Response to Referees

Manuscript wes-2026-98 — Wind Energy Science Discussions

Impact of wind farms on the occurrence and characteristics of low-level jets

Coleman Moss, Rupert Storey, Tanmoy Chatterjee and Giacomo Valerio Iungo

https://doi.org/10.5194/wes-2026-98 · Corresponding author: valerio.iungo@utdallas.edu

Letter to the Editor

Dear Dr. Arnqvist,

Thank you for handling our manuscript and for your comment dated 24 September 2026, which identifies the two issues you regard as crucial. We are grateful to both referees for the time they invested, and in particular to Referee #1, whose assessment was unsparing but, in its central diagnosis, correct.

Both referees independently concluded that the manuscript does not pose a clear scientific question and reads as a sequence of methods rather than an argument. We accept that. The revision is therefore structural rather than cosmetic: the manuscript has been rebuilt around one research question, three hypotheses derived from the physics of Great Plains low-level jets, and analyses presented as tests of those hypotheses.

On your two crucial points:

> *Physical mechanisms.* A new Introduction subsection covers LLJ formation in the Southern Great Plains, e.g., including frictional decoupling and the inertial oscillation, diurnal thermal forcing over the sloping terrain, and their combination, followed by three falsifiable hypotheses for how a wind-farm cluster could perturb them. We should say that some of this reasoning was in the submitted manuscript already, at L39–49, L447–452 and L620–626; what it lacked was any account of how LLJs form, and any statement of the hypotheses before the results that bear on them. The revision, therefore, reorganizes this material to the front rather than simply adding to it.
>
> *Radar sampling uncertainty*. We set out four independent arguments: that profiling radar is standard for LLJ work and the only instrument offering a homogeneous record across the construction era; that the Great Plains nose lies predominantly between about 300 and 800 m at this site, which our window resolves with roughly 3 to 11 gates beneath it, and that although a radiosonde resolves the same layer some six times more finely, it drifts 2–4 km during the ascent and so samples a slanted path rather than a column; that our detection rates are of the same order as independent estimates here, which we offer as context rather than as validation; and that a sampling bias constant in time largely cancels in the pre/post contrast, since the instrument, its siting and its processing are unchanged across the periods compared. We have accordingly presented our absolute rates as conditional on the sampling, while showing that the contrast we draw conclusions from is not sensitive to it.

We should be straightforward about the limits of that second answer. We have not carried out the comparison against the degraded high-resolution soundings that Referee #1 proposes, and we do not claim that our arguments are equivalent to those. They establish that our conclusion is robust to the vertical resolution; they do not establish that the absolute climatology is unbiased. We state that distinction explicitly in the revised Conclusions, identify the comparison as the natural next step, and note the form it would need to take, since a straightforward sonde-versus-radar comparison would conflate instrument differences with sampling differences.

We have restructured for readability rather than length: one table and four figures are moved to the appendices, and the case B/C and clustering-background discussions are condensed, so that the core argument runs uninterrupted by method verification. Every referee comment is answered individually in the pages that follow. Where we decline a request or cannot meet it in full, we say so and give our reasons rather than pass it over.

Yours sincerely, Giacomo Valerio Iungo, on behalf of the authors

How to read this document

Each referee comment is reproduced verbatim in grey, followed by our response and the specific change we propose to make. Comments are numbered RC1.x (Referee #1), RC2.x (Referee #2) and EC1 (Editor). Major remarks from Referee #1 are numbered RC1.M1–M4. All figure, table and line numbers in this document refer to the submitted manuscript unless stated otherwise. Because four figures moved to appendices, the main-text figures have been renumbered from Fig. 7 onward.

Referee #1 (Anonymous, 10 July 2026)

We have taken their criticism seriously and treated the review as a structural brief rather than a list of corrections. Our responses follow the order in which the comments were given: the overall assessment, the four major remarks, and then the minor remarks.

RC1.0 Overall assessment and recommendation

Referee comment.

> *This manuscript reports on a study that investigates the change of low level jet characteristics between a pre-wind farm and a wind farm era in the region. Long term radar data are used to study wind profile statistics to answer this question. I have several concerns about this manuscript in terms of clearness which research question is precisely studied, how the applied methods serve the goal, the maturity of the text and especially maturity of figure captions. The text reads as a bumpy ride where the reader is not taken by the hand to learn about the outcomes, but it feels like a chain of detours. I gave up reading after Fig 12…… (I think for the first time in 20 years of reviewing papers). Recommendation: reject.*

Response.

We thank Referee #1 for a frank assessment. The criticism that the manuscript does not lead the reader through a clearly posed question, and instead reads as a sequence of methodological detours, is fair. We have therefore treated this as a structural revision. The revised manuscript states a single testable research question at the end of the Introduction and returns to it in the Discussion and Conclusions; introduces the physical mechanisms of LLJ formation before the analysis rather than alluding to them afterward; moves method-verification material out of the main text and into appendices, so that the core argument can be read without interruption; and rewrites every figure caption to be self-contained.

We would add one observation, offered in the spirit of clarification rather than rebuttal. Several of the ideas the referee finds missing are present in the manuscript but arrive too late to help the reader: the momentum and mixing arguments at L39–49, the inertial-oscillation reasoning that motivates our choice of time windows at L447–452, and the discussion of disrupted LLJ formation at L620–626. That they were not visible to the reader before Fig. 12 is a failure of organization on our part, and the revision addresses it by moving this material forward rather than by adding to it.

RC1.M1 Major 1 — Include a researchable research question

Referee comment.

> *Include a researchable research question in the paper, that is based on a knowledge gap that is identified from the literature review in the Introduction. It is currently too vague.*

Response.

Agreed. The Introduction moves directly from motivation to the paper's structure without stating what is being asked. Knowledge gap is now stated explicitly: simulations predict that wind farms suppress LLJ occurrence and raise the nose height (Quint et al., 2025, offshore), but this prediction has never been tested observationally, and no observational test exists for onshore farm clusters.

Research question, now stated at the end of the Introduction: “Does the construction of a large cluster of onshore wind farms measurably change the rate at which low-level jets occur, or the morphology of the profiles that do occur, at a Southern Great Plains site — and can any such change be distinguished from background climatological variability?”

