The School of Mathematics

THE UNIVERSITY
of EDINBURGH

Quantifying Wind Turbine-Induced Seismic Noise via Generalized Additive Modeling

by

Tilak Pundalik Heble
s2563426

Dissertation Presented for the Degree of
MSc in Statistics with Data Science

May 26th – June 27th 2025

Supervised by
Ioannis Papastathopoulos and Lily Li

Executive Summary

How much seismic energy do nearby wind turbines contribute, and does this impact the national monitoring capability at Eskdalemuir? This project investigated the seismic effects of wind turbine operations near the Eskdalemuir Seismic Array (EKA), a critical component of the UK's international nuclear monitoring commitments. The key question was whether turbine activity contributes measurable seismic energy — and if so, whether it poses any risk to detection thresholds enforced by national regulation.

Using a large dataset of seismic recordings, wind measurements, and turbine operational status, we developed a robust statistical model to isolate turbine-related seismic energy from natural environmental variation. This model accounted for weather patterns, daily changes, and turbine activity, allowing us to identify even subtle sources of vibration.

The results clearly show that turbine operation is associated with a consistent and statistically significant increase in seismic energy. However, this increase is modest in physical terms and does not exceed the UK's detection threshold of 0.336 nanometres (RMS displacement). At peak, turbine-induced ground motion reached only 53% of the allowed limit, suggesting that turbine installations in this region are currently compliant — and that further expansion may still be feasible if carefully managed.

Three effects were particularly important:

- **Operational Status** was the most significant factor, with turbines producing a clear uplift in seismic energy when active. This demonstrates the detectability of turbine signals even under variable weather conditions.
- **Wind Direction** influenced energy levels in a distinct pattern, with south-westerly winds producing higher readings — likely due to the physical alignment between turbines and the seismic sensor. This highlights the importance of considering turbine orientation and prevailing wind when assessing impact.
- **Daily Effects** captured background fluctuations in seismic energy not explained by weather or turbine status, ensuring the model remained robust to unobserved variation like regional noise or atmospheric changes.

The modeling approach used — a Generalized Additive Model (GAM) — allowed us to reflect these relationships transparently and flexibly, using real-world physical units and avoiding complex black-box methods.

Some limitations apply. The analysis was restricted to a small area and frequency band (0.5–8 Hz) relevant to EKA, and certain site-specific data (e.g., turbine telemetry, soil moisture) were unavailable. These factors could influence generalisability and the accuracy of forecasts under changing conditions. Moreover, the cumulative effects of large-scale turbine expansion remain uncertain and merit further study.

Nonetheless, this study establishes a solid, scalable framework for ongoing seismic monitoring. It equips regulators and energy developers with a practical tool to ensure wind energy deployment remains compatible with national security requirements and environmental oversight.

Acknowledgments

I would like to thank my academic supervisors, **Ioannis Papastathopoulos** and **Lily Li**, for their expert guidance, encouragement, and constructive feedback throughout this project. I am also sincerely grateful to Mr. **Brett Marmo** and **Mr. Josh Walton** at **Xi Engineering** for their invaluable technical insights and for providing access to the seismic datasets. Special thanks to **Sergio Gomez Anaya** for his mentorship and timely support during key stages of the analysis.

I would also like to acknowledge the **School of Mathematics** at the **University of Edinburgh** for the opportunity to work on a project that blends statistical modelling with real-world environmental impact. Lastly, I am deeply thankful to my family and friends for their unwavering support during a difficult year, marked by personal loss and emotional challenges. Their belief in me helped make this work possible.

University of Edinburgh – Own Work Declaration

This sheet must be filled in, signed and dated - your work will not be marked unless this is done.

Name: **Tilak Pundalik Heble**

Matriculation Number: **s2563426**

Title of work: **Quantifying Wind Turbine-Induced Seismic Noise via Generalized Additive Modeling**

I confirm that all this work is my own except where indicated, and that I have:

- Clearly referenced/listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Not sought or used the help of any external professional academic agencies for the work
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Complied with any other plagiarism criteria specified in the Course handbook

I understand that any false claim for this work will be penalised in accordance with the University regulations (`https://teaching.maths.ed.ac.uk/main/msc-students/msc-programmes/statistics/data-science/assessment/academic-misconduct`).

Sincerely,

Signature: Tilak Pundalik Heble

Date: 25 June 2025

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background and Motivation

As the global energy sector transitions toward sustainable sources, *onshore wind farms* have emerged as a key technology for meeting national Net Zero targets. In regions such as southern Scotland, abundant wind resources and suitable terrain offer an ideal setting for renewable energy development. However, this progress introduces new challenges where modern infrastructure intersects with *legacy scientific systems*. One such case is the *Eskdalemuir Seismic Array (EKA)* — a critical node within the *International Monitoring System (IMS)*, used to detect underground nuclear tests under the *Comprehensive Nuclear Test-Ban Treaty (CTBT)* [1].

EKA is located in a relatively remote and seismically quiet area, chosen deliberately to reduce ambient seismic noise. It is designed to detect minute ground motions, particularly within the *0.5–8 Hz frequency band*, which overlaps with the *low-frequency vibrations generated by wind turbines* [9]. These vibrations arise from two main mechanisms: aerodynamic forces acting on the blades, and resonant oscillations of the tower and foundation.

(a) Map showing Eskdalemuir Seismic Array site.

(b) Eskdalemuir Wind Farm, Scotland

Figure 1: Location of Eskdalemuir Seismic Array and Wind farm

As wind energy deployment has increased, so too has concern over potential *seismic interference* from turbines, particularly within close proximity to monitoring stations.

To protect the integrity of EKA's detection capabilities, the UK has imposed *strict planning constraints*: no turbines may be installed within 10 km of the array, and any turbines within 50 km must not collectively exceed a specified seismic noise threshold — typically 0.336 nm root-mean-square ground displacement in the 0.5–8 Hz band [7]. These restrictions significantly limit turbine deployment potential in areas otherwise suitable for renewable energy development. Yet, the *true contribution of operational turbines to seismic noise remains difficult to isolate*, particularly because seismic signals are strongly affected by environmental factors such as wind speed, direction, and local terrain.

This project addresses that challenge by applying statistical methods to *quantify the contribution of wind turbine activity to ground motion*, separating it from natural background seismic noise. Specifically, it focuses on analyzing two parallel seismic datasets collected near a wind farm site — one captured *before* turbine construction and the other *after* the turbines became operational. These data span high-resolution 10-minute blocks and allow direct comparison under different turbine states. By controlling for meteorological factors and using signal processing techniques that emphasize the sensitive frequency range, the analysis aims to deliver a robust, interpretable estimate of turbine-induced seismic energy.

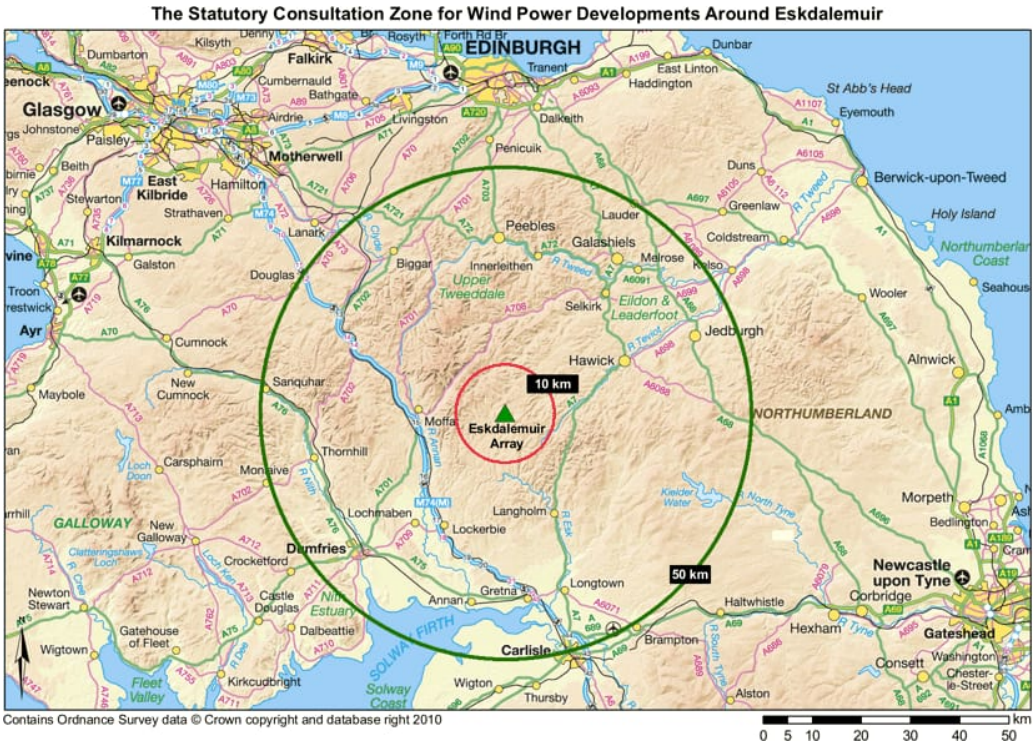


Figure 2: Statutory wind development consultation zones surrounding the Eskdalemuir Seismic Array.

1.2 Project Objectives

The core aim of this project is to determine how much of the observed seismic energy near the EKA site can be attributed directly to wind turbine operation. This involves:

- Constructing a physically meaningful metric of seismic energy by integrating *frequency-weighted displacement power spectral density (PSD)* over the sensitive 0.5–8 Hz range;
- Analyzing how this energy measure varies under different wind and operational conditions;
- Quantifying the *marginal effect of turbine activity* on seismic energy, accounting for environmental confounders.

To accomplish these goals, several statistical modeling strategies were considered, including *linear regression*, *quantile regression*, and *spline-based smoothing models*. Ultimately, *Generalized Additive Models (GAMs)* were selected for their capacity to capture *nonlinear relationships and smooth interactions* without requiring rigid parametric forms [10]. GAMs also provide interpretable visual summaries of how seismic energy changes with wind speed and direction under different operational states.

