REVIEWER 1

Bodini et al. provide a manuscript which describes the observed conditions for a benchmark which was performed in the context of an IEA Wind Task. Since the outcome and interpretation of the benchmark strongly depends on the understanding of the condition during the benchmark time period, this manuscript is of high relevance. The manuscript describes the wind park and wind turbine configuration and puts a focus on the meteorological conditions, respectively the flow features at the experimental site. Although the manuscript is well written in general, I think there are parts which are too vague and require some more explanation. Since the manuscript is clearly put into the context of the modeling benchmark, some more explanation should be given how the described curiosities in the flow features are incorporated in the benchmark and how data is processed and provided within the benchmark. I suggest major revisions in this direction before publication of the manuscript in WES. Below, I list general and specific comments.

We thank the Reviewer for their thorough and constructive review. We have addressed the specific comments in the rest of this document.

General comments:

- The description of the meteorological conditions is very much limited to a local description of the wind. I think it would be important for this Part 1 paper of the benchmark description to include a bigger picture synoptic scale description of the meteorological situation and put it in context with statistics at this site. Is this a "typical" condition, or a rather special one?

Regarding the lack of a synoptic-scale description: we have addressed the synoptic description in response to the later "double LLJ" comment. That revision adds an explicit description of the broad high-pressure ridge, weak pressure gradients, and absence of frontal systems, as well as mechanistic explanations for both LLJ features observed during the benchmark day. We refer the Reviewer to that revised text, which now opens Section 4.

Regarding the absence of any site climatological context to judge whether the selected day is "typical" or unusual: the benchmark day is representative of the typical summer southerly-flow regime at this site. The long-term climatology of the ARM Southern Great

Plains site (extensively described in Krishnamurthy et al., 2021) provides the relevant statistics, which are now described in Section 3:

"Based on these constraints, we selected 24 August 2023 as the benchmark case study. The selected day is representative of the summer southerly-flow regime that dominates wind plant operation at this site. Drawing on the long-term boundary layer climatology documented by Krishnamurthy et al., 2021 for the ARM SGP Central Facility, which is situated approximately 60 km north of the AWAKEN domain and shares the same regional climate regime, several characteristics of the benchmark day fall within climatological norms. First, the nocturnal low-level jet is a persistent feature at this site, with spring and summer together accounting for approximately 60% of LLJ occurrences. In summer, the LLJ nose is predominantly below 600 m a.g.l. (modal height ~300–400 m) -- precisely the geometry observed on 24 August 2023. Second, the strong diurnal stability cycle -- stable stratification at night transitioning to a well-mixed convective boundary layer during the day -- is observed over 80% of the time at this site. Third, the southerly wind direction regime that prevailed during the benchmark day is the dominant wind direction at hub height, and summer months feature predominantly southerly flow of moderate intensity consistent with Region 2 turbine operation. Finally, the interannual variability of hub-height wind speed at the SGP site is relatively low (~3.4%), confirming that conditions like those of 24 August 2023 recur reliably."

- How is the complex dataset and meteorological information finally provided for the benchmark? What were the recommendations for its use?

We agree that the data provision workflow was not sufficiently described in this Part 1 paper. We have added a paragraph at the end of Section 2 describing the packaging and phased release of data, with a forward pointer to Table 1 in the companion Part 2 paper, which details exactly which variables were released in each benchmark phase:

"The observational data and associated products described in this section were released to benchmark participants in three progressive phases. Phase 1 provided only a minimal upwind dataset (wind speed profiles at Site A1 up to 240 m a.g.l., surface-layer Obukhov length, and near-surface temperature), along with the open-source turbine models and farm layouts described in Section 2.1. In Phases 2 and 3, progressively richer inflow and wake observations were released -- including additional

upwind lidar sites, thermodynamic profiles, turbulence data, and ultimately the full King Plains SCADA dataset -- to enable iterative model improvement. The specific variables released in each phase are summarized in Table 1 of the companion Part 2 paper (Bodini et al., 2025a). Participants were provided with recommended topographic data and land-use data, but no mandatory prescription was imposed on how to process or ingest the observations, in order to faithfully replicate the range of academic and industrial modeling best practices. All benchmark input data are publicly available on Zenodo (Bodini et al., 2025b).”

- Without having participated in the benchmark, I do not completely understand the context and the relevance of the analyses of the flow features for the benchmark. For example: Was the terrain index used in the data processing and a map with "corrected" elevation values provided? Based on all the measurements, what are the boundary conditions that models used to set up their simulations? Was there a recommendation or best practice advice?

