The School of Mathematics

THE UNIVERSITY
of EDINBURGH

Seismic Noise from Eskdalemuir Wind Turbines:

An XGBoost Modeling Approach

by

Cyrus Seyrafi

June 2025
Supervised by Ioannis Papastathopoulos

Executive Summary

The Eskdalemuir Seismic Array (EKA) is the UK's primary contribution to the Comprehensive Nuclear-Test-Ban Treaty (CTBT), monitoring global seismic activity in the 0.5-8 Hz range to detect underground nuclear tests. However, modern wind turbines—which are in demand across southern Scotland—also emit energy within this frequency band, raising concerns that continued wind development could interfere with the EKA's detection capabilities.

To mitigate this risk, the Ministry of Defence (MoD) has imposed strict planning constraints: a 10 km exclusion zone around the array and a cumulative seismic noise cap of 0.336 nanometers (nm) in displacement amplitude for all turbines within a 50 km radius.

Initial impact assessments were based on a conservative 2014 model developed by Xi Engineering. By 2018, this model indicated that the seismic noise budget had been fully allocated, effectively halting further turbine development in the region. However, more recent field data suggest that actual turbine emissions may be lower than originally estimated, highlighting the need for a more accurate and adaptive model.

This study addresses that need by developing a machine learning model to quantify the seismic uplift caused by turbine operation. Using ground motion data collected before and after the construction of a wind farm in southern Scotland, it applies eXtreme Gradient Boosting (XGBoost) to predict the EKA-weighted power spectral density (PSD) of seismic displacement. Separate models are trained on background and operational periods, with their difference isolating the uplift attributable to turbine activity.

Key findings include:

- Frequency-specific uplift: Peaks occur at 5 Hz and 8–10 Hz—within or adjacent to the EKA's primary band—where turbine-induced uplift reaches as high as 17 dB.
- The environment is a significant factor: Uplift varies with wind direction, with northeast winds producing the highest increases.
- High model performance: XGBoost achieved $R^2 > 0.998$ and offered tight uncertainty bounds, outperforming traditional methods.

Although results are expressed in decibels of PSD, they directly relate to the MoD's displacement threshold: elevated PSD at key frequencies implies higher seismic energy at EKA and greater risk of breaching the amplitude cap.

To enhance policy relevance, the following extensions are proposed:

1. Propagation modeling to estimate uplift at the EKA itself, allowing direct comparison with the 0.336 nm limit.
2. Multi-site training to assess generalizability across turbine types and terrains.
3. Expanded environmental data (e.g. wind speeds, turbine logs) to enable real-time forecasting under variable conditions.

These improvements would greatly enhance future, data-driven assessments of seismic compatibility—upholding the UK's CTBT obligations while further enabling Scotland's renewable energy goals.

Despite limitations such as single-site data and restricted wind conditions, the present study offers a scalable, empirically grounded framework. It demonstrates that seismic monitoring and wind development need not be in conflict and provides a robust foundation for modernizing regulation in a way that balances international security with clean energy growth.

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: Cyrus Seyrafi
Matriculation Number: s2766504
Title of work: Seismic Noise from Eskdalemuir Wind Turbines: An XGBoost Modeling Approach

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`).

Signature: Cyrus Seyrafi

Date: 27 June 2025

Contents

1 Introduction

The Eskdalemuir Seismic Array (EKA) is the UK's contribution to the International Monitoring System (IMS) established under the Comprehensive Nuclear-Test-Ban Treaty (CTBT). Since 1962, the array has played a critical role in detecting underground explosions from nuclear testing in foreign nations, and to this day remains an essential component of the UK's global monitoring obligations [5].

The array comprises of a set of nuclear seismometers that are particularly sensitive to ground motions in the 0.5-8 Hz frequency range. This makes it well suited for identifying seismic signals from nuclear tests. However, this frequency range also captures vibrations generated by wind turbines, leading to growing concern in recent decades that large-scale wind energy development in the area could compromise the array's monitoring capabilities [7].

In 2005, the Ministry of Defence (MoD) introduced planning regulations to protect the EKA from such threats. These include prohibitions on the construction of wind turbines within 10 kilometers of the array and a cap on the total seismic contribution from all wind turbines within 50 kilometers of the array at 0.336 nanometers (nm) in amplitude [5, 7].

This consultation zone covers approximately 10% of Scotland's land area and includes several high-potential sites for renewable energy [7]. To assess whether new turbines fall within the allowable seismic budget, the MoD relies on a predictive model developed by Xi Engineering in 2014 (as of 2020) [6]. This model estimates seismic emissions conservatively based on turbine parameters such as hub height, rotor diameter, and distance from the array, and inflates predictions by 20% to account for uncertainty in turbines' models and makes [6].

By 2018, this model indicated that the seismic budget had been fully allocated [5, 6], halting new wind developments within the consultation zone. However, more recent field measurements have shown that the actual seismic emissions from operational wind farms are often significantly lower than predicted. This suggests that model tuning may permit further wind development without compromising the EKA's objective [6, 7]

This project aims to contribute to that refinement. Using field data provided by Xi Engineering, I model EKA-sensitivity-adjusted power spectral density (PSD) of seismic displacement. This analysis accounts for variation across frequency, wind direction, and time, with the goal of isolating the seismic contribution of wind turbines from background noise. My findings are intended to inform ongoing efforts to maximize wind energy development relative to the UK's nuclear commitments [8].

2 Data

The data provided by Xi Engineering comprise ground motion measurements collected near a wind farm site in southern Scotland. Recordings were made using a seismometer installed in a dedicated seismic pit, both before and after the construction of nearby wind turbines. This setup allows for direct comparison between background and operational conditions under matched environmental settings [8].

Two datasets were initially provided. The first is a velocity time series representing vertical ground motion recorded at 100 Hz across 394 ten-minute intervals (hereafter referred to as "chunks"). Each chunk contains 60,000 samples of ground velocity (in meters per second), along with associated metadata: date and time, wind speed (filtered to a range of 11.5–12.5 m/s), wind direction, and turbine operational status.

The second dataset is in the frequency domain. It includes 47 chunks that have been converted into displacement power spectral densities (PSDs) using Welch's method. Each entry contains both the raw PSD and an EKA-sensitivity adjusted PSD, which applies a frequency weighting curve tailored to EKA's detection capabilities. This adjustment emphasizes energy in the 0.5–8 Hz range, where turbine-generated signals are most likely to intersect with the detection array's sensitivity.

Of the 47 frequency-domain records, 42 directly correspond to entries in the time series dataset. PSDs are reported for 8,192 frequency bins, uniformly distributed between 0 Hz and the Nyquist frequency of 50 Hz (based on the 100 Hz sampling rate). These records also include the same metadata as the time series.

To improve sample size and ensure uniform preprocessing, I generated an expanded frequency-

domain dataset by reapplying Welch's method (`gsignal::pwelch`) to all 394 chunks in the time series dataset. Using a Hamming window with 50% overlap and 28 segments per chunk—conditions similar to the original methodology—I estimated the velocity PSDs for each chunk at 8,192 frequency bins, converted them to displacement, and applied EKA's sensitivity weighting using a spline-interpolated version of the provided weighting function.

As shown in Figure 1, this yielded a dataset structurally similar to the original frequency-domain data but covering the entire set of available time chunks.

