Evaluating effects of the terrain on modelled winds in multiple atmospheric model datasets

Reviewer's comments

1 General comment

The manuscript by Pogumirskis et al. presents a principal component analysis (PCA) of wind speed and wind direction biases in numerical atmospheric model outputs at wind turbine hub heights. The analysis is based on a large-scale comparison with more than 500 observational campaigns across Europe, with the aim of identifying coherent spatial, diurnal, and seasonal patterns of model bias and relating them to discrepancies in terrain height representation.

The topic is within the scope of Wind Energy Science and is of broad international interest. The dataset is substantial, and the methodological approach is generally sound. The manuscript is clearly structured and addresses a relevant scientific question. However, the research question could be formulated more precisely. At present, the title appears much broader than the actual scope of the analysis, which focuses predominantly on the PCA-based identification of bias patterns. A clearer and more focused statement of the research objective would improve the overall coherence of the paper. In its current form, the manuscript has several substantial shortcomings that, in my opinion, warrant major revision.

1. The manuscript states that it performs a "large-scale validation" of modelled winds. However, the term validation may be too strong in this context. The locations of the masts and lidars cannot be disclosed, and no information is provided regarding the processing, filtering, or quality control of the measurement data. While this may be due to commercial constraints and may not be entirely within the scope of the manuscript, the absence of this information limits the transparency and reproducibility of the analysis. I therefore suggest reconsidering the terminology (e.g., comparison or large-scale bias assessment rather than validation), unless additional methodological details can be provided.

2. The manuscript identifies elevation differences between model terrain and actual terrain as a major source of bias, which is a reasonable conclusion. However, the underlying causes of these differences are not sufficiently examined. In mesoscale models, terrain is routinely smoothed because native elevation data are downsampled to ensure numerical stability, and commonly used digital elevation models (e.g., SRTM or Copernicus DEM/TanDEM-X) have vertical uncertainties on the order of tens of meters. Terrain smoothing and slope representation have been recognized since the 1980s as key sources of uncertainty in numerical weather prediction and wind resource assessment (e.g. Mahrer, 1984). Although the manuscript clearly demonstrates a relationship between elevation differences and wind speed

bias, it remains unclear what novel practical guidance is provided beyond this established understanding.

3. The manuscript states on multiple occasions that models underestimate wind speed in complex orography. However, this statement is contradicted by several studies that are not cited (Jiménez et al., 2011; Gómez-Navarro et al., 2015; Solbakken et al., 2021; Cheynet et al., 2025). I assume this may be an oversight, but it should be corrected. A more accurate statement is that both underestimation and overestimation are reported in the literature, depending on location, flow conditions, and model resolution. A clear example is provided by Solbakken et al. (2021), where the sign and magnitude of the bias depends on the model and location. This issue indicates that the literature review is incomplete in its current form. A comprehensive and balanced review of prior findings is a fundamental component of academic work and is necessary to properly position the present study.

4. I believe the discussion section should be substantially revised in this regard. In its current form, it reads more like a narrative review of known issues than a critical reflection on the specific findings of the present study. The discussion would benefit from a stronger focus on interpreting the results, clarifying what is genuinely new, and explaining how the findings advance, or refine existing understanding. At present, the connection between the presented results and their broader implications remains insufficiently developed.

Despite the substantial number of comments provided below, I would like to emphasise that the manuscript is based on an impressive observational dataset and addresses a highly relevant question for wind energy applications onshore and offshore. The large number of sites and the attempt to extract coherent spatial and temporal bias structures across multiple atmospheric models represent a considerable effort. With clearer positioning in the literature, more explicit discussion of methodological choices, and a strengthened critical reflection in the discussion and conclusions, this work has the potential to make a valuable contribution to the field. I provide additional specific comments below that I hope will assist the authors in improving the manuscript.