Three subsidiary questions follow and map onto the analysis sections: (Q1) are aggregate LLJ occurrence rates altered, and is the answer sensitive to the detection method used? (Q2) is the vertical structure of the profile altered in ways a binary LLJ flag cannot capture? (Q3) given the background climatological variability at this site, what effect size is detectable at all, and does the observed effect exceed it?

Q3 reframes what currently reads as an inconclusive result into a quantitative statement about detectability, which we believe is the most durable contribution of the work.

RC1.M2 Major 2 — Justification that the radar vertical resolution is sufficient

Referee comment.

> *The study builds on radar based wind profiles, but the paper never justifies clearly that this vertical resolution is sufficient to detect LLJs adequately, especially in relation to the 2 detection methods used.*

Response.

This is a fair criticism of the manuscript as written: we state the gate spacing and the analysis window, but never argue that they are adequate. We address it with four independent lines of argument, which we set out in a new subsection.

1. The instrument is standard for this purpose. Profiling radars have been used for LLJ studies for decades, including at this site (Poulos et al., 2002; Lundquist, 2003). Studies resolving jets below about 500 m now generally use profiling lidar, but instruments probing the full depth to 1 km and above are comparably coarse (Vanderwende et al., 2015), and no finer-resolution instrument provides a continuous record spanning the pre-construction era, which the comparison this study makes requires.

2. The jet at this site sits well inside the resolved range. The product is delivered on a fixed-height grid whose lowest gate lies near 63 m and whose spacing is approximately 62 m, so the 15 heights we retain range from 125 m to about 993 m. Whiteman et al. (1997), using enhanced rawinsondes in this same region, place the wind maximum predominantly between about 300 and 800 m AGL. Our window resolves that range with roughly 3-11 gates below the nose. The referee's concern is well-founded for shallow jets — the Cabauw nose sits at 140–260 m (Baas et al., 2009) and offshore North Sea jets are shallower still (Kalverla et al., 2017) — but the Great Plains jet is a deeper phenomenon. It is worth saying explicitly what a finer instrument would and would not give us, since the radiosonde is the natural comparison. Sondes at this site sample every 2 s and ascend at roughly 5 m/s, with a spacing of about 10 m and some six times finer than our gate: over the same 300–800 m nose layer, a sonde returns on the order of 20 to 70 samples, whereas we have 3 to 11. That advantage in vertical detail is real, but it does not make the sonde a drop-in reference

for the measurement we make. It is launched four times daily, and each ascent is instantaneous, whereas ours is an hourly consensus average on a continuous record; the balloon drifts, reaching 1 km in about 200 s, traveling some 2 to 4 km downwind under the 10–20 m/s winds typical of an LLJ. The sonde, therefore, samples a slanted path rather than a vertical column, and any horizontal gradient along that path is aliased into the apparent vertical structure. At 2–4 km, that displacement is negligible against the O(100) km extent of the jet, but it is comparable to the 5–15 km footprint of the farms whose influence we are trying to detect. The two instruments answer different questions, and neither is a reference standard for the other without careful matching in time, in space, and along the ascent.

3. Our detection rates are of the same order as independent estimates at this site — roughly 40 % (shear) and 36 % (drop-off) against the 45–47 % of soundings reported by Whiteman et al. (1997), and we offer this as context rather than as validation. The two statistics are not commensurable: theirs is a fraction of soundings taken up to eight times daily over two years, ours a fraction of hourly profiles over twenty-seven years, and because the Great Plains jet is strongly nocturnal, the two depend on quite different diurnal sampling. We are also conscious that it would be circular to argue in one breath that the sonde record is unsuited to the measurement we make and in the next to treat a sonde climatology as our validation standard. The position we actually hold is the one set out above: the two instruments answer different questions. Agreement at this level is reassuring, but it is not a controlled comparison, and we rest no conclusion on it.

4. A sampling bias that is constant in time largely cancels in the pre/post contrast. The instrument, its siting, its scan strategy, and the consensus-averaging procedure are unchanged across the periods compared, so whatever fraction of jets the 62 m sampling fails to resolve, it fails to resolve in much the same way before and after construction. What we report is a difference between periods, and a time-invariant bias enters both terms equally.

This last argument is supported by the threshold sensitivity reported in Sect. 3: halving the shear threshold raises the absolute detection rate by about 25 %, while varying it between −50 % and +50 % moves the pre/post difference by about 0.5 %. It demonstrates that the conclusion is insensitive to where the decision boundary is placed, at the resolution we have. Reading it as a statement about resolution requires the further assumption that coarse sampling acts as an approximately uniform attenuation of measured shear, so that under-resolution is equivalent to a stricter threshold, since smoothing attenuates sharp noses more than broad ones. We therefore present the sweep as a robustness result for the detection method rather than as a direct test of vertical resolution. We note in passing that the shear criterion is the one at risk, being a finite difference between adjacent gates, whereas the drop-off criterion compares two well-separated levels and is correspondingly less sensitive to gate spacing.
We are explicit about what this does and does not establish. It is an indirect argument: threshold variation emulates shear attenuation to first order but does not reproduce the vertical smoothing of the nose itself, and it says nothing about jets below our 125 m floor. A direct comparison with degraded high-resolution soundings, as the referee proposes at Ln 122, would be more conclusive, and we have identified this as the natural next step in the Conclusions. What we can state on the present evidence is that our absolute occurrence rates should be read as conditional on 62 m sampling over 125 m–1 km, while the pre/post contrast that the paper actually concludes from is not sensitive to that sampling.

RC1.M3 Major 3 — The paper is too long, with too many figures

Referee comment.

> *The paper is too long, with too many figures for the message that it conveys (of what I understood of it).*

Response.

Agreed. Reductions made, several of which the referee requests individually below: remove the case-transition analysis (Table 2) entirely; move the detection-method agreement analysis (Fig. 7 and its metrics) to an appendix, retaining one sentence in the main text; move the dendrogram (Fig. 10) and the relabelling confusion matrix (Fig. 11) to Appendix B alongside the clustering algorithm; move the background-condition histograms (Fig. 16) to an appendix; shorten the discussion of cases B and C, and condense the Introduction road map from eighteen lines to three.

We have treated the underlying problem as one of readability rather than of page count, and acted on it structurally. One table and four figures leave the main text for appendices, and the case B and C discussion and the clustering background in Sect. 4 are condensed. The effect is that a reader of the main text is no longer required to work through method verification in order to reach the result: the core argument (i.e., research question, hypotheses, detection results, cluster results, interpretation), now runs without that interruption, while the verification material remains available in full to anyone wishing to reproduce the procedure.

