoom windows to better illustrate the characteristic of the AM tone in the two important frequency ranges. The tones have a bell shape with side bands spaced at the blade pass frequency.

TABLE 5.2. AM tone characteristics

Wind farm	Location	Distance, km	AM center frequency, Hz	SPL, dB	Hearing threshold, dB
A	H1	0.8	80.0	41 [31 51]	32 [25 38]
B	H2	1.3	25.0	52 [38 57]	69 [62 76]
	H3	3.8	25.0	42 [33 49]	69 [62 76]
C	H4	3.5	23.3	43 [34 50]	72 [65 79]
			28.1	42 [32 51]	64 [58 71]
			46.6	42 [30 52]	46 [40 53]
	H6	8.8	56.0	41 [31 51]	41 [34 48]
			23.3	40 [26 50]	72 [65 79]
			28.1	40 [25 51]	64 [57 71]
			46.6	38 [32 51]	46 [40 53]
			56.0	36 [30 48]	41 [34 48]

5.3.4 Hearing threshold variability

The variability in the hearing threshold between subjects was quantified using one standard deviation ([Figure 5.6a](#)). The variability was slightly higher at 31.5 Hz and lower at 250 Hz. To investigate whether the variability of the hearing threshold was dependent on frequency, I applied a third order polynomial regression fit to the data. However, the association was not statistically significant ($R^2 = 0.003$, $P = 0.313$). Therefore, I used the mean value of 5.35 dB (95% CI 5.12 to 5.58 dB) as the standard deviation of the hearing threshold across frequencies from 3.15 to 1,000 Hz. To examine the variability of the hearing threshold, I plotted the ISO 226:2003 ([ISO226:2003, 2003](#)) hearing threshold curve (for frequencies > 20 Hz) and the infrasound hearing threshold curve in [Møller and Pedersen \(2004\)](#) (for frequencies < 20 Hz) and its 1 and 3 standard deviations as shown in ([Figure 5.6b](#)).

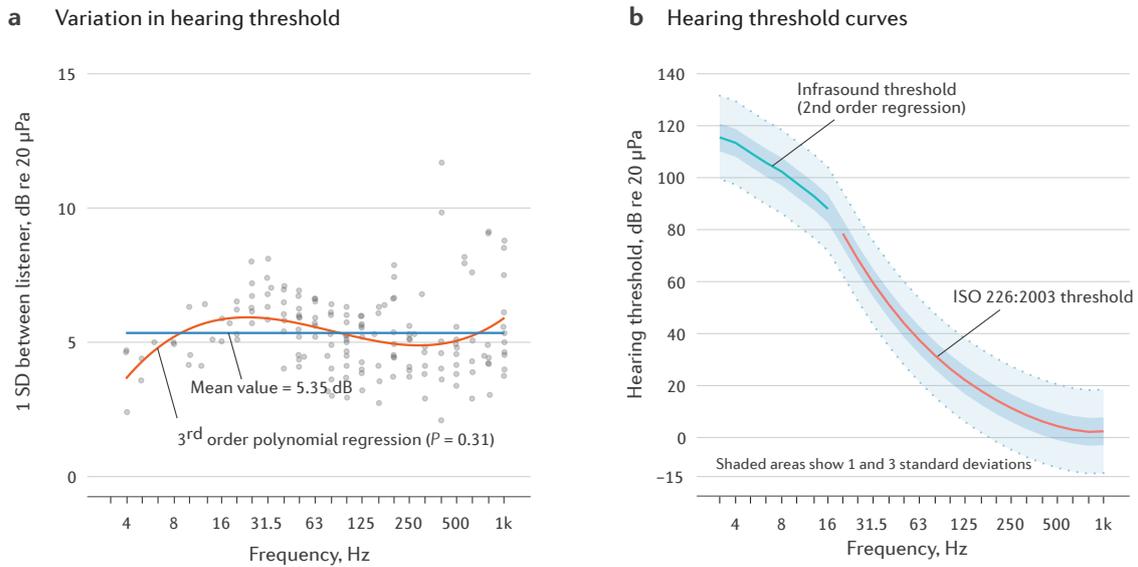


FIGURE 5.6. Hearing threshold variability. (a) One standard deviation of the hearing threshold. Dot points represent one standard deviation of the hearing threshold from different studies. (b) The variability of the hearing threshold (1 and 3 standard deviations) is incorporated in the standard hearing threshold curves. The mean value of the hearing threshold was based on the ISO 226:2003 (ISO226:2003, 2003) hearing threshold curve (for frequencies > 20 Hz) and Møller and Pedersen (2004) (for frequencies < 20 Hz).

5.3.5 Audibility of wind farm infrasound

I used computer simulations to determine the audibility of wind farm infrasound (see Methods section for details). Convergence analysis showed that the probability was stable after 500,000 simulations (Figure 5.7a). To be conservative, I used 1 million simulations throughout this study. These simulations showed that the chance for wind farm noise infrasound to be audible was extremely low with a probability of $< 5 \times 10^{-5}$ (Figure 5.7b). A higher probability was observed at H1 located approximately 1 km from the wind farm, but remained unlikely to be audible to residents living at these distances. As expected, the probability increased at higher frequencies as the hearing threshold decreases with frequency. Visual assessment also revealed this very low chance of audibility as shown in Figure 5.7c. The distributions of infrasound levels in 1/3-octave bands measured over one year was plotted together with the hearing threshold. These two distributions were well separated, further supporting that the infrasound measured at locations greater than 1 km is inaudible.

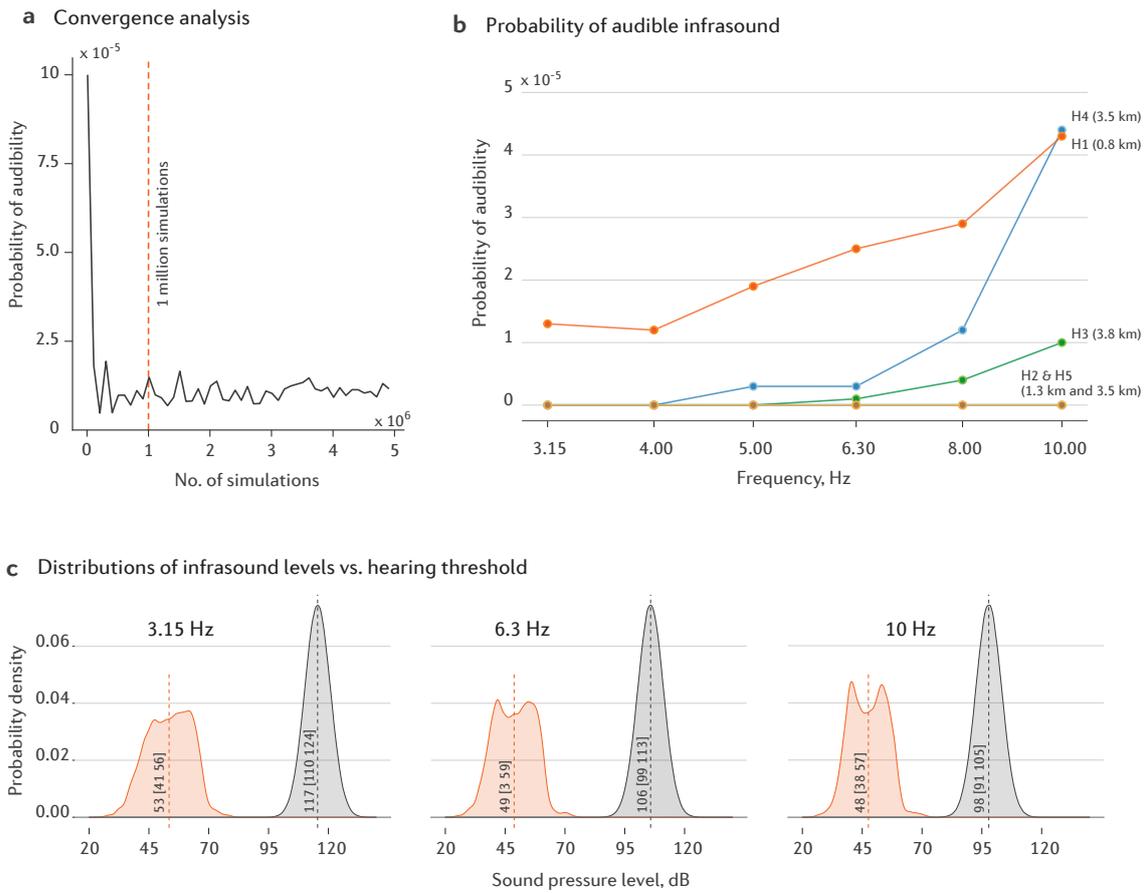


FIGURE 5.7. Audibility of infrasound. (a) Convergence analysis of computer simulation. (b) Probability of audibility of infrasound. (c) Distributions of 1/3-octave SPL measured at H1 in wind farm A and the corresponding hearing threshold.

5.3.6 Audibility of amplitude modulated tones

Amplitude modulated tone of 80 Hz at location H1 near wind farm A

The AM tone produced by wind farm A occurred at 80 Hz and was measured at location H1 at a distance of 0.8 km. The distributions of the measured SPLs and hearing thresholds are shown in [Figure 5.8a](#). The hearing threshold at the 80 Hz 1/3-octave frequency was assumed to have a normal distribution with a mean of 31.5 and a standard deviation of 5.35 dB. A significant overlap between these distributions was observed, indicating that the AM tone was likely to be audible at the measurement location to a significant fraction of the community. The probability of audibility increased with wind farm power output (Spearman's rank correlation, $R^2 = 0.99$, $P = 3.95 \times 10^{-6}$, [Figure 5.8b](#)). At power outputs $> 80\%$, the AM tone was audible to anyone with a normal hearing threshold (Probability = 1). I also observed a high probability at low wind farm power output, but this probability was likely due to background noise only rather than WFN.

I found that wind direction was likely to have an effect on the audibility of the AM tone ($P = 0.014$, [Figure 5.8c](#)). The probability of audibility for the North and East wind directions was higher than that for the West and South directions ($P = 0.010$). The highest probability was observed for the North direction, followed by the East, South and West. Higher probability of audibility for the North and East directions was expected as the residence was downwind from the majority of wind turbines in these directions ([Figure 5.8c](#)).

The probability of audibility during March to May (Autumn season in Australia) was slightly but significantly lower than it was for other months ($P = 0.025$, based on a linear regression of probability for month and day/night). The noise was likely to be more audible during the Summer season (December to February). This seasonal pattern was likely due to the higher wind farm power output during these seasons. I did not find a difference in the audibility between daytime and nighttime at this location ($P = 0.587$).

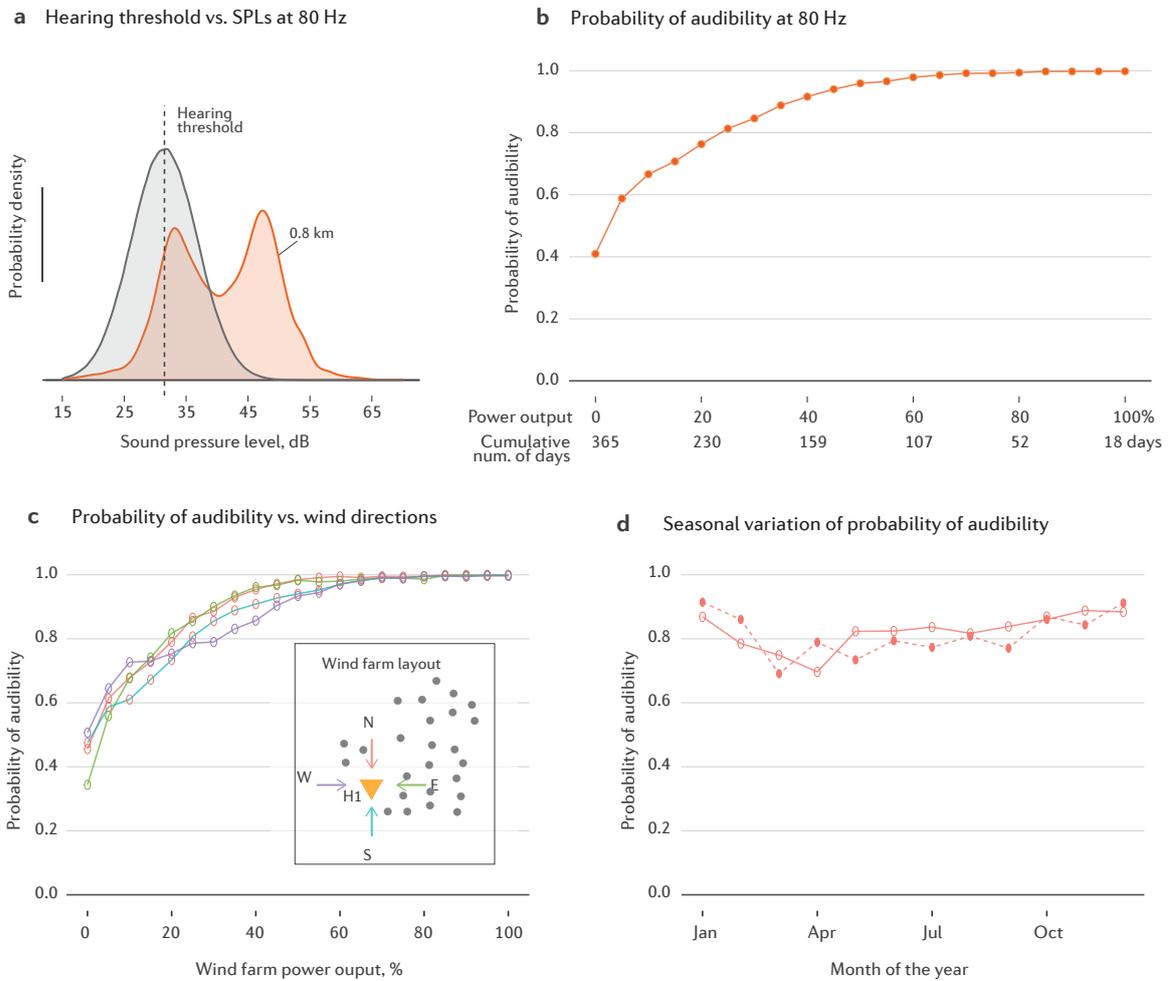


FIGURE 5.8. AM tone at 80 Hz at wind farm A. (a) Distributions of measured SPL and hearing thresholds. (b) Probability of audibility at different wind farm power outputs. On the x-axis, cumulative number of days refers to the number of days that the wind farm power output was at a particular value. (c) Relationship between audibility and wind direction. Inside the rectangle is the wind farm layout, where black dot points represent nearby wind turbines, the orange triangle represents the measurement location and the arrows indicate wind direction. (d) Seasonal pattern of audibility.