We thank the Reviewer for raising this important contextual question. To clarify: the terrain index (Δz_{rel}) is a purely diagnostic metric computed within this paper to characterize and explain the observed relationship between local topography and turbine power variability; it was not provided to benchmark participants as an input or processing product.
The boundary conditions used by each modeling team are described in detail in the companion Part 2 paper (Bodini et al., 2025a), where Section 3 documents the specific inflow data, roughness parameters, topographic inputs, and model setups adopted by each participant.
As also mentioned in the previous response, participants received recommended topographic data (USGS 1/3 arc-second DEM) and land-use data (released alongside observational inputs in all three phases; see Table 1 of Part 2), but no mandatory protocol was imposed for how to process or ingest the observations, in order to faithfully replicate the diversity of academic and industrial modeling practices. No “corrected” elevation map was provided, as the terrain index analysis is an observational diagnostic in Part 1, not a modeling input.

p.2, l.45f: instead of RANS "or" actuator disk, i suggest to say "RANS incorporating the actuator disk method" if that is what is meant here.

We have revised the sentence to read:
"Computational fluid dynamics models, particularly those based on Reynolds-averaged Navier-Stokes (RANS) equations coupled with actuator disk representations of wind turbines..."

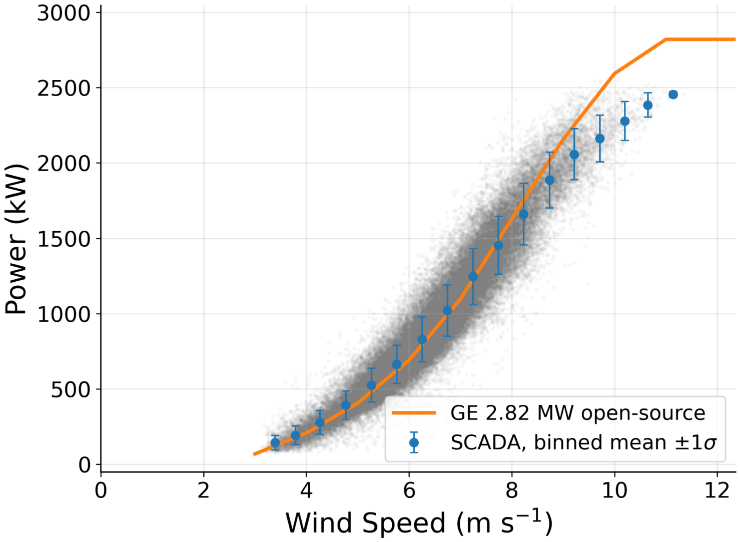
p.2, l.54: I am not convinced that WRF really bridges the gap between the above-mentioned methods. It actually serves the purpose to connect wake modeling with numerical weather prediction and thus atmospheric features, thus closing an open flank in most other models.

We agree with the Reviewer's more precise characterization. We have revised this passage to:
"Connecting wake modeling with numerical weather prediction and mesoscale atmospheric features, the Weather Research and Forecasting (WRF) model..."

p.5, l.118 and Fig.2: Is there any way to quantify how much these generated thrust and power curves differ from the actual wind turbines in the field? this could be helpful to understand uncertainties, even if it is only for a dedicated operation point which occured during the chosen case study.

We agree that a quantitative characterization is more informative. Confidentiality constraints prevent us from sharing the proprietary OEM power and thrust curves, or a formal long-term validation against manufacturer data. However, a direct comparison against SCADA-recorded power (for King Plains only) on the benchmark day is permitted.

For every King Plains turbine, we compared the 1-min SCADA power against the open-source GE 2.82 MW power curve as a function of density-normalized nacelle wind speed (WS_norm = WS (ρ/ρ_ref)^(1/3), using measured air density at site A1), after removing curtailed periods. Aggregated across the minimally-curtailed turbines (<1% curtailment), the open-source curve reproduces the production-weighted binned SCADA power with a bias of +0.1% and an RMSE of 2.2% of rated power (new Fig. 2b):



Because the benchmark day remained below rated wind speed, this spans the full Region 2 operating range that characterized the majority of the day; the largest deviations are confined to the sparsely sampled bins approaching rated. Individual turbines show a spread, which does not follow a systematic upwind/waked ordering and is therefore most consistent with per-turbine differences in the nacelle wind-speed measurement (nacelle transfer function and anemometer calibration) rather than with error in the open-source curve.