2 Specific comments

Point 1

Introduction, paragraph 1: The 37 GW value refers to the estimated annual installation rate required to meet EU climate targets, rather than a target in itself. This distinction should be clarified. In addition, the term "total cumulative capacity" (or "cumulative installed capacity") would be more precise than "total capacity" in the sentence: "To reach the goal, wind turbines with a total capacity of at least 37 GW have to be installed each year (European Commission, 2023)."

Point 2

Introduction, paragraph 2: The statement that wind farms aim to maximize profits appears overly simplified. In practice, developers often prioritize stable and predictable revenue streams, particularly in offshore wind projects. This issue is currently discussed in industry reports, including recent publications by the Global Wind Energy Council (GWEC). The authors may wish to nuance this statement and provide an appropriate reference.

Point 3

Introduction, lines 26–34: The statements in this paragraph require supporting references; in their current form, they read as assertions rather than established facts. Claims related to wind resource assessment should be supported by relevant guidelines (e.g., IEC standards), and statements concerning wind speed measurements at 100 m or higher could be referenced to representative studies such as Knoop et al. (2021), Ramon et al. (2020), and Cheynet et al. (2025). Adding these references would strengthen the scientific basis of the introduction and clearly distinguish established knowledge from interpretation.

Point 4

Introduction: The distinction between synoptic, mesoscale, and microscale processes should be introduced more systematically and earlier in the manuscript. While synoptic and mesoscale models are briefly described, the microscale is not defined, despite the fact that terrain representation and elevation differences are central to the analysis. Since terrain-induced flow features (e.g., slope flows, recirculation, wake effects) occur at microscale and are closely linked to how terrain is numerically represented, this scale separation is conceptually important. Mesoscale models typically rely on terrain-following coordinate systems, whereas microscale models often use unstructured meshes or immersed boundary methods that better resolve steep slopes and complex topography. Given that terrain representation is a core theme of the manuscript, a clearer review of modelling scales and their implications (with appropriate reference) for wind bias would strengthen the theoretical framing of the study.

Point 5

Introduction, lines 52–58: I understand that the exact mast and/or lidar locations cannot be disclosed due to commercial restrictions. However, the current presentation of country-aggregated locations (Fig. 1) is overly coarse and limits the reader's ability to assess spatial representativeness. Do you think it would be possible to display approximate locations with sufficiently large markers or spatial aggregation (e.g., within ±100 km) that preserve confidentiality? Improving the spatial transparency of the dataset would strengthen the interpretability of the large-scale analysis.

Point 6

Line 62: The statement that wind speed bias generally shows a negative correlation with elevation appears overly general. As noted in my major feedback, both underestimation and overestimation have been reported in complex terrain, depending on model resolution and flow conditions. Moreover, the assertion that the physical causes of the bias are "still unclear" is too strong. Several mechanisms contributing to bias in mountainous regions are well documented, including ridge-top acceleration, flow separation, sub-grid orographic drag parameterizations, and terrain smoothing effects. The sentence, and possibly the surrounding paragraph, should therefore be revised to provide a more balanced and physically grounded discussion.

Point 7

Introduction, final paragraph: The last paragraph does not clearly outline the structure and organisation of the manuscript. Instead, it reiterates a general statement about terrain representation. It would improve clarity if the authors briefly summarised the main components of the study (datasets, bias analysis, PCA approach, and evaluation strategy) and indicated how the paper is structured. This would provide the reader with a clearer roadmap of the work.

Point 8

Section 2.1: The distribution of measurement heights is described numerically, but a graphical representation (e.g., a histogram) would be more informative. Given that vertical representativeness is central to hub-height validation, a histogram would allow the reader to immediately assess how strongly the dataset is weighted toward sub-100 m observations and how many sites represent modern turbine hub heights.

Point 9

Section 2.1: The manuscript states that "Wind speed was interpolated using the power law, i.e., assuming that the logarithm of wind speed is a linear function of the logarithm of the height." This formulation may cause confusion. The power-law profile and the logarithmic wind profile are distinct approaches. It would be helpful to clarify whether a fixed exponent was assumed, how it was determined, and to ensure consistent terminology throughout.