1.3 Scientific and Practical Significance

This study sits at the interface of environmental geophysics, statistical modeling, and energy infrastructure planning. Scientifically, it contributes to a growing body of research focused on *anthropogenic sources of seismic noise*, a subject of increasing relevance as renewable energy projects expand into quieter rural areas. From a practical perspective, the findings offer *evidence-based insight* into how seismic monitoring can coexist with wind energy development, helping policymakers balance national energy goals with treaty compliance and geophysical integrity.

Technically, the project demonstrates how *modern statistical tools* — such as GAMs — can be combined with *domain-specific physical constraints* (e.g., frequency weighting) to yield transparent, reproducible, and context-aware analyses. Unlike traditional regression, GAMs accommodate the reality that the relationship between wind speed and seismic response is *nonlinear* and may vary depending on turbine status. By leveraging high-resolution time-series data and meteorological inputs,

the model can adapt to complex ground motion patterns that are otherwise difficult to represent analytically.

This work also has broader implications for *automated seismic filtering*, *early-warning systems*, and *site assessment frameworks*. For example, understanding which wind speeds are most associated with harmful frequency components could guide turbine siting or lead to targeted signal cleaning in EKA's operational pipeline.

1.4 Report Structure

Report Section Overview

The remainder of this report is structured as follows:

- **Section 2: Data and Preprocessing** — Introduces the seismic and environmental datasets, defines the energy metric, and constructs relevant features.
- **Section 3: Modeling Methodology** — Details the GAM framework, variable selection, interaction terms, and model evaluation.
- **Section 4: Results** — Presents fitted effects, interprets turbine-related uplift, and visualizes RMS compliance.
- **Section 5: Discussion** — Reflects on findings, model assumptions, and industrial implications.
- **Section 6: Conclusion** — Summarizes contributions and highlights future directions.

2 Data Description

2.1 Site Context and Seismic Monitoring Relevance

This study investigates the influence of wind turbine activity on ground motion recorded near the **Eskdalemuir Seismic Array (EKA)**, a critical seismic station situated in southern Scotland. EKA forms a central component of the **International Monitoring System (IMS)** and plays a key role in enforcing the **Comprehensive Nuclear-Test-Ban Treaty (CTBT)** by detecting low-frequency ground vibrations that may signal clandestine nuclear tests [1].

The seismic array is deliberately located in a seismically quiet region to minimise background noise and maximise sensitivity in the **0.5–8 Hz** frequency band [9]. However, this band coincides with vibrations generated by wind turbines, such as aerodynamic blade loading and tower resonance, making them a potential source of anthropogenic seismic noise.

The wind farm examined in this study is located within the CTBT-designated 50 km buffer zone surrounding EKA. To evaluate the seismic influence of wind turbines, data were collected over two operational phases:

- **Background (Pre-Operational)** phase: Turbines were erected but remained inactive;
- **Operational** phase: Turbines were actively generating electricity.

This temporal design facilitates a controlled before-and-after comparison, holding natural environmental conditions approximately constant.

2.2 Overview of Data Sources and Structure

The analysis integrates three core data types, each synchronised to 1-second resolution and spatially co-located:

2.2.1 Time-Domain Seismic Data

Raw vertical ground velocity data were collected from two CSV files: `Background_Time_Series_WS12.csv` and `Operational_Time_Series_WS12.csv`. Each row represents a 1-second time frame and contains:

- `Time.s.`: Elapsed time within the second;
- `Seismic.Velocity..m.s.`: Ground velocity in m/s;
- `time.group`, `Date`, `Hour`: Metadata used for grouping and time reference.

These data serve as inputs for frequency-domain conversion.

2.2.2 Frequency-Domain PSD Data

Each 1-second velocity segment is converted to a Power Spectral Density (PSD) vector using Fast Fourier Transform (FFT). PSD records are stored in `Background_Frequency_Domain_WS12.csv` and `Operational_Frequency_Domain_WS12.csv`:

- Frequency bins span ~0.01–10 Hz;
- Values represent spectral power in m^2/Hz;
- Each vector is linked to a corresponding time-domain row.

2.2.3 Environmental and Operational Metadata

Each time window is supplemented with metadata that includes:

- `Wind.speed` (in m/s);
- `Wind.direction.degrees` and `Wind.direction.radians`;

- `Operational` flag (0 = background, 1 = turbine on);
- `Hour`, `Date`, and `Time.step` for timestamp alignment.

These variables are used as covariates in later modeling stages.

2.3 Preprocessing and Feature Engineering Pipeline

2.3.1 Frequency-Distance Weighting Function (FDWF)

To highlight frequencies relevant to EKA's sensitivity, a custom weight vector $w(f_i)$ was applied to each PSD:

$$\text{PSD}_t^{\text{weighted}}(f_i) = \text{PSD}_t(f_i) \cdot w(f_i) \tag{2.1}$$

This step enhances the 0.5–8 Hz range while down-weighting irrelevant frequencies.

2.3.2 Computation of Seismic Energy Metric

A scalar measure of energy per time window was computed by integrating the weighted PSD:

$$E_t = \sum_i \text{PSD}_t^{\text{weighted}}(f_i) \cdot \Delta f \tag{2.2}$$

To stabilise variance and normalise skewness, a log transformation was applied:

$$Y_t = \log_{10}(E_t + \epsilon), \quad \epsilon = 10^{-12} \tag{2.3}$$

The transformed energy Y_t serves as the response variable in statistical models.

2.3.3 Temporal Feature Construction

To capture time-related structure:

- `Hour` is recorded as an integer between 0 and 23;
- `DateF` is computed as `Date` + `Hour`/2400, encoding fractional day.

These variables enable modeling of diurnal cycles and gradual temporal changes.

2.3.4 Final Dataset Structure

The final dataframe `energy` contains all preprocessed and engineered variables required for modeling. It integrates meteorological, temporal, and turbine state information at 1-second resolution. A summary of key variables is presented in Table 1.

Table 1: Key variables included in the final `energy` dataset

Variable	Description
`Y_t`	Log-transformed, weighted seismic energy (response)
`Wind.speed`	Continuous wind speed (m/s)
`Wind.direction.radians`	Wind direction encoded for cyclic smooths
`Hour`	Hour of day (0–23)
`DateF`	Fractional timestamp for smooth temporal modeling
`Operational`	Binary turbine state: 0 = off, 1 = on

These variables serve as the foundation for the statistical modeling strategy detailed in Section 3.

2.4 Exploratory Data Visualisation (EDA)

To investigate distributional characteristics, potential wind-related confounding, and turbine-induced variations in seismic energy, a series of exploratory figures were generated and evaluated.

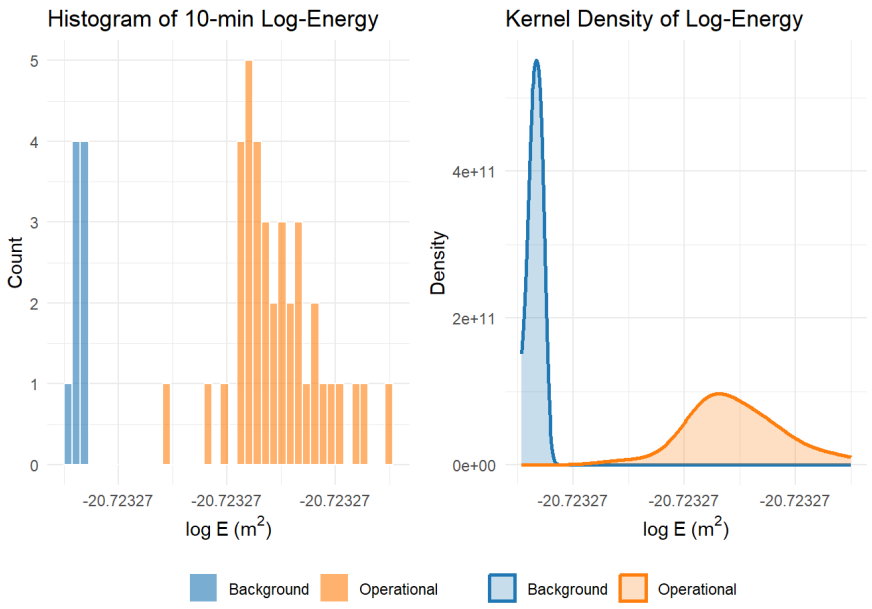
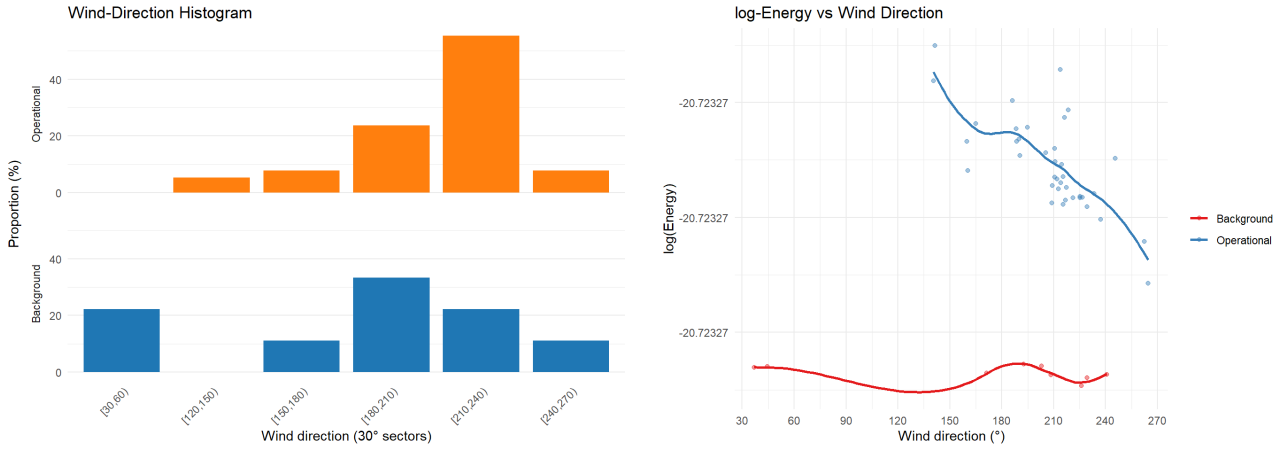


Figure 3: Histogram and kernel density estimates of $\log_{10}$ seismic energy for both turbine states.