We should be equally straightforward about the page count: the revised manuscript is longer overall. What has been removed is outweighed by what the two Reviewers asked us to add: the physical mechanisms of LLJ formation and the hypotheses derived from them (RC1.M4, RC2.0), the justification of the vertical sampling (RC1.M2), the research question and its framing (RC1.M1), and the self-contained figure captions requested in several comments, which are necessarily longer than the cross-referring ones they replace.

RC1.M4 Major 4 — Link atmospheric physics to the question

Referee comment.

> *The paper does not try to link atmospheric physics to the question whether LLJ occurrence and wind profiles change after windfarm installation. E.g. wind farms take out momentum but create turbulence downwind, which means I expect less sharp wind gradients near the surface. Also I expect a later (in the course of the night) decoupling of the surface from the atmosphere. This type or reasoning is not at all present in the paper, but would generate a lot of insight. Anyway the paper does not discuss any physical mechanism of LLJ occurrence, like inertial oscillation or contrasting surface heating along the slope of the Great Plains.*

Response.

We accept the substance of this criticism and have used it to restructure the paper. We should, however, note for the record that the reasoning is not entirely absent: momentum extraction and wake mixing are discussed at L39–49; the inertial oscillation is what motivates our choice of evening and night windows at L447–452; and the possibility that farms retard surface decoupling and thereby disrupt jet formation is argued at L620–626 on the basis of a stronger night-time than evening effect. We recognize that material arriving on page 21 cannot do the work the referee needs it to do on page 2, and that is the real problem.

What is genuinely missing is any account of how LLJs form. The revised Introduction adds a subsection covering the frictional decoupling and inertial oscillation mechanism (Blackadar, 1957; Van de Wiel et al., 2010; Baas et al., 2012) and diurnal thermal forcing over the sloping terrain of the Great Plains (Holton, 1967), together with their combination and the role of broad baroclinic zones (Shapiro et al., 2022). Most

of these references are already in our bibliography, but were never used in the Introduction, which is an oversight.

A second new subsection states, before any result is shown, three hypotheses for how a farm cluster could perturb these mechanisms — adopting the referee's own reasoning, which sharpens them considerably: (H1) wake turbulence enhances mixing, weakens near-surface stratification and so delays or weakens decoupling and the resulting inertial oscillation; (H2) momentum extraction reduces the ageostrophic departure available to oscillate, weakening the jet; (H3) enhanced mixing and effective roughness reduce near-ground shear directly. The existing results at L620–626 then read as evidence bearing on H1, and the near-ground shear reduction as evidence bearing on H3, rather than as observations rationalized after the fact.

We do not overstate what this buys us. The farms lie 17–67 km away, so these are aggregate, farm-cluster-scale pathways rather than wake impingement; the shear reduction appears in every sector examined, including the sector with no upstream farms, so it cannot be claimed as confirmation of H3 without qualification; and our statistical power is not sufficient to discriminate among the three. Our contribution is to establish whether there is observational evidence of such a mechanism at all, and to bound the size of the effect that would be required for this data set to detect it. We have said so explicitly in a scope statement, rather than leaving the reader to expect a mechanistic study we have not done.

RC1.1 Ln 1 — “super-geostrophic”

Referee comment.

> *Ln 1: is an occurrence of super-geostrophic wind speeds. This is only the case if it is an inertial oscillation, but there are other mechanisms to create wind maxima.*

Response.

Correct. Super-geostrophic speed is a consequence of the inertial-oscillation mechanism, not a defining property of a low-level jet; thermally and terrain-forced maxima need not be super-geostrophic. The same wording appears in the abstract and at L25–26. Both are reworded to “a local maximum in the vertical profile of wind speed in the lower atmosphere, typically below 1 km”, with the mechanism-dependence of super-geostrophy then developed in the new Sect. 1.2.

RC1.2 Figure 1 — panel labels

Referee comment.

> *Figure 1: label the panels a,b,c, and refer to them in the fig captions.*

Response.

Agreed. The three panels are now labelled (a), (b) and (c) and referred to individually in the caption, and we have applied the same convention to every multi-panel figure.

RC1.3 Ln 47 — Fig. 1 shows a hypothesis, not an interaction

Referee comment.

> *Ln 47: Figure 1 illustrates schematically how this interaction might look. I must say I do not see any interaction in the graph, I only see a hypothesis how profiles and histograms look like before and after wind farm implementation. It would strengthen the paper if more of the mechanisms are being discussed. E.g. till this point in the paper I did not read about atmospheric stability. Wind turbines create turbulence in their wakes which will mix the atmospheric profiles and thus make the flow less stably stratified, which should reduce or delay the de-coupling between surface and atmosphere.*

Response.

We accept the terminological point: the figure shows a hypothesized outcome, and calling it an interaction overstates it. The text and caption have been changed to “hypothesized modification”.

The center panel is annotated “turbulent mixing and momentum transport”, and these mechanisms are introduced immediately before the figure at L39–49, so the mechanism is indicated rather than absent. But the referee's underlying point stands: atmospheric stability is nowhere mentioned before this figure, and the connection from mixing to weakened stratification to delayed decoupling is not made. The new Sect. 1.2 now precedes the figure and makes exactly that chain of reasoning, following the referee's own formulation.

RC1.4 Ln 59 — turbine shutdowns under strong shear

Referee comment.

> *Ln 59: is also considered how often turbines need to be switched off between cases, i.e. when LLJ and wind shear below the LLJ become too large/strong?*

Response.

We have no operational data for the neighboring farms, so we cannot quantify shutdown frequency, and we would rather decline than speculate. A sentence has been added to the Introduction noting that strong rotor-layer shear under LLJ conditions drives fatigue loading and can trigger curtailment (Paulsen et al., 2025; Chatterjee et al., 2025), and a sentence to the Discussion noting that the reduction in near-rotor shear we observe would, if real, tend to reduce rather than increase such events.

RC1.5 Ln 69 — “local maximum” in the vertical versus the horizontal

Referee comment.

> *Ln 69: the LLJ is a local maximum wind speed. This needs to be reworded. If you mean locally in the vertical, then yes I agree, but horizontally I would disagree. Decoupling of the surface from the atmosphere occurs over large areas, and these areas are also advected by the mean wind. So you cannot think in vertical columns only. Please comment.*

Response.