Point 10

Section 2.1: The description of the vertical procedure is conceptually unclear. The power-law wind profile is a parametric model, not inherently an interpolation method. Whether the procedure constitutes interpolation or regression depends on how the exponent is obtained (prescribed, derived from two levels, or fitted from multiple levels). The manuscript should clarify how the exponent is determined and use consistent terminology.

Point 11

Section 2.1: The statement that ERA5 native model-level data are "labelled for expert users" and therefore not used may be misleading. ERA5 provides geopotential, allowing pressure- or model-level data to be converted to geometric height in a straightforward manner. While this requires additional processing compared to directly using 10 m and 100 m products, it would not generally be considered "expert-level" usage. The authors may wish to rephrase this justification and clarify whether the choice was made for consistency and simplicity.

Point 12

Section 2: The manuscript provides little information about the digital elevation models (DEMs) used in the terrain representation of the atmospheric models and does not define basic terminology (e.g., DEM vs. digital terrain model, DTM). For example, when WRF is used, the terrain data are typically derived from global DEM sources such as the Shuttle Radar Topography Mission (SRTM). Other models or datasets may use alternative elevation products (e.g., CopernicusDEM, TanDEM-X) that differ in source, spatial resolution, and vertical characteristics. Since the choice of DEM, its vertical accuracy, and whether a bare-earth terrain model (DTM) or surface model (DSM) is used directly affect terrain representation and model bias, the manuscript should clearly state which elevation products are used for each model dataset (including source satellite mission and vertical resolution) and distinguish between terrain model types.

Point 13

Section 2.1 / Table 1: The manuscript uses both “HARMONIE-ALADIN” and “HARMONIE-AROME”. It is unclear whether this reflects two different configurations or inconsistent terminology. The authors should clarify the modelling system used and adopt consistent naming throughout the manuscript.

Point 14

Section 2.2.4: The terrain resolutions are reported using “s” or “seconds”, which may be confused with temporal units. Since these refer to angular resolution, the term “arcseconds” (or “arc-minutes” where applicable) should be used consistently to avoid ambiguity.

Point 15

Table 1: Units should be included in the column headers wherever possible (e.g. temporal resolution). Providing units directly in the header improves clarity and avoids ambiguity.

Point 16

Section 2.2.4: The conversion of angular resolution to metric units is latitude dependent. While 1 arcsecond in latitude corresponds to approximately 30 m globally, the longitudinal spacing decreases with cos(latitude). Since the WRF domain is located in Sweden, the effective zonal resolution differs substantially from the equatorial approximation. It would be clearer to either report the resolution in angular units only or clarify that the metric conversion is approximate and direction-dependent.

Point 17

Section 2.1 / Figures: Kriging is mentioned briefly as a method to present spatial patterns while masking exact locations, but it is not described further. The manuscript should clarify how the variogram was defined and whether cross-validation was performed. Since kriged fields are used to interpret spatial bias patterns, the interpolation methodology should be documented.

Point 18

Section 2.1: It is stated that exact observation locations cannot be disclosed. It would be helpful to clarify whether all locations are confidential or whether a subset could be shown at higher spatial resolution. This would improve transparency regarding the spatial representativeness of the dataset.

Point 19

Section 2.2.1: Model values are taken from the closest horizontal grid cell to the observation location without considering land properties. This corresponds to nearest-neighbour sampling. In complex terrain, such an approach may introduce additional representativeness error. Scattered linear interpolation is computationally inexpensive and would generally provide a smoother and more physically consistent estimate of model values at the observation location. The choice of nearest-neighbour sampling should therefore be justified or replaced with a more appropriate method.

Point 20

Section 2.2.1: The vertical interpolation methodology requires further clarification. The power-law profile depends on the shear exponent, yet it is not stated whether a fixed value was assumed or derived from model levels. Moreover, surface roughness length is not discussed, although it directly influences vertical shear. If higher-order polynomial fitting was applied between two levels, this would risk overfitting. The vertical matching procedure should be described more explicitly to ensure reproducibility and to assess its impact on the diagnosed bias.