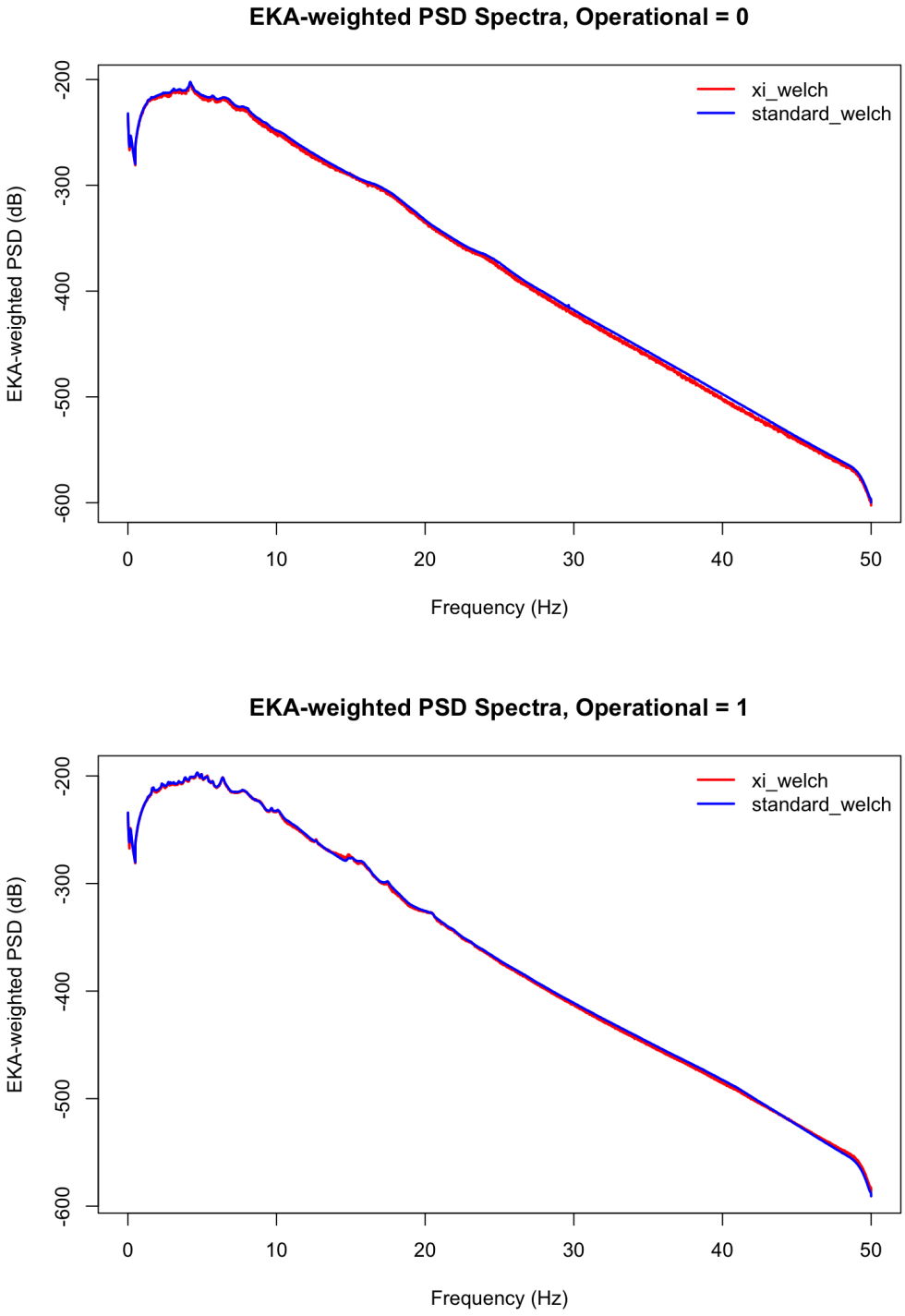

Figure 1: EKA-adjusted power spectral densities (in dB) by operational status across Xi dataset and reconstructed Welch's method dataset.

I also explored the use of debiased Welch method (`dwelch::dwelch`), a recent technique developed

by Astfalck et al. (2024) that aims to correct for bias inherent in standard spectral estimates [1]. This method fits the PSD in a basis expansion framework via constrained weighted least squares and offers theoretical advantages in low-power settings. In this context, however, it performed poorly: it consistently overestimated PSD levels, showed erratic fluctuations across frequencies, and proved computationally expensive, requiring several seconds per time chunk.

For these reasons, the debiased method was excluded from the primary analysis, though it is retained for visual comparison in Figure 2.

The final expanded dataset, for each time chunk and frequency bin, includes:

- frequency (Hz),
- raw and EKA-adjusted displacement PSDs (in $\frac{m^2}{\text{Hz}}$),
- wind speed,
- wind direction,
- chunk date and time, and
- operational status.

Exploratory data analysis, as shown in the Appendix, revealed minimal effects from wind speed and chunk date and time on EKA-adjusted displacement PSD. As a result, modeling efforts focused on frequency, wind direction, and operational status as primary features.

3 Models

3.1 XGBoost

To model EKA-adjusted displacement PSD, I employ eXtreme Gradient Boosting (XGBoost), a high-performance implementation of gradient-boosted decision trees [2]. XGBoost was selected after comparison with generalized additive models (GAMs) and quantile GAMs. While the latter approaches offered interpretability or theoretical appeal, XGBoost yielded superior empirical performance in predicting log-transformed PSD values and allowed flexible modeling of complex nonlinear relationships across frequency and environmental conditions.

3.1.1 Model Structure

XGBoost builds a predictive model by adding together a series of decision trees. Each tree attempts to improve upon the errors made by previous ones, forming a powerful ensemble that can capture complex patterns in the data—a process known as gradient boosting.

Each observation i represents a frequency bin in a given 10-minute time chunk. For easy conversion to decibels, the model predicts the base-10 log-transformed EKA-adjusted PSD for that bin. Formally, the prediction is:

$$\hat{y} = \sum_{k=1}^{K} f_k(x_i) \quad f_k \in \mathcal{F} \tag{3.1}$$

where:

- $\hat{y}_i$ is the predicted $\log_{10}$ PSD,
- f_k is the k-th decision tree (of K),
- $\mathcal{F}$ is the space of regularized regression trees,
- x_i is the feature vector:

$$x_i = \begin{bmatrix} \nu_i \\ \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix}$$