Agreed, and the definition has been reworded to “a local maximum in the vertical profile of wind speed”. The wider point deserves more than a wording fix, because it bears on the experimental design and, we think, supports our interpretation. The jet forms over an area far larger than any individual wind farm and

is advected past the measurement column, so a profiling radar samples an Eulerian slice through a structure whose horizontal scale greatly exceeds the perturbation we are looking for.

Two consequences follow, and we have added both. First, treating the problem in vertical columns is defensible precisely because the decoupling region is large compared with the farm footprint. Second, this is a plausible reason why our no-farm control sector shows near-ground shear changes similar to the farm-affected sectors: if the modification acts on a spatially extensive, advected structure, then wind-direction sectors are weaker controls than one would hope. The revised Discussion offers this as an interpretation of a result that currently reads as an unexplained inconsistency.

RC1.6 Ln 68–80 — detection paragraph unclear; missing references

Referee comment.

> *Ln 68-80: this paragraph needs to be tidied, since it is very unclear. I feel you try to explain how jet detection may occur, but at the same time basic references to works that explain and illustrate this detection as Baas et al. and Kalverla et al. and others are missing.*

Response.

Agreed. The paragraph accumulates definitional features — nose, nose height, drop-off, intensity, shear — without an organizing logic. It has been restructured around the two families of criteria the paper actually uses, amplitude-based and gradient-based, and cut by roughly half. Baas et al. (2009) and Kalverla et al. (2017) have been added and cited here; the first is a standard long-term observational climatology with an explicit detection criterion, the second a wind-energy LLJ detection study with the associated height caveats. Both are also relevant to the height discussion at L139–140.

RC1.7 Ln 92 — “To achieve these goals”

Referee comment.

> *Ln 92: To achieve these goals. I did not read anything about goals or objectives before. Reword…*

Response.

Correct — no goals had been stated at that point. This is resolved by the research question and objectives Q1–Q3, added under RC1.M1, which now immediately precede this sentence; the sentence now explicitly refers back to them.

RC1.8 Ln 92–109 — the road map is too long

Referee comment.

> *Ln 92-109: this piece can be drastically shortened to 3-4 sentences. It is a TOC that is too long. Better use the word count for formulating a proper and researchable research question.*

Response.

Agreed, including on how to reallocate the space. The eighteen-line road map is cut to three sentences, and the recovered length is used for the research question (RC1.M1) and the mechanism subsection (RC1.M4).

RC1.9 Ln 116 — suitability of radar; scattering mechanism; comparison with soundings and lidar

Referee comment.

> *Ln 116: it is not clear here how a radar is a suitable instrument to detect wind speed profiles, and specifically wind maxima as LLJs. I would say a radar needs larger particles in the atmosphere to detect displacements than that are available in the atmosphere. The authors should also make clear why a radar is better for this purpose than a radio sounding or a wind profiler.*

Response.

We should have explained the measurement principle and have done so. It may help to note that the instrument is not a precipitation radar and is, in fact, the same class of instrument as the "wind profiler" the referee contrasts it with; some of the ambiguity is ours, since the manuscript calls it a "profiling radar" throughout. At 915 MHz, the dominant clear-air return is Bragg scattering from refractive-index irregularities at half the radar wavelength (≈16 cm), produced by turbulent fluctuations in humidity and temperature. Hydrometeors are not required; when present, they contribute to Rayleigh scattering, which is one reason the ARM consensus-averaging procedure is applied. Radial velocities from the five-beam scan are combined into horizontal wind components. Two or three sentences to this effect have been added with an instrument reference, and the term "radar wind profiler (RWP)" has been used consistently.

On instrument choice: radiosondes at this site are launched four times daily, which is too sparse for hourly, seasonally resolved statistics over a 27-year record; Doppler lidar covers only the recent part of the record and thus cannot span the pre-construction era required for the comparison. Profiling radar is the only instrument here that provides a continuous, homogeneous hourly record spanning the construction period, and it has a long history of use in LLJ studies (Poulos et al., 2002; Lundquist, 2003).

RC1.10 Ln 122/124 — vertical resolution too coarse; comparison against degraded high-resolution data

Referee comment.

> *Ln 122: one should indicate what is the vertical resolution of the radar output? Is it fine enough to detect wind maxima? Ln 124 says it is 60 m vertical spacing. In my view that is way too coarse to detect LLJs. I would like to see a comparison of LLJ detection between high-resolution tower data or radio sounding data on the one hand and the radar at 60 m on the other. I expect that when the first two methods are degraded to 60 m, then the LLJ climatology will alter to less LLJs, more vertically diluted LLJs and weaker LLJs.*

Response.

We have not carried out the degradation comparison, and we would rather say so plainly than imply otherwise. Our response to RC1.M2 sets out the four arguments we offer in its place. The most directly

relevant here is the fourth — that a sampling bias constant in time largely cancels in the pre/post contrast — together with the threshold sensitivity we report in support of it: halving the shear threshold raises the absolute detection rate by about 25 %, while varying it between −50 % and +50 % moves the pre/post difference by about 0.5 %. We are explicit under RC1.M2 about why that sweep supports the argument without establishing it.

We also think the referee's prediction is probably right in direction, and we have said so rather than contest it: degrading high-resolution profiles to the 62 m grid would likely yield somewhat fewer, vertically smoother and weaker jets. What that would change is our absolute occurrence rates, which we have accordingly presented as conditional on the sampling. What would be unlikely to change is the pre-/post-contrast, because a sampling bias that is constant across the record largely cancels out a difference, and the instrument, its siting, and its processing are unchanged across the periods compared.

We propose the comparison as future work in the Conclusions, and we note that it is not entirely decisive even in principle: radiosondes drift horizontally through the jet layer; they are instantaneous, whereas the radar is an hourly consensus average; and a sonde-versus-radar comparison conflates instrument differences with sampling differences. Degrading soundings against themselves avoids the last of these, and that is the form we would recommend.

RC1.11 Ln 122 — measurement accuracy of speed, shear and nose

Referee comment.

> *Ln 122, same section please add what is the accuracy of the wind speed profile and its shear as detected with the radar, and for the jet nose in particular.*

Response.