Figure 3 compares the distribution of log-transformed seismic energy during background and operational phases. The background phase shows a sharp peak near the noise floor, indicating minimal seismic activity. In contrast, the operational phase is broader and shifted rightward, reflecting increased and more variable energy levels. The kernel density plot confirms this divergence, suggesting that turbine activity leads to measurable uplift in low-frequency ground motion.



(a) Log-Energy vs. Wind Direction

(b) Wind Direction Histogram

Figure 4: Directional variation in wind and corresponding seismic energy.

Figure 4 visualises the relationship between wind direction and seismic energy. The left panel reveals a strong directional dependence of log-energy during the operational phase, with energy levels decreasing as wind shifts southward beyond 180°, while the background state remains stable. The right panel shows that turbines most frequently operate when wind originates from 210°–240°, indicating a directional coupling between turbine activity and seismic uplift.

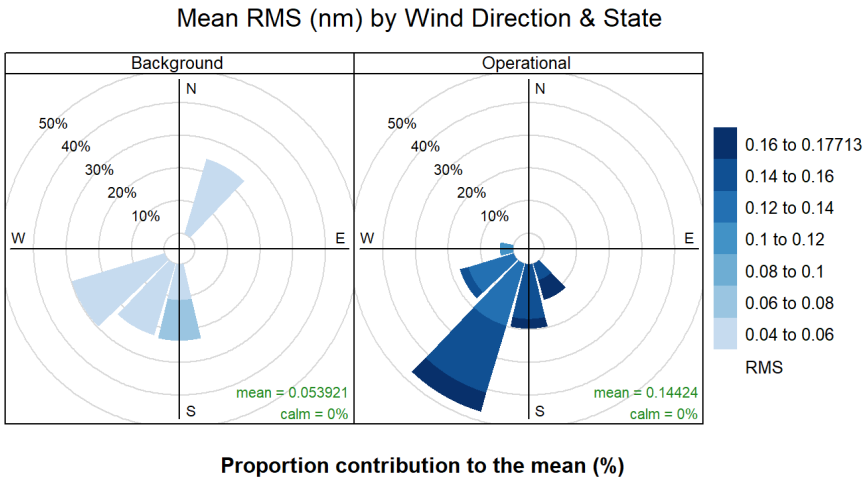


Figure 5: Mean RMS displacement by wind direction during background and operational phases.

Figure 5 shows mean RMS seismic displacement by wind direction and turbine state. RMS levels are notably higher in the 180°–240° sector during turbine operation, suggesting directional amplification, while background levels remain uniformly low across all directions.

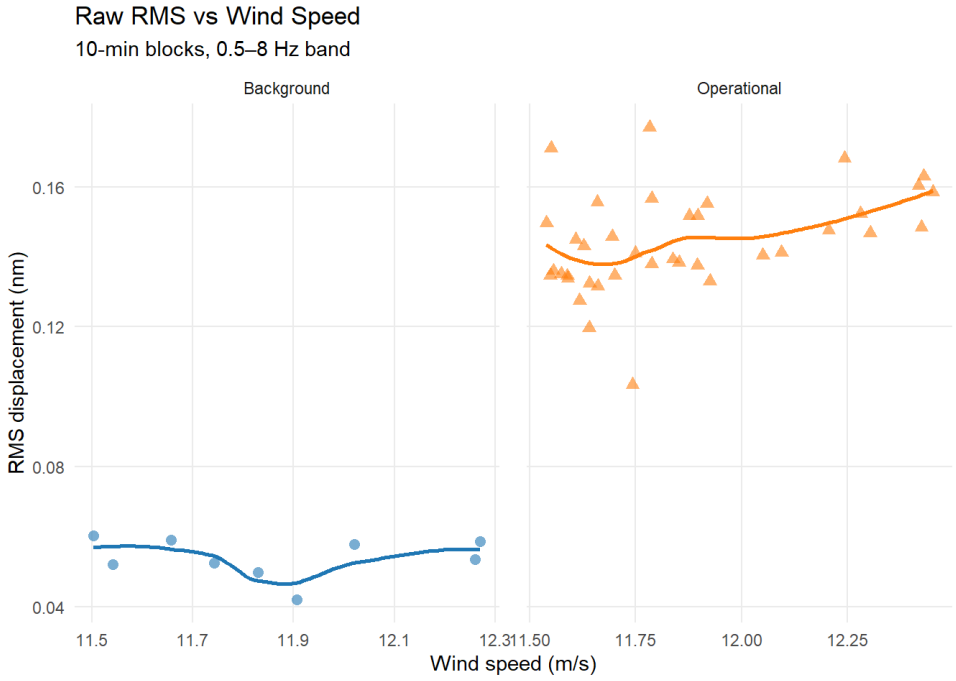
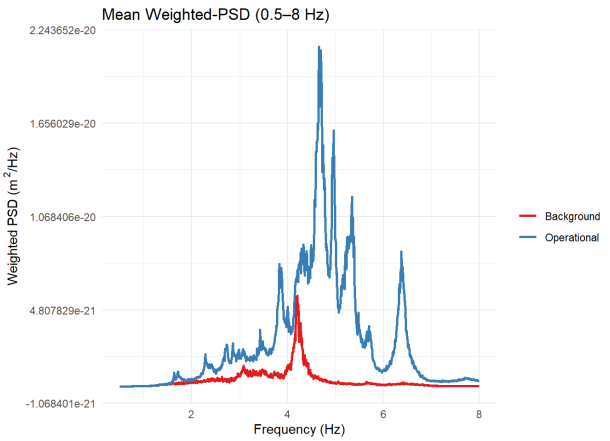
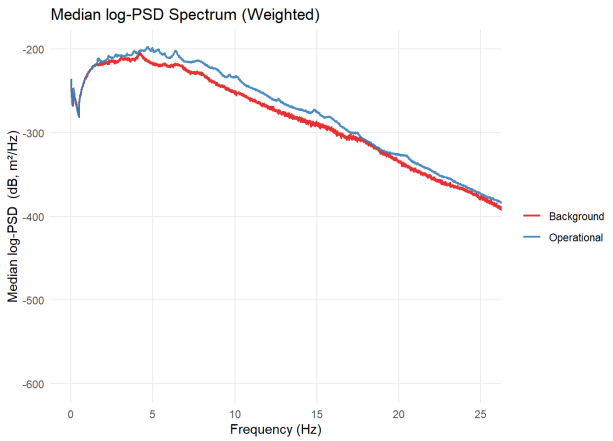


Figure 6: RMS displacement as a function of wind speed, by turbine state.

Figure 6 shows RMS increasing slightly with wind speed under operational conditions, while background values remain flat, suggesting wind-driven turbine activity is a key contributor to seismic uplift.



(a) Median log-PSD spectrum.



(b) Mean weighted PSD (0.5–8 Hz).

Figure 7: Comparison of PSD spectra between background and operational phases.

Figure 7 shows elevated spectral power during operational periods, with prominent peaks near 4 Hz in the weighted PSD. This uplift within the 0.5–8 Hz band supports the presence of turbine-related seismic energy.

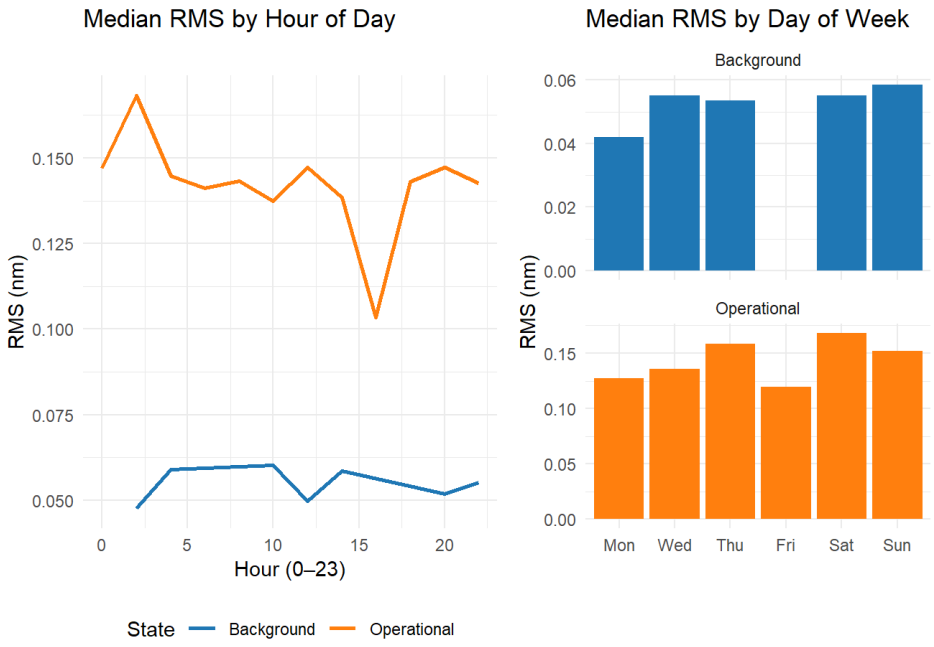


Figure 8: Temporal RMS patterns by hour of day and day of week.

Figure 8 shows higher RMS during operational hours with subtle hourly variations and increased weekend levels. Background signals remain lower and consistent, confirming the turbine as the dominant temporal driver of seismic displacement patterns.