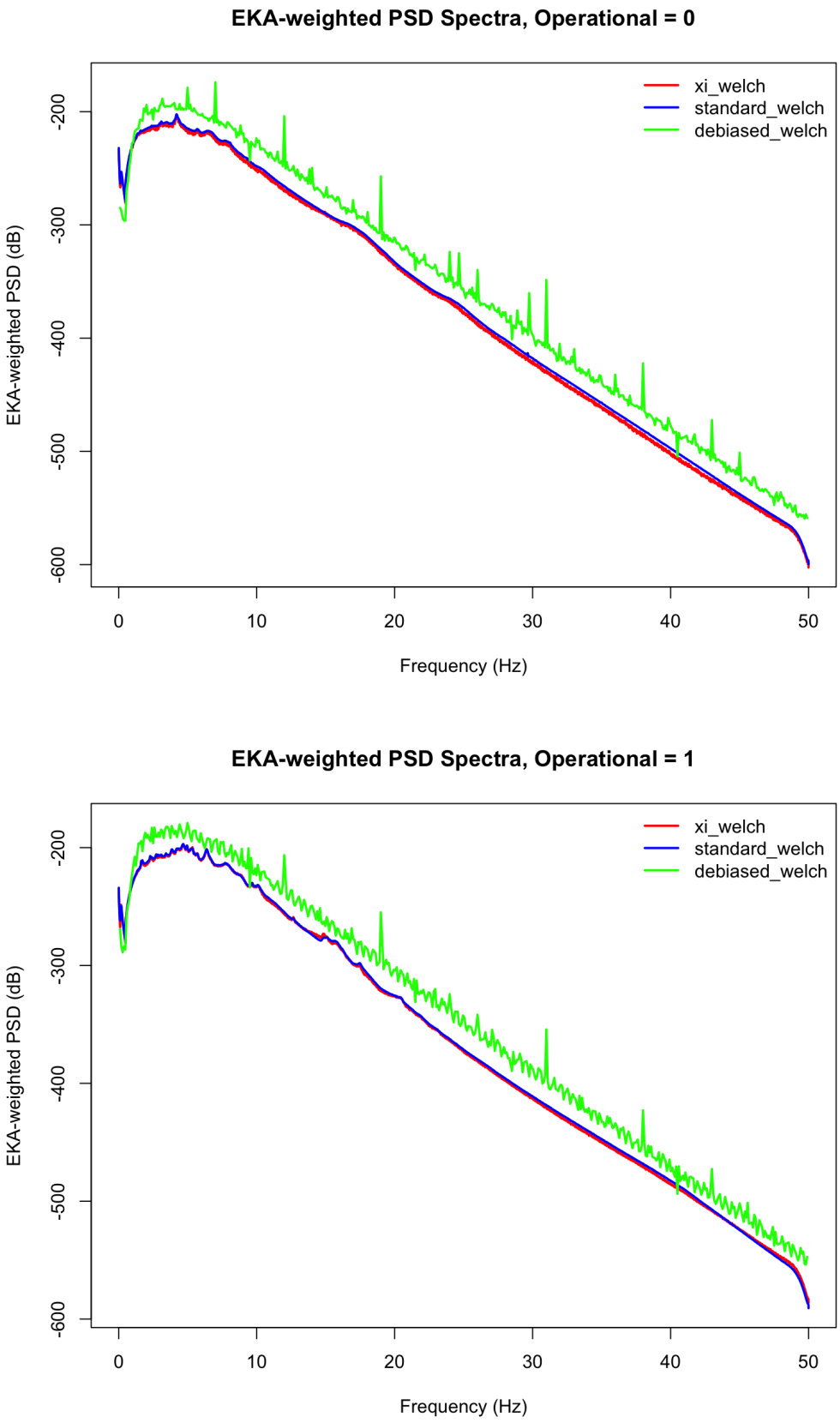

Figure 2: EKA-adjusted power spectral densities (in dB) by operational status across the Xi, reconstructed standard Welch, and debiased Welch datasets. The debiased Welch dataset proves unstable.

with ν_i representing frequency (Hz) and θ_i representing wind direction (radians).

To train these trees, XGBoost minimizes a regularized objective function that balances predictive

accuracy with model simplicity:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{k=1}^{K}\Omega(f_k) \tag{3.2}$$

This consists of:

1. A loss term $(y_i - \hat{y}_i)^2$, representing squared prediction error, and
2. A regularization term $\Omega(f)$, which penalizes complex trees to prevent overfitting.

Each decision tree partitions the feature space into regions, represented as the leaves of the tree. Each leaf corresponds to a unique set of feature conditions (e.g. specific ranges of frequency and wind directions) and assigns a constant prediction value to all observations that fall into it.

The regularization term is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \tag{3.3}$$

where:

- T is the number of leaves (i.e. prediction regions) in the tree,
- w_j is the predicted value assigned to leaf j,
- γ penalizes trees with more leaves, and
- λ penalizes large prediction values through L2 regularization.

Together, these regularization terms discourage overly complex trees and promote models that generalize well to new data. As a practical benefit, they also improve numerical stability and training speed.

3.1.2 Training Design

Separate XGBoost models were trained for the operational and background periods. As previously mentioned, the target variable in each case was the base-10 logarithm of the EKA-adjusted displacement PSD. The predictor variables were:

- Frequency ν: the spectral frequency bin; and
- Wind direction: $\sin\theta$, $\cos\theta$, where θ is wind direction in radians (cyclical encoding).

To avoid data leakage, the dataset was split into testing and training sets on a per-chunk basis: 80% of chunks were assigned to training and 20% to testing. This ensured that frequency bins from the same chunk were not split across sets, which would otherwise inflate performance metrics.

3.1.3 Uplift Estimation

To isolate the seismic contribution of turbines, I generated predictions under fixed environmental conditions and computed the uplift as the difference in predicted EKA-adjusted PSD between operational and background models. Wind direction was held constant at its overall mean (167.23°).

The resulting uplift curve was scaled by a factor of 10 to convert from log-transformed PSD to decibels (dB):

$$\mathcal{U}(\nu) = 10 \cdot [\hat{y}_{\omega=1}(\nu) - \hat{y}_{\omega=0}(\nu)] \tag{3.4}$$

where $\hat{y}_\omega(\nu)$ represents the model-predicted log-PSD values at frequency ν for operational status $\omega \in \{0, 1\}$

3.2 Generalized Additive Models

Before adopting XGBoost, I explored a series of generalized additive models (GAMs) and quantile GAMs (qGAMs) to model log-EKA-adjusted PSD. GAMs generalize linear regression by replacing fixed coefficients with smooth functions of predictors, allowing for flexible, nonlinear relationships. This structure is useful for uncovering gradual trends without assuming a strict parametric form. However, the GAMs were unable to capture turbine-specific uplift with sufficient resolution, which led me to replace them with XGBoost.

3.2.1 Model Structure

The GAMs modeled log-transformed EKA-adjusted PSD as smooth functions $s(\cdot)$ of frequency ν, wind direction θ, and turbine operational status $\omega \in \{0, 1\}$ using `mgcv::bam`. The two strongest formulations were:

Model A

$$y = s(\nu) + s(\nu|\omega = 0) + s(\nu|\omega = 1) + s(\theta) + s(c) \tag{3.5}$$

where:

- $s(\nu)$ is the overall PSD-frequency trend,
- $s(\nu|\omega = 0)$ and $s(\nu|\omega = 1)$ account for turbine-off and turbine-on spectral differences,
- $s(\theta)$ models the cyclic effect of wind direction (treating 0 and 2π radians equivalently), and
- $s(c)$ is a random-effect smooth for chunk-level variation (each date-time interaction).

This model separated turbine-on and turbine-off conditions explicitly, while accommodating environmental and temporal variation.

Model B

$$y = s(\nu) + t(\nu, \omega) + t(\nu, \theta) + s(c) \tag{3.6}$$

where:

- $s(\nu)$ is the overall PSD-frequency trend,
- $t(\nu, \omega)$ is an additive tensor interaction between frequency and operational status,
- $t(\nu, \theta)$ is a full tensor interaction between frequency and wind direction, and
- $s(c)$ is a random-effect smooth for chunk-level variation (each date-time interaction).

Both models were fit using cubic regression splines (typically with $k = 20$) and estimated using penalized restricted maximum likelihood. Wind direction was encoded using cyclic splines to respect its circular structure.