Amplitude modulated tone of 25 Hz at locations H2 and H3 near wind farm B

The distributions of the 1/3-octave band SPLs centered at 25 Hz and hearing thresholds are shown in [Figure 5.9a](#). I ran simulations at different wind farm power outputs ([Figure 5.9b](#)), and as expected, the probability of audibility of the AM tone increased with power output at both locations ($R^2 = 0.98$, Spearman's rank correlation). The probability of audibility at H2 was much higher than H3 ([Figure 5.9b](#)) as H2 is closer to the wind farm. The highest probability was around 0.015. This probability could be interpreted as 1.5% of normal hearing residents living around this distance could hear the noise when the wind farm is operating at power outputs $> 80\%$. However, the AM tone was only audible to 0.2% of people when the wind farm was operating at its median annual power output (30%).

At locations H2 and H3 near this wind farm, the audibility was highly dependent on wind direction ($P = 3.326 \times 10^{-11}$), as shown by the regression model of audible probability based on wind direction and wind farm power output ([Figure 5.9c](#)). I observed the highest probability of audibility for the West direction, while the lowest probability was for the East direction. This was expected as the West and East wind directions correspond to the downwind and upwind directions, respectively. In the worst case scenario (i.e., 100% power output and West direction), the AM tone at 25 Hz was audible to 2% of people with normal hearing living at this distance.

I also observed an association between the audibility and month of the year ($P = 1.628 \times 10^{-6}$, linear regression of probability on month and day/night) ([Figure 5.9d](#)). The probability of audibility during Autumn and Winter (April - August) was higher compared to Spring and Summer (September - February). The higher probability during the Winter was very likely due to the higher associated wind farm power output. The overall probability of audibility of the AM tone was not different for the daytime versus the nighttime ($P = 0.459$).

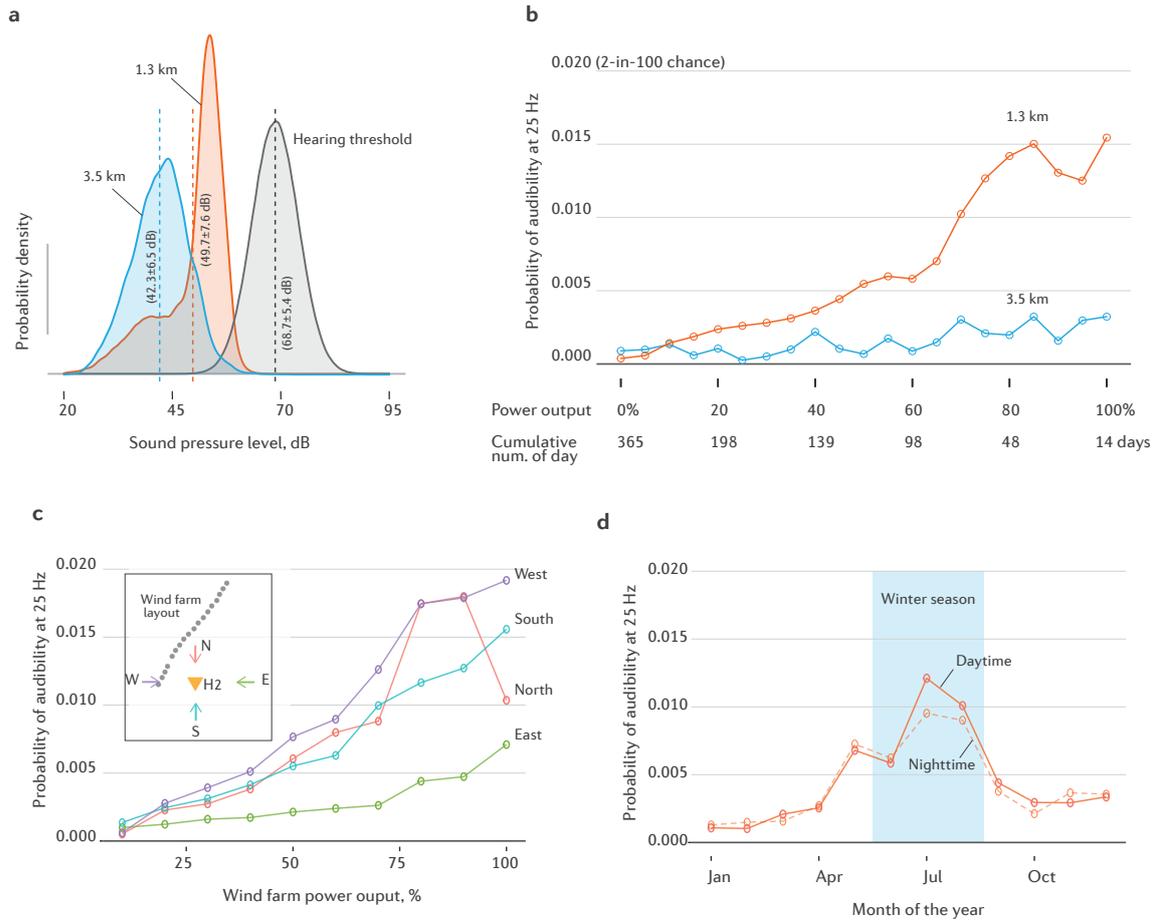


FIGURE 5.9. AM tone of 25 Hz at wind farm B. (a) Distributions of measured SPL at locations H2 and H3 and hearing threshold. (b) Probability of audibility at locations H2 and H3 for different wind farm power outputs. (c) Relationship between audibility and wind direction for location H2. (d) Seasonal pattern of audibility for location H2.

Amplitude modulated tones 46 and 56 Hz at wind farm C

Wind farm C is a special case, as multiple AM tones were observed to occur at different frequencies (Table 5.2). In this section, I only investigated AM tones at 46 and 56 Hz. The hearing threshold at 46 Hz was interpolated linearly between the thresholds at 1/3-octave bands centered at 45 and 56 Hz. The distributions of the measured SPLs and hearing thresholds are shown in Figure 5.10a.

The probability of audibility increased with wind farm power output at both locations (Figure 5.10b). At location H4, the probability increased from approximately 0.2 to over 0.5 when the wind farm power output increased from 0 to 100% ($R^2 = 0.98$, Spearman's rank correlation). Interestingly, the probability of audibility at location H6 (8.8 km) was even higher than that at H4 (3.5 km), particularly at power outputs $> 70\%$. At location H6, the probability of audibility was above 0.7 at 100% power output. In other words, at least 70% of people with normal hearing would be expected to perceive this noise at 8.8 km when the wind farm is operating at 100% power output.

I observed similar results for the AM tone at 56 Hz (Figure 5.10c, d). The probability of audibility was slightly higher compared with the AM tone at 46 Hz. This was expected, as the hearing threshold is lower for higher frequencies. At the annual median power output (approximately 40%), the probability was approximately 0.6 and 0.3 at H4 (3.5 km) and H6 (8.8 km), respectively. At 100% power output, the probability was nearly 0.8 for both locations, indicating that 80% of the normal hearing population would be expected to hear the AM tone at 56 Hz at distances above 3.5 km.

At location H4, I found that the probability of audibility was highly directional. The probability of audibility for the East wind direction was much higher compared to other wind directions (Figure 5.10e). On average, the probability for the West direction was approximately 0.25 higher than other directions ($P = 4.14 \times 10^{-13}$). In other words, the percentage of people who could hear the noise increased by 25% in the downwind direction. The probability of audibility in the East direction increased from 0.2 to approximately 1.0 when power output increased from 0 to 100%. In the worst case scenario (wind farm operating at 100% power output and receivers in the downwind direction), almost all people

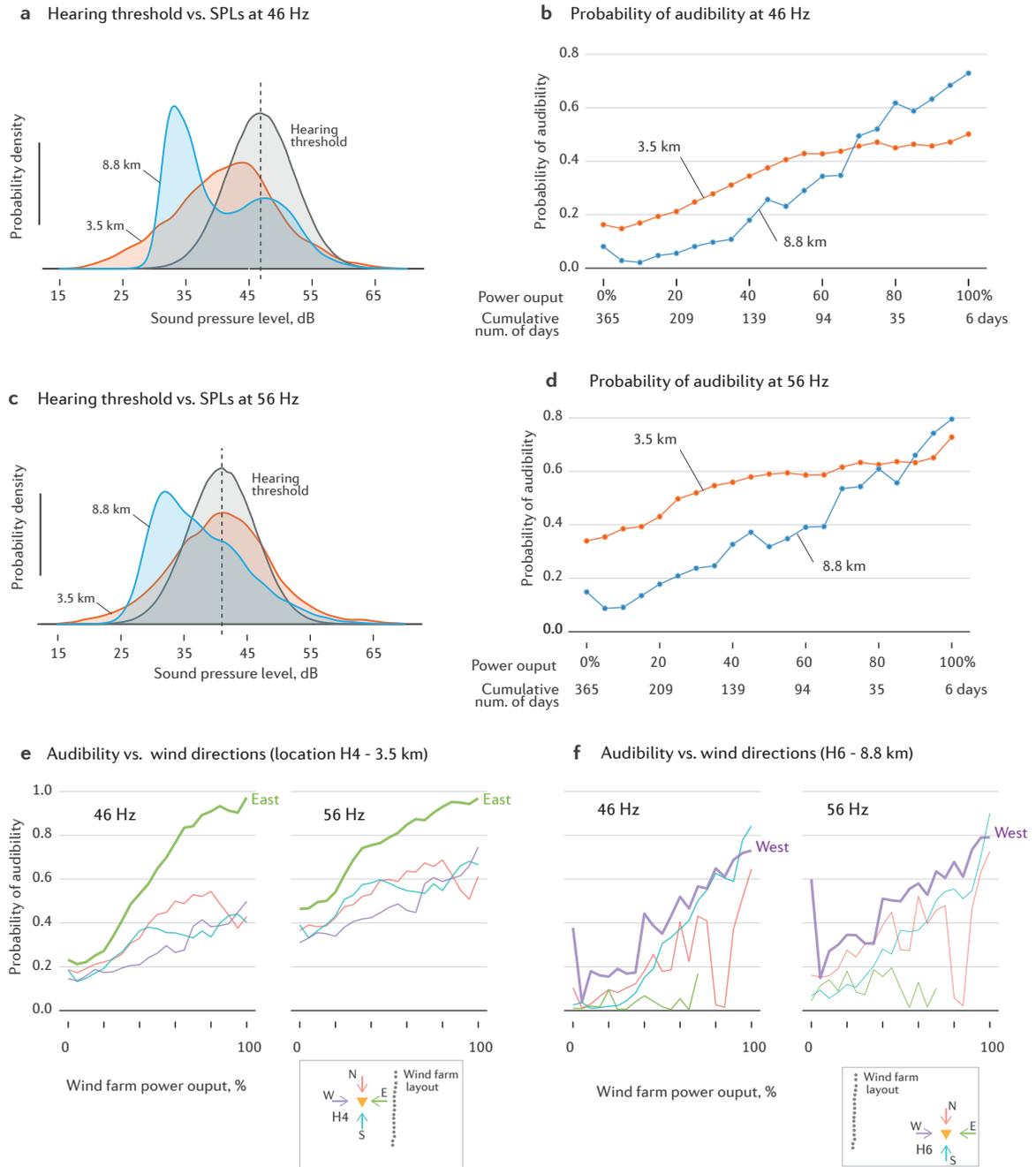


FIGURE 5.10. Audibility of AM tones at wind farm C.

with a normal hearing threshold would be expected to hear this AM tone at 3.5 km. A similar pattern was also observed for the AM tone at 56 Hz.