Point 21

Section 2.2.2: Would it be more appropriate to describe PCA as identifying dominant "modes" or "patterns of variability" rather than "features"? The term "feature" has been widely used in machine learning, but it may be somewhat ambiguous here, as principal components represent directions of maximum variance rather than discrete physical features.

Point 22

Section 2.2.2: The description of PCA appears to conflate statistical variance decomposition with physical causal decomposition. PCA identifies orthogonal modes that explain variance in the dataset, but it does not in itself decompose the "causes" of model bias. The interpretation of principal components in terms of physical mechanisms is a subsequent step and should be described more cautiously. Similarly, the statement that the decomposition allows estimation of how model output would improve if certain issues were fixed may be overstated unless explicitly demonstrated. Clarifying this distinction would strengthen the manuscript.

Point 23

Section 2.2.2: The description of normalisation in the PCA procedure would benefit from clarification. It appears that the analysis was performed using covariance-based PCA (data centred but not standardised), and that scaling was applied only afterwards for interpretability of scores and loadings. This choice is not necessarily incorrect, but it should be justified, as using unscaled variables implies that time periods with larger variance dominate the principal components. Moreover, the statement that the magnitude of principal component values is affected by the number of dimensions is not clear to me. A clearer explanation of the scaling and its implications for interpretation would improve methodological transparency.

Point 24

Section 2.2.3: The terrain representation section requires substantial clarification and restructuring. First, the statement that wind speed bias in complex terrain is generally negative is overly general (see my general feedback above). Both underestimation and overestimation have been reported in the literature, depending on model resolution and flow regime, and this should be reflected more accurately.

Second, the physical mechanisms discussed (e.g., unresolved terrain features, orographic drag, speed-up effects) are introduced, but the manuscript then narrows the hypothesis to elevation differences without clearly distinguishing between: (i) horizontal terrain smoothing, (ii) vertical elevation mismatch, (iii) sub-grid orographic parameterisations, (iv) flow acceleration over ridges and slope flows.

These mechanisms operate differently and should not be conflated. In particular, horizontal smoothing alters terrain gradients and slope angles, thereby modifying flow acceleration and deceleration processes, whereas vertical elevation mismatch primarily shifts the reference height of the comparison. A clearer separation of these mechanisms supported by appropriate reference to the scientific literature would strengthen the physical interpretation of the results.

Point 25

Figures 4 and 11: The colourbar label may be misleading. The plotted quantity is described as "standard deviations", yet both positive and negative values are shown. A standard deviation itself is non-negative; what is displayed appears to be a standardised principal component score (i.e., the PC value divided by its standard deviation, equivalent to a z-score). It would therefore be clearer to label the colourbar as "standardised PC1 score" or "PC1 (in units of the standard deviation)" and to revise the caption accordingly to avoid confusion.

Point 26

Section 3.1: The manuscript suggests that PC1 of the wind direction bias may be caused by improper calibration of the reference wind direction or by local obstacles near the mast. However, potential measurement-related uncertainties are not discussed in Section 2, where the observation dataset and quality control procedures are described. If sensor calibration issues, mast shadowing, flow distortion, or nearby obstacles are considered plausible contributors to the diagnosed bias, the applied observation quality assurance procedures should be described more explicitly. Given the large number of observation sites (>500), it is understandable that individual manual inspection is not feasible. Nevertheless, automated quality-control measures (e.g., sector-wise filtering, detection of flow distortion, or flagging of sites with nearby obstacles) could potentially be implemented or at least discussed.

Point 27

Furthermore, the reported standard deviation of wind direction bias (8–10°) and spatial variations on the order of 5° might be comparable to typical wind vane installation and alignment uncertainties. This raises the interesting question of whether part of PC1 reflects measurement uncertainty rather than model deficiency? Explicitly discussing the expected observational accuracy would help contextualise the magnitude of the diagnosed bias and strengthen the interpretation of the results.