Revised text:

"We acknowledge that the open-source models are numerical approximations of the proprietary turbines installed at King Plains and Armadillo Flats, and that some discrepancy between the open-source and true performance curves is unavoidable. A formal validation against proprietary manufacturer data is not possible given confidentiality constraints. We can, however, compare the King Plains open-source power curve against SCADA-recorded power on the benchmark day. For each King Plains turbine, the 1-min SCADA power was compared against the open-source curve as a function of density-normalized nacelle wind speed (IEC 61400-12-1, with air density measured at site A1), after removing curtailed periods and binning in 0.5 m s-1 intervals. Aggregated across the minimally-curtailed (<1% curtailment) turbines, the open-source curve reproduces the production-weighted binned power with a bias of +0.1% and an RMSE of 2.2% of rated power (Fig. 2b), spanning the Region 2 operating range that characterized the majority of the benchmark day. Individual turbines deviate

more widely, with the best-agreeing turbines distributed throughout the array. We therefore attribute this scatter to differences in the local inflow sampled by each nacelle anemometer and per-turbine calibration, rather than to a systematic error in the open-source curve.”

p.7f, Tab.2&3: It is great to have such a detailed overview of instrumentation, but how was it prepared for the benchmark participants. They just choose what they find suitable to set up their models? Maybe you could separate it between input data and validation data?

The distinction between input and validation data depends on the benchmark phase: variables that served as blind model validation in one phase were released to the participants later. Rather than modifying Tables 2 and 3 -- which are designed as a comprehensive instrument inventory -- we have added (in Section 2.2) a forward reference in the body text directing the reader to Table 1 of the companion Part 2 paper (Bodini et al., 2025a), which provides a complete phase-by-phase breakdown of exactly which variables were released as inputs in each phase. This table effectively separates input from withheld validation data for each of the three benchmark phases. Participants were free to use whatever combination of the released data they found appropriate for their modeling approach; no mandatory ingestion protocol was imposed.

p.9, l.139ff: The criteria for the selection are mostly availability criteria, with the exception of wind direction. In the next section it is written that there are no "adverse synoptic conditions", but were there no criteria for a minimum wind speed, a wind speed range, or similar during the selection process already? I assume that some variety of atmospheric stability conditions was desired as well, wasn't it?

The Reviewer is correct that the four stated criteria are dominated by availability constraints. We have revised the passage to clarify the two-step nature of the selection: an initial automated filter on availability and wind direction, followed by qualitative assessment of meteorological character:

“Given the extensive duration of the AWAKEN field campaign and the large volume of data collected, an automated search algorithm was developed to identify a one-day reference case satisfying specific operational and observational criteria. In particular,

we focused on periods when: (i) more than 80% of turbines in both the King Plains and Armadillo Flats wind farms were operating normally the whole day; (ii) nacelle-mounted lidar data from the northernmost instrumented turbine in King Plains (Turbine 41) were available; (iii) ground-based lidar data from sites A1 and H were available; and (iv) the wind direction at 90 m (measured at Site A1) ranged between 135° and 225° (dominant conditions) throughout the day. Approximately 700 hours met all four criteria, but only a small number of days maintained them consistently over a complete 24-hour period -- a requirement essential for capturing the full diurnal cycle of wake behavior. From this reduced pool, candidate days were then assessed qualitatively against additional meteorological criteria: the absence of synoptic disturbances (e.g., frontal passages or convective activity), a clean diurnal stability cycle representative of the region (stably stratified conditions at night transitioning to a well-mixed, convective boundary layer during the day), and wind speeds falling predominantly within Region 2 of the turbine power curve -- in which the rotor thrust coefficient is highest and wake deficits are therefore expected to be most pronounced. Together, these qualitative criteria ensured that the selected day would serve as a canonical reference case for studying the full range of atmospheric stability effects on wind farm wakes under conditions that are both representative of the site climatology and dynamically significant for wake evolution.”

p.9, l.153: what is a "double LLJ"? Maybe rephrase to "two distinct LLJ features" or similar. Please also give a short explanation what is the forcing mechanism for those LLJ. I assume the first is a nighttime LLJ due to inertial oscillation. The second maybe similarly so, but originating from a different area and being advected over the area of interest after a shift of wind direction? Since you are presenting the meteorological conditions in this Part 1 paper, please include some synoptic and mesoscale description of the situation.