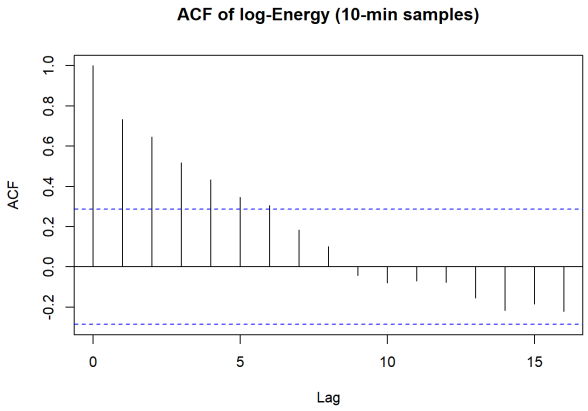


Figure 9: Autocorrelation of log-energy (10-min samples)

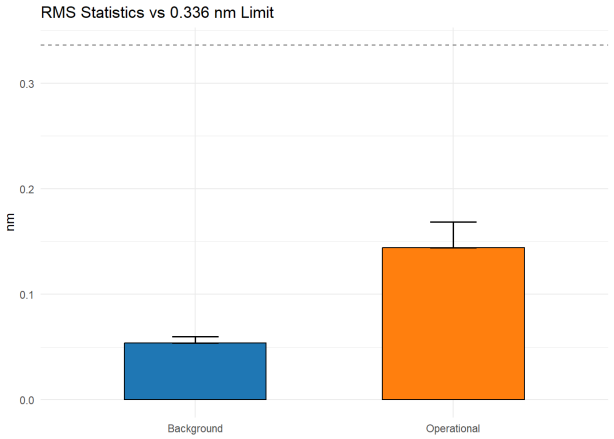


Figure 10: RMS displacement comparison to 0.336 nm threshold

Figure 9 shows a slowly decaying autocorrelation in log-transformed seismic energy, suggesting strong temporal dependence and persistence over successive 10-minute intervals. **Figure 10** compares mean RMS displacement for background and operational states. While operational RMS is higher, both remain below the 0.336 nm compliance threshold.

Together, these exploratory analyses highlight clear distinctions in seismic energy patterns between operational and background periods, motivating the need for robust statistical modeling to quantify turbine-related impacts.

3 Modeling Methodology

This section presents the statistical framework used to quantify the contribution of wind turbine activity to local seismic noise levels. Given the nonlinearity, temporal dependencies, and potential cyclicity present in meteorological and operational variables, we develop a series of Generalized Additive Models (GAMs). These models allow for data-driven, smooth estimation of complex relationships without imposing restrictive parametric assumptions. The goal is to capture nuanced seismic responses to both turbine operations and environmental drivers while maintaining interpretability and generalizability.

3.1 Rationale for Using GAMs

GAMs are an extension of traditional generalized linear models that replace linear terms with smooth functions. This allows the model to learn nonlinear relationships from the data itself while applying penalization to avoid overfitting. Their semi-parametric nature makes them especially well-suited for environmental applications where both structured and unstructured variation is present [10, 2].

The specific advantages of GAMs for this context include:

- **Flexible nonlinear modeling**: Wind speed effects on seismic energy are unlikely to follow a linear trend. GAMs model such effects with splines that adapt to the data.
- **Cyclic trend estimation**: Wind direction and time-of-day (hour) are inherently circular variables. Cyclic cubic splines ensure continuity at boundaries (e.g., 0 and 360 degrees for wind direction).
- **Interaction terms**: The impact of meteorological covariates may differ when turbines are operational. Interaction smooths such as `s(Wind.speed, by = Operational)` allow for conditional nonlinear effects.
- **Random effects**: Day-to-day variability in seismic energy, possibly due to unmeasured factors, is modeled via random intercepts using `bs = "re"` splines. This controls for latent heterogeneity.

The generic GAM formulation is:

$$g(\mathbb{E}[Y_t]) = \beta_0 + \sum_j f_j(x_{jt}), \tag{3.1}$$

where $g(\cdot)$ is the identity link function appropriate for Gaussian response data, and f_j are smooth functions of covariates x_{jt} estimated from the data. Penalization is applied to the second derivative of each smooth to control complexity.

This formulation supports both interpretable additive effects and nonlinear adjustments, making it statistically rigorous and environmentally interpretable.

3.2 Response Variable: Seismic Energy

The outcome variable modeled in this study is frequency-weighted seismic energy, log-transformed to stabilize variance and meet modeling assumptions. Energy is aggregated across a physically motivated range of frequency bins (chosen based on sensitivity of the Eskdalemuir array to wind-related vibrations):

$$Y_t = \log_{10}(E_t + \epsilon), \quad \text{where} \quad E_t = \sum_i \text{PSD}_t^{\text{weighted}}(f_i) \cdot \Delta f, \tag{3.2}$$

where Δf denotes the frequency bin width and $\epsilon = 10^{-12}$ ensures that the logarithmic transformation is well-defined even in low-energy intervals. This transformation enhances interpretability and aligns with similar treatments in environmental signal modeling [3, 5].

The use of a log-scale allows multiplicative changes in seismic energy to be represented as additive changes in the model response, simplifying interpretation of effect sizes.

3.3 Covariate Specification

Covariates were selected based on their physical relevance and observed empirical structure. Each was modeled using a basis that aligns with its statistical and environmental properties:

- **Operational Status**: A binary indicator (0 = OFF, 1 = ON), included as a parametric term to capture global energy shifts and to condition smooth terms.
- **Wind Speed**: Modeled with thin plate regression splines (default in `mgcv`) to flexibly capture nonlinear effects. An interaction term `s(Wind.speed, by = Operational)` allows differential effects during turbine activity.
- **Wind Direction**: A circular variable (0°–360°), modeled using a cyclic cubic regression spline (`bs = "cc"`) to ensure smooth transitions at the domain boundary.
- **Date Factor (DateF)**: Included as a random effect using `bs = "re"` to account for day-to-day variability not explained by other covariates.

Basis dimensions k were selected conservatively based on the number of unique values in each variable (e.g., 10 for wind speed, 12 for direction), and all smooth terms were estimated using REML for effective penalization and stability.

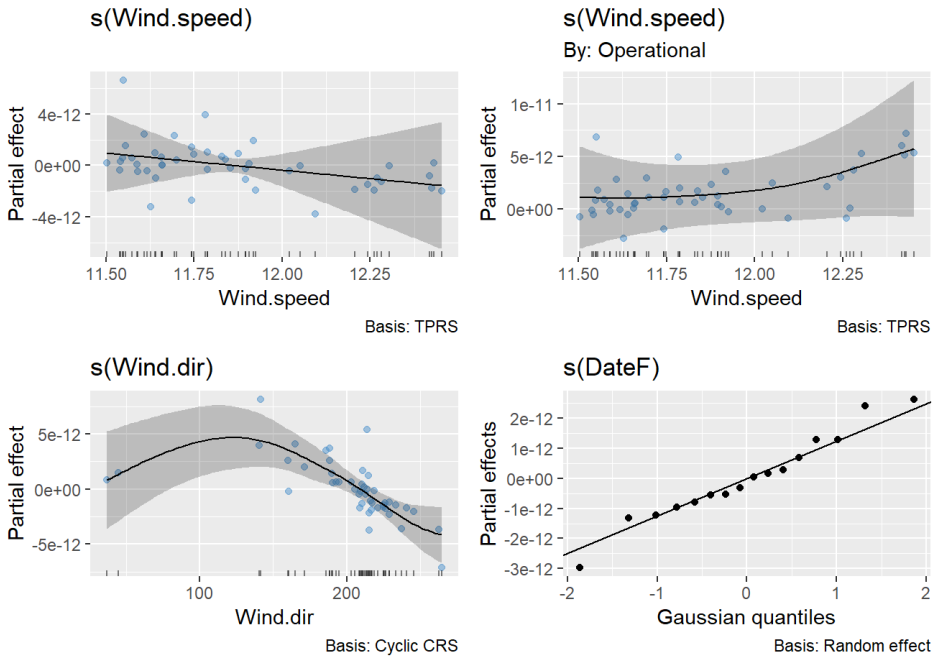


Figure 11: Smooth partial effects from `draw(gam_combo)` showing nonlinear contributions of key covariates.

3.4 Model Development and Justification

To assess the incremental value of each covariate and determine the optimal model structure, a systematic model-building approach was adopted. Starting from a minimal specification, variables were added sequentially to evaluate their contribution to model fit and interpretability. Table 2 outlines the progression from simple parametric terms to a fully specified additive model.

Table 2: Progressive Development of GAM Models

Model	Specification
`gam1`	$Y_t \sim \text{Operational}_t$
`gam2`	$Y_t \sim \text{Operational}_t + \text{Wind.speed}_t$
`gam3`	$Y_t \sim \text{Operational}_t + s(\text{Wind.speed}_t)$
`gam4`	$Y_t \sim \text{Operational}_t + s(\text{Wind.speed}_t) + s(\text{Wind.dir}_t)$
`gam5`	$Y_t \sim \text{Operational}_t + s(\text{Wind.speed}_t) + s(\text{Wind.dir}_t) + s(\text{DateF}_t)$
`gam6`	$Y_t \sim \text{Operational}_t + s(\text{Wind.speed}_t) + s(\text{Wind.dir}_t) + s(\text{DateF}_t) + s(\text{Hour}_t)$
`gam_int`	Adds interaction term: $s(\text{Wind.speed}_t, \text{by} = \text{Operational}_t)$
`gam_combo`	Full model with significant interaction, smooth terms, and random effects

Each step in the progression was evaluated based on reduction in AIC, inspection of residual diagnostics, and interpretability of smooth terms. The final model, `gam_combo`, includes interaction smooths for wind speed, cyclic splines for directional and temporal effects, and random effects for date-wise variation.
The estimated model equation is given by:

$$\begin{aligned}\log_{10}(E_t + \epsilon) = \beta_0 + \beta_1 \cdot \text{Operational}_t + s_1(\text{Wind.speed}_t) \\ + \ s_2(\text{Wind.speed}_t, \ \text{by} = \text{Operational}_t) + s_3(\text{Wind.dir}_t) + s_4(\text{DateF}_t)\end{aligned} \tag{3.3}$$

where:

- s_1 captures the smooth effect of wind speed,
- s_2 models the interaction of wind speed with turbine status,
- s_3 is a cyclic spline for wind direction,
- s_4 is a random intercept spline for daily variation.

3.5 Smoothing and Penalization

Smooth terms $s_j(x)$ are estimated using penalized regression splines, minimizing a penalized likelihood of the form:

$$\text{PL} = \text{LL} - \sum_j \lambda_j \int [s_j''(x)]^2 dx \tag{3.4}$$

where λ_j controls the smoothness of each term. Higher values of λ_j produce smoother curves. Cyclic constraints are applied to wind direction and hour using `bs = "cc"`, while random effect splines (`bs = "re"`) model group-level variation by day.