At location H6, the probability of audibility was highest in the West wind direction. This direction also corresponds with the downwind direction. The probability of audibility for the South wind direction was also high compared to North and East. This was expected, as there are no wind turbines East of H6 (Figure 5.10f). At the highest power output (100%) with a west wind direction, around 80% of people (probability of audibility 0.8) with normal hearing would be expected to hear both AM tones at 46 and 56 Hz at a distance of 8.8 km from the wind farm.

5.4 Discussion

In summary, this work using a computational approach considering uncertainties associated with WFN measurement and variability in human hearing thresholds demonstrates that infrasound is highly unlikely to be audible to people living greater than 1 km from a wind farm. However, AM tones occurring at low frequencies are likely to be readily perceivable to most people with normal hearing up to, and likely beyond, 9 km from a wind farm.

Consistent with previous findings (Jakobsen, 2005; Turnbull et al., 2012; Van den Berg, 2005), it is not surprising that infrasound generated from wind farms is inaudible given levels far below the average human hearing threshold. I confirmed this finding with a more comprehensive approach and large data set. However, whether or not inaudible WFN such as infrasound has adverse effects on humans through mechanisms other than hearing is still debated (Knopper and Ollson, 2011; Tonin, 2018). AM tones occurring within the low frequency range could be audible to people with normal hearing living up to approximately 9 km from a wind farm, especially at high power output and downwind conditions, and could potentially be misinterpreted by some people as infrasound. Preliminary results have shown that AM tones may be more annoying than pure tones (Hansen et al., 2019b). Thus, AM tones that are audible to many under unfavourable conditions could be a significant factor contributing to annoyance at long-range locations. This characteristic of WFN usually

occurs within the infrasonic and low-frequency range. Thus, assessment of WFN based on a common metric such as the A weighted-SPL is likely to underestimate the impacts of WFN on surrounding communities. At long-range locations, low-frequency (but not infrasonic) AM tones could potentially be the only feature of WFN that is problematic because other characteristics of WFN such as infrasound, low-frequency dominance or pure tonality are unlikely to be audible. Therefore, wind farms could be more acceptable to surrounding communities if problematic AM features can be successfully mitigated.

There are several limitations of our study which could be addressed in the future. I was not able to separate WFN from ambient noise. As a result, the measured SPL comprised both wind farm and local ambient noise. Although I attempted to remove samples with extraneous noise, the estimated probability of audibility could be overestimated. Although there is currently no validated method to isolate WFN from ambient noise ([Hansen and Hansen, 2020](#)), recent studies ([Bigot and Hochard, 2019](#); [Gloaguen et al., 2021](#)) using machine learning have shown promising results to address this challenge. A further limitation is that simulations were based on the hearing threshold curve in the ISO 226:2003 for frequencies greater than 20 Hz. This curve is established using data tested on young subjects in the age range from 18 to 25 years. Therefore, results from this study regarding the probability of audibility are more relevant to young populations unlikely to be representative of rural communities exposed to WFN. Data were only available up to a distance of approximately 9 km, so AM audibility at greater distances remain unknown. Future studies, especially in areas associated with noise complaints, should consider that WFN AM could remain audible at greater distances, particularly downwind during high power output conditions.

5.5 Conclusion

I conclude that a computational approach that considers the uncertainty associated with WFN measurement and hearing thresholds is an appropriate and very useful approach to assess the audibility of WFN. I confirmed that infrasound at distances greater than 1 km is highly unlikely to be audible to most individuals, but that low-frequency AM tones are audible to

many individuals at distances up to and likely beyond 9 km from a wind farm. At long-range locations, AM tones could potentially be the only noise problem associated with wind farm operation. Thus, wind farms could be more acceptable to surrounding communities if particularly prominent WFN can be addressed.

Chapter 6

Beyond traditional wind farm noise characterisation with deep acoustic features

This chapter presents work that I published as first author in (Nguyen et al., 2022b). My coauthors were primarily involved in an advisory role so this chapter predominantly reflects my own work.

Contributions: This is an exploratory chapter that presents a innovative approach to characterise WFN. My contributions to this chapter are as follows:

- I showed that deep acoustic features contain a meaningful representation of WFN.
- I showed that combining low- and high-level deep acoustic features works better than using high-level acoustic features alone.
- I showed that deep acoustic features can reveal an improved spatial and temporal representation of WFN compared to traditional acoustic features.

6.1 Introduction

The global wind industry has undergone rapid expansion in power generation capacity over the past ten years, reaching to over 22,000 wind farms and 1,110 GW in 2021 (WindPower, 2021). This fast growth is expected to continue, along with ongoing concerns regarding social (Krohn and Damborg, 1999; Wolsink, 2007), ecological (Schuster et al., 2015; Thaker et al., 2018) and environmental impacts such as noise generated by wind turbines (Liebich et al., 2021a; Micic et al., 2018). Multiple guidelines and standards (AS4959, 2010; NZS6808, 2010; Søndergaard et al., 2019) have been developed to help mitigate the effects of WFN on surrounding communities. Although these guidelines and standards have been updated regularly, the potential impact of WFN is still based on common traditional noise metrics such as A- or C-weighted SPL (Hansen et al., 2017). These aggregate metrics are clearly important indicators related to the human perception of noise. However, there remains no consensus agreement or firm evidence to support which metrics are most strongly related to human impacts and are thus most suitable for WFN assessment (Hansen and Hansen, 2020). Moreover, prominent characteristics of WFN, such as amplitude modulation (AM), also appear to importantly contribute to annoyance (Ioannidou et al., 2016; Lee et al., 2011; Schäffer et al., 2016) and possible sleep disturbance (Smith et al., 2020). Consequently, it is unlikely that any single simple noise metric can adequately encapsulate both physical and psychological aspects of WFN impacts on humans and more comprehensive and evidence-based approaches remain needed. This problem is not unique to WFN research but has also been identified as an issue in other research areas such as sonic boom research for which there is no internationally agreed-upon standard noise metric (DeGolia and Loubeau, 2017).

Recent advances in deep learning in acoustics (Bianco et al., 2019) hold significant promise for improving WFN noise assessment. In particular, Sethi et al. (2020) recently used a pretrained deep convolutional neural network (CNN), namely VGGish (Hershey et al., 2017), to extract feature patterns in spectrograms. Hereafter these features are referred to as deep acoustic features. These authors showed that deep acoustic features can be used to accurately quantify variations in the natural environment across locations and time. Using

a pretrained deep model (DEEP-Hybrid DataCloud project), [Clark et al. \(2021\)](#) were able to characterise details of diurnal and spatial community-generated sound and noise sources. Other useful applications of machine learning models are presented in many aspects of environmental acoustics, such as outdoor sound propagation ([Hart et al., 2016, 2021](#)) and sound emergence of wind turbine noise (the difference in A-weighted SPLs between wind farm operating and non-operating condition) ([Gloaguen et al., 2021](#)).

The purpose of this study was to investigate if deep acoustic features can be used as an alternative to traditional acoustic features for WFN characterisation. I first sought to understand the degree of correlation between deep acoustic features and traditional acoustic features such as spectral shape and commonly derived noise indicators. I then explored the ability of deep acoustic features to reveal the variations in WFN characteristics across locations and time (spatial and temporal characteristics of WFN). I also compared the performance of deep acoustic features and traditional acoustic features to reveal WFN characteristics.

6.2 Methods

This section presents key information regarding measurement locations and data sets. Other details such as equipment, wind farm characteristics and measurement setup are provided in more detail in [Chapter 2](#).

6.2.1 Data sets

All data sets used in this study were collected from four locations in the mid-north region of South Australia ([Figure 6.4a](#)). The first data set (data set 1) was a benchmark data set ([Nguyen et al., 2021b](#)), which contains 6,000 10-second audio files of WFN with 40% of audio samples containing AM. These data were extracted from a data bank containing continuous data measured over one year at locations 1 and 2 ([Figure 6.4b](#)). In this study, the data set 1 was used to evaluate the accuracy of AM detection algorithms trained with deep acoustic features. The second data set (data set 2) included a combination of the first data

set and samples extracted from data measured at locations 3 and 4 (Figure 6.4a). I note that the locations 3 and 4 in this Chapter are previously indicated in Figure 2.1 in Chapter 2 as locations 4 and 6. The data set 2 contained data measured near three wind farms and four residences located between approximately 1 and 9 km from the wind farms. The data set 2 was used to investigate if deep acoustic features can reveal the spatial characteristics of WFN. The third data set (data set 3) was extracted from data measured at location 2 over one year. To reduce the computational time, I extracted 10-second audio files from 10-minute samples. In total, the third data set contained over 50,000 10-second audio samples. The third data set was used to evaluate if deep acoustic features can reveal the temporal characteristics of WFN. More details regarding equipment, measurement setup and wind farm characteristics can be found in my previously published work (Nguyen et al., 2021a).

Regarding data cleaning, although I removed all data containing rain contamination and farming machine noise, I was unable to separate WFN from ambient noise. This is a current issue in the WFN research area and there is no current validated method to address this problem (Hansen and Hansen, 2020). A recent study conducted by Gloaguen et al. (2021) showed promising results to address this challenge and this approach could improve WFN characterisation in future studies.

6.2.2 VGGish

I used the VGGish model, which is a CNN model that has been pretrained on more than 2 million YouTube videos to predict over 600 audio event classes (Hershey et al., 2017). The architecture of the VGGish model is shown in (Figure 6.1a). The VGGish model includes a single channel input layer, followed by four convolutional (CONV) layers and three fully connected layers. VGGish is a variant of the well-known VGG model, in particular Configuration A with 11 weight layers. To prepare the input for the VGGish model, the audio signal was framed into zero overlapping windows of 0.96 seconds. Each window was converted into a spectrogram using a Short-Time Fourier Transform with a window size of 25 ms, a hop length of 10 ms, and a periodic Hanning window. A Mel spectrogram was then computed by mapping the spectrogram to 64 Mel bins covering the range of 125-7500 Hz.

To avoid calculating a logarithm of zero, a log Mel spectrogram was computed by adding a log offset value of 0.001. The result of each window was a 2D spectrogram image 96×64 pixels (i.e., 96 frames by 64 Mel bands). This image was used as the input to the VGGish model. For other details regarding the VGGish model, I refer the reader to the source code provided by the TensorFlow team¹.

A pretrained model was used to save computational and human resources involved with the training and validation of a new model. Although the VGGish model was not trained to identify WFN indicators specifically, it was trained using audio files that contain unique features of noise such as swoosh, swish and thump widely associated with WFN (Hansen et al., 2021). Application of this pretrained model thus could be particularly useful when applied to characterise WFN. Another benefit of using the pre-trained VGGish model is that it is unlikely to be over fitted because it has been trained using a massive data set and is thus capable of classifying a wide range of noise types not necessarily represented within our potentially somewhat location-specific data set that may differ compared to other regions. In other words, this model has the potential to perform well when applied to different data sets containing WFN with a wider range of noise characteristics. Also, the VGGish model was trained using Mel spectrogram images which better reflect how humans perceive sound compared to conventional spectrograms calculated using a short-time Fourier transform. The Mel spectrogram represents a psychoacoustic scale, and thus the resulting deep acoustic features are expected to show stronger relationships with human responses (i.e., annoyance and sleep disruption) compared to conventional spectrograms, although further investigation is needed to verify this. The limitation of VGGish model is that the upper bound frequency of the input Mel spectrogram was high for WFN which typically contains frequencies below 2,000 Hz. Removing these frequencies could thus potentially improve model performance. Also, the frequency content of WFN in the infrasonic and very low frequency ranges (i.e., below 125 Hz) is removed when using the VGGish model. However, this could still be acceptable because wind farm infrasound is normally well below the human hearing threshold (Jakobsen, 2005) and no studies to date have shown that it has adverse

¹<https://github.com/tensorflow/models/tree/master/research/audioset/vggish>

effects on humans. Also, the most commonly identified problematic component of WFN is AM which often occurs within the frequency range between 200 and 800 Hz (AMWG, 2015). This information is thus well captured by the VGGish model.

6.2.3 Dimensionality reduction methods

I used principal component analysis (PCA) (Jolliffe, 2002) to reduce from higher- to lower-dimensional acoustic features. This process can efficiently remove highly-correlated acoustic features which are redundant, resulting in improved machine learning algorithm performance and a reduced risk of overfitting. In our analysis, the number of features extracted from the CONV1 and embedding layers was reduced from 32 to 10 and 128 to 20, respectively (Figure 6.1b). The number of principal components were chosen such that most (95%) of the variance within the high-dimensional data was preserved in the lower-dimensional data (i.e., the cumulative explained variance was above 0.95 in this study). Thus, the total number of combined low-level and high-level deep acoustic features was 30 (Figure 6.1b). To visualise deep acoustic features on a two-dimensional plot I used the uniform manifold approximation and projection (UMAP)(McInnes et al., 2018) to further reduce dimensionality from 30 deep acoustic features to 2 features. The benefit of UMAP compared to PCA is that it can preserve both global and local geometry of the data so is expected to provide superior visualisation of WFN components. In addition, UMAP is non-linear and can therefore represent non-linear factors more efficiently than linear PCA.