We agree that “double LLJ” is non-standard terminology. We have rephrased to refer to “two distinct elevated wind speed maxima” and expanded the meteorological description to provide both synoptic context and mechanistic explanation:

“A typical diurnal cycle is apparent, with low wind shear and enhanced turbulence during the local afternoon and a stably stratified boundary layer overnight. The synoptic setting on this day -- characterized by a broad high-pressure ridge anchored over the southern Great Plains, weak horizontal pressure gradients over northern Oklahoma,

and no proximate frontal disturbances -- was particularly conducive to the development of the Great Plains nocturnal LLJ. Two distinct elevated wind speed maxima are observed during the nocturnal and early-morning period. The first, occurring between approximately 04 UTC and 09 UTC (roughly 11 pm to 4 am local CDT), is consistent with a classic Great Plains nocturnal LLJ driven by inertial oscillation: following the collapse of the daytime convective boundary layer near sunset, surface friction decouples from the flow aloft, allowing the ageostrophic wind component to undergo an inertial oscillation about the geostrophic wind vector and accelerate to supergeostrophic speeds at low levels (Blackadar, 1957; Krishnamurthy et al., 2021). The jet nose of this feature descends to heights that partially intersect turbine rotor and hub levels and therefore has a direct potential impact on power production and wake structure, as discussed below. Comparing observations at the three lidar sites (A1, A2, and H), we find considerable spatial variability in the LLJ structure (Rai et al., 2026). The first LLJ is lowest at Site A2 (upwind and far from the turbines), slightly higher at Site A1 (upwind but adjacent to the southernmost King Plains turbines), and highest at Site H (downwind of the King Plains wind farm). The second wind speed maximum, observed between approximately 10 UTC and 15 UTC (5 am to 10 am local CDT), is located higher in the atmosphere (approximately 400-500 m a.g.l.) and is therefore unlikely to significantly affect turbine operation directly. Its origin is less straightforward: it may reflect the slow southward progression of a weak frontal boundary evident in the synoptic analyses to the north of the site, which, as the wind direction gradually veers through the morning hours, could advect a distinct elevated residual layer or a decoupled air mass from a different source region over the area of interest. Alternatively, this feature may represent the remnant of an earlier LLJ from a different fetch that has been displaced upward as the boundary layer begins to recover after dawn. Together, the two features illustrate how complex the vertical wind structure over the site can be even on a synoptically quiescent day.”

p.9, l.155: From Fig. 3 it seems that only at site A2 and in the weakening phase of the LLJ does the "nose" intersect hub height, but at A2 there are no wind turbines, so why do you claim this could effect power production?

The Reviewer raises a valid point that requires clarification. Site A2 is indeed upwind of the King Plains turbines and therefore captures the least-perturbed inflow condition. The relevance to power production is comparative, not direct: the LLJ nose being lowest

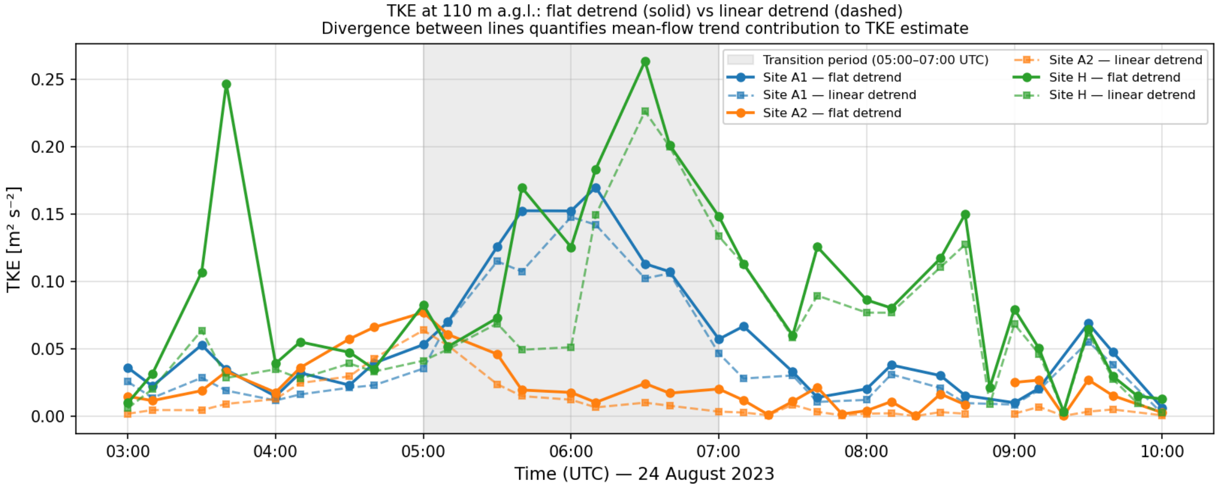
at A2 (where hub-height wind speeds are consequently highest) reflects the undisturbed inflow state, while the progressive lifting of the LLJ nose from A2 to A1 (on the southern edge of the farm) and further to H (downwind) is associated with progressively reduced hub-height wind speeds over and downstream of the King Plains wind plant. We have revised the paper text to make this comparative mechanism explicit and to clarify that the power impact is mediated by the site-to-site contrast in LLJ nose height rather than the absolute nose position at any single site:

"The first LLJ develops a nose between 140 and 300 m above ground level -- a range that engages the upper portion of the turbine rotor layer -- and its structure is initially similar across all three sites around 3-4 UTC. Pronounced site-specific differences then emerge between 5 and 7 UTC. The key pattern in this period is a progressive lifting of the LLJ nose from the undisturbed upwind reference (A2) through the farm-adjacent site (A1) to the downwind site (H). Site A2, which lies approximately 5 km south of the southernmost King Plains turbines and is unaffected by their wakes under southerly winds, exhibits the lowest jet nose during this period -- one that descends to heights partially intersecting the King Plains hub height (89 m a.g.l.) and the upper rotor layer. This alignment of the jet core with the rotor at the undisturbed inflow location represents the maximum kinetic energy available for power production. However, this resource is not fully realized at the King Plains turbines: the LLJ nose rises by roughly 100 m as the flow progresses from A2 northward over and through the wind plant to Site H. Consequently, hub-height wind speeds are highest at A2 and substantially lower at A1 and H during this period, while wind direction remains relatively constant across all three sites. The practical implication for the benchmark is direct: a model initialized solely from A2 observations that does not include the physics associated with the LLJ lifting -- implicitly treating the A2 wind profile as representative of inflow conditions at the turbines -- would substantially overestimate hub-height wind speeds and power production at King Plains during this period."

p.9, l.161 and 169: How exactly is TKE derived from the lidar? The peak in TKE is during a transitional phase. Have you checked stationarity during the averaging period for TKE estmation? I wonder if this is rather an instationarity effect than actual TKE. Also, there is a data gap in the vertical profiles exactly during that period. How does that relate?

TKE is derived following Letizia et al. (2024): for the six-beam scanning lidars, TKE is computed as half the trace of the Reynolds stress tensor obtained by deprojection of 10-minute line-of-sight velocity variances; a flat detrend (10-min mean removal) is applied, consistent with wind energy industrial standards. This formulation includes contributions from all atmospheric scales without separation of turbulent and mesoscale components.

To assess non-stationarity, we compared flat- vs linear-detrended TKE (at 110 m a.g.l.) for all 10-min windows during 03:00-10:00 UTC. We have included this comparison plot here for your evaluation, but to maintain manuscript brevity, we have opted to summarize these findings in the text rather than adding a new figure.



For Site H, the inflation fraction (flat minus linear, normalized) reaches ~60% in one transition window around 05:40 UTC, suggesting a partial mean-flow trend contribution at the onset of the TKE rise; however, at the main peak (~06:20 UTC) inflation is only ~10%, indicating the peak itself is not predominantly a trend artifact. At Site A1, inflation remains 10-30% throughout. An additional Site H peak appears earlier, around 03:40 UTC; here the flat-minus-linear inflation fraction reaches ~400%, the largest in the record, indicating this feature is predominantly a mean-flow trend artifact (consistent with a rapid wind-speed change at that time) rather than a turbulence event, and it is therefore not discussed in the manuscript.

Regarding the data gap in the vertical profiles in Figure 3: Regarding the data gap in the vertical wind speed profiles during this period (visible in Fig. 3): this gap corresponds to periods when the lidar carrier-to-noise ratio fell below the quality-control threshold of Beck and Kühn (2017), most likely due to low aerosol backscatter in the stable boundary layer associated with the LLJ, and we cannot speculate on a direct connection with the TKE peak at site H.

We have added a note to the manuscript clarifying the TKE derivation and acknowledging that the 10-min flat-detrend estimate includes mesoscale contributions, consistent with the industrial standard used.

Revised text:
"We note that the lidar-derived TKE is computed following the methodology described in Letizia et al., 2024, using a 10-minute flat-detrend window consistent with wind energy industry standards, and thus includes variability across all atmospheric scales without separating turbulent from mesoscale contributions."