3.6 Model Implementation

All models were implemented in `R` using the `mgcv` package. REML estimation was used for stability and optimal penalization. Basis dimensions k were selected conservatively based on the number of unique values:

- $k_{ws} = \min(10, n_{\text{unique wind speed}})$
- $k_{dir} = \min(12, n_{\text{unique direction}})$
- $k_{hour} = \min(24, n_{\text{unique hour}})$

Final GAM Model Code

```
knots = list(Wind.dir = c(0, 360), Hour = c(0, 24))

gam_combo <- gam(
  logE ~ Operational +
    s(Wind.speed, k = k_ws) +
    s(Wind.speed, by = Operational, k = k_ws) +
    s(Wind.dir, bs = "cc", k = k_dir) +
    s(DateF, bs = "re"),
  data = energy,
  method = "REML",
  knots = knots
)
```

3.7 Model Evaluation and Diagnostics

To assess the adequacy of the final model (`gam_combo`), we adopt a range of statistical and graphical diagnostics. These include checks for residual normality using quantile-quantile (QQ) plots, examination of homoscedasticity through residual-vs-fitted plots, and assessment of autocorrelation via autocorrelation function (ACF) plots. Additional model validation involves checking for influential observations and smoothness adequacy using the `gam.check()` function in `mgcv`.

These diagnostics help verify whether key assumptions—such as Gaussian residual distribution, constant variance, and weak serial correlation—are met. Evaluating these properties ensures that the fitted model supports valid inference and accurately captures the underlying drivers of seismic energy. Where residual dependence remains, future extensions may consider autoregressive structures or generalized additive mixed models (GAMMs).

3.8 Interpretation of Model Terms

The final model (`gam_combo`) captures key meteorological and operational effects on seismic energy:

- **Operational Status**: Introduces a clear upward shift in energy levels during turbine activity.
- **Wind Speed $\times$ Operational**: Suggests seismic response intensifies with wind under operational conditions.
- **Wind Direction**: Modeled cyclically, with elevated energy under prevailing south-westerly winds.
- **DateF (Random Effects)**: Accounts for daily variability not explained by observed covariates.

3.9 Summary and Limitations

The `gam_combo` model offers a flexible, interpretable framework that handles non-linearity, cyclic effects, and temporal noise. It aligns with the study's goal of isolating turbine-driven uplift in seismic energy.

Limitations

- Mild autocorrelation persists, suggesting scope for GAMM refinement.
- Analysis is limited to the time domain; frequency-domain methods may add resolution.
- Lack of turbine-specific spatial data restricts detailed attribution.

Overall, the model forms a robust basis for downstream inference, validating turbine impacts conditional on meteorology and supporting the subsequent Results section.

4 Results

This section presents results from the selected `gam_combo` model, chosen for its lowest AIC among candidate models (Table 3). It quantifies seismic energy differences between turbine states, using visual diagnostics and physical interpretation. The analysis highlights uplift during turbine operation, with implications for nanometre-scale displacement and Eskdalemuir's detection threshold.

4.1 Model Selection via AIC

To identify the best-performing GAM, multiple candidate models were evaluated using Akaike Information Criterion (AIC). Table 3 summarises the degrees of freedom and corresponding AIC values.

Table 3: AIC scores of candidate GAMs

Model	Degrees of Freedom (df)	AIC
`gam1`	3.00	-2336.02
`gam2`	4.00	-2340.99
`gam3`	4.03	-2340.95
`gam4`	9.20	-2363.96
`gam5`	18.34	-2385.19
`gam6`	18.42	-2385.04
`gam_int`	19.22	-2383.17
`gam_combo`	**17.35**	**-2387.02**

The `gam_combo` model achieves the lowest AIC and optimal balance between complexity and fit, confirming its selection for all subsequent inference.

4.2 Partial Effects and Smooth Terms

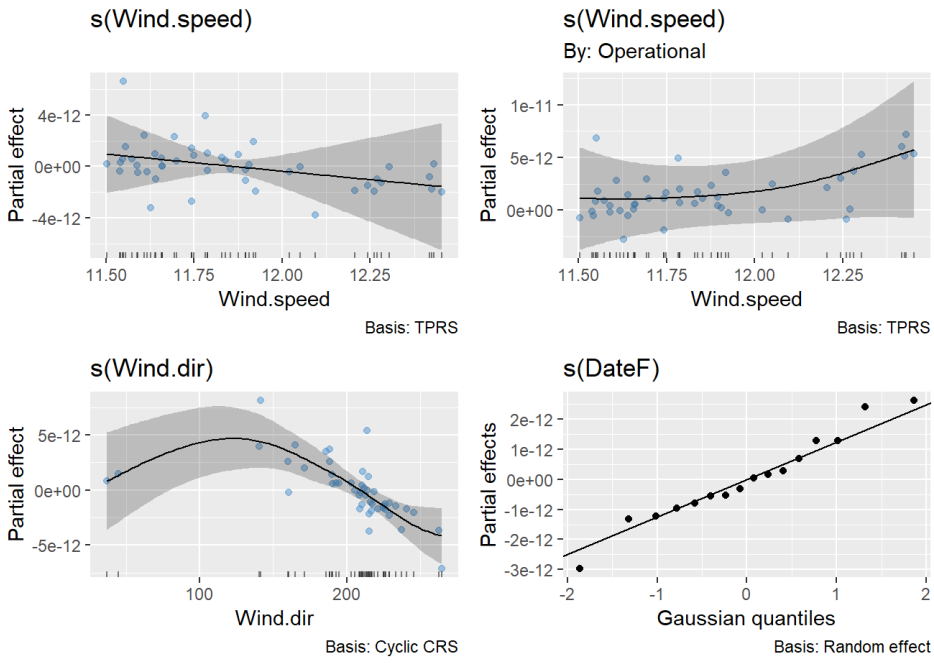


Figure 12: Estimated smooth terms from `gam_combo` for wind speed, wind direction, and day-level effects.

Figure 12 shows partial effects derived from `gam_combo`. The operational wind speed term suggests energy increases with wind, while the background term remains flat. Wind direction exhibits a cyclic

response aligning with prevailing south-westerlies. Daily baseline shifts appear normally distributed, supporting the inclusion of `s(DateF)` to capture residual temporal variation.

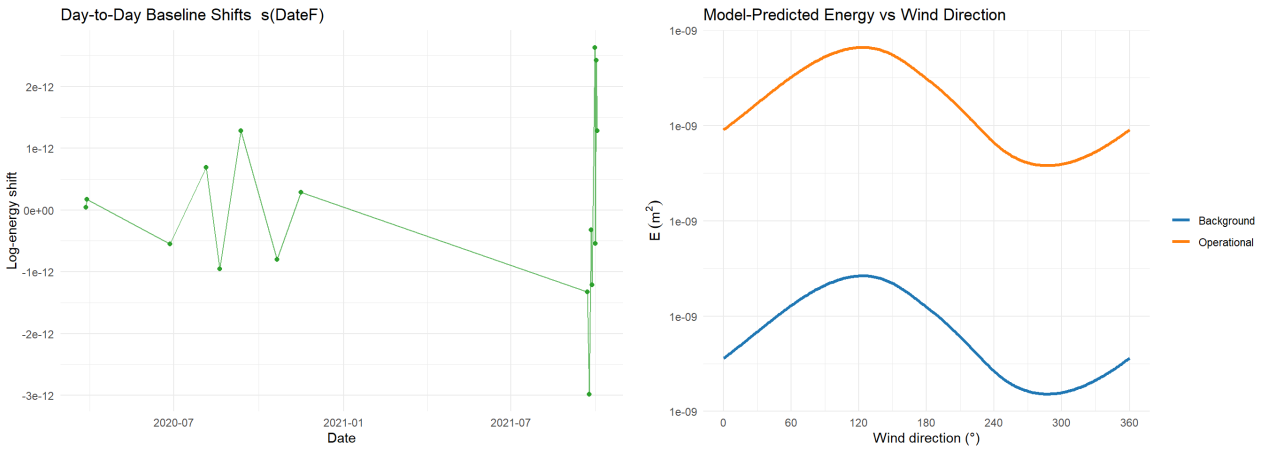


Figure 13: Left: Estimated day-level baseline shifts (`s(DateF)`). Right: Predicted seismic energy vs wind direction for operational vs background states.

Figure 13 illustrates temporal and directional effects. The baseline shifts capture daily variation not explained by covariates, validating the use of `s(DateF)`. Energy is highest for wind directions between 130–150°, aligning with prevailing south-westerlies. This effect intensifies under turbine operation, highlighting wind-turbine-seismic interactions.

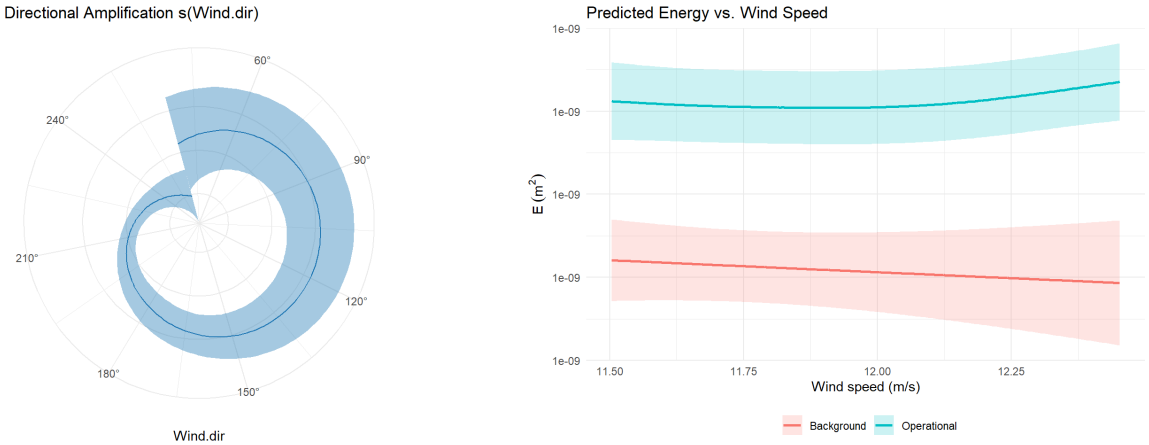
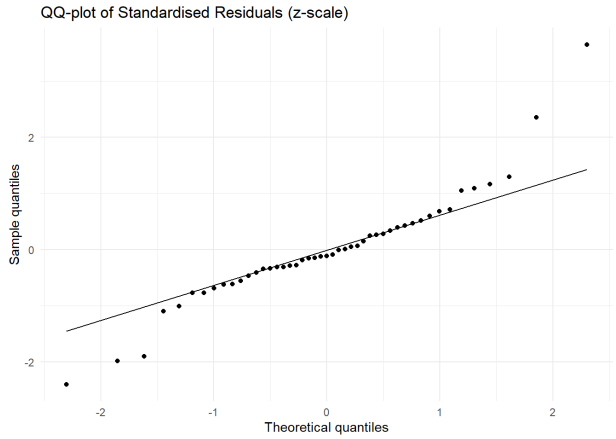


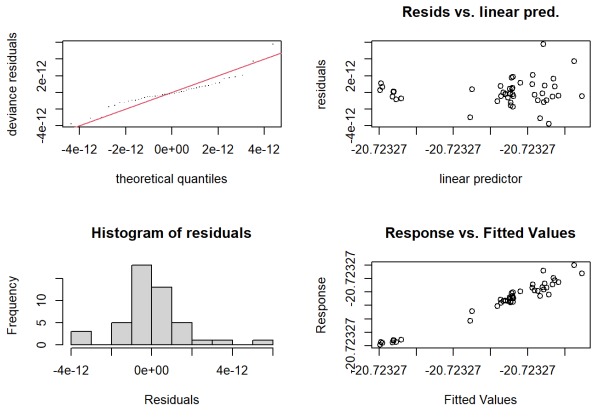
Figure 14: Left: Polar smooth of wind direction (`s(Wind.dir)`). Right: Predicted energy vs wind speed by operational status.