Agreed; no uncertainty figures appear anywhere in the manuscript, which is an omission given that both detection methods apply fixed numerical thresholds. We have added the nominal instrument uncertainties from the ARM instrument documentation and derived the implied uncertainties for the quantities the paper actually uses from them. The nominal wind-speed uncertainty of the consensus-averaged product is 1 m/s, with the achieved accuracy depending on the signal-to-noise ratio and atmospheric conditions. Two consequences follow. The uncertainty in nose height is bounded below by the gate spacing, that is, by approximately 62 m. The uncertainty in shear requires more care, and we state the awkward number rather than leave it to be inferred: propagating 1 m/s through a difference between two adjacent gates gives $\sqrt{2}$ m/s over 62 m, or 0.023 s^{-1}, which is more than twice the 0.01 s^{-1} detection threshold. Taken at face value, this would place the shear criterion below its own noise floor. Two considerations reconcile it. First, the quoted figure is a per-sample uncertainty, whereas each profile is an hourly consensus average over many independent samples, thereby substantially reducing the random component. Second, and more importantly, the criterion is not evaluated on a single gate pair but across the layer between the lowest valid gate and the nose, which for the nose heights characteristic of this site spans some 300 m; over that depth, a shear of 0.01 s^{-1} corresponds to a speed difference of about 3 m/s, comfortably above the noise. We state the corollary too: for jets with noses close to the lower limit of our window, the available depth is smaller and the margin narrows, so detections in that regime carry more uncertainty than the aggregate figures suggest. Nose-height uncertainty is bounded below by the gate spacing itself, which is the more important number for this referee's concern, and we now state it directly.

RC1.12 Ln 130 — justification for 1-hour profiles

Referee comment.

> *Ln 130: 1-hour profiles. Please justify why 1h profiles are suitable for studying LLJ dynamics. LLJs can evolve relatively quickly, i.e. they originate from a sudden decoupling of the surface from the atmosphere. The fact it happens suddenly already means short time scales are relevant.*

Response.

Three reasons are now stated in the text.

First, the constraint imposed by the record is that the legacy windconC1 product covering the entire pre-construction era is an hourly consensus average, so a homogeneous 27-year record can only be built at an hourly cadence, and we vector-average the post-2019 five-minute product back to 1 h to match. Homogeneity across the construction boundary matters more for this study than resolution within it.

Second, hourly or coarser sampling is standard in LLJ climatology: Whiteman et al. (1997) used soundings up to eight times daily, and Baas et al. (2009) half-hourly data.

Third, LLJ events at this site persist longer than the averaging window: Debnath et al. (2023) find that the great majority of events at the SGP site last at least an hour. Hourly sampling, therefore, resolves the occurrence statistics that this paper reports.

We accept the limitation the referee points out, which is real: hourly averaging cannot resolve the sharpness of the evening transition itself, so our data can characterize how often jets occur and what they look like, but not the dynamics of their onset. We have stated that the boundary is explicit rather than letting the reader assume otherwise.

RC1.13 Ln 139–140 — the 125 m lower limit excludes low jets

Referee comment.

> *Ln 139-140: We restrict the data to heights between 125m and 1 km. This is tricky in the sense that earlier studies have shown that substantial number of LLJs occur below the 125 m level (e.g. Baas et al. 2009), while Kalverla et al, (2017) remarks that "It is noteworthy that most LLJs are observed at or around 100 m altitude". It seems the current authors sample only a subset of LLJ potential. Please comment.*

Response.

We acknowledge that we sample a subset, and this has been stated as a limitation rather than left to inference. In the comparison, the two studies cited are a Dutch inland site and an offshore North Sea platform, both with shallower boundary layers and lower jets than those in the Great Plains. Whiteman et al. (1997), in this same region, place the nose predominantly between roughly 300 and 800 m AGL, so the 125 m floor excludes a considerably smaller fraction here than it would at Cabauw or IJmuiden.

We would also argue that the restriction is not purely a loss. Modulation of the jet within the rotor layer is the expected and relatively uninteresting consequence of direct momentum extraction. By confining the analysis to elevations above the rotor layer, we are testing for the subtler question the paper is actually about — whether the farms alter jet formation and morphology beyond the layer they directly act on. We

have made that argument explicitly, since at present the height restriction reads as an apology rather than a choice. We should also be candid about why the floor exists at all, which the manuscript does not address: the lowest gates of a 915 MHz profiler are subject to ground clutter and near-field contamination, so it is a data-quality limit rather than an analysis preference.

RC1.14 Ln 142 — the PCA appears out of the blue

Referee comment.

> *Ln 142: the PCA appears completely out of the blue here. The reader is not guided why this analysis is performed, neither what is the precise purpose of it. Need to be revised.*

Response.

Agreed. The PCA serves one narrow purpose, i.e., identifying residual outliers that survive the ARM quality control; however, the text introduces the machinery before the motivation. A motivating sentence has been added before it, and the main text now states plainly what Appendix A currently says only in passing: the retained components are used to flag deviant gates, and the original rather than the reconstructed values are kept, so the procedure rejects data but does not smooth it.

RC1.15 Figure 5 caption

Referee comment.

> *Figure 5, caption: refer to each panel in the caption. Also expand MBB, so the reader does not need to go back to the main text for clarification. Also be more precise in the caption about the precise location of these data.*

Response.

Agreed on all three. On the third: we read the referee as asking that the caption name the measurement site rather than requiring the reader to recall it from Sect. 2, and we have stated “ARM SGP Central Facility (C1), Lamont, Oklahoma” together with the time range of the data shown. We have adopted a uniform caption policy in response to this and the referee's subsequent caption comments: each caption identifies its panels individually, expands abbreviations on first use within the caption, names the site and time range, and states the period–color mapping for periods plotted.

RC1.16 Ln 222–225 — justify the threshold values for this site

Referee comment.

> *Ln 222-225: justify why the mentioned threshold values are suitable for your specific study location.*

Response.

Two justifications, neither currently in the text.

First, site-specific provenance: the drop-off and intensity thresholds follow Debnath et al. (2023), who studied the same experimental domain, and are not transplanted from a different climate. The shear criterion

follows Hallgren et al. (2023). Both are adopted unmodified and deliberately so, to keep our rates comparable with published work.