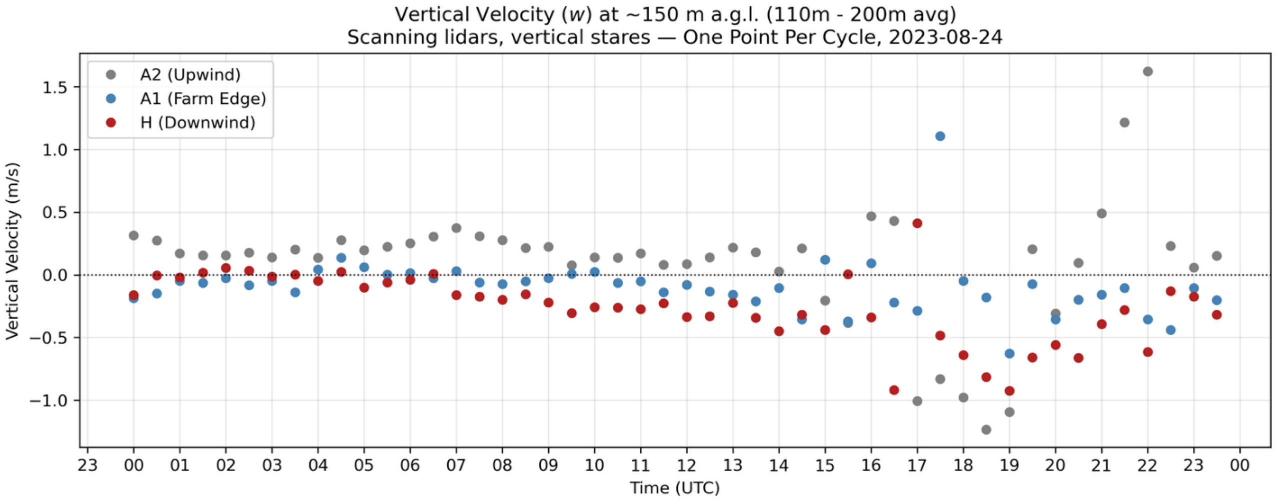
and:
"A comparison of flat- and linear-detrend TKE estimates for all 10-minute windows during this period (not shown) confirms that at the main TKE peak (~06:20 UTC) the inflation attributable to mean-flow trends is small (~10% at Site H), while a single onset window (~05:40 UTC) shows larger inflation (~60%), suggesting a partial non-stationarity contribution at the beginning of the transition."

Reference:

- ***Letizia, S., Robey, R., Bodini, N., Sanchez Gomez, M., Lundquist, J. K., Krishnamurthy, R., & Moriarty, P. J. (2024). Tilted lidar profiling: Development and testing of a novel scanning strategy for inhomogeneous flows. Journal of Renewable and Sustainable Energy, 16(4), 043310. https://doi.org/10.1063/5.0209729***

p.10, l.170ff: There is a lot of speculation about a hydraulic jump and wave breaking. If you include this in the discussion I would expect some more evidence. You may have this looking at the vertical velocity of the lidar measurements, but it is not shown or explained.

We retrieved vertical velocity (w) from the scanning lidar vertical stares at Sites A2, A1, and H for 24 August 2023, computing one mean w value per 10-minute scan cycle (repeated every 30 min) by averaging over the 110-200 m range layer. We use the vertical stare data specifically because radial velocity in this mode equals w directly, without the horizontal-homogeneity assumption required by DBS or VAD profiling/scanning lidar retrievals. The resulting time series is shown in the new figure below:



During the 03-09 UTC window coinciding with the LLJ, the observations reveal a coherent spatial gradient in w, sustained throughout the six-hour period: Site A2 (upwind reference) exhibits modest but consistent upward motion of approximately +0.1 to +0.3 m s^{-1}; Site A1 (farm edge) shows near-zero vertical motion; and Site H (downwind) shows persistent downward motion of approximately −0.1 to −0.3 m s^{-1}. This upwind-ascent/downwind-descent pattern is consistent with wave-like behavior in the stably stratified nocturnal boundary layer. Terrain-induced vertical displacement of the LLJ has been independently documented at this same site by Radünz et al. (2025) using multiscale WRF-LES simulations. We acknowledge, however, that these w observations alone cannot uniquely identify a hydraulic jump versus broader gravity wave activity, and we have revised the text accordingly.

The revised passage now reads:

"To examine this more directly, we retrieved the vertical wind speed component w from the scanning lidar vertical stares at Sites A2, A1, and H. During the 03-09 UTC LLJ period, the stare-averaged w in the 110-200 m layer reveals a coherent spatial gradient: Site A2 (upwind) shows modest upward motion of +0.1 to +0.3 m s^{-1}, Site A1 (farm edge) is near-zero, and Site H (downwind) shows persistent downward motion of −0.1 to −0.3 m s^{-1}. This upwind-ascent/downwind-descent pattern, sustained throughout the six-hour window, is consistent with wave-like behavior in the stably stratified boundary layer and supports the interpretation of terrain-induced stable-layer displacement. Terrain-induced vertical displacement of the LLJ has been documented at this same site for an independent LLJ event by Radünz et al. (2025) using multiscale WRF-LES simulations, providing an independent dynamical basis for this pattern. While a hydraulic jump -- as described by Peña and Santos (2021) for analogous terrain -- remains another plausible

mechanism, and the TKE peak at Site H around 06 UTC is consistent with wave breaking, the scanning lidar w observations alone cannot distinguish a hydraulic jump from broader gravity wave activity. Farm-induced flow deformation also cannot be excluded: the 'deep array effect' documented by Krishnamurthy et al. (2025); Rai et al. (2026) could contribute to the observed upward motion at A2."

p.12, l.197ff: Please give more details on how the EOF was implemented. The database for it is the single, but full day of 23 August?