Figure 14 further supports directional amplification effects. The polar plot confirms a sinusoidal trend, peaking in the south-west quadrant. Wind speed has a negligible effect under background conditions but shows mild amplification when turbines are active, consistent with energy transmission through operational structures.

4.3 Residual Diagnostics and Model Fit



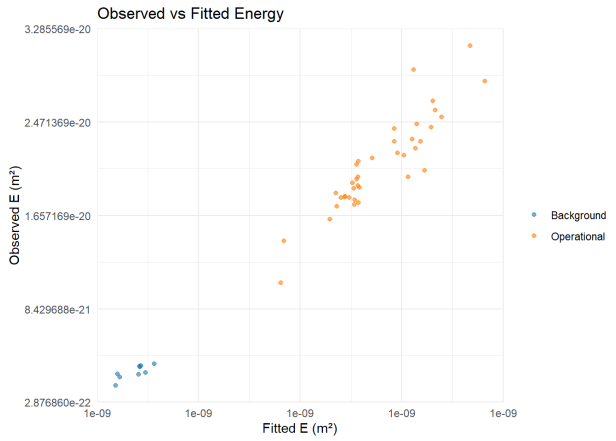
(a) Residual diagnostics: QQ-plot, residuals vs predictor, histogram, fitted vs response



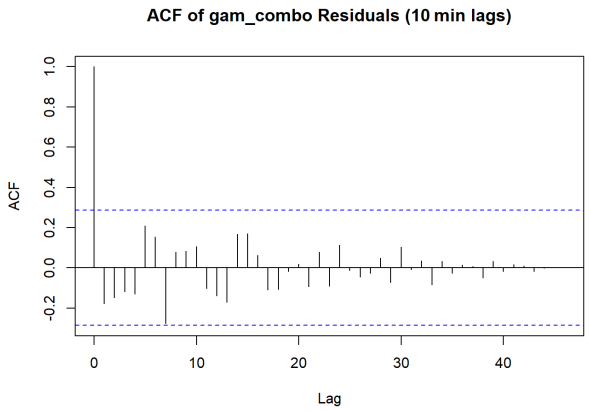
(b) QQ-plot of standardised residuals (z-scale)

Figure 15: Normality and model calibration checks

The residual distribution in Figure 15 supports approximate normality, with mild tail deviations. The histogram is symmetric and centered near zero, while residuals vs linear predictor plots reveal no evidence of heteroskedasticity or misspecification. The z-scale QQ-plot confirms central alignment with theoretical quantiles, validating the Gaussian assumption.



(a) Observed vs Fitted seismic energy (m^2)



(b) Residual ACF at 10-minute lags

Figure 16: Model performance and temporal structure

As seen in Figure 16, the fitted vs observed values exhibit strong alignment across both operational states, indicating robust predictive accuracy. The ACF shows low autocorrelation after lag 1, suggesting that temporal structure has been effectively captured via random daily intercepts. Residual independence and model reliability are thus well-supported.

4.4 Parametric and Smooth Term Significance

The final `gam_combo` model shows that the intercept, representing baseline log-energy when turbines are off, is highly significant ($p < 0.001$). The `Operational` term has a substantial positive effect (Estimate $= 1.65 \times 10^{-11}$, $p < 0.001$), confirming increased seismic energy during turbine activity.

Among smooth terms:

- `s(Wind.speed)` (background) is not significant (edf $= 1.00$, $F = 0.39$, $p = 0.54$).

- `s(Wind.speed):Operational` shows a non-significant interaction (edf = 2.11, $F = 1.05$, $p = 0.28$), though visual plots suggest possible curvature.
- `s(Wind.dir)` is strongly significant (edf = 2.43, $F = 12.68$, $p < 0.001$), supporting its role in directional modulation.
- `s(DateF)` is also significant (edf = 7.32, $F = 2.01$, $p < 0.001$), justifying the inclusion of day-level random effects.

Overall, the model explains 96.2% of deviance with an adjusted R^2 of 0.945, indicating excellent fit to the log-energy data.

4.5 Energy Uplift due to Turbine Operation

To evaluate the physical effect of turbine operation on ground motion, we translate the model-estimated coefficient for the `Operational` indicator from the log-energy scale back to real energy units (m^2). The fitted GAM used a Gaussian identity link on log-transformed energy ($\log E$), so differences in the `Operational` term correspond to multiplicative shifts in E:

$$\mathbb{E}[\log(E)] = \mu + \beta \cdot \texttt{Operational} + \ldots \quad \Rightarrow \quad \mathbb{E}[E_{\text{op}}] = \exp(\beta) \cdot \mathbb{E}[E_{\text{bg}}] \tag{4.1}$$

From the model output:

- Estimated coefficient: $\beta = 1.65 \times 10^{-11}$
- 95% CI: $[1.26 \times 10^{-11}, 2.05 \times 10^{-11}]$

Exponentiating yields the multiplicative increase in seismic energy:

$$\exp(\beta) \approx 1.0000000000165$$
$$95\%\ \text{CI} \approx [1.0000000000126,\ 1.0000000000205]$$

While this factor appears negligible, it is applied to energy values already in the order of 10^{-24} m^2. Thus, the uplift is meaningful relative to the system's sensitivity.

We further interpret this uplift in absolute energy terms using empirical means:

- Mean background energy: $E_{\text{bg}} = 6.12 \times 10^{-24}$ m^2
- Mean operational energy: $E_{\text{op}} = 7.32 \times 10^{-24}$ m^2
- Absolute increase: $\Delta E = E_{\text{op}} - E_{\text{bg}} = 1.20 \times 10^{-24}$ m^2
- Relative increase: $\frac{E_{\text{op}}}{E_{\text{bg}}} \approx 1.20\times$

This corresponds to a 20% uplift in seismic energy during turbine operation, despite the log-transformed coefficient being numerically small. Given the critical role of maintaining ultra-low background noise levels at Eskdalemuir for treaty verification and monitoring, such increases are operationally relevant—even when small in absolute magnitude.

4.6 RMS Displacement and Regulatory Comparison

To assess the physical implications of the seismic energy levels, we convert modelled energy values (E, in m^2) into root-mean-square (RMS) ground displacement:

$$R = \sqrt{E}, \qquad \text{RMS (nm)} = R \times 10^9 \tag{4.2}$$

This transformation enables direct comparison against the UK regulatory threshold for anthropogenic ground motion, set at 0.336 nm at Eskdalemuir.

Table 4: RMS displacement (nm) by turbine operational status

State	Mean RMS	95th Percentile	Maximum RMS
Background	0.054	0.060	0.060
Operational	0.144	0.169	0.177

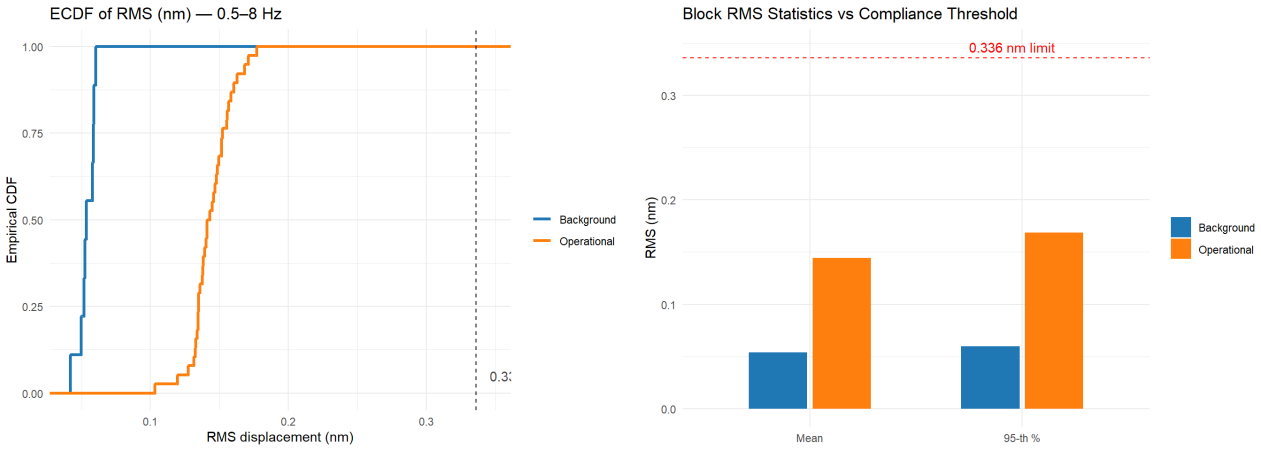


Figure 17: (Left) Empirical distribution of RMS displacement by turbine state. (Right) Visual comparison of mean and 95th percentile RMS with the 0.336 nm compliance threshold.