3.2.2 Model Training

Additional models tested alternative interaction structures and included excluded features like wind speed and time of day, but Models A and B consistently outperformed these variants capturing 99.6%+ of deviance and having the lowest AIC scores. To address heteroskedasticity and skew in PSD values, I also evaluated quantile GAMs using `qgam::qgam`, which estimate conditional quantiles (e.g., median, 75th percentile) instead of means. However, both standard and quantile GAMs underperformed.

3.2.3 Limitations and Rejection

Several limitations led to the abandonment of the GAM-based models:

- GAMs tended to oversmooth, obscuring frequency-specific effects.
- Wind direction and turbine operation were partially confounded, complicating the separation of their effects on PSD.
- In models allowing frequency-specific operational effects, the uplift curve was often regularized away, frequently collapsing to zero.

Thus, despite their strengths in interpretability, GAMs and qGAMs were ultimately set aside in favor of XGBoost.

4 Results

4.1 Model Performance and Diagnostics

The final XGBoost models demonstrated exceptionally strong predictive performance on the test set. Summary metrics are shown in Table 1:

Model	RMSE	MAE	R^2
Background	0.430	0.314	0.9986
Operational	0.196	0.137	0.9997

Table 1: Performance metrics for background and operational XGBoost models. High fit for both models.

These values indicate near-perfect accuracy with very low residual error. Predicted versus observed log-PSD plots (Figure 3) confirm these results, showing tight alignment along the 1:1 reference line.

Feature importance analysis (Figure 4) reveals that spectral frequency overwhelmingly dominates model predictions. Wind direction—cyclically encoded via sine and cosine transforms—contributes only marginally.

4.2 Uplift Curve

To quantify uplift attributable to turbine operation, I applied a chunk-level bootstrap procedure across 100 resampled datasets. In each iteration, separate XGBoost models were trained on background and operational data, and uplift was computed across the 0.5–50 Hz frequency range. Wind direction was held constant at its mean value (167.23°) to isolate the effect of operational status. Confidence intervals at the 50%, 80%, 90%, and 95% levels were derived from empirical quantiles of these bootstrap replicates.

The resulting uplift curve is presented in Figure 5. The model-predicted median uplift appears in black, with nested blue ribbons representing confidence intervals, and the uplift derived from raw data using Welch's method overlaid in red for comparison.

Several distinct peaks emerge:

- The global maximum occurs at 10 Hz, where the median uplift reaches 18 dB.
- Prominent secondary peaks appear at 5 Hz (15.7 dB), 8–9 Hz (13.4–14.5 dB), 15-16 Hz (14.8-15.9 dB) and 43–45 Hz (17.2–17.6 dB).

Substantial uplift at high frequencies (above 35 Hz) is particularly noteworthy. Despite the EKA-weighting function strongly attenuating these frequencies, the models consistently detected high PSD uplift due to turbine operation, suggesting wind turbines may affect EKA readings even in this range.

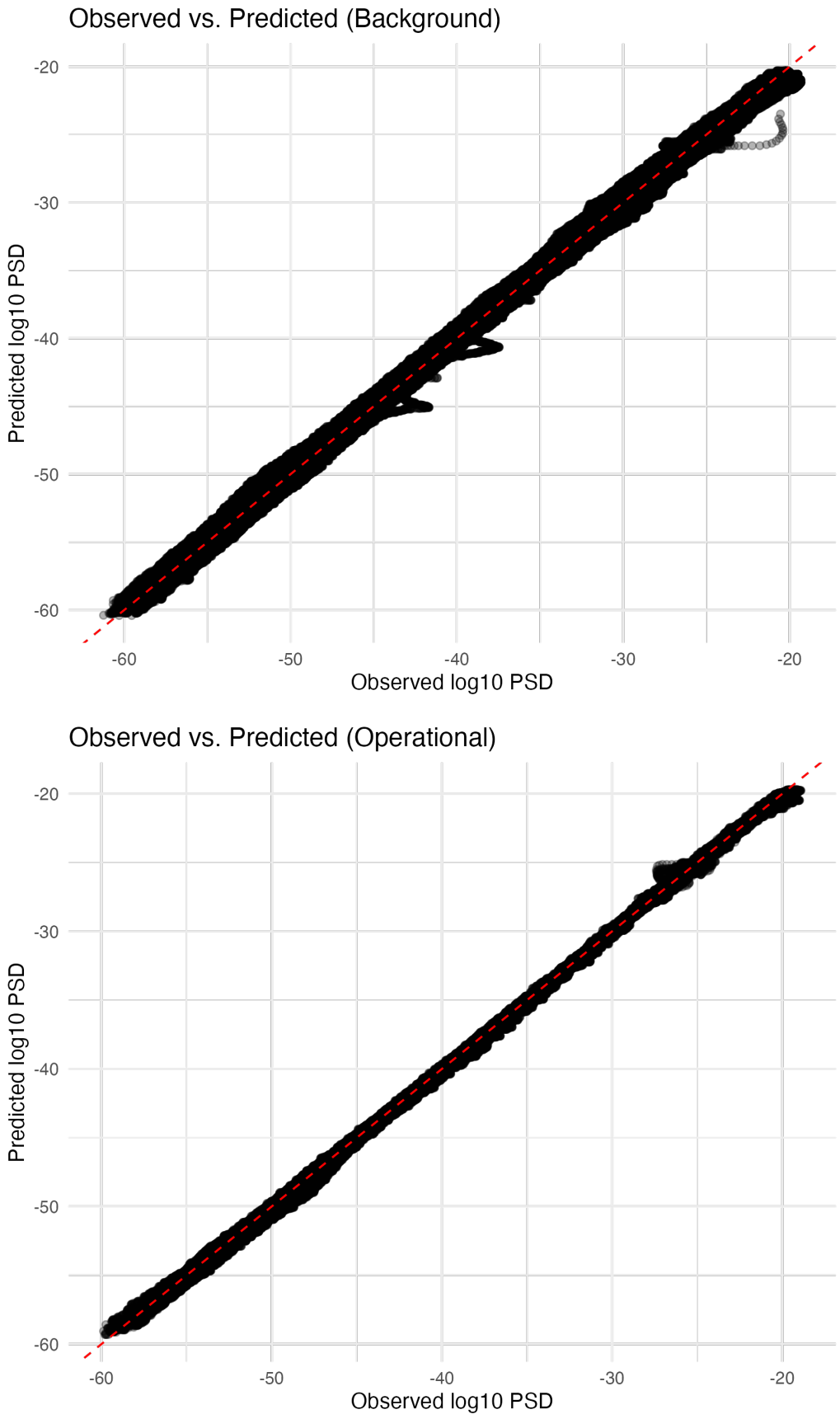

Figure 3: Predicted vs. Observed log-PSD plots for background and operational XGBoost models. High fit for both models.