I also compared UMAP with another dimensionality reduction method called t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008), which has a favourable running speed and the ability to preserve the global structure of the data. t-SNE is a well-established dimensionality reduction method, but this method does not preserve the global structure of the data very well. Specifically, the distances between paired data points in the high dimensional space (original data) is not correlated well with these distances in the low dimensional space. As a result, although clusters can be identified in the low-dimensional space, how much these clusters differ based on their distances in the low-dimensional space remains unclear. For example, larger or shorter distances between two

clusters in the low-dimensional space does not imply that these clusters are very or slightly different. In other words, the relative distances between clusters in the low-dimensional space are not meaningful. On the other hand, UMAP is expected to address this problem of t-SNE. The fundamental background of UMAP is similar to t-SNE. However, UMAP can preserve both the global and local structure of the data. To demonstrate this quantitatively using my data, I first calculated the distances between paired points in both high- and low-dimensional spaces (see Section 6.2.5). I then did a correlation analysis to quantify the level of correlation between these distances. I ran this analysis for five methods including PCA, t-SNE, and UMAP, applying PCA before using t-SNE (PCA-tSNE), and applying PCA before using UMAP (PCA-UMAP). The results are shown in Section 6.3.3. Note that to facilitate comparisons between methods, I used the default settings of these methods. Software versions and parameters used are listed in Table 6.1.

TABLE 6.1. Software versions and parameters.

Method	Version	Parameter	Availability
t-SNE	scikit-learn 0.24.1	perplexity 30, n_iter 1000, metric: 'euclidean'	https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html
UMAP	0.5.1	min_dist 0.1, n_neighbors 15, metric: 'euclidean'	https://github.com/lmcinnes/umap
PCA	scikit-learn 0.24.1	N/A	https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

6.2.4 Computational time

To estimate the run time of each dimensionality reduction method, I calculated the average over five run times. The run time was estimated with *time* function in Python. The code was

run on a single node of a Deepthought high performance computer² with 128 threads and 500G RAM.

6.2.5 Correlation of pairwise distances

I randomly selected two data points on the deep acoustic feature space. The Euclidean distance between these two points (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n) on n -dimensional space is calculated as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^{i=n} (x_i - y_i)^2} \quad (6.1)$$

The above process was repeated for 10,000 times to estimate 10,000 pairwise distances on the deep acoustic feature space. This process was also applied to embedding space to estimate 10,000 corresponding pairwise distances. To quantify the preservation of distances, I computed the Pearson correlation coefficient between pairwise distances in the deep and the embedding space.

6.2.6 Performance of deep acoustic features

To evaluate the ability of deep acoustic features to identify WFN AM, I used a recent advanced and successful machine learning method called Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016) to identify AM in data set 1. The XGBoost algorithm uses multiple decision trees to capture nonlinear relationships between input variables (acoustic features) and output (AM vs. no AM) to make an ensemble prediction. I also used XGboost classifiers to evaluate the ability of deep acoustic features to reveal spatial and temporal characteristics of WFN. All XGboost classifiers were trained on 80% and evaluated on 20% of the data. Hyperparameters of XGboost classifiers such as *learning_rate* and *max_depth* were optimised with the Bayesian hyperparameter tuning method (Snoek et al., 2012). The searching space and optimised hyperparameters of the AM classifiers are provided in

²<https://doi.org/10.25957/FLINDERS.HPC.DEEPThought>

TABLE 6.2. Hyperparameter tuning.

Parameter	Searching space	Distribution
max_depth	{1, 20}	discrete uniform (step=1)
learning_rate	{0.05, 1}	continuous uniform
gamma	{1, 20}	continuous uniform
reg_alpha	{0, 30}	discrete uniform (step=1)
reg_lambda	{0, 1}	continuous uniform
colsample_bytree	{0.1, 1}	continuous uniform
min_child_weight	{1, 50}	discrete uniform (step=1)

TABLE 6.3. Optimised Hyperparameters.

Parameter	AM detection (traditional features & AM features)	AM detection (deep features)
max_depth	13	19
learning_rate	0.07	0.13
gamma	1.65	1.1
reg_alpha	1	5
reg_lambda	0.64	0.79
colsample_bytree	0.7	0.92
min_child_weight	50	31

Table 6.2 and Table 6.3. I report the performance of the classifiers on the test sets throughout this study.

6.2.7 Statistical analysis

Statistical analyses including Pearson's correlations and receiver operating characteristic (ROC) curve analyses were performed using R³ version 4.0.0. ROC area under curve (AUC) was used as the main classifier performance metric throughout this study. All visualisation was implemented in R using *ggplot2* and *circlize* packages. All grouped data are reported as mean \pm standard deviation (SD).

³<http://www.r-project.org/>

6.2.8 Data and code availability

The data sets used in this study are available in [Nguyen \(2021b\)](#). The source code used to extract deep acoustic features and generate the main figures is published at [GitHub](#)⁴.

6.3 Results

6.3.1 Deep acoustic features

I proposed a different approach to previous studies ([Sethi et al., 2020](#)), in which features in convolutional and embedding layers were extracted instead of features in the embedding layer only. I hereafter refer to features extracted from convolutional layers as low-level acoustic features, while features extracted from the embedding layer as high-level acoustic features ([Figure 6.1b](#)). While high-level acoustic features can capture global acoustic content of an input spectrogram such as energy distribution as a function of frequency and time, I also expected that low-level acoustic features could indicate local patterns in the input spectrogram such as intermittent acoustic features represented by vertical and horizontal lines. Combining these features was expected to capture both the general and detailed characteristics of the noise. To investigate the meaning of each layer in the deep model, and to test if low-level acoustic features indeed capture detailed patterns of the input spectrogram, I show a case study in [Figure 6.1b](#). In this case study, a spectrogram of a 0.96-second audio recording was input to the deep model. The input spectrogram was convoluted with 64 filters to create 64 new images in the CONV1 layer. Each filter was trained to detect different patterns of the input spectrogram. To visualise this, I highlighted two fundamental filters and showed the corresponding images in the CONV1 layer ([Figure 6.1b](#)). The filters successfully detected vertical (distribution of acoustic energy over frequency bands) and horizontal (intermittent acoustic energy over time) patterns of the spectrogram. There were 64 filters in the CONV1 layer that can thus detect other detailed patterns of the input spectrogram. These new spectrogram images representing local patterns of the input were then passed to deeper

⁴<https://github.com/ducphucnguyen/TransferLearningWFN>

layers (i.e., CONV2-4 layers), where the number of filters were doubled after each layer, resulting in much more detailed and comprehensive patterns of the input spectrogram. I reasoned that this combination of low- and high-level acoustic features (see [Figure 6.1b](#)) can comprehensively capture both detailed and general information about the input noise sample. However, features extracted from deeper CONV layers were very localised rather than global. In this study, I only used the low-level acoustic features extracted from the CONV1 layer ([Figure 6.1b](#)). The explanations for this are analysed in details in [Section 6.3.2](#). Finally, the deep acoustic features include 10 low-level features and 20 high-level features.

6.3.2 Deep acoustic features reveal noise characteristics

To investigate potential relationships between deep and traditional acoustic features and WFN unique features (i.e., AM characteristics), I ran a pairwise correlation analysis, in which all possible pairs between these features were determined, and their corresponding Pearson's correlation coefficients, r , were then calculated. The traditional acoustic features included spectral shape features (i.e., spectral slope, spectral spread, spectral flux) and environmental noise indicators (i.e., A-, C-weighted SPLs). The AM features included AM fundamental frequency and AM depth. A full list of these features and their physical meaning is provided in [Table 3.2](#) (or Table 3 in ([Nguyen et al., 2021b](#))). The pairwise correlation coefficients were presented in a Chord diagram⁵ as shown in [Figure 6.2a](#). I observed moderate to strong correlations between these features. Both low- and high-level acoustic features were correlated with traditional acoustic features and AM features, indicating that both low- and high-level acoustic features are useful for capturing information about the noise character. Deep acoustic features were correlated with the most important AM features, such as the AM fundamental frequency, *peakloc*; AM strength (depth), *DAM*, *AMfactor*; rising and decay slopes of AM peaks, *pos_slope*, *neg_slope*; and variations of unweighted SPL in octave bands centred at 1,000 Hz, *L_1000*. Furthermore, deep acoustic features were also correlated with common traditional acoustic features such as A-, C-, G-weighted SPLs and spectral shape features such as spectral spread and spectral slope. This is striking as the correlation

⁵<http://circos.ca/>

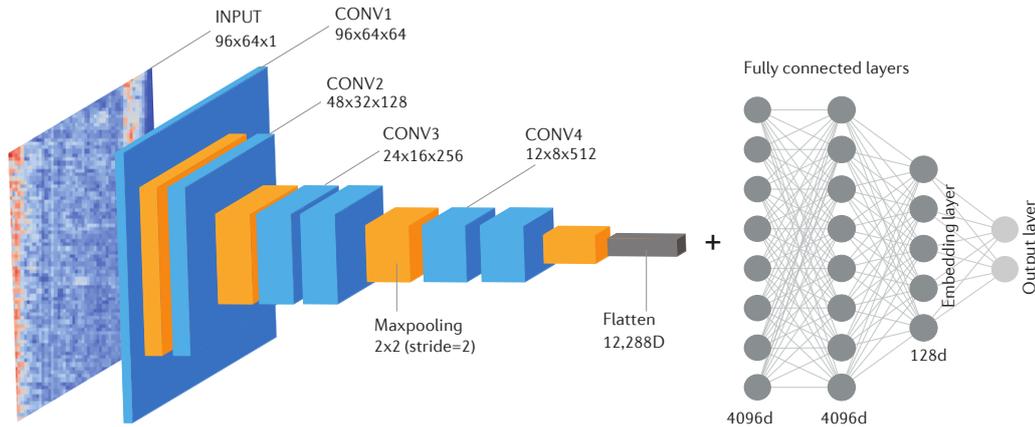
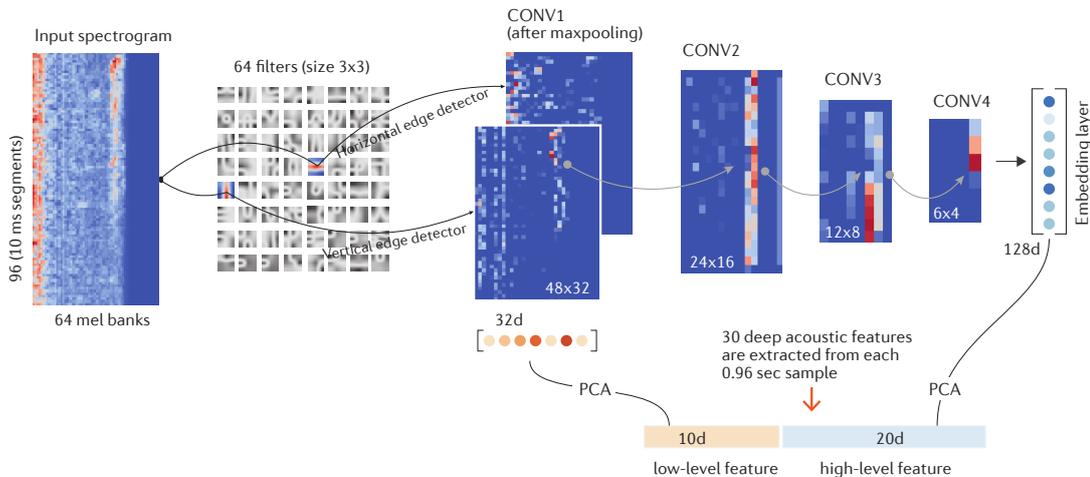
a VGGish deep convolutional neural network architecture**b** Illustration of VGGish feature extraction

FIGURE 6.1. VGGish and deep acoustic features. **a** The architecture of the VGGish model. The general architecture is that the image size after each convolutional layer was reduced to half, while the number of filters were doubled. The embedding layer is the layer before the output layer. **b** An illustration of the VGGish feature extraction. Curved paths represent how the input spectrogram is convolved with the filters to create new images in the CONV1 layer. I draw these paths for only two fundamental filters, which are horizontal and vertical detectors. I also show random images in the CONV2-4 layers for illustrative purposes. I finally combine the features in the CONV1 and embedding layers to represent the acoustic information of the input noise sample. These combined features are referred to as deep acoustic features.