We thank the Reviewer for pointing out the need for further methodological details regarding the EOF implementation. The Reviewer's question prompted us to critically re-evaluate our approach, resulting in a more robust and physically representative analysis.
We confirm that the EOF analysis is based exclusively on the single 24-hour period of the benchmark day, 24 August 2023. Restricting the EOF to this 24-hour window allows us to rigorously isolate the specific diurnal and mesoscale drivers acting on the wind plant during this canonical case study.
Also, we updated the methodology to resample the 1 Hz SCADA data to 1-minute averages prior to the decomposition. This mitigates high-frequency turbulent noise and autocorrelation while providing 1,440 temporal snapshots for the 88 turbines, ensuring a mathematically robust singular value decomposition.
We also transitioned to a covariance-based EOF (mean-centered, but not standardized to unit variance). This change ensures that the physical magnitude of power fluctuations -- which differs significantly between free-stream and wake-affected turbines -- is preserved in the spatial loadings. We have updated the text in Section 4 and the corresponding figures to reflect this improved methodology:

"The main patterns of spatial variability in energy production are systematically assessed by computing empirical orthogonal functions (EOFs) of turbine-level power production across the King Plains wind farm for the benchmark day (Fig. 7 top). The EOF analysis was applied to the power output from all 88 King Plains turbines recorded by the SCADA system over the full 24-hour benchmark day. To reduce high-frequency turbulent noise and focus on physically meaningful timescales, the 1 Hz SCADA data were first resampled to 1-minute averages, yielding 1,440 temporal snapshots. Each turbine time series was mean-centered by subtracting its 24-hour mean; unlike a

correlation-based approach, the time series were not normalized to unit variance. This covariance-based formulation preserves the physical magnitude of power fluctuations: turbines with higher intrinsic variance contribute to the decomposition in proportion to their actual variability, rather than being artificially equalized with chronically wake-affected or curtailed turbines whose power fluctuations are structurally suppressed. Turbines that were offline for the entire benchmark day produce a zero mean-centered time series and are therefore excluded from the variance pool, appearing as zero loadings in Fig. 7. The decomposition was performed via singular value decomposition. EOF signs are arbitrary, but relative phase differences between turbines are preserved to maintain physical interpretability."

p.13, l.209: How are the stability classes defined? Simply positive or negative Obhukov length for unstable, resp. stable ABL? Please let the readers know who may not be familiar with the concept.

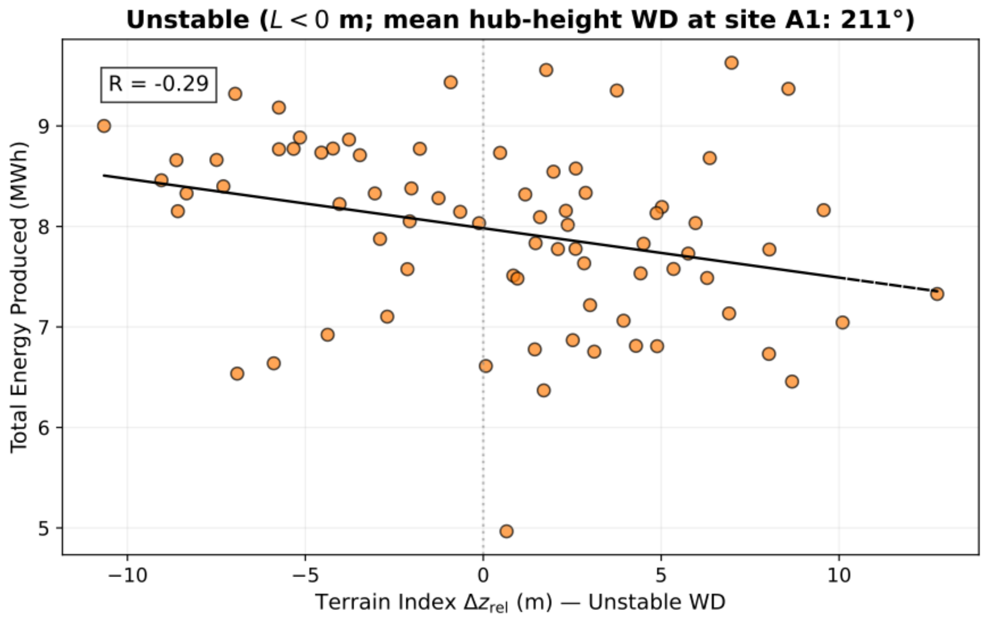
We have added an explicit definition in the text (and in the title of each panel in Figure 8):
"For every stability class (determined using the 30-min average Obukhov length, L, calculated from sonic anemometer measurements collected at Site A1; stable: L > 0 m; unstable: L < 0 m) …"