That said, confidence intervals widen considerably at higher frequencies, indicating increased model

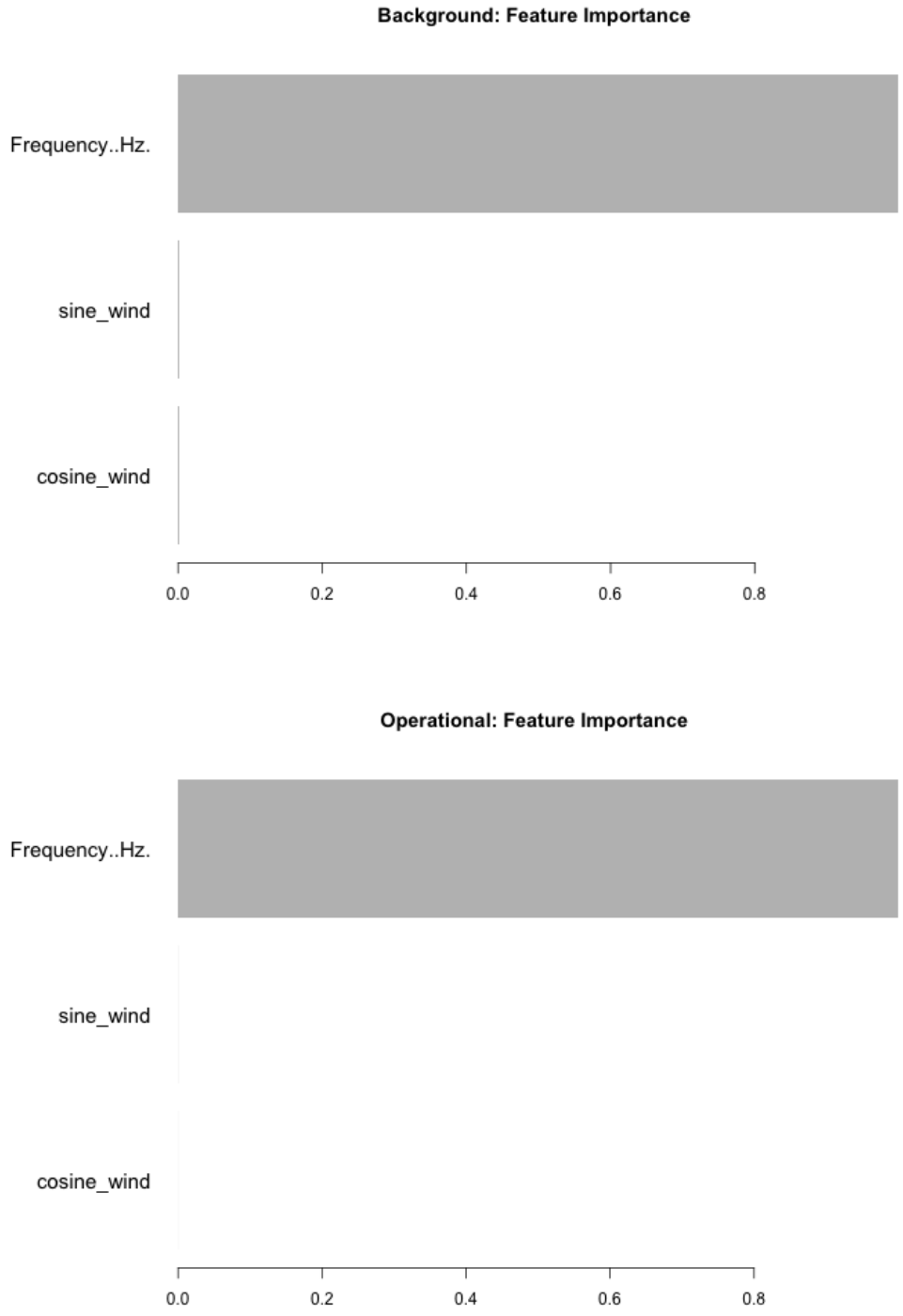

Figure 4: Feature importance analysis for background and operational XGBoost models. Frequency dominates.

uncertainty.

In contrast, the tightest confidence intervals are observed between 0.5 and 10 Hz, which conveniently aligns with the EKA's most sensitive frequency band. Across this 0.5–10 Hz range, the mean half-width of the 95% confidence interval is 1.08 dB, with a standard deviation of 0.300 dB—indicating well-bounded uncertainty.

Table 2 summarizes the median uplift and confidence intervals at key frequencies.

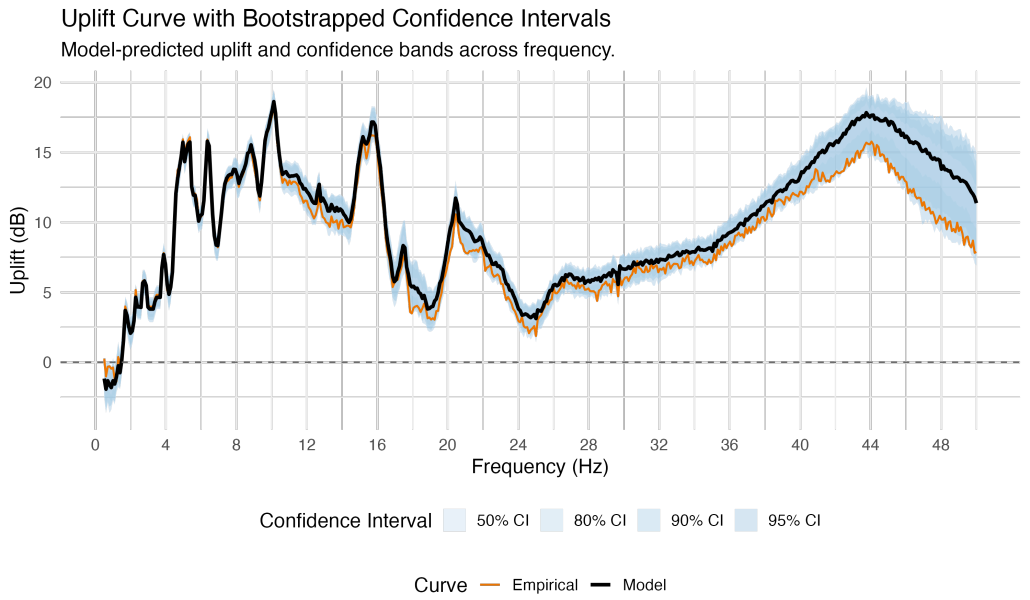

Figure 5: Model-predicted uplift curve (black) with confidence intervals (blue), and standard Welch dataset (red) for comparison.

Frequency (Hz)	Median	50%	80%	90%	95%
2	2.07	[1.63, 2.42]	[1.39, 2.73]	[1.24, 2.88]	[0.90, 3.18]
3	3.94	[3.68, 4.35]	[3.33, 4.57]	[3.12, 4.64]	[3.02, 4.79]
4	6.99	[6.60, 7.24]	[6.27, 7.50]	[6.14, 7.76]	[5.97, 7.93]
5	15.7	[15.28, 15.93]	[14.98, 16.10]	[14.79, 16.24]	[14.57, 16.35]
6	10.5	[10.03, 10.76]	[9.80, 11.00]	[9.69, 11.13]	[9.46, 11.19]
7	9.47	[9.03, 9.72]	[8.61, 9.99]	[8.31, 10.34]	[8.16, 10.43]
8	13.4	[12.93, 13.84]	[12.16, 14.21]	[12.07, 14.40]	[11.76, 14.52]
10	18.0	[17.59, 18.38]	[17.17, 18.69]	[17.02, 18.84]	[16.70, 18.89]
15	14.8	[14.25, 15.28]	[13.67, 15.56]	[13.35, 15.74]	[13.19, 15.98]
20	8.89	[8.33, 9.16]	[7.82, 9.43]	[7.26, 9.67]	[7.03, 9.78]
30	6.66	[6.31, 7.07]	[5.81, 7.45]	[5.48, 7.69]	[5.34, 7.76]
40	13.2	[12.60, 13.68]	[12.07, 14.04]	[11.86, 14.32]	[11.46, 14.44]
45	16.9	[15.74, 17.65]	[14.52, 18.20]	[13.87, 18.35]	[13.57, 18.57]

Table 2: Model-predicted uplift curve at key frequencies with confidence intervals.