Second, we have tested the sensitivity of the shear criterion. Halving its threshold raises the absolute detection rate by about 25 %; varying it between −50 % and +50 % moves the pre-minus-post difference by about 0.5 %. We confine the test to the shear criterion because it is the one evaluated as a finite difference between adjacent gates, and so is exposed both to the threshold choice and to the vertical sampling; the drop-off criterion, which compares the nose speed with the local minimum above it, is structurally less sensitive to both. The conclusions of the paper, therefore, do not depend on the choice of threshold. We have added this result, which also answers the resolution concern under RC1.M2, and we have noted that Aird et al. (2021) — already cited — is exactly why such a check was warranted.

RC1.17 Figure 7 — purpose of the quadrant populations and MCC

Referee comment.

> *Figure 7: "Populations of the four quadrants" does not say much to the reader, who is forced to go back to the text. And then learns that the analysis is not between obs and model, for which MCC is usually used, but between two models. What is exactly the purpose of this aspect of the study?*

Response.

The purpose is narrow, and we state it in one sentence: to establish that the null result is not an artifact of the detection method. The two detectors agree over 90% of the time, so the finding of unchanged occurrence rates does not depend on which detector is used.

One clarification: both quantities compared are observations, not models — two detection criteria applied to the same measured profiles. That the referee read them as models suggests the framing is genuinely unclear, and we accept the terminological criticism about MCC, which is conventionally a skill score against truth, where here neither criterion is truth. We have presented the comparison as method agreement, retained the raw agreement fraction and quadrant populations, and dropped the MCC framing from the main text.

The figure and metrics move to an appendix; one summarizing sentence stays in Sect. 3; the caption defines the four cases without requiring the main text.

RC1.18 Table 2 — the case-transition analysis

Referee comment.

> *Table 2, and related text: it is completely unclear to me why this analysis is performed. Why do we need to know how often and when a case A transitions into case B or vv. I.e. we talk about nothing that is related to the physics, only about whether two detection methods agree or not. I do not see what we learn about this analysis in answering whether LLJ statistics change with the implementation of a wind farm or not. It is a big de-tour.*

Response.

The intent was slightly different from the referee's reading: cases B and C correspond to different nose-elevation regimes, so transitions between them and into case A were meant to test whether a jet-formation

sequence is modified across the construction periods, rather than to track detector disagreement for its own sake. That said, the referee is right about the outcome. The analysis is an indirect route to a question that the rest of the paper answers directly; it consumes space that the length criticism does not permit, and its intent evidently did not survive contact with a reader. We have removed Table 2 and its accompanying text.

RC1.19 Section 4 up to Ln 361 — wordy; motivation unclear

Referee comment.

> *Section 4 till ln 361: this section is very wordy and not necessarily clarifying. Please add more context why this analysis is needed to answer the research question. Of course you can enter all data into such an algorithm, and get output, but why would one perform such an analysis if it does not contribute to the research question. It also reconfirms that the research question in the Introduction is not well stated.*

Response.

Agreed that the section opens with the machinery and never explains why it is needed. The motivation is specific and is now stated in a short paragraph at the head of Sect. 4. A binary LLJ flag can only answer whether the number of jets changed, and it inherits whatever threshold was used to define it — the very sensitivity the referee probes at Ln 222–225. Hypothesis H3 of the new Sect. 1.3 predicts a change in profile shape that could leave the jet count untouched, and no threshold-based detector can see it. The clustering exists to test that hypothesis without committing to a definition of an LLJ, which is the point of a data-driven approach. Regarding length, the algorithmic description is pared back, with details left in Appendix B, and the dendrogram and confusion matrix are moved there as well.

RC1.20 Figure 10 — discussed in only four lines

Referee comment.

> *Figure 10: is discussed in the text in only 4 lines. So is it really necessary to be included?*

Response.

The figure earns its place for a reader unfamiliar with dendrogram-based clustering, since it shows at a glance how the nine clusters were arrived at, and we would prefer to retain it. We accept that it is method provenance rather than result, and that four lines of discussion cannot justify a main-text figure in a paper both referees find too long. We have therefore moved it to Appendix B alongside the clustering algorithm, where a reader wanting to reproduce the procedure will look for it, rather than delete it.

RC1.21 Figure 10 caption — imprecise

Referee comment.

> *Figure 10, caption: … of subclustered results. Of what? Be more precise in the formulation of the fig captions. The caption lacks the time range of the data involved, and the location where data have been taken. The difference between green and red boxes is not indicated in the caption. Also red/green choice makes the figure not very colourblind proof. Meaning of C1…C9 is not indicated in the caption either.*

Response.

Agreed on the caption: it now states the site and full-time range defines what is being sub-clustered, explains what the green and red boxes represent, and defines the cluster labels C1…C9.

On color, we accept the criticism. The green/red distinction has been described in the caption, so it does not rely on color alone, and we have noted the accessibility limitation.

One further point the comment surfaces, which is our error: the cluster labels C1…C9 collide with C1, the name of the ARM Central Facility used throughout Sect. 2. We have disambiguated by writing "cluster 1"…"cluster 9" throughout the text and reserving "site C1" for the ARM Central Facility. Two places retain the compact labels C1–C9 because the artwork and the table headers have no room for the longer form — the dendrogram of Fig. B1 and the column headings of Table 2 — and in both, the caption now states the distinction explicitly, so the label cannot be misread in either place.

RC1.22 Figure 11 — axis orientation and caption

Referee comment.

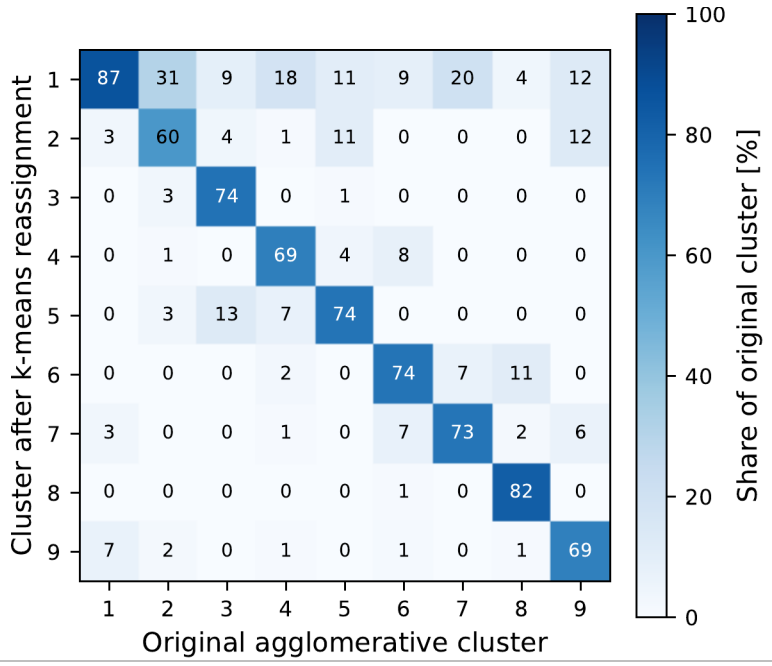
> *Figure 11: x and y axes should be swapped since the original cluster is the independent variable in this analysis and the reclustered one the dependent variable. Figure 11, caption: Confusion matrix showing initial agglomerative labels and relabelled results from the final k-means pass. The caption does not say much. I read it several times, and tried to find it in the main text what I should learn from this Fig, but it is unclear. If it is the % of data that stick to their original cluster after implementation of wind farm, then please say it more clearly and more straightforward. If this is not the message, then I do not see why this "statistics sauce" (sorry to say) is needed.*