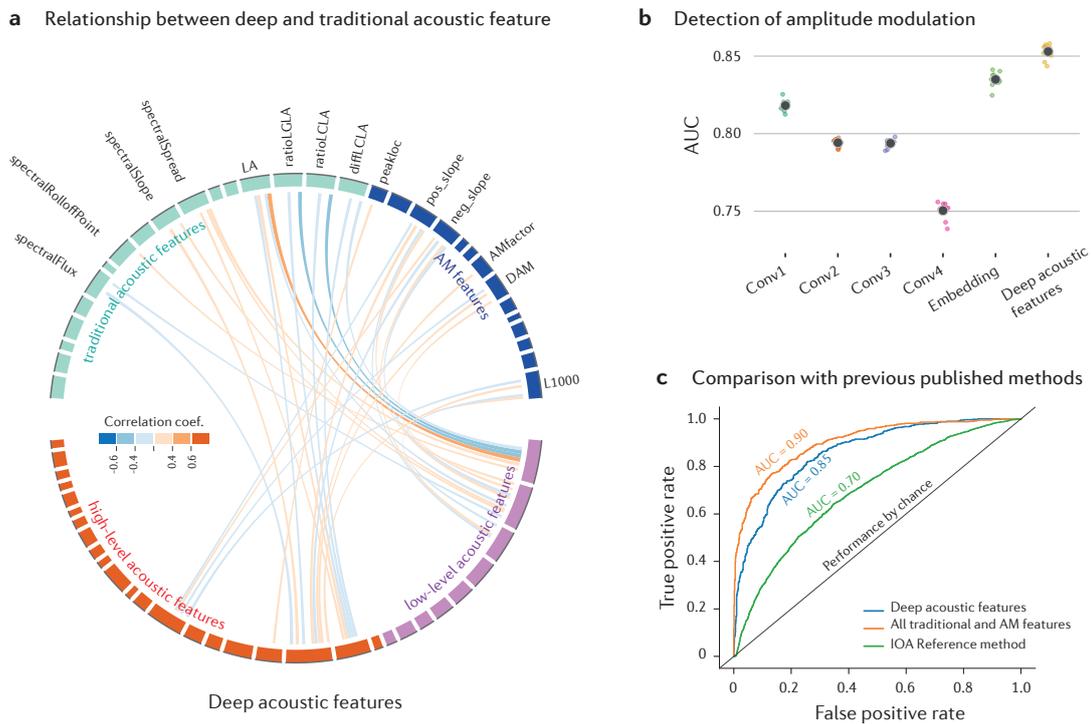


FIGURE 6.2. Deep acoustic features reveal unique noise characteristics. **a** Chord diagram describing the relationship between deep acoustic features, traditional acoustic and AM features. The colour of the curves connected between the two features indicates the degree of correlation. I only show curves representing correlation levels ranging from moderate to very strong ($r > 0.39$ or $r < -0.39$). **b** Performance of models trained with features extracted from CONV1-4 layers, embedding layers and deep acoustic features. The black dot points indicate the mean value of AUC, while the small coloured dot points indicate AUC values from 10 repeated runs. **c** Performance of the model trained with deep acoustic features is compared with previously published methods.

analysis reveals that the deep acoustic features are not only able to represent traditional acoustic features, but are also related to unique characteristics of WFN.

To further investigate the ability of deep acoustic features to detect AM, I trained an XGBoost classifier to detect AM in data set 1 and measured the performance using AUC (see Methods section). I also compared the performance of deep acoustic features and features extracted from CONV and embedding layers. The models trained with features extracted from CONV1 and embedding layers had higher performance compared with those trained with features extracted from CONV2-4 layers (Figure 6.2b). The deep acoustic features showed the best performance. This was expected, as the deep acoustic features were a combination of features extracted from CONV1 and embedding layers (Figure 6.1b).

Finally, I compared the performance of the deep acoustic features method with previously published methods for identifying AM, including the Reference method developed by the Institute of Acoustics UK (AMWG, 2015) and our previously published method (Nguyen et al., 2021b) using all traditional acoustic and AM features. The performance of the deep acoustic features method was high and better than the Reference method. However, the deep acoustic features method had a lower performance than the model trained with all traditional acoustic and AM features. This was expected, as these features were designed specifically for AM detection, including features from other AM detectors such as *DAM*, *AMfactor*, and prominence ratio. Furthermore, all of these features were carefully selected by acoustic experts for the AM detection task. It is thus not surprising that its performance was the best, but the main advantage of deep acoustic features is that it had high performance without involving an acoustic expert. I expected that deep acoustic features could be used to detect other unique characteristics of WFN such as tonal and impulsive characteristics.

6.3.3 Performance of dimensionality reduction methods

I measured the computational costs of five dimensionality reduction methods, and as expected I found that PCA was the fastest method (Figure 6.3a). UMAP was also a very efficient method that was much faster than t-SNE. Using PCA initialisation reduced run times of t-SNE and UMAP. The run time of each method was 11 secs (PCA), 2.2 hrs (t-SNE), 477 secs (UMAP), 1.4 hrs (PCA-t-SNE) and 344 secs (PCA-UMAP). I found that UMAP was much better than t-SNE in preserving global data structure (Figure 6.3b). Using PCA initialisation could improve the ability of t-SNE and UMAP to preserve the global data structure. The global structure of the data is especially important in the case where distances in the two-dimensional data are used as a new metric for assessing environmental noise. Our results here suggest that UMAP with PCA initialisation appeared to be a suitable method for environmental noise data. Specifically, UMAP with PCA initialisation is efficient for analysing large data sets, shows comparable performance to other nonlinear dimensionality reduction methods for classification tasks and has the highest ability to preserve the global data structure.

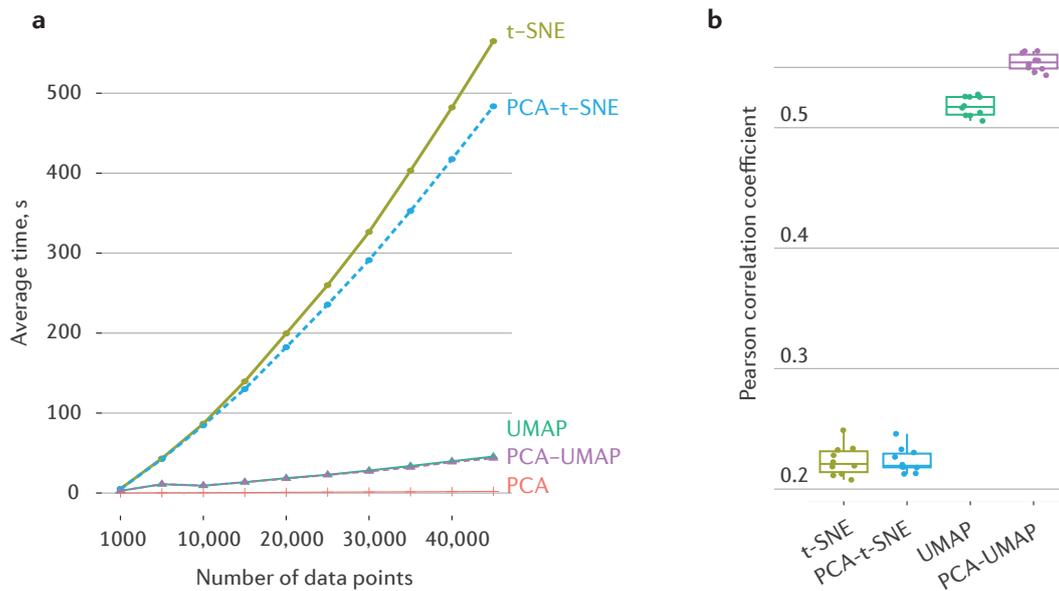


FIGURE 6.3. Dimensionality reduction methods. **a**, Average computational time of five dimensionality reduction methods. t-SNE was much slower than UMAP. Although using PCA to initialise data can slightly improve the speed of t-SNE, it is still much slower compared to UMAP and PCA-UMAP. The speed is important because environmental noise monitoring was usually carried over months or years. **b**, Preservation of global structure of data. The correlation analysis clearly illustrated that UMAP is significantly better than t-SNE in terms of preserving the global structure of data. Using PCA to initialise the data can improve the global structure of UMAP, but not t-SNE.

6.3.4 Deep acoustic features reveal spatial and temporal structures of wind farm noise

Motivated by the ability of deep acoustic features to represent acoustic information and unique characteristics of WFN, I further explored if I could use deep acoustic features to reveal the spatial and temporal characteristics of WFN. I refer to the spatial and temporal noise characteristics as unique characteristics of the noise within a given area or particular time. I observed four clear clusters as shown in [Figure 6.4b](#), corresponding with noise measured at four locations 1-4 as shown in [Figure 6.4a](#). This indicates that the characteristics of the noise at these locations are different from one another. The distances between these clusters could be used to assess the degree by which the noise differs between locations. To evaluate the performance of deep acoustic features, I trained an XGboost model to classify noise at different locations. I found that the performance of deep acoustic features was

remarkable ($AUC = 0.98 \pm 0.003$). The performance of traditional and AM features for this task was 0.76 ± 0.008 , and even lower $AUC = 0.71 \pm 0.01$ if only A- and C-weighted SPLs (two commonly used metrics for WFN assessment) were used. These results indicate that the traditional acoustic features used here are not as sensitive to spatial variations in WFN as deep acoustic features. Although I observed differences in noise characteristics measured at different locations using the deep acoustic features, the underlying reason for these differences remain unclear and the results could be affected by local ambient noise, distance to the wind farm and the number of wind turbines. Larger and more suitable data sets are needed to investigate if the deep acoustic features are sensitive to these changing conditions.

Daytime and nighttime noise characteristics were expected to be different due to atmospheric stability which changes between night-time and daytime and that is also what was revealed using deep acoustic features (Figure 6.4c). The diurnal pattern corresponds to the UMAP1 axis where the nighttime data points were clustered on the left while the daytime data points were clustered on the right. Interestingly, I also observed a transition between the daytime and nighttime noise dependencies. For example, the centroid of data points at clock times close to midnight (22:00 to 2:00) and midday (10:00 to 15:00) were well separated into two groups (Figure 6.4c). The transition hours were distributed between these two groups. These findings indicate that the deep acoustic features are very sensitive to small changes in the diurnal noise dependencies. Although the characteristic of the noise at different seasons was also expected to be different, due to changes in weather conditions, this was however not observed (Figure 6.4d), except for a slightly higher standard deviation in winter. This observation is likely due to specific seasonal effects. It is possible that the weather conditions were not sufficiently different across seasons at our measurement locations to create detectable difference in noise features. For example, there is no snow cover during winter, resulting in a reduced change in the ground impedance compared to other locations in colder climates. Furthermore, the average temperature difference between winter and summer during the measurement period was approximately 15 degrees (Nguyen et al., 2021a). Finally, the wind farm power output between months was not significantly

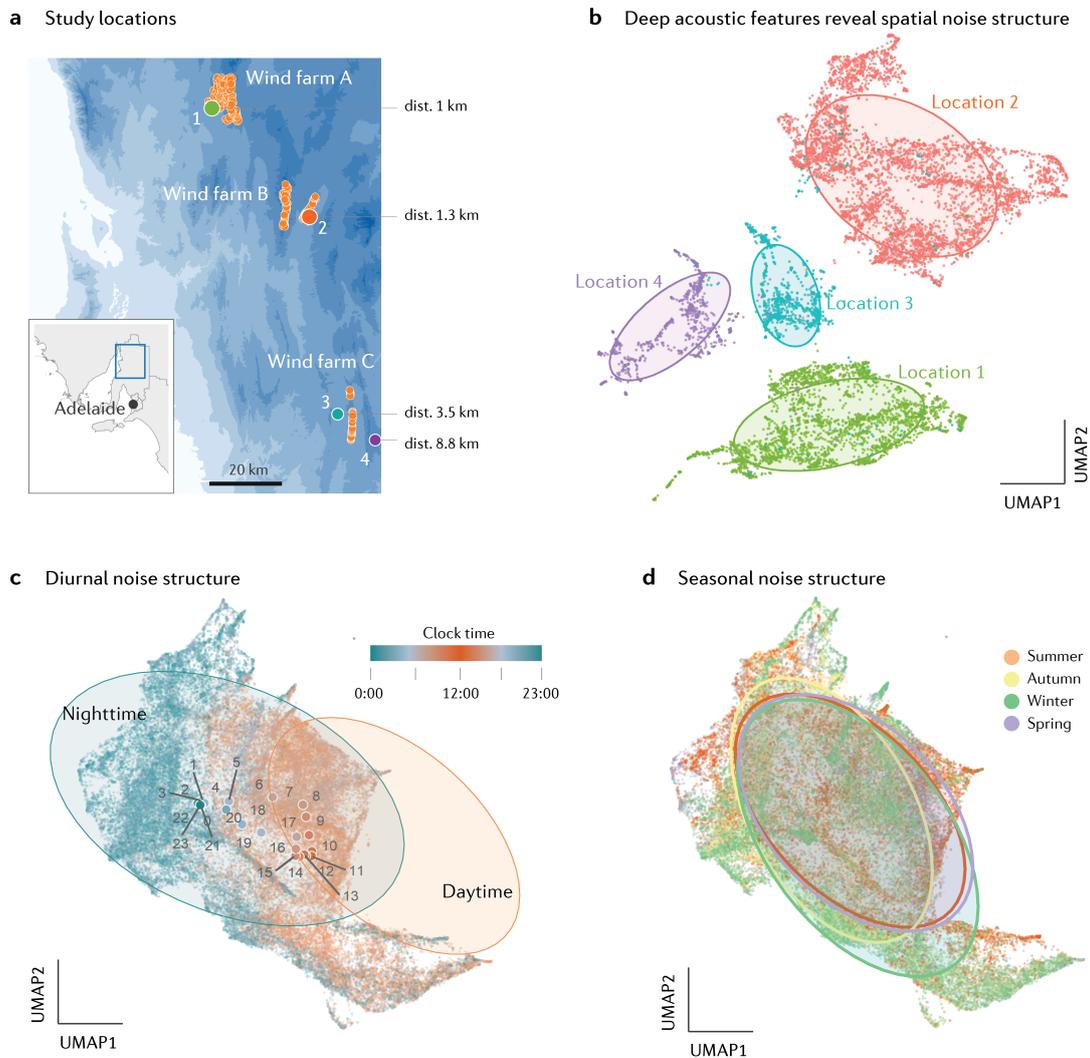


FIGURE 6.4. Spatial and temporal structure of WFN. **a** Noise measured at four locations in South Australia. **b** The number of data points at locations 1-4 are $N_{1-2}=3,000$ and $N_{3-4} = 1,000$. The elliptical shape shows one standard deviation of the data in the cluster. **c** Colours of data points indicate clock time. The numbers indicate the centroid of data points at different clock times from 0 to 23. The number of data points are 56,356. **d** Seasonal characteristic of the noise with Summer: Dec-Feb; Autumn: Mar - May; Winter: Jun-Aug; Spring: Sep-Nov.

different. These factors could explain the small changes in noise characteristics between seasons.