4.3 Frequency Structure of Seismic Uplift at Low Frequencies

The spectral features of the uplift curve reflect well-established physical mechanisms by which wind turbines generate ground motion. In particular, the uplift peaks at 5 Hz and 8–9 Hz are consistent with turbine-related energy observed in field studies. For example, Hu et al. (2018) identified spectral peaks at 2.5, 4.5, 6, and 8.5 Hz during turbine operation, attributing them to blade-passing harmonics and structural resonances [3]. These signals, detectable up to 500 meters from the turbine base, were found to increase with wind speed.

Although the wind speed in the present study is fixed at approximately 12 m/s, the location and strength of the observed uplift peaks remain physically meaningful. The 5 Hz and 8–9 Hz features likely correspond to mechanical modes tied to blade passage and tower response, both of which fall within the 1–10 Hz range where Hu et al. consistently observed enhanced seismic power [3].

Hu et al. also noted that low-frequency seismic energy—especially near 6–7 Hz—can arise from non-turbine environmental factors such as wind gusts, pressure variations, and shear stress [4]. However, since the uplift curve represents the modeled difference between operational and background seismic

conditions, such non-turbine contributions are effectively removed. What remains is a net signal attributable to turbine activity.

4.4 Wind Direction

To assess the model's adaptability to wind direction, uplift curves were generated across four quadrants—NE, NW, SE, and SW—and compared with curves from raw data using Welch's method (Figure 6). The model consistently predicts the highest EKA-adjusted PSD when winds originate from the northeast and the lowest when from the southwest.

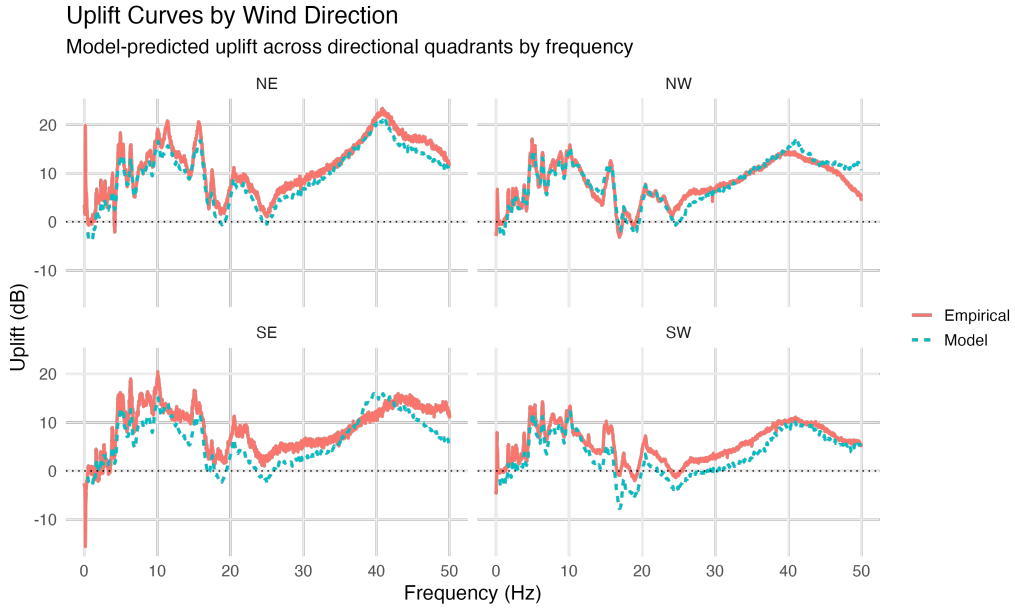

Figure 6: Model-predicted uplift curves by wind direction: NE, NW, SE, SW.

At low frequencies, this pattern aligns reasonably well with the observed data. At high frequencies, however, some divergence appears—especially in the neighborhood of 40 Hz. This reflects the already noted increased model variability beyond 35 Hz.

5 Discussion

This study set out to quantify the seismic impact of wind turbine operation near the Eskdalemuir Seismic Array (EKA) using high-resolution frequency-domain data and advanced statistical modeling. The XGBoost model met this objective, revealing that uplift is real, frequency-localized, and statistically significant—especially in the 0.5–8 Hz band, where the EKA is most sensitive. These findings have direct implications for seismic monitoring policy and wind energy planning in southern Scotland.

5.1 High Frequencies

The uplift curve shows that turbine-induced seismic energy is concentrated in sharply defined frequency bands. The strongest and most statistically confident peaks lie within EKA's 0.5–8 Hz sensitivity range, particularly at 5 Hz and 8–9 Hz. A notable global maximum also appears at 10 Hz—just outside EKA's formal sensitivity band—but emerges even after applying the EKA sensitivity-weighting function, suggesting possible relevance. Future work should explore the implications of uplift in these higher-frequency bands.

5.2 Limitations

While the model performs well and reveals plausible uplift patterns, several limitations constrain its precision and generalizability:

1. The dataset was collected near a single wind farm, limiting geographic and environmental robustness. As a result, uplift trends may not generalize across different terrains or turbine types.

2. Key operational details—such as turbine count, coordinates, hub heights, and schedules—were unavailable. This prevents per-turbine attribution and limits the model's immediate use for planning or simulation.

3. Wind speed was tightly filtered (11.5–12.5 m/s) to reduce variability, which enhanced internal validity but restricts generalizability. Wind speed variability, as highlighted by Hu et al. (2020), may significantly influence uplift [4].

4. While Welch's method effectively converts time-series data to the frequency domain, Astfalck et al. (2024) note it often overestimates real PSD [1]. Although the proposed debiased method was ineffective here, standard Welch preprocessing may have led to PSD overestimation, potentially biasing model outputs.

5. At high frequencies, the model's confidence intervals widen considerably—by the high 40s Hz exceeding 5 dB at the 95% level. This limits interpretability; however, these frequencies fortunately lie outside EKA's core sensitivity range.

6. The background and operational datasets are not fully balanced, and neither is either subset among wind direction. Despite efforts to control for chunking and environmental factors, these imbalances could introduce bias, particularly in high-variance frequency regions.

Despite these constraints, the model provides a solid foundation for future work and could be expanded to support turbine-level assessments, site comparisons, and policy development.

5.3 The Seismic Budget

The Ministry of Defence (MOD) imposes a cumulative seismic ground vibration (SGV) threshold of 0.336 nanometres (nm) displacement amplitude for all wind turbine activity within 50 km of the Eskdalemuir Seismic Array (EKA) [5]. On the other hand, the uplift estimated in this study is expressed in decibels of displacement power spectral density (PSD), weighted by EKA's frequency sensitivity.

While these units are not directly comparable, their physical relationship allows meaningful interpretation: increases in PSD at any frequency translate into greater total ground motion energy, raising the likelihood of exceeding the seismic budget. Accordingly, the uplift curve produced here can be read as a frequency-by-frequency draw on the national seismic budget.

The most substantial uplift in the EKA's range occur at 5 Hz, 6.5 Hz, and 8–10 Hz, with power levels rising roughly 15 dB above background. Even without precise turbine counts, the consistency of these peaks across operational data suggests that individual turbines may contribute materially to total SGV.