Point 28

Section 3.1: The manuscript states that the physical interpretation of PC1 of the wind direction bias is outside the scope of the paper. However, since PCA is the central methodological framework of the study and PC1 represents the dominant mode of variability, leaving its interpretation largely unresolved weakens the overall contribution. It would therefore be preferable to either provide a more thorough discussion of PC1 within the present manuscript or clearly justify why it cannot be meaningfully interpreted with the available data. Given that PC1 appears to explain a substantial portion of the variance, its role should be addressed more explicitly rather than deferred.

Point 29

Section 3.2, paragraph 1: The manuscript states that wind speed and direction biases are correlated with elevation differences and that the WRF sensitivity experiments provide "more robust support for the hypothesis." While the correlation is clearly demonstrated, the causal interpretation should be stated more cautiously. The WRF experiments modify terrain input resolution, but other mechanisms associated with complex terrain (e.g., slope gradients, orographic drag parameterisations) are not isolated. It would therefore be helpful to clarify what aspect of terrain representation is being tested and to avoid conflating statistical correlation with causal attribution.

Furthermore, the statement that the two WRF simulations were hypothesised to show differences "similar in nature" to those between models and observations is somewhat vague. Maybe this sentence can be reprhased more clearly?

Point 30

Lines 482–483: The sentence "Correlations between values of principal components of model differences and elevation differences can be interpreted similarly to the case when models and observations are compared" is unclear to me. Maybe can it be reformulated?

Point 31

Section 4: The discussion currently reiterates several results already presented in Section 3 rather than critically reflecting on their implications. The section "Discussion" should not serve as an extended interpretation of the results, but rather as a critical reflection on the assumptions and limitations of the present work. Reducing repetition and strengthening the critical synthesis would improve the scientific depth of the manuscript.

Point 32

Section 4: The repeated statement that wind speed underestimation in complex terrain has "unclear" causes is somewhat misleading. Several mechanisms are well documented in the literature. In particular, terrain smoothing is often required for numerical stability and accuracy in atmospheric models, regardless of whether terrain-following or other grid formulations are used. Excessively steep slopes can degrade numerical stability, introduce pressure-gradient errors, or distort flow representation, which motivates the use of terrain filtering (e.g., Gaussian smoothing; see Schmidli et al. (2018)). Thus, terrain smoothing is not merely a modelling deficiency but also a practical numerical constraint. This distinction should be discussed more explicitly when interpreting bias in complex terrain.

Point 33

Section 4: The manuscript primarily attributes bias to elevation differences between model and real-world terrain. However, horizontal terrain smoothing alters slope gradients and curvature, which affect flow acceleration, separation, and exposure independently of simple vertical elevation mismatch. These mechanisms operate differently and should be discussed separately. A clearer distinction between elevation mismatch and slope-induced flow dynamics would strengthen the physical interpretation.

Point 34

Section 4: The discussion of microscale modelling is introduced only briefly in the final paragraph. This topic deserves a more balanced and technically grounded treatment. Computational fluid dynamics (CFD) includes both RANS and LES approaches, which differ substantially in computational cost. RANS-based methods have been successfully applied in highly complex terrain (e.g., Norwegian fjords), capturing flow features unresolved by mesoscale models. Operational tools such as WindNinja also incorporate fast CFD modules (Wagenbrenner et al., 2019). WAsP has also a CFD module. Expanding this discussion would strengthen the practical implications of the study.

Point 35

Section 4: The discussion would benefit from a clearer reflection on the assumptions and limitations of the study. For example: (i) the reliance on elevation difference as a proxy for terrain representation, (ii) the neglect of roughness and land-use effects, (iii) the use of covariance-based PCA, and (iv) the assumption that local bias correction propagates non-locally.