6.4 Discussion

This work shows that deep acoustic features can produce similarly meaningful representations of noise as traditional acoustic features such as A-, C-, G-weighted SPL and spectral shape indicators, but can also represent unique spatial and temporal characteristics such as AM.

Although the current approach to assess WFN is mainly based on common metrics such as A- or C-weighted SPLs, there is still debate around these metrics as they do not capture unique noise characteristics associated with WFN (Hansen and Hansen, 2020). The advantages of traditional acoustic features are that they are already widely implemented, accepted and interpretable. However, these weighted measures have largely evolved around measurement and interpretation convenience with uncertain relationships with human impacts. Given recent advances in computational resources, publicly available deep models, and especially the ability of deep acoustic features to provide more detailed information, more systematic and evidence-based measures beyond traditional approaches for assessing WFN should be considered. For example, the baseline characteristic of the noise can be established using deep acoustic features extracted from data measured during the pre-construction phase of wind farms. The operational characteristic of the noise during the operational phase of a wind farm could also be monitored. The difference in the characteristics of the noise during the pre-construction and operational phases of a wind farm could be estimated and used as an overall indicator to quantify how a wind farm alters the noise characteristics with potential impacts on amenity in surrounding communities specific to the localised environment. A similar approach has indeed been successfully applied in other systems such as for civil (Khoa et al., 2018) and mechanical structural health monitoring (Stetco et al., 2019) systems. I thus anticipate that this approach will have significant utility for more comprehensive evaluation of environmental noise impacts.

A limitation of this study is that using deep acoustic features is a useful approach for characterising WFN and environmental noise in general, but the WFN frequency content below 125 Hz was removed automatically by the VGGish model. This configuration of the model resulted in infrasound and very low-frequency content being missed. The main reason is that the VGGish model was not able to include this frequency content as it was trained using YouTube audio that is unlikely to contain infrasound and very low frequency contents. Also, as mentioned in previous chapters, background noise contamination is a major issue in wind farm noise assessment given the fact that the deep learning method cannot separate WFN and ambient noise, and thus investigating approaches to remove it from wind farm noise remains an active area of research (Hansen and Hansen, 2020). Recent studies have proposed some promising approaches, but these methods have not been tested using real wind farm noise data.

6.5 Conclusion

I conclude that using deep acoustic features is a useful approach for characterising WFN and environmental noise in general. I showed that deep acoustic features represent both overall physical properties and characteristics unique to WFN. Deep acoustic features can clearly reveal the spatial and temporal characteristic of WFN, providing more detailed insight into the noise character than traditional acoustic features. Future use of deep acoustic features holds major promise for comprehensive assessment of the overall character of environmental noise in regions surrounding wind farms.

Chapter 7

Conclusions and Recommendations for Future Work

In this chapter, I first reiterate the main aims of this PhD thesis. I then summarise the findings and contributions to the field. Finally, conclusions and recommendations for future work are also included at the end of this chapter.

7.1 Overview of thesis aims

The general aim of this PhD thesis was to quantify and characterise WFN at long-range locations via long-term noise monitoring. To achieve this general aim, I divided the work into several studies that began with (1) developing and validating a comprehensive model to detect AM; then (2) quantifying and characterising AM at long-range locations in the vicinity of South Australian wind farms; (3) investigating the audibility and characteristics of infrasound and AM tones using a probabilistic approach; and finally (4) exploring a new approach for WFN characterisation using a transfer learning method.

7.2 Summary findings

Chapter 1 introduced the major recent developments relating to the wind industry and the ongoing concerns regarding associated noise issues. A brief overview was also provided of the characteristics of WFN, the underlying noise generation mechanism and possible impacts on humans. Also, promising applications of machine learning to environmental noise was introduced. Within this contextual framework, the major aims were presented.

Chapter 2 provided important knowledge on the technical aspects of WFN measurement. Following a description of the study region and wind farm properties an overview was provided of the key technical aspects that warrant consideration to ensure the collection of high quality acoustic data relevant to measurements of WFN. This included details regarding the selection of microphone types, microphone positions, calibration, wind screen design, data acquisition details and system power supply. This chapter provided a detailed description of methods relevant to Chapters 3-6.

Chapter 3 developed a comprehensive AM detection method based on a machine learning algorithm. The human scored data set was then used to establish benchmark AM characteristics. The results in this chapter showed that this novel method could detect AM as accurately as an expert listener. This advanced AM detection method substantially outperformed traditional AM detection methods which rely on a single noise indicator. The work described in Chapter 3 also showed that human scoring is a feasible and valuable approach to identify prominent features of AM readily perceived by humans. This approach is invaluable for detecting unique characteristics of wind farm noise in cases where the performance of automated detectors is low or not validated against features most obvious to humans. Benchmark characteristics of AM such as the fundamental frequency, AM depth and its relationship with SPL and meteorological conditions were also established.

Chapter 4 leveraged the advanced AM detection method developed in Chapter 3 and a large data set to investigate long-term characteristics of AM at long-range locations. The results showed that AM occurs 2 to 5 times more often during the nighttime compared to the daytime. Indoor AM occurred 1.1 to 1.7 times less often than outdoor AM. A diurnal variation in AM prevalence was also clearly observed. Also, AM prevalence was strongly

associated with sunset and sunrise time. AM occurred most often during both downwind and crosswind conditions. Other characteristics of AM such as AM depth, fundamental frequency and AM event duration were also revealed through application of this novel and comprehensive AM detection approach.

Chapter 5 investigated other characteristics of WFN such as infrasound and AM tones, which are dominant at long-range locations. This chapter used a comprehensive computational approach to evaluate the audibility of infrasound and low frequency noise at these locations. The results confirmed that infrasound is highly unlikely to be audible to normal hearing people living at long-range locations. However, the work presented in Chapter 5 also showed that AM tones occurring at low-frequencies were likely to be audible to most normal hearing people living up to, and likely beyond, 9 km from a wind farm. This work also showed that downwind conditions and high power output were consistently associated with a higher probability of audibility of AM tones.

Finally, Chapter 6 presented an exploratory investigation of the successful use of a transfer learning technique in machine learning to characterise WFN. I demonstrated that deep acoustic features extracted from deep CNN models can very usefully represent a physical meaning for acoustic signals. The deep acoustic features revealed an improved temporal and spatial structure of WFN compared to traditional noise features such as spectral shape or commonly used summary noise metrics. Although suitable data are needed to further validate the results presented in this chapter, this novel approach shows major promise towards use as a new and more comprehensive framework for WFN assessment than is currently practical in field measurements.

7.3 Conclusions

In this thesis I characterised and quantified WFN at long-range locations in a region with a dry climate. I successfully developed an AM detection method which has a predictive power close to the practical limit set by human scoring. Applying this algorithm to the long-term data set revealed strong associations between AM prevalence and meteorological

conditions. The nighttime AM prevalence was much higher than the daytime prevalence. On average, indoor AM occurred less often than outdoor AM, but the indoor AM depth was higher than that measured outdoors. There was also an association between AM prevalence and sunset and sunrise. AM occurred more often during downwind and crosswind conditions. Regarding the impacts of WFN characteristics, AM tones were likely audible at distances up to 9 km, depending on wind direction and wind farm power output. In an exploratory study, I demonstrated that deep acoustic features extracted from a deep CNN model reveal the temporal and spatial structure of WFN better than traditional noise indicators, which could give rise to a new WFN assessment framework. Taken together, these novel methods and findings make an important contribution needed to improve our understanding of WFN. Further application of these methods is likely to facilitate an improved understanding of noise generation and impacts on humans needed to help guide improved design of wind turbines, more comprehensive WFN assessment guidelines and a better framework for WFN assessment, ultimately making wind energy more acceptable to surrounding communities.

7.4 Future directions

Given advances in WFN research in recent years, there are numerous opportunities for future work on this topic since there are many research questions for which answers are still uncertain.

Mechanisms of WFN characteristics. The noise generation mechanisms of some WFN characteristics such as infrasound, ‘swishing’ noise or broadband low frequency noise are well understood. Due to sound propagation loss, these characteristics of WFN have less impact compared to ‘enhanced AM’ or AM tones at long-range locations. Although a considerable amount of research effort has been expended on attempts to understand the cause of these characteristics, causal mechanisms remain uncertain. A greater understanding of the mechanisms responsible for these characteristics would be very useful to help guide wind turbine design improvements and potentially other noise control approaches that could

reduce noise associated with these specific components. AM tones occurring within the low frequency range could be audible to people living at long-range distances. This characteristic of WFN could potentially impact on wind energy acceptance and thus future work is needed to mitigate this problematic AM feature. Quantifying the prevalence of this phenomenon is also important for future studies, especially at wind farms located in other parts of the world.

Reducing noise impacts of wind farms. It is impractical to design inaudible wind turbines, and thus the impacts of WFN on residents living nearby wind farms should be quantified both subjectively and objectively to help facilitate acceptance in neighbouring communities exposed to WFN. It is useful to investigate noise impacts in a field-based setting and using population-based studies to identify the key components of WFN that contribute most to disturbance. Potential mitigation strategies should then be explored to reduce the impacts of WFN on local residents. Due to the remarkably rapid expansion of the wind industry and long-term nature of noise exposure, even modest reductions in noise impacts have major potential to benefit both a wide population and the wind farm industry.

Probabilistic noise prediction. A key stage in the wind farm development process is environmental noise modelling. WFN predictions are usually carried out following procedures outlined in the International Standard, ISO 1996-231 as this model is considered adequate by many practitioners. However, discrepancies between measured and predicted SPLs has been reported in the literature and thus more accurate models, including detailed topography and more accurate calculations of ground and meteorological effects should be developed and validated. In fact, the accuracy of these models is strongly dependent on the reliability of input data and each model has unique strengths and weaknesses. Similar challenges exist in other fields, and while remarkable progress has been made using multi-model ensembles for weather forecasting, crop yield predictions and climate projections, this has not yet translated into the field of acoustics and thus warrants immediate application and advances. It would be useful to explore new multi-model noise prediction methods with improved uncertainty estimates to allow for better evidence-based optimisation of wind farm layouts, reduced instances of non-compliance, and improved public amenity and acceptance of wind energy.

New framework for wind farm noise assessment. WFN measurement is required to follow high standards with regards to equipment, inclusion of a range of weather conditions and measurement durations. However, the assessment of WFN is typically based on a single indicator such as L_{Aeq} or L_{A90} . Other unique characteristics of WFN such as AM, tonality, imbalanced spectrum or other psychoacoustic-related indicators are not usually considered and may well contribute to annoyance. It would be valuable to develop a more comprehensive framework for WFN assessment. Although an early step towards an improved assessment approach using machine learning techniques was demonstrated in Chapter 6, future work is necessary to improve these algorithms and to validate the performance of these machine learning algorithms on a wide range of data sets. With recent advances in computational resources and publicly available deep models, the next step beyond current approaches for assessing WFN should be considered. The new framework should consider the most important features of noise most relevant to human impacts, including both physical and psychoacoustic characteristics. This new framework should examine multiple acoustic indicators, thus minimising bias towards any particular acoustic indicator until sufficient evidence becomes available to guide which WFN features are most closely related to negative impacts on humans.