With access to turbine specifications—such as location, count, and operational profiles—the model could be extended to estimate per-turbine contributions in absolute terms. This would enable simulation of proposed layouts, evaluation of retrofits, and retrospective analysis of past projects, directly supporting MoD efforts to move beyond conservative assumptions toward data-driven, site-specific regulation that aligns with Scotland's renewable energy goals.

5.4 Future Work

This study arrives at a key moment: the MOD's 2025 consultation calls for more robust, empirical approaches to seismic uplift estimation [5]. The methodology presented here—combining operational data, frequency-specific modeling, and machine learning—offers a flexible and scalable alternative to static, worst-case turbine models.

Several extensions could strengthen its policy relevance:

1. Currently, uplift is measured at or near the turbine site. To assess compliance with the MOD's seismic threshold at Eskdalemuir, future work should incorporate propagation modeling—simulating how uplifted seismic energy travels through the ground to EKA. This would allow output in nanometres at the array, enabling direct comparison with the 0.336 nm cap.

2. Retraining the model on data from multiple wind farms across varying terrains would test generalizability and help identify turbine-specific or region-specific uplift patterns.

3. Including broader wind speed ranges, environmental parameters, and turbine operational logs would enable the model to forecast uplift under dynamic, real-world conditions.

Taken together, these enhancements would enable a transition from fixed exclusion zones to continuous, data-driven assessments of seismic compatibility—supporting both the UK's CTBT commitments and Scotland's renewable energy ambitions.

6 Conclusion

This study demonstrates that turbine-induced seismic uplift is real, frequency-specific, and detectable with high statistical confidence—particularly within the 0.5–8 Hz band where the Eskdalemuir Seismic Array is most sensitive. By leveraging operational data and machine learning, the XGBoost model provides a robust, data-driven estimate of seismic power attributable to wind turbines, offering a meaningful improvement over worst-case models. While limitations in site diversity, turbine metadata, and wind speed coverage remain, the framework developed here lays a strong foundation for more precise seismic budgeting. With extensions such as site-to-array propagation modeling and regional retraining, this approach could support the MOD in refining its seismic safeguards while unlocking new opportunities for wind development within the consultation zone—advancing both national security obligations and renewable energy targets.

References

[1] L. C. Astfalck, A. M. Sykulski, and E. J. Cripps. Debiasing welch's method for spectral density estimation. *Biometrika*, 111(4):1313–1329, December 2024.

[2] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pages 785–794, San Francisco, CA, USA, 2016. ACM.

[3] W. Hu, R. J. Barthelmie, F. Letson, and S. C. Pryor. A new seismic-based monitoring approach for wind turbines. *Wind Energy*, 22(4):473–486, 2019.

[4] W. Hu, R. J. Barthelmie, F. Letson, and S. C. Pryor. Seismic noise induced by wind turbine operation and wind gusts. *Seismological Research Letters*, 91(1):427–437, 2020.

[5] Ministry of Defence. Consultation on the ministry of defence's approach to safeguarding the eskdalemuir seismological array. Web, March 2025. Closed consultation updated 5 March 2025.

[6] Xi Engineering Consultants. Eskdalemuir wind turbine seismic vibration: Assessment of headroom. Technical Report SGV_201_Tech_Report_v04, Xi Engineering Consultants Ltd., CodeBase, Argyle House, 3 Lady Lawson Street, Edinburgh, EH3 9DR, United Kingdom, February 2020. Presented to The Scottish Government.

[7] Xi Engineering Consultants. Eskdalemuir wind turbine seismic vibration: Calculations to confirm maximum turbine seismic level to deploy minimum of 1gw deployment. Technical Report SGV-205-LimitSet-TechReport-v13, Xi Engineering Consultants Ltd., CodeBase, Argyle House, 3 Lady Lawson Street, Edinburgh, EH3 9DR, United Kingdom, January 2023. Report presented to the Scottish Government.

[8] Xi Engineering Consultants. Wind turbine and the eskdalemuir seismic array. Technical report, Xi Engineering Consultants Ltd., June 2025. Dataset given to MSc student at the University of Edinburgh.

Appendix

A Exploratory Data Analysis

Exploratory data analysis evaluated whether any chunk metadata features warranted inclusion in the models. The following plots illustrate relationships between these features, frequency, and EKA-adjusted log-PSD. Scatterplots for wind speed, hour, and date (within operational status groups) show little variation in log-PSD across frequencies, offering no strong justification for their inclusion. However, a polar plot of wind direction reveals a modest directional effect, supporting its use in the final models.

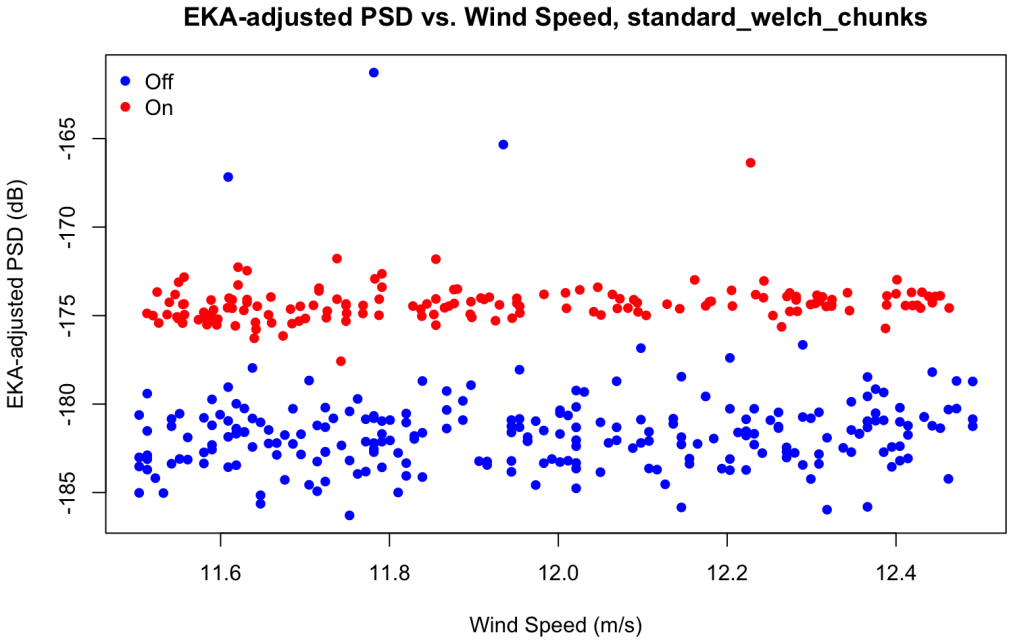

Figure 7: EKA-adjusted log-PSD by Wind speed. No clear relationship is found.

