RC1: 'Comment on egusphere-2025-4174', Anonymous Referee #1, 07 Nov 2025

The manuscript represents an outstanding contribution to scientific progress by demonstrating a scalable, autonomous system for wind speed estimation using profiling floats and passive acoustic monitoring (PAM).

The core novelty is the successful deployment and retrieval of wind data from deep parking depths (500–1000 m) and the development of a residual learning framework. This framework is critical because it addresses the core limitation of acoustic wind retrieval in remote regions: the lack of local calibration data. By combining global reanalysis (ERA5) with sparse in-situ observations to correct systematic biases, the authors achieved a major quantitative improvement: a 37% reduction in RMSE and an increase in R^2 from 0.85 to 0.91. This work provides a practical path to integrate wind forcing observations with the BGC-Argo float array.

The scientific approach and applied methods are valid and well-justified. The use of established empirical models (like Nystuen et al., 2015) and the innovative application of the XGBoost algorithm for residual learning are appropriate for handling the non-linear relationship between acoustics and wind. The necessary preprocessing steps, such as depth correction and noise mitigation, are included.

The discussion is appropriate and balanced, explicitly acknowledging the systematic bias of the ERA5-fitted model in high-wind regimes (>10 m/s) and the need for the residual correction. It accurately situates the findings within the framework of prior moored and mobile PAM research.

The authors present the scientific results and conclusions in a clear, concise, and well-structured manner. The manuscript adheres to a standard, logical flow, and the technical language is precise. The figures are of high quality and clearly show the main findings, especially when comparing the unoptimized, ERA5-fitted, and ML-corrected time series (Figure 8) and the scatter plots (Figure 6 Tables clearly present all necessary methodological details, including frequency band integration times and model requirements.

Given the groundbreaking nature of the results and the high overall quality, the manuscript should be accepted subject to minor revisions.

We thank the reviewer for their very positive and constructive assessment of the manuscript.

The "minor revisions" category is suggested to meet the need for a clearer explanation of the current validation strategy. Suggested minor revisions focus on enhancing the discussion (Section 3.3.1, paragraph 5) and the conclusions by explicitly stating that while the framework works, the performance metrics may represent the upper bound of expected accuracy due to using the same short-duration deployment for both training and validation.

Following the reviewer's suggestion, we have revised Section 3.3.1 (paragraph 5) and the Conclusions to explicitly acknowledge this limitation and to clarify that the reported metrics likely represent an upper bound:

Section 3.3.1 – added sentence: "Taken together, these factors imply that the reported performance metrics likely represent an upper bound of the framework's accuracy for long-duration or multi-region deployments."

Conclusions – added sentence: "Nevertheless, our results stem from a single short-duration deployment. Broader validation across regions, seasons, and acoustic environments is needed, and performance estimates likely represent an upper bound. Recent benchmarking efforts (e.g., Gros-Martial et al., 2025) already demonstrate the value of assembling multi-site acoustic–meteorological datasets and highlight the challenges of model transferability across diverse soundscapes."

Expand the need for future work to validate the model's generalizability using spatially or temporally distinct training-validation splits to confirm the framework's robustness for global, remote deployment.

L691-693 (Conclusions) – added sentence: "Future missions should employ independent training–validation–test partitions to rigorously evaluate generalizability, following best practices established in recent WOTAN studies that explicitly address temporal correlation and multi-site validation requirements (e.g., Cauchy et al., 2018; Taylor et al., 2020; Trucco et al., 2022; Trucco et al., 2023)."