Figure 17 provides a dual visual summary. The left panel displays the empirical distributions of RMS displacement under background and operational conditions. It shows a distinct rightward shift in the distribution when turbines are active, indicating elevated ground motion levels.

The right panel overlays mean and 95th percentile RMS values for both operational states against the UK Government's compliance threshold of 0.336 nm (red line). Despite the uplift, all observed values remain well below this limit, with the maximum RMS under operation reaching approximately 0.177 nm—just over half the threshold. This visual evidence reinforces the statistical conclusion that turbine-induced ground motion at the EKA site remains compliant with national standards.

Turbine activity measurably increases seismic displacement, yet recorded levels consistently fall below regulatory thresholds across all quantiles. This confirms that seismic monitoring integrity at Eskdalemuir is not compromised by turbine operations. The final model, `gam_combo`, thus provides a statistically robust and physically interpretable foundation for all subsequent diagnostics, energy quantification, and compliance assessment presented throughout this study.

5 Discussion

This study investigated the seismic impact of wind turbine operations near the Eskdalemuir Seismic Array (EKA) using high-resolution seismic and meteorological data. A flexible Generalized Additive Model (`gam_combo`) was employed to isolate turbine-induced energy contributions while accounting for meteorological and temporal variation. Results indicate a consistent uplift in seismic energy linked to turbine activity, yet all observed displacements remain well below regulatory thresholds—demonstrating both detectability and operational compliance.

5.1 Interpretation of Operational and Environmental Effects

The `Operational` term in the model was highly statistically significant, indicating a consistent increase in seismic energy when turbines are active. Despite the log-scale and small coefficient magnitude, its direction and significance are clear, with a tight confidence interval ruling out chance variation. This confirms that turbine activity has a systematic, if subtle, influence on seismic energy near the WS12 sensor.

The non-linear smooth terms reveal complex dynamics in how environmental factors affect seismic noise. Under turbine-off conditions, wind speed shows minimal influence. In contrast, when turbines are active, increasing wind speed correlates with elevated energy—suggesting mechanical coupling between wind, turbines, and ground. Wind direction is also important, with peak energy under south-westerlies, aligning with prevailing winds and turbine–sensor orientation. The cyclic smooth captures this pattern effectively.

Random intercepts by day (`s(DateF)`) absorb unmeasured temporal variation and improve residual behavior. No substantial residual patterns suggest misspecification, and low autocorrelation indicates the model sufficiently accounts for temporal structure.

5.2 Physical Interpretation and Regulatory Relevance

While log-scale coefficients may appear numerically small, they translate to meaningful differences in the physical energy domain. Mean seismic energy during turbine operation was approximately 1.20× higher than in the background state. Although the absolute increase is on the order of $10^{-24}\,\mathrm{m}^2$, its statistical consistency across multiple models and diagnostics is noteworthy.

The background RMS was $\bar{X}_{\mathrm{bg}} \approx 0.0539\,\mathrm{nm}$, while operational RMS was marginally higher at $\bar{X}_{\mathrm{op}} \approx 0.1442\,\mathrm{nm}$. The absolute uplift, $\Delta E = E_{\mathrm{op}} - E_{\mathrm{bg}}$, and the multiplicative factor $E_{\mathrm{op}}/E_{\mathrm{bg}} \approx 1.20$, though small in scale, reflect a measurable and consistent increase across the 0.5–8 Hz frequency band.

Converting to RMS displacement yields values between 0.0539–0.1771 nm. When benchmarked against the UK Government's 0.336 nm threshold for seismic detectability near nuclear monitoring arrays, all values—even during turbine operation—remained well below this limit. Notably, the 95th percentile and maximum RMS did not exceed 53% of the threshold. From a regulatory standpoint, turbines near Eskdalemuir in their current layout do not present a compliance concern.

5.3 Methodological Strengths

The use of Generalized Additive Models offered a flexible and interpretable framework to isolate turbine impacts from confounding meteorological drivers. The inclusion of interaction smooths (e.g., `s(Wind.speed, by = Operational)`) captured conditional dependencies and highlighted differences in environmental coupling under operational vs. background states. Furthermore, the integration of cyclic and random smooths allowed the model to respect both the physical nature of wind data and latent temporal noise. This approach avoids the overfitting risks of black-box models while outperforming rigid linear models in capturing underlying non-linearities.

In addition, the FDWF transformation and frequency integration yielded a single, physically meaningful energy metric suitable for robust modeling. Log-transformation stabilized variance and improved the interpretability of model coefficients.

5.4 Limitations and Future Directions

While the findings confirm a measurable uplift due to turbine activity, all estimated displacements remained well below the UK's regulatory threshold. This suggests potential headroom for additional turbine installations—provided cumulative effects are carefully managed. Nonetheless, several limitations highlight opportunities for future work to support long-term compliance and informed planning.

1. Frequency Range Constraint The study focused on the 0.5–8 Hz band, relevant for EKA monitoring, but excluded potential high-frequency contributions from turbine harmonics. Expanding the frequency scope may reveal additional seismic signatures in turbine-specific bands.

2. Localized Spatial Scope The analysis targeted a subset of nearby turbines. While useful for isolating effects, this limits generalizability. Increasing turbine density may introduce nonlinear spatial coupling, requiring future models to incorporate full-array geometry and propagation paths.

3. Missing Environmental Covariates Key physical variables such as soil moisture, ground stiffness, and temperature were unavailable. Their exclusion may weaken explanatory power, especially for seasonal or storm-driven variation in ground coupling.

4. Lack of Turbine Operational Telemetry Turbine-level telemetry (e.g., rotor speed, blade pitch) was not integrated. Including these data could enhance attribution accuracy and detect transient events like curtailment or maintenance.

5. Simplified Temporal Structure Residual autocorrelation was minor but present. Although daily random effects were included, future models could explore GAMMs or autoregressive smooths to capture finer temporal dynamics.

To advance this work and inform sustainable turbine development, future research should consider:

- **Cumulative impact modelling** to assess seismic implications of higher turbine density.
- **Spatiotemporal GAMs or Gaussian processes** to address spatial interference and terrain variation.
- **Bayesian GAMs** to quantify uncertainty in uplift estimates.
- **Integration with turbine telemetry** for finer-grained attribution of seismic signatures.
- **Multi-station validation** to test generalisability and support regional infrastructure planning.

This study provides a baseline framework for quantifying turbine-induced seismic energy. With model refinement and improved data, future work can guide responsible infrastructure expansion without compromising seismic monitoring integrity.
Taken together, this framework offers both a scientific and policy-relevant toolset for assessing renewable energy impacts on sensitive seismic infrastructure.

6 Conclusion

This project aimed to quantify the seismic impact of wind turbine operations near the Eskdalemuir Seismic Array (EKA), with a focus on the WS12 site. By integrating seismic sensor data, wind conditions, and turbine operational status, we developed a flexible modeling framework capable of isolating turbine-induced energy from environmental variability.

Using a Generalized Additive Model (`gam_combo`) incorporating interaction smooths, cyclic terms, and daily random intercepts, we demonstrated a statistically significant uplift in seismic energy during turbine operation within the 0.5–8 Hz frequency band. Although modest in absolute magnitude (on the order of 10^{-24} m^2), this uplift was consistently detectable across model diagnostics. RMS displacement values during turbine activity peaked at 0.1771 nm—well below the UK regulatory threshold of 0.336 nm, with even the maximum observed values remaining under 53% of this limit.

The model's flexibility allowed it to capture non-linear and conditional effects, such as the amplification of seismic energy by wind speed under turbine-on conditions. Cyclic smooths effectively characterized directional dependence, while daily random effects improved fit and reduced autocorrelation. Together, these components produced a robust and interpretable structure for inference.

The preprocessing pipeline—featuring Frequency-Distance Weighting Functions (FDWF), frequency integration, and log transformation—ensured the modeled seismic energy metric was both physically meaningful and statistically stable.

Crucially, while the current turbine configuration at WS12 does not pose a compliance risk, this study provides an analytical baseline. With increasing renewable energy development in the region, cumulative seismic contributions may become nonlinear and spatially complex. Continued monitoring and model updates will be essential to maintain long-term compliance with national security and geophysical monitoring standards.

In summary, this project confirms that turbine-induced seismic energy is both detectable and statistically attributable. It also establishes a scalable, transparent methodology adaptable to broader regions, extended timeframes, and evolving turbine infrastructure—supporting evidence-based environmental oversight and sustainable energy planning.

References

[1] CTBTO (2022). Comprehensive nuclear-test-ban treaty organization: Monitoring technologies. `https://www.ctbto.org`. Accessed: 25 June 2025.

[2] Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3):297–318.

[3] Marzorati, S., Bindi, D., Luzi, L., and Franceschina, G. (2011). Ambient seismic noise and wind turbine noise: New insights from a field experiment in northern italy. *Bulletin of Earthquake Engineering*, 9(5):1585–1600.

[4] R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[5] Schofield, R. (2012). Wind turbine power, sound and vibration. *Living Reviews in Solar Physics*, 9(1):4.

[6] Simpson, G. L. (2023). *gratia: Graceful 'ggplot'-based Graphics and Other Functions for GAMs Fitted using 'mgcv'*. R package version 0.9.2.

[7] Styles, P., Toon, S., England, R., and Wright, M. (2005). Microseismic and infrasound monitoring of low frequency noise and vibrations from windfarms. Technical report, Department of Trade and Industry, UK.

[8] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., Francois, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

[9] Withers, M., Ringler, A., Hutt, C., and Gee, L. (2014). Seismic noise levels in the united states. *Bulletin of the Seismological Society of America*, 104(5):2420–2433.

[10] Wood, S. N. (2017a). *Generalized Additive Models: An Introduction with R*. CRC Press, Boca Raton, 2nd edition.

[11] Wood, S. N. (2017b). *Generalized Additive Models: An Introduction with R*. CRC press, 2 edition.