Response.

The axes have been swapped as requested, and the cells shown as percentages rather than counts.

On the message: the figure has nothing to do with wind farms, and we understand how the caption invited that reading. It documents a methodological step, i.e., how many profiles changed cluster label during the single k-means refinement applied after agglomerative clustering, so that a reader concerned that the refinement might have reorganized the clusters can see that it did not. That is worth recording for reproducibility, but it is not a result, and we agree it does not belong in the main text. It moves to Appendix B with a caption stating exactly this in one sentence.

We have also revisited the description of the relabeling as "minor" at L366 and replaced it with the actual retention figures, which are both more informative and more honest than a qualitative adjective.



Regenerated Fig. 11 (Fig. B2 in the revision). Confusion matrix of agglomerative labels against labels after the single k-means pass, with the axes in the orientation the referee requests. Each column is expressed as a percentage of the original agglomerative cluster, so the columns sum to 100 %. Shown here for the referee's convenience; it is not a new analysis.

RC1.23 Figure 12 — legend overlaps the data

Referee comment.

> *Figure 12: In many panels text/legend is plot over the key lines, which is not a proper way of doing.*

Response.

Agreed, and Figure 12 has been redrawn. The per-period occurrence labels, which in several panels sat over the profiles, have been moved into clear space, with the position chosen panel by panel so that no label now overlies a curve. The plotted profiles themselves are unchanged. We have checked the other multi-panel figures for the same defect.

RC1.24 Figure 12 caption — colour/period mapping

Referee comment.

> *Figure 12, caption: Color indicates the period. Yes they do, but as reader I first need to go back to earlier figures to find which color relates to which period.*

Response.

Agreed. Each caption using the period colors now names the mapping explicitly rather than referring back. A consistent scheme across figures spares the reader learning it more than once, but it does not spare a reader who arrives at Fig. 12 first, or who returns to a single figure later, and the cost of repeating it in the caption is two words.

RC1.25 Closing remark — nine clusters chosen without a physical basis

Referee comment.

> *I am sorry to say but I gave up reading from here, since it is unclear what the paper heads towards, what is the precise goal, how the applied methods serve that goal, why 9 clusters have been chosen without a physical basis….*

Response.

Regarding the number of clusters, our wording invited criticism: Appendix B says the threshold was set "somewhat arbitrarily". That phrase has gone, replaced by the actual justification. The basis is statistical rather than physical, and we have said so plainly: sub-clustering proceeds only while the separation between candidate sub-cluster mean profiles is large relative to the within-cluster variance, subject to a minimum cluster population of 1 %. Nine is the outcome of that rule at the chosen threshold, not a preselected number.

The physical interpretation the referee finds missing can be supplied from results already in the paper, and we have added it: the manuscript establishes at L393–395, from the detector cross-tabulation, that clusters 1–5 are non-LLJ and clusters 6–9 are LLJ clusters, with 7–9 consistently so at over 80 %. We have characterized each retained cluster against known Great Plains profile types — nocturnal jet profiles, well-mixed daytime profiles, and strongly sheared stable profiles — and connected that to the mechanisms in the new Section. 1.2. Referee #2 makes the same point from the other direction, and our response to RC2.11 develops it.

We regret that the manuscript became hard to follow before the end, and we recognize that the restructuring described under RC1.0 and RC1.M1 is the only real answer to that.

Referee #2 (Anonymous, 12 September 2026)

We thank Referee #2 for a detailed and constructive review. Several comments coincide with Referee #1's; where that is so we cross-reference rather than repeat.

RC2.0 General comments

Referee comment.

> *Overall, the paper is well written, and it presents an interesting analysis of wind profiler data that span nearly 2 decades during which significant wind farm development happened in the study domain. However, the paper is rather long and reads more like a summary report of different analysis methods rather than a paper motivated by specific scientific questions.*

Response.

We thank Referee #2 for a detailed and constructive review, and for the positive assessment of the data set.

The central criticism coincides with Referee #1's, and we have treated it as the organizing problem of the revision: the paper is restructured around a stated research question and hypotheses derived from LLJ physics, with the analyses presented as tests of them. Please see RC1.M1, RC1.M3 and RC1.M4, which we do not repeat here.

RC2.1 Introduction — SGP LLJ climatology and formation mechanisms; motivation for including daytime

Referee comment.

> *The introduction could be improved by adding more information about the climatology and formation mechanisms about the LLJ in the Southern Great Plains (SGP). Many studies have focused on the SGP LLJ and it is well known that the SGP LLJ is primarily a nocturnal phenomenon with seasonal differences in strength and prevailing wind directions. These findings from previous studies should inform the analyses conducted by the authors. The motivation for including daytime periods into their analysis is e.g. a bit unclear to me. While daytime jets can occur the dynamics of these jets and possible interactions with mixing induced by wind farms are going to be different from the processes of nocturnal jets.*

Response.

Agreed on the climatology and mechanisms; the new Sect. 1.2 described under RC1.M4 covers both, including the nocturnal character of the SGP jet, its warm-season maximum, and prevailing southerly nose directions (Whiteman et al., 1997; Klein et al., 2016; Smith et al., 2019; Shapiro et al., 2022).