In addition, future research directions could be broadened. Beyond physics-based microscale modelling approaches, recent advances in machine learning have shown potential for accelerating microscale flow prediction and bias correction (e.g. Dujardin and Lehning, 2022). Addressing such developments would better position the study within current research trends.

Point 36

Section 5 (Conclusions): The conclusions are largely qualitative and would benefit from a clearer quantitative synthesis of the main findings. For example, the manuscript reports correlations between principal components and elevation differences, as well as bias magnitudes expressed in $\mathrm{m\,s^{-1}}$ per 100 m elevation difference. Summarising these key quantitative results explicitly in the conclusion would strengthen the clarity of the study.

Point 37

Section 5 (Conclusions): Since complex terrain in Scandinavia is highlighted as a key region of systematic bias, the conclusions could be better contextualised within existing mesoscale and microscale modelling studies conducted in Norway and Sweden. Several studies have evaluated both mesoscale and CFD-based approaches in highly complex Nordic terrain. Positioning the present findings relative to this body of work would clarify whether the identified elevation-driven bias is consistent with, complementary to, or distinct from previously reported results.

Point 38

Section 5 (Conclusions): The conclusion could more explicitly reflect on the study's limitations and assumptions, particularly the reliance on elevation difference as a proxy for terrain representation and the inability to disentangle the contributions of slope effects, roughness, and sub-grid parameterisations. A concise statement of limitations and implications for future work would improve the balance and completeness of the section.

References

Mahrer, Y.. An improved numerical approximation of the horizontal gradients in a terrain-following coordinate system. Monthly Weather Review 1984;112(5):918–922. doi:10.1175/1520-0493(1984)112<0918:AINAOT>2.0.CO;2.

Jiménez, P.A., Dudhia, J., Navarro, J.. On the surface wind speed probability density function over complex terrain. Geophysical Research Letters 2011;38(22):L22404. doi:10.1029/2011GL049597.

Gómez-Navarro, J.J., Raible, C.C., Dierer, S.. Sensitivity of the wrf model to pbl parametrisations and nesting techniques: Evaluation of wind storms over complex terrain. Geoscientific Model Development 2015;8(10):3349–3363. doi:10.5194/gmd-8-3349-2015.

Solbakken, K., Birkelund, Y., Samuelsen, E.M.. Evaluation of surface wind using wrf in complex terrain: Atmospheric input data and grid spacing. Environmental Modelling & Software 2021;145:105182. doi:10.1016/j.envsoft.2021.105182.

Cheynet, E., Diezel, J.M., Haakenstad, H., Breivik, O., Peña, A., Reuder, J.. Tall wind profile validation of era5, nora3, and newa datasets using lidar observations. Wind Energy Science 2025;10(4):733–754. doi:10.5194/wes-10-733-2025.

Knoop, S., Bosveld, F.C., De Haij, M.J., Apituley, A.. A 2-year intercomparison of continuous-wave focusing wind lidar and tall mast wind measurements at Cabauw. Atmospheric Measurement Techniques 2021;14(3):2219–2235.

Ramon, J., Lledó, L., Pérez-Zanón, N., Soret, A., Doblas-Reyes, F.J.. The tall tower dataset: a unique initiative to boost wind energy research. Earth System Science Data 2020;12(1):429–439.

Schmidli, J., Böing, S., Fuhrer, O.. Accuracy of simulated diurnal valley winds in the Swiss Alps: Influence of grid resolution, topography filtering, and land surface datasets. Atmosphere 2018;9(5):196.

Wagenbrenner, N.S., Forthofer, J.M., Page, W.G., Butler, B.W.. Development and evaluation of a Reynolds-Averaged Navier–Stokes solver in WindNinja for operational wildland fire applications. Atmosphere 2019;10(11):672.

Dujardin, J., Lehning, M.. Wind-topo: Downscaling near-surface wind fields to high-resolution topography in highly complex terrain with deep learning. Quarterly Journal of the Royal Meteorological Society 2022;148(744):1368–1388.