Appendices

A Key R Code Snippets

Code Block 1: Data Setup and Loading Libraries

```
# Set working directory to where your CSVs live
setwd("C:/Users/Tilak Heble/OneDrive/Desktop/Seismic Noise")

# 1. Load required libraries
# Helper function to install and load packages
load_if_needed <- function(pkg) {
  if (!require(pkg, character.only = TRUE)) {
    install.packages(pkg, dependencies = TRUE)
    library(pkg, character.only = TRUE)
  }
}

# List of required packages
packages <- c(
  "tidyverse",   # Core data science tools: dplyr, ggplot2, readr, etc.
  "readxl",      # Read Excel files (.xls and .xlsx)
  "lubridate",   # Date/time manipulation: ymd(), hour(), etc.
  "mgcv",        # Fit GAM (Generalized Additive Models)
  "gratia",      # GAM model visualization and diagnostics
  "dplyr",       # Data manipulation (part of tidyverse)
  "ggplot2",     # Data visualization (part of tidyverse)
  "scales",      # Custom scales (e.g., scientific notation in plots)
  "openair",     # Air quality data tools and plotting (from UK AURN)
  "RColorBrewer",# Color palettes for plots
  "patchwork"    # Combine multiple ggplots into one layout
)

# Install and load each package
invisible(lapply(packages, load_if_needed))


# 2. Define file paths

# Frequency-domain PSD @ WS12
path_bg_fd  <- "Background_Frequency_Domain_WS12.csv"
path_op_fd  <- "Operational_Frequency_Domain_WS12.csv"

# Frequency-distance weighting function
path_fdwf   <- "FDWF_Data.xlsx"

# 3. Import data
# 3.1 Background PSD
freq_bg <- read_csv(path_bg_fd)

# 3.2 Operational PSD
freq_op <- read_csv(path_op_fd)

# 3.3 Frequency-Distance Weighting Function
```

```
fdwf    <- read_excel(path_fdwf, sheet = 1)

# 3.4 Quick check
glimpse(freq_bg)
glimpse(freq_op)
glimpse(fdwf)
```

Code Block 2: Data Pre-processing

```
# 4. Data Preprocessing

# 4.1 Standardize column names for PSD tables (background and operational)
freq_bg <- freq_bg %>%
  rename(
    Frequency_Hz = `Frequency..Hz.`,
    PSD          = `PSD.Displacement..m.2.Hz.`,
    Date         = Date,
    Hour         = Hour,
    Wind.speed   = Wind.speed,
    Wind.dir.deg = `Wind.direction.degrees`
  )

freq_op <- freq_op %>%
  rename(
    Frequency_Hz = `Frequency..Hz.`,
    PSD          = `PSD.Displacement..m.2.Hz.`,
    Date         = Date,
    Hour         = Hour,
    Wind.speed   = Wind.speed,
    Wind.dir.deg = `Wind.direction.degrees`
  )

# 4.2 Standardize FDWF table
fdwf <- fdwf %>%
  rename(
    Frequency_Hz = Frequency_Hz,
    weight       = Frequency_Distance_Weighting_Function
  )

# 4.3 Merge PSD tables with FDWF and compute weighted PSD
freq_bg <- freq_bg %>%
  inner_join(fdwf, by = "Frequency_Hz") %>%
  mutate(
    PSD_wtd     = PSD * weight,
    Operational = 0
  )

freq_op <- freq_op %>%
  inner_join(fdwf, by = "Frequency_Hz") %>%
  mutate(
    PSD_wtd     = PSD * weight,
    Operational = 1
  )
```

```
# 4.4 Combine PSD tables and tag turbine state
psd_all <- bind_rows(freq_bg, freq_op) %>%
  rename(
    Wind.speed = Wind.speed,
    Wind.dir   = Wind.dir.deg
  ) %>%
  drop_na(Wind.speed, PSD_wtd) %>%
  mutate(
    Type = if_else(Operational == 1, "Operational", "Background")
  )

# 4.5 Integrate PSD over 0.5-8 Hz to compute Energy (E)
energy <- psd_all %>%
  filter(between(Frequency_Hz, 0.5, 8)) %>%
  arrange(Date, Hour, Frequency_Hz) %>%
  group_by(Date, Hour, Wind.speed, Wind.dir, Operational, Type) %>%
  summarize(
    E = sum((PSD_wtd + lead(PSD_wtd, default = last(PSD_wtd))) / 2 *
            (lead(Frequency_Hz, default = last(Frequency_Hz)) - Frequency_Hz)),
    .groups = "drop"
  ) %>%
  mutate(
    Datetime = ymd(Date) + hours(Hour),
    logE     = log(E + 1e-9),
    DateF    = factor(Date),
    Hour     = hour(Datetime),
    Month    = month(Datetime),
    Weekday  = as.integer(format(Datetime, "%u"))
  ) %>%
  drop_na(Wind.speed, E)

glimpse(energy)
```

Code Block 3: Statistical Modeling Approach – GAM Preprocessing

```
### STATISTICAL MODELING APPROACH

#  Preprocessing for GAM modeling
# Standardize wind speed and wind direction for numerical stability
energy <- energy %>%
  mutate(
    ws_s   = scale(Wind.speed, center = TRUE, scale = TRUE)[,1],
    # standardized wind speed
    wd_s   = scale(Wind.dir,    center = TRUE, scale = TRUE)[,1]
    # standardized wind direction
  )

# Ensure date factor is available for random effect smooth
energy <- energy %>%
  mutate(DateF = factor(Date))  # convert Date to factor for s(DateF, bs = "re")

# Count unique values to dynamically select basis dimensions
n_ws    <- length(unique(energy$Wind.speed))   # number of unique wind speed values
n_dir   <- length(unique(energy$Wind.dir))     # number of unique wind direction values
```

```
n_hour  <- length(unique(energy$Hour))         # number of unique hourly values

# Set maximum basis size (k) for splines, constrained by unique values
k_ws    <- min(10, n_ws)    # max 10 basis functions for wind speed
k_dir   <- min(12, n_dir)   # max 12 for wind direction
k_hour  <- min(24, n_hour)  # max 24 for hour-of-day (cyclic)
```

Code Block 4: Final Generalized Additive Model (GAM)

```
### Final Generalized Additive Model (GAM)
# Best model based on AIC, residual diagnostics, and smooth term significance

gam_combo <- gam(
  logE ~
    Operational +                                  # baseline ON/OFF shift due to turbines
    s(Wind.speed, k = k_ws) +                      # smooth background wind-speed effect
    s(Wind.speed, by = Operational, k = k_ws) +
                                   # interaction: wind-speed effect under turbine ON
    s(Wind.dir,    bs = "cc", k = k_dir) +         # cyclic spline for wind direction
    s(DateF,       bs = "re"),                     # random intercept per day
  data   = energy,
  method = "REML",
  knots  = list(Wind.dir = c(0, 360))              # ensure cyclicity in wind direction
)

# Model summary (parametric + smooth terms)
summary(gam_combo)
```

Code Block 5: GAM Diagnostic Checks – Basis Dimension & Residual Inspection

```
# Visualize partial effects of wind speed (overall and by operational state)
draw(gam_combo, select = c("s(Wind.speed)", "s(Wind.speed):Operational"))

# Visualize all smooths for interpretability
draw(gam_combo, select = c("s(Wind.speed)", "s(Wind.speed):Operational",
                           "s(Wind.dir)", "s(DateF)"))

### Residual Diagnostics

# Extract Pearson residuals for autocorrelation check
resid_combo <- resid(gam_combo, type = "pearson")

# Autocorrelation function (ACF) plots at 10-min lag intervals
acf(resid_combo, lag.max = 48, main = "ACF of gam_combo Residuals (10 min lags)")
acf(resid(gam_combo), main = "ACF of GAM_combo residuals")

# Standardized residuals (deviance scale) for QQ plot
z_resid <- scale(resid(gam_combo, type = "deviance"))[,1]
qq_df <- data.frame(sample = z_resid)

# QQ-plot to assess normality of residuals
ggplot(qq_df, aes(sample = sample)) +
  stat_qq() + stat_qq_line() +
```

```
  labs(title = "QQ-plot of Standardised Residuals (z-scale)",
       x = "Theoretical quantiles", y = "Sample quantiles") +
  theme_minimal()


### GAM Diagnostic Checks: Basis Dimension & Residual Inspection

# 1. Check adequacy of basis dimension (k-index):
#    - A k-index < 1.2 indicates that the chosen basis dimension is sufficient.
#    - A value > 1.2 may suggest underfitting or the need for higher k in s() terms.
gam.check(gam_combo)

# 2. Visual diagnostic: Residuals plotted over each smooth term
#    - Uses 'gratia' package to draw smooth terms with residual clouds.
#    - Helps visually assess model fit and detect systematic bias.
#    - Grey points represent residuals; patterns could indicate poor fit.
draw(gam_combo, residuals = TRUE)
```

Code Block 6: AIC Comparison Across Competing GAMs

```
### AIC Comparison Across Competing GAMs
# Compare Akaike Information Criterion for model selection
AIC(gam1, gam2, gam3, gam4, gam5, gam6, gam_int, gam_combo)
```

B Final Model Summary Output

```
Family: gaussian
Link function: identity

Formula:
logE ~ Operational + s(Wind.speed, k = k_ws) +
       s(Wind.speed, by = Operational, k = k_ws) +
       s(Wind.dir, bs = "cc", k = k_dir) +
       s(DateF, bs = "re")

Parametric coefficients:
               Estimate   Std. Error   t value  Pr(>|t|)
(Intercept)    -2.072e+01   1.046e-12 -1.981e+13  < 2e-16 ***
Operational     1.650e-11   2.043e-12  8.077e+00  3.09e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                              edf  Ref.df       F   p-value
s(Wind.speed)                1.004   1.006    0.391   0.536308
s(Wind.speed):Operational    2.110   2.579    1.054   0.278081
s(Wind.dir)                  2.432  10.000   12.677   0.000104 ***
s(DateF)                     7.320  14.000    2.012   0.000659 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Rank: 46/47
R-sq.(adj) = 0.945    Deviance explained = 96.2%
-REML = -1056.8       Scale est. = 3.5994e-24    n = 47
```

C Rmd Code and Knitted Pdf Files

The **R Markdown source code** (`Seismic_Noise_v7.Rmd`) and its **knitted pdf output** (`Seismic_Noise_v7.pdf`) used to generate all figures, tables, and statistical models in this report are available at the following GitHub Repo:

`https://github.com/TilakHeble/Seismic-Noise`

This decision was made to avoid cluttering the appendix with long code blocks. All scripts are reproducible and commented for transparency.