References

- A. Kjellberg, M. Goldstein, F. G. (1984). An assesment of dB(A) for predicting loudness and annoyance of noise containing low frequency components. *Journal of Low Frequency Noise, Vibration and Active Control*, pages 10–16.
- Alías, F., Socoró, J. C., and Sevillano, X. (2016). A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Applied Sciences*, 6(5):143.
- AMWG (2015). Discussion document: Methods for rating amplitude modulation in wind turbine noise. Technical report, IOA Noise Working Group (Wind Turbine Noise): Amplitude Modulation Working Group.
- ANSI/ASA (1993). Quantities and procedures for description and measurement of environmental sound - part 3: Short-term measurements with an observer present. Technical report, ANSI.
- AS4959 (2010). Acoustics - measurement, prediction and assessment of noise from wind turbine generators.
- Bakker, R., Pedersen, E., van den Berg, G., Stewart, R., Lok, W., and Bouma, J. (2012). Impact of wind turbine sound on annoyance, self-reported sleep disturbance and psychological distress. *The Science of the total environment*, 425:42–51.
- Barlas, E., Zhu, W. J., Shen, W. Z., Dag, K. O., and Moriarty, P. (2017). Consistent modelling of wind turbine noise propagation from source to receiver. *The Journal of the Acoustical Society of America*, 142(5):3297–3310.
- Bass, J. (2011). Investigation of the “Den Brook” amplitude modulation methodology for wind turbine noise. *Acoustics Bulletin*, 36(6):18–24.
- Bass, J., Cand, M., Coles, D., Davis, R., Irvine, G., Leventhall, G., Levet, T., Miller, S., Sexton, D., and Shelton, J. (2016). INSTITUTE OF ACOUSTICS IOA Noise Working Group (Wind Turbine Noise) amplitude modulation working group final report a method for rating amplitude modulation in wind turbine noise version 1. *INSTITUTE OF ACOUSTICS IOA*.
- Baumgart, J., Fritzsche, C., and Marburg, S. (2021). Infrasound of a wind turbine reanalyzed as power spectrum and power spectral density. *Journal of Sound and Vibration*, page 116310.

- Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1):281–305.
- Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., and Deledalle, C.-A. (2019). Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, 146(5):3590–3628.
- Biau, G. and Scornet, E. (2016). A random forest guided tour. *Test*, 25(2):197–227.
- Bies, D. A., Hansen, C., and Howard, C. (2017). *Engineering noise control*. CRC press.
- Bigot, A. and Hochard, G. (2019). Is it possible to predict background noise levels from measured meteorological data with machine learning techniques. In *Proceedings of the 8th International Conference on Wind Turbine Noise, Lisbon, Portugal*, pages 12–14.
- Bowdler, D. (2008). Amplitude modulation of wind turbine noise: a review of the evidence. *Institute of Acoustics Bulletin*, 33(4):31–41.
- Bowdler, R. and Leventhall, G. (2011). *Wind Turbine Noise*. Multi-Science.
- Breiman, L. (1996). Out-of-bag estimation. *Technical report, Statistics Department, University of California Berkeley*.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Broner, N. (2010). A simple criterion for low frequency noise emission assessment. *Journal of low frequency noise, vibration and active control*, 29(1):1–13.
- Brownlee, J. (2017). A gentle introduction to transfer learning for deep learning. *Machine Learning Mastery*, 20.
- Bruel, P. V. (2001). Is a-weighting of noise correct? *InterNoise2001*.
- Chawla, N. V., Lazarevic, A., Hall, L. O., and Bowyer, K. W. (2003). SMOTEBoost: Improving prediction of the minority class in boosting. In *European conference on principles of data mining and knowledge discovery*, pages 107–119. Springer.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.
- Chicco, D. and Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1):6.
- Ciaburro, G., Iannace, G., Puyana-Romero, V., and Trematerra, A. (2021). Machine learning-based tools for wind turbine acoustic monitoring. *Applied Sciences*, 11(14):6488.
- Clark, S. N., Alli, A. S., Nathvani, R., Hughes, A., Ezzati, M., Brauer, M., Toledano, M. B., Baumgartner, J., Bennett, J. E., Nimo, J., et al. (2021). Space-time characterization of community noise and sound sources in Accra, Ghana. *Scientific reports*, 11(1):1–14.

- Coles, D., Bass, H. J., and Cand, M. (2017). IOA AM code - implementation of the core routine for AM analysis from the IOA AMWG. *Sourceforge*.
- Conrady, K., Bolin, K., Sjöblom, A., and Rutgersson, A. (2020). Amplitude modulation of wind turbine sound in cold climates. *Applied Acoustics*, 158:107024.
- Cooper, J. and Evans, T. (2013). Automated detection and analysis of amplitude modulation at a residence and wind turbine. *Proceedings of Acoustics 2013*.
- DeGolia, J. and Loubeau, A. (2017). A multiple-criteria decision analysis to evaluate sonic boom noise metrics. *The Journal of the Acoustical Society of America*, 141(5):3624–3624.
- Department of Industry, Science, E. and Resources (2021). Australian energy statistics.
- Doolan, C. (2013). A review of wind turbine noise perception, annoyance and low frequency emission. *Wind Engineering*, 37(1):97–104.
- Doolan, C. J., Moreau, D. J., and Brooks, L. A. (2012). Wind turbine noise mechanisms and some concepts for its control. *Acoustics Australia*, 40(1).
- DPA (1997). Information no. 9/1997. orientering om lavfrekvent støj, infralyd og vibrationer i eksternt miljø (In Danish). Technical report, Danish Ministry of Environment.
- Dykes, K., Veers, P., Lantz, E., Holttinen, H., Carlson, O., Tuohy, A., Sempreviva, A. M., Clifton, A., Rodrigo, J. S., Berry, D., et al. (2019). Results of IEA Wind TCP workshop on a grand vision for wind energy technology. *IEA Wind TCP*.
- Energy, R. R. (2014). Written scheme relating to Condition 21 Den Brook wind farm. implementation of condition 20 for the identification of greater than expected amplitude modulation. Technical report, RES Developments Ltd.
- Fukushima, A., Yamamoto, K., Uchida, H., Sueoka, S., Kobayashi, T., and Tachibana, H. (2013). Study on the amplitude modulation of wind turbine noise: Part 1—physical investigation. In *Internoise 2013*.
- Gemmeke, J. F., Ellis, D. P., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M., and Ritter, M. (2017). Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 776–780. IEEE.
- Gloaguen, J.-R., Ecotièrre, D., Gauvreau, B., Finez, A., Petit, A., and Le Bourdat, C. (2021). Automatic estimation of the sound emergence of wind turbine noise with nonnegative matrix factorization. *The Journal of the Acoustical Society of America*, 150(4):3127–3138.
- GWEC (2019). Global wind report: Annual market update 2019. <http://gwec.net/global-figures/graphs/>. [Accessed Feb 14, 2021].
- Hansen, C. and Hansen, K. (2020). Recent advances in wind turbine noise research. In *Acoustics*, volume 2, pages 172–207. Multidisciplinary Digital Publishing Institute.
- Hansen, C. H., Doolan, C. J., and Hansen, K. L. (2017). *Wind Farm Noise: Measurement, Assessment and Control*. John Wiley & Sons Ltd, 1 edition.

- Hansen, K., C., H., and Zajamsek, B. (2015). Outdoor to indoor reduction of wind farm noise for rural residences. *Building and Environment*, 94:764–772.
- Hansen, K., Zajamsek, B., and Hansen, C. (2014a). Comparison of the noise levels measured in the vicinity of a wind farm for shutdown and operational conditions. In *INTER-NOISE and NOISE-CON congress and conference proceedings*, volume 249, pages 5192–5202. Institute of Noise Control Engineering.
- Hansen, K., Zajamsek, B., and Hansen, C. (2014b). Identification of low frequency wind turbine noise using secondary windscreens of various geometries. *Noise Control Engineering Journal*, 62(2):69–82.
- Hansen, K., Zajamšek, B., and Hansen, C. (2014c). Comparison of the noise levels measured in the vicinity of a wind farm for shutdown and operational conditions. In *Internoise2014*, Melbourne, Australia.
- Hansen, K. L., Nguyen, P., Micic, G., Lechat, B., Catcheside, P., and Zajamšek, B. (2021). Amplitude modulated wind farm noise relationship with annoyance: A year-long field study. *The Journal of the Acoustical Society of America*, 150(2):1198–1208.
- Hansen, K. L., Nguyen, P., Zajamšek, B., Catcheside, P., and Hansen, C. H. (2019a). Prevalence of wind farm amplitude modulation at long-range residential locations. *Journal of Sound and Vibration*, 455:136–149.
- Hansen, K. L., Nguyen, P., Zajamsek, B., Micic, G., and Catcheside, P. (2019b). *Pilot study on perceived sleep acceptability of low-frequency, amplitude modulated tonal noise*. Universitätsbibliothek der RWTH Aachen.
- Hart, C. R., Reznicek, N. J., Wilson, D. K., Pettit, C. L., and Nykaza, E. T. (2016). Comparisons between physics-based, engineering, and statistical learning models for outdoor sound propagation. *The Journal of the Acoustical Society of America*, 139(5):2640–2655.
- Hart, C. R., Wilson, D. K., Pettit, C. L., and Nykaza, E. T. (2021). Machine-learning of long-range sound propagation through simulated atmospheric turbulence. *The Journal of the Acoustical Society of America*, 149(6):4384–4395.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., et al. (2017). Cnn architectures for large-scale audio classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 131–135. IEEE.
- Hubbard, H. (1990). *Wind Turbine Acoustics*. NASA technical paper. National Aeronautics and Space Administration, Office of Management, Scientific and Technical Information Division.
- Hubbard, H. and Shepherd, K. (1991). Aeroacoustics of large wind turbines. *Journal of the Acoustical Society of America*, 89(6):2495–2508.

- Iannace, G., Ciaburro, G., and Trematerra, A. (2019). Wind turbine noise prediction using random forest regression. *Machines*, 7(4):69.
- IEC61400-11 (2012). Wind turbines - part 11: Acoustic noise measurement techniques.
- Ingielewicz, R., Zagubień, A., et al. (2014). Infrasound noise of natural sources in environment and infrasound noise of wind turbines. *Polish Journal of Environmental Studies*, 23(4):1323–1327.
- Ioannidou, C., Santurette, S., and Jeong, C.-H. (2016). Effect of modulation depth, frequency, and intermittence on wind turbine noise annoyance. *The Journal of the Acoustical Society of America*, 139(3):1241–1251.
- ISO226:2003 (2003). Acoustics: Normal equal-loudness-level contours.
- Jakobsen, J. (2005). Infrasound emission from wind turbines. *Journal of low frequency noise, vibration and active control*, 24(3):145–155.
- Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.
- Jurado, C., Gordillo, D., and Moore, B. C. (2019). Effect of amplitude fluctuations on the loudness of low-frequency sounds. *The Journal of the Acoustical Society of America*, 146(4):3047–3047.
- Keith, S. E., Michaud, D. S., Feder, K. P., Soukhovtsev, V., Voicescu, S. A., Denning, A. R., Tsang, J., Broner, N., and Richarz, W. G. (2019). Wind turbine audibility calculations inside dwellings. *The Journal of the Acoustical Society of America*, 145(4):2435–2444.
- Kelley, N. (1987). A proposed metric for assessing the potential of community annoyance from wind turbine low-frequency noise emissions. Technical report, Solar Energy Research Inst., Golden, CO (USA).
- Khoa, N. L. D., Alamdari, M. M., Rakotoarivelo, T., Anaissi, A., and Wang, Y. (2018). Structural health monitoring using machine learning techniques and domain knowledge based features. In *Human and machine learning*, pages 409–435. Springer.
- Kitzing, L., Jensen, M. K., Telsnig, T., and Lantz, E. (2020). Multifaceted drivers for onshore wind energy repowering and their implications for energy transition. *Nature Energy*, pages 1–10.
- Knopper, L. D. and Ollson, C. A. (2011). Health effects and wind turbines: A review of the literature. *Environmental health*, 10(1):1–10.
- Krohn, S. and Damborg, S. (1999). On public attitudes towards wind power. *Renewable energy*, 16(1-4):954–960.
- Kurakata, K. and Mizunami, T. (2008). The statistical distribution of normal hearing thresholds for low-frequency tones. *Journal of low frequency noise, vibration and active control*, 27(2):97–104.
- Kusiak, A. (2016). Renewables: Share data on wind energy. *Nature*, 529(7584):19–21.