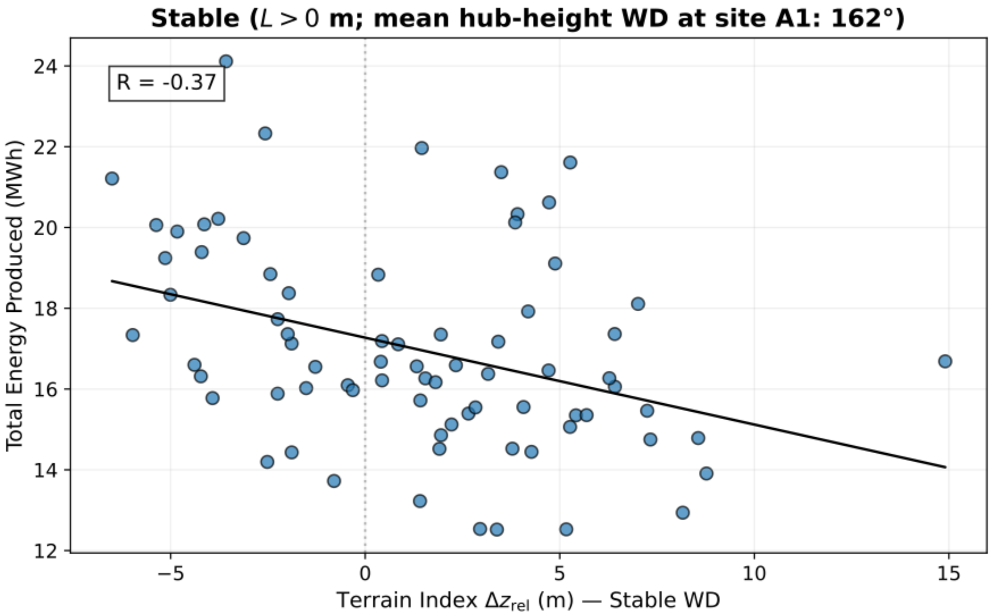
p.13, l.217f and Fig.8: the terrain index is equal to z_{rel} and presented in units of meters? Please explain and use the units in the figure.

We agree the physical meaning and units of Δz_{rel} were not stated explicitly. We have updated both the figure axis labels and the paper text:

"This Δz_{rel} value, expressed in meters of elevation difference and derived from the United States Geological Survey 1/3 arc-second (approximately 10 m horizontal resolution) Digital Elevation Model, represents the turbine's elevation relative to the mean terrain it 'sees' immediately upwind."

Revised Figure 8:





p.16, Fig10: Wouldn't it make sense to make a bias correction for the lidars to remove the discontinuity in the profile?

We appreciate this suggestion. In principle, an offset correction at the 200 m transition could reduce the visual discontinuity. However, a robust bias correction is difficult for two reasons: (1) the profiling and scanning lidars have different measurement geometries, sampling volumes, and temporal averaging periods, so a difference at 200 m may reflect genuine sampling differences (e.g., representativeness of the averaging

volume) rather than a calibration offset; and (2) the instruments do not always provide simultaneous measurements at a common height, making a statistically robust overlap comparison limited to a small number of concurrent data points. Rather than applying an offset that could introduce artifacts, we have updated the figure caption to make explicitly clear that the discontinuity is an instrumental artefact at the hand-off height.

The revised caption reads:
"Profiles are constructed using a composite of profiling lidar data for the rotor layer (z ≤ 200 m; solid lines) and scanning lidar data for the upper boundary layer (z ≥ 200 m; dashed lines); the discontinuity at 200 m reflects measurement uncertainty at the instrument hand-off height (different sampling volumes and averaging periods) and does not correspond to a physical flow feature."

p.17, l.242: I think the meteorological representativeness remains to be shown. The authors did not give any context of the site statistics.

We agree that the original text asserted representativeness without substantiation. We have now grounded that claim explicitly with the new paragraph at the end of Section 3 (see text in response to the first general comment above), and we also updated the relevant sentence in the Conclusions to read:
"...selected through an automated search process to ensure high data availability, consistent turbine operations, and meteorological conditions representative of the summer southerly-flow, nocturnal-LLJ regime that characterizes wind plant operation at this site (Krishnamurthy et al., 2021)."