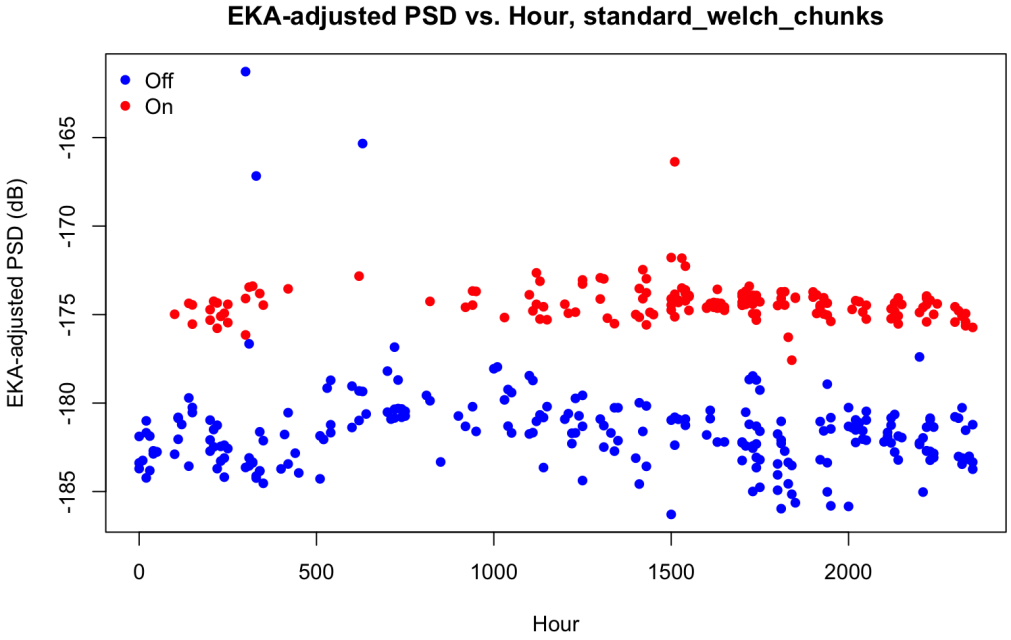

Figure 8: EKA-adjusted log-PSD by Hour. No clear relationship is found.

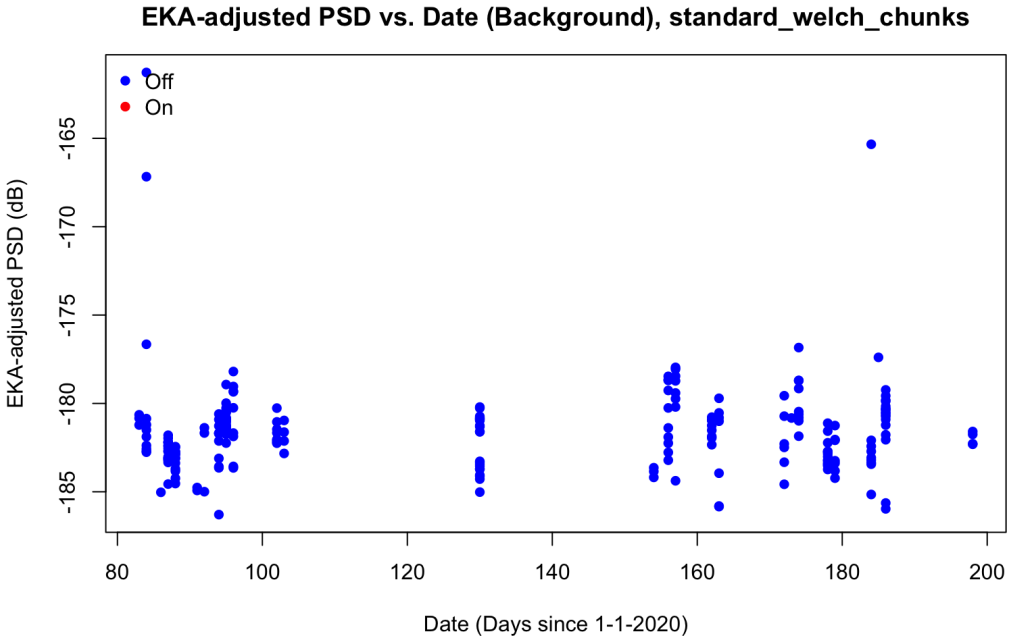

Figure 9: EKA-adjusted log-PSD by Date (inactive turbines). No clear relationship is found.

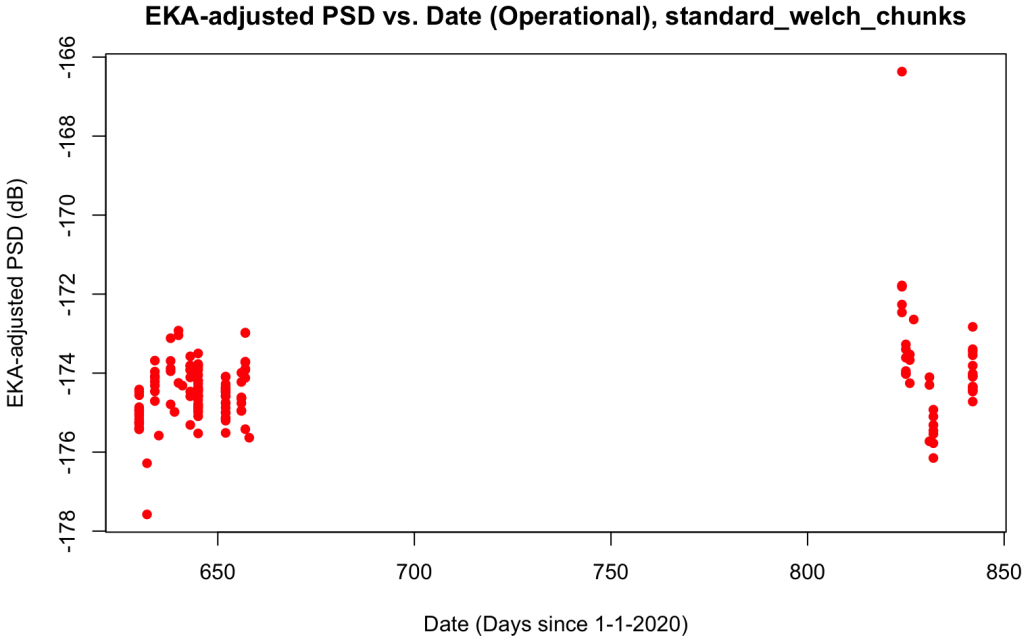

Figure 10: EKA-adjusted log-PSD by Date (active turbines). No clear relationship is found.

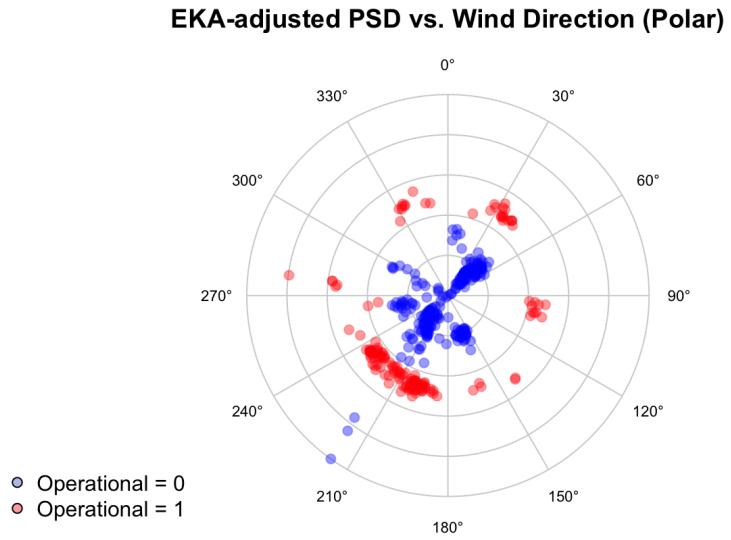

Figure 11: Polar graph of EKA-adjusted log-PSD by Wind Direction. There is heavy clustering in the SW quadrant showing marginally greater background-operational differences in the NE, NW, and SE. Thus it was included in the final model.