- Larsson, C. and Öhlund, O. (2012). Variations of sound from wind turbines during different weather conditions. *Inter Noise 2012*.
- Larsson, C. and Öhlund, O. (2014). Amplitude modulation of sound from wind turbines under various meteorological conditions. *Journal of the Acoustical Society of America*, 135(1):67–73.
- Lee, G.-S., Cheong, C., Shin, S.-H., and Jung, S.-S. (2012). A case study of localization and identification of noise sources from a pitch and a stall regulated wind turbine. *Applied Acoustics*, 73(8):817–827.
- Lee, S., Kim, K., Choi, W., and Lee, S. (2011). Annoyance caused by amplitude modulation of wind turbine noise. *Noise Control Engineering Journal*, 59(1):38.
- Lever, J., Krzywinski, M., and Altman, N. (2016). Classification evaluation.
- Liebich, T., Lack, L., Hansen, K., Zajamšek, B., Lovato, N., Catcheside, P., and Micic, G. (2021a). A systematic review and meta-analysis of wind turbine noise effects on sleep using validated objective and subjective sleep assessments. *Journal of Sleep Research*, 30(4):e13228.
- Liebich, T., Lack, L., Micic, G., Hansen, K., Zajamsek, B., Dunbar, C., Lechat, B., Scott, H., Lovato, N., Decup, F., et al. (2021b). The effect of wind turbine noise on polysomnographically-measured and self-reported sleep latency in wind turbine noise naïve participants. *Sleep*.
- Liu, X., Bo, L., and Veidt, M. (2012). Tonality evaluation of wind turbine noise by filter-segmentation. *Measurement*, 45(4):711–718.
- Liu, Y., Cheng, Z., Liu, J., Yassin, B., Nan, Z., and Luo, J. (2019). AI for earth: rainforest conservation by acoustic surveillance. *arXiv preprint arXiv:1908.07517*.
- Lundmark, G. (2011). Measurement of swish noise: a new method. In *Fourth International Meeting on Wind Turbine Noise*, Rome, Italy.
- Macmillan, N. A. and Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
- Maffei, L., Masullo, M., Di Gabriele, M., Votsi, N.-E. P., Pantis, J. D., and Senese, V. P. (2015). Auditory recognition of familiar and unfamiliar subjects with wind turbine noise. *International Journal of Environmental Research and Public Health*, 12(4):4306–4320.
- Makarewicz, R. and Golebiewski, R. (2019). The influence of a low level jet on the thumps generated by a wind turbine. *Renewable and Sustainable Energy Reviews*, 104:337–342.
- McCabe, J. (2011). Detection and quantification of amplitude modulation in wind turbine noise. In *Fourth International Meeting on Wind Turbine Noise*, Rome, Italy.
- McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- Merlin, T., Newton, S., Ellery, B., Milverton, J., and Farah, C. (2013). *Systematic review of the human health effects of wind farms*. National Health & Medical Research Council.

- Michalopoulou, Z.-H., Gerstoft, P., Kostek, B., and Roch, M. A. (2021). Introduction to the special issue on machine learning in acoustics. *The Journal of the Acoustical Society of America*, 150(4):3204–3210.
- Micic, G., Zajamsek, B., Lack, t. L., Hansen, K., Doolan, C., Hansen, C., Vakulin, A., Lovato, N., Bruck, D., Ching, t., Chai-Coetzer, L., Mercer, J., and Catcheside, P. (2018). A review of the potential impacts of wind farm noise on sleep. *Acoustics Australia*.
- Miller, G. A. (1947). Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *The Journal of the Acoustical Society of America*, 19(4):609–619.
- Møller, H. and Pedersen, C. S. (2004). Hearing at low and infrasonic frequencies. *Noise & Health*, 6(23):37–57.
- Moreau, S., Roger, M., and Christophe, J. (2009). Flow features and self-noise of airfoils near stall or in stall. In *30th AIAA Aeroacoustics Conference*, Miami, Florida.
- Mungiole, M. and Wilson, D. K. (2006). Prediction of outdoor sound transmission loss with an artificial neural network. *Applied Acoustics*, 67(4):324–345.
- Ngo, T. T. and Letchford, C. W. (2009). Experimental study of topographic effects on gust wind speed. *Journal of wind engineering and industrial aerodynamics*, 97(9-10):426–438.
- Nguyen, P. D. (2021a). Benchmark wind farm noise data set. <https://doi.org/10.25451/flinders.19618023.v1>.
- Nguyen, P. D. (2021b). Long-term quantification of wind farm noise - audio data. https://open.flinders.edu.au/articles/dataset/dataset_for_JASA_express_letter/17451830.
- Nguyen, P. D., Hansen, K. L., Catcheside, P., Hansen, C., and Zajamsek, B. (2021a). Long-term quantification and characterisation of wind farm noise amplitude modulation. *Measurement*, page 109678.
- Nguyen, P. D., Hansen, K. L., Lechat, B., Catcheside, P., Zajamsek, B., and Hansen, C. H. (2021b). Benchmark characterisation and automated detection of wind farm noise amplitude modulation. *Applied Acoustics*, 183:108286.
- Nguyen, P. D., Hansen, K. L., Lechat, B., Hansen, C., Catcheside, P., and Zajamsek, B. (2022a). Audibility of wind farm infrasound and amplitude modulated tonal noise at long-range locations. *Applied Acoustics*, 201:109106.
- Nguyen, P. D., Hansen, K. L., Lechat, B., Zajamsek, B., Hansen, C., and Catcheside, P. (2022b). Beyond traditional wind farm noise characterisation using transfer learning. *JASA Express Letters*, 2(5):052801.
- Nordtest (2002). NT-ACOU-112: Prominence of impulsive sounds and for adjustment of LAeq.
- NZS6808 (2010). Acoustics - wind farm noise.
- Oerlemans, S. (2015). Effect of wind shear on amplitude modulation of wind turbine noise. *International Journal of Aeroacoustics*, 14(5-6):715–728.

- Oerlemans, S., Schepers, J., Guidati, G., and Wagner, S. (2001). Experimental demonstration of wind turbine noise reduction through optimized airfoil shape and trailing-edge serrations. *National Aerospace Laboratory NLR*.
- Oerlemans, S. and Schepers, J. G. (2009). Prediction of wind turbine noise and validation against experiment. *International journal of aeroacoustics*, 8(6):555–584.
- Oerlemans, S., Sijtsma, P., and López, B. M. (2007). Location and quantification of noise sources on a wind turbine. *Journal of sound and vibration*, 299(4-5):869–883.
- Ostashev, V. E. and Wilson, D. K. (2015). *Acoustics in moving inhomogeneous media*. CRC Press.
- Paulraj, T. and Välisuo, P. (2017). Effect of wind speed and wind direction on amplitude modulation of wind turbine noise. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 255, pages 5479–5489. Institute of Noise Control Engineering.
- Paulraj, T. and Välisuo, P. (2020). A method to generate a database of source labelled environmental noise samples using open noise data and to quantify wind turbine noise in it. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 261, pages 2145–2156. Institute of Noise Control Engineering.
- Perkins, R., Lotinga, M., Berry, B., Grimwood, C., and Stansfeld, S. (2016). A review of research into the human response to amplitude modulated wind turbine noise and development of a planning control method. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 253, pages 5222–5233. Institute of Noise Control Engineering.
- Ramachandran, R. C., Raman, G., and Dougherty, R. P. (2014). Wind turbine noise measurement using a compact microphone array with advanced deconvolution algorithms. *Journal of Sound and Vibration*, 333(14):3058–3080.
- Rohatgi, A. (2017). Webplotdigitizer. <https://automeris.io/WebPlotDigitizer/>.
- SA-EPA (2021). Wind farms: environmental noise guidelines.
- Schäffer, B., Schlittmeier, S. J., Pieren, R., Heutschi, K., Brink, M., Graf, R., and Hellbrück, J. (2016). Short-term annoyance reactions to stationary and time-varying wind turbine and road traffic noise: A laboratory study. *The Journal of the Acoustical Society of America*, 139(5):2949–2963.
- Schubel, P. J. and Crossley, R. J. (2012). Wind turbine blade design. *Energies*, 5(9):3425–3449.
- Schuster, E., Bulling, L., and Köppel, J. (2015). Consolidating the state of knowledge: a synoptical review of wind energy’s wildlife effects. *Environmental management*, 56(2):300–331.
- Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., and Napolitano, A. (2009). Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(1):185–197.

- Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., Orme, C. D. L., Wrege, P. H., and Ewers, R. M. (2020). Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proceedings of the National Academy of Sciences*, 117(29):17049–17055.
- Sharma, G., Umamathy, K., and Krishnan, S. (2020). Trends in audio signal feature extraction methods. *Applied Acoustics*, 158:107020.
- Smith, M. G., Ögren, M., Thorsson, P., Hussain-Alkhateeb, L., Pedersen, E., Forssén, J., Ageborg Morsing, J., and Persson Waye, K. (2020). A laboratory study on the effects of wind turbine noise on sleep: results of the polysomnographic WiTNES study. *Sleep*.
- Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25.
- Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., and Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable energy*, 133:620–635.
- Stull, R. B. (2012). *An introduction to boundary layer meteorology*, volume 13. Springer Science & Business Media.
- Søndergaard, L. S., Thomsen, C., and Pedersen, T. H. (2019). Prominent tones in wind turbine noise - Round-robin test of the IEC 61400-11 and ISO/PAS 20065 methods for analysing tonality content. In *8th International Conference on Wind Turbine Noise*. Lisbon, Portugal.
- Thaker, M., Zambre, A., and Bhosale, H. (2018). Wind farms have cascading impacts on ecosystems across trophic levels. *Nature ecology & evolution*, 2(12):1854–1858.
- Thorsson, P., Persson Waye, K., Smith, M., Ögren, M., Pedersen, E., and Forssén, J. (2018). Low-frequency outdoor–indoor noise level difference for wind turbine assessment. *The Journal of the Acoustical Society of America*, 143(3):EL206–EL211.
- Tokita, Y., Oda, A., and Shimizu, K. (1984). On the frequency weighting characteristics for evaluation of infra and low frequency noise. In *Proceedings of INTER-NOISE 84 and NOISE-CON 84*, pages 917–920. Institute of Noise Control Engineering.
- Tonin, R. (2018). A review of wind turbine-generated infrasound: source, measurement and effect on health. *Acoustics Australia*, 46(1):69–86.
- Turnbull, C., Turner, J., and Walsh, D. (2012). Measurement and level of infrasound from wind farms and other sources. *Acoustics Australia*, 40(1):45–50.
- Valente, D. (2013). Data-driven prediction of peak sound levels at long range using sparse, ground-level meteorological measurements and a random forest. *The Journal of the Acoustical Society of America*, 134(5):4159–4159.
- Välisuo, P. O. (2017). Automated wind turbine noise analysis by machine learning. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 255, pages 5667–5678. Institute of Noise Control Engineering.

- Van den Berg, G. (2004). Effects of the wind profile at night on wind turbine sound. *Journal of sound and vibration*, 277(4):955–970.
- Van den Berg, G. (2005). The beat is getting stronger: the effect of atmospheric stability on low frequency modulated sound of wind turbines. *Journal of Low Frequency Noise, Vibration and Active Control*, 24(1):1–24.
- Van den Berg, G. and van den Berg, P. (2006). The sounds of high winds. *The effect of atmospheric stability on wind turbine sound and microphone noise*.
- Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
- Vautard, R., Thais, F., Tobin, I., Bréon, F.-M., De Lavergne, J.-g. D., Colette, A., Yiou, P., and Ruti, P. M. (2014). Regional climate model simulations indicate limited climatic impacts by operational and planned european wind farms. *Nature communications*, 5:3196.
- Veggeberg, K. and Zheng, A. (2008). Real-time noise source identification using field programmable gate array (FPGA) technology. In *Proceedings of Meetings on Acoustics 156ASA*, volume 5, page 055005. Acoustical Society of America.
- Von Hünenbein, S., King, A., Piper, B., and Cand, M. (2013). Wind turbine amplitude modulation: Research to improve understanding as to its cause and effect—Work Package B (2): Development of an AM dose-response relationship. *RenewableUK*, <http://www.renewableuk.com/>(Last viewed April 11, 2016).
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G., Carrillo, O., Sorensen, H. B., Jennum, P., Peppard, P. E., Perona, P., and Mignot, E. (2014). Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nature methods*, 11(4):385.
- Wendt, S. L., Welinder, P., Sorensen, H. B., Peppard, P. E., Jennum, P., Perona, P., Mignot, E., and Warby, S. C. (2015). Inter-expert and intra-expert reliability in sleep spindle scoring. *Clinical Neurophysiology*, 126(8):1548–1556.
- WindPower (2021). Wind farms databases. <https://www.thewindpower.net/>. [Accessed Feb 11, 2021].
- Wolsink, M. (2007). Wind power implementation: the nature of public attitudes: equity and fairness instead of ‘backyard motives’. *Renewable and sustainable energy reviews*, 11(6):1188–1207.
- Yokoyama, S., Sakamoto, S., and Tachibana, H. (2013). Study on the amplitude modulation of wind turbine noise: part 2-auditory experiments. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 247, pages 3136–3145. Institute of Noise Control Engineering.
- Zagubień, A. (2018). Analysis of acoustic pressure fluctuation around wind farms. *Polish Journal of Environmental Studies*, 27(6).
- Zagubień, A. and Wolniewicz, K. (2019). The impact of supporting tower on wind turbine noise emission. *Applied Acoustics*, 155:260–270.

-
- Zajamsek, B., Hansen, K. L., Doolan, C. J., and Hansen, C. H. (2016). Characterisation of wind farm infrasound and low-frequency noise. *Journal of Sound and Vibration*, pages 1–15.
- Zajamsek, B., Yauwenas, Y., Doolan, C. J., Hansen, K. L., Timchenko, V., Reizes, J., and Hansen, C. H. (2019). Experimental and numerical investigation of blade–tower interaction noise. *Journal of Sound and Vibration*, 443:362–375.
- Zhou, L., Tian, Y., Roy, S. B., Thorncroft, C., Bosart, L. F., and Hu, Y. (2012). Impacts of wind farms on land surface temperature. *Nature Climate Change*, 2(7):539–543.