Regarding the treatment of daytime, we note that the cluster analysis is already stratified along exactly the reasoning the referee describes: at L447–452, we restrict to two windows: 18:00–24:00 local as the evening period during which inertial-oscillation jets form, and 00:00–06:00 local as the established nocturnal period. The nocturnal focus the referee asks for is therefore present in the analysis, but it is introduced in Sect. 4 as a technical choice, with no explanation of why those hours and not others, so a reader cannot see that it is physically motivated. That is a presentational failure, and we have fixed it by motivating the windows from the formation mechanisms in Sect. 1.2 and by foregrounding the nocturnal results.

Where daytime data do enter the aggregate detection-rate analysis of Sect. 3, we have reframed them as a control rather than as part of a pooled sample: a mechanism that predicts an effect at night predicts a weaker or absent effect by day, so finding no effect where none is expected constrains the false-positive rate, in the same way the sector with no upstream farms serves as a spatial control. We think this is more informative than discarding the daytime data.

RC2.2 Vertical and downstream area of turbine influence versus the measurement layer

Referee comment.

> *Similarly, the introduction could be strengthened by more clearly summarizing the results from previous studies about the vertical and downstream area of influence of wind turbines and how the selected data sets will be able to resolve possible wind turbine effects. The lowest height of the radar wind profiler data sets used ended up being 125 m with a vertical resolution of 60 m, while the typical rotor heights in the study domain are between 30 m -150 m. Given these rotor heights, over which layer would we expect to see impacts and how many data points fall into this layer?*

Response.

This is the sharpest formulation of the sampling question, and we have answered it directly in the Introduction. We have summarized the published vertical and streamwise extent of wind-farm influence from work already in our bibliography (Frandsen et al., 2006; Calaf et al., 2010; Markfort et al., 2012, 2018; Stevens et al., 2015; Segalini and Chericoni, 2021; Chatterjee et al., 2022), and state the rotor layer explicitly in terms of the tallest machines in the region, for which we have reliable figures.

The referee asks over which layer an impact should be expected, which is a question about vertical extent; the answer depends on streamwise distance, because a wind farm wake deepens with distance downstream and then decays as it mixes out, so how much of the profile can still be affected at the measurement point is set by how far away the farm is. The farms lie 17–67 km from the radar, which at these rotor diameters is of order 150 to 550 rotor diameters — far beyond the distance over which an individual wake persists as a coherent momentum deficit. By the time the flow reaches the radar, any residual wake-affected layer has collapsed back toward the rotor layer itself, so only the lowest one to three of our fifteen gates lie within it. The mechanism under test is therefore not wake impingement on the measurement column but an aggregate, farm-cluster-scale modification of boundary-layer structure, which is also what the simulations we compare against describe (Quint et al., 2025). Stating this reframes both the hypothesis and the interpretation of the null result, and we are grateful to the referee for forcing the point.

The referee's rotor heights also prompt us to re-examine a statement in our Discussion (L578) that the lowest gate lies “just above top-tip”, and we withdraw it. The tallest machines in the region are those at the King Plains farm, with a hub height of 89 m and a tip height of approximately 152 m. For those turbines, the lowest valid gate at ~125 m lies within the rotor layer, so exactly one of our fifteen gates samples the rotor layer directly; for the older and smaller machines, the top tip falls below 125 m and no gate samples it at all. We state this in place of the original claim and note that hub heights are not available for every farm in Table 1, so we report the bound set by the tallest machines rather than a farm-by-farm tabulation.

RC2.3 Label C1 on the star in Fig. 2

Referee comment.

> *It would be good to add the label C1 to the star in Fig. 2*

Response.

Agreed. The caption explicitly identifies the black star as the ARM SGP Central Facility (C1), and the label has been added to the figure itself. We are also removing the clash between the site name C1 and the cluster labels C1…C9 elsewhere in the paper (see RC1.21).

RC2.4 Figure 4 — definition of the frequency normalization

Referee comment.

> *In Fig. 4, it is unclear how the data availability in panel a) and the frequency in panels b) and c) are computed. Panel a appears to show the percentage of data available relative to the yearly record, but what exactly does the frequency in panels b) and c) show? How is N, the total sample size relative to which the frequency is calculated, defined in this case? It looks like N may be the total number of data points over each of the 3 time periods (Pre, During, Post). The value of ~ 0.075 for the blue bar in January does then indicate that 7.5% of the data from the pre-construction were observed in January, and all the blue bars add up to a value of 1 (100%)? Would it not be more informative to define N as the number of data points that should be available each month (or for each time block in panel c) for the 3 time periods such that the bar plots in panels b) and c) also show the percentage of data available for each months/time block. In any case, the caption should be expanded to add clarity and the formatting of the tick marks of the axes should be better aligned.*

Response.

The referee's reconstruction is exactly right: in panels (b) and (c), N is the total number of accepted samples within each period, so the bars for each period sum to one and show how the available data are distributed across months and hours. That this had to be reconstructed is a failure of the caption; the caption now explicitly defines N for each panel, along with the site and time range. On the referee's second suggestion, we have kept N as the number of accepted samples rather than the number that should have been available: the two normalizations answer different questions, and panel (a) already reports availability against the full record, so redefining N in panels (b) and (c) would duplicate it while removing the distribution the panels exist to show. We have also realigned the tick marks and axis formatting across the three panels, as requested.

RC2.5 Line 160 — climatological versus farm-specific splits

Referee comment.

> *The arguments around line 160 on p. 8 are a bit unclear. Is one analysis targeting the effects of multiple wind farms on average wind fields, while the ad hoc selection targets more local effects of specific wind turbines?*

Response.

That is precisely the distinction, and the referee has put it more clearly than we did. The climatological split uses fixed 2012 and 2020 boundaries to assess the aggregate effect of the whole farm cluster on site climatology; the ad hoc split places the pre/post boundary at one farm's commissioning date and restricts wind direction to the sector containing it, to isolate that farm.

We have rewritten the passage in the referee's framing and added the trade-off: the farm-specific analysis is more targeted but has smaller sample sizes and a shorter baseline, which is why both are reported.

RC2.6 Seasonal quarters should follow the LLJ climatology

Referee comment.

> *In line 165, the authors define 4 calendar quarters. Shouldn't the selection of seasonal quarters be informed by what is known about the LLJ climatology?*

Response.

We would respectfully push back on the premise, while agreeing that the manuscript failed to explain itself. The quarters are not arbitrary, and they do not cut across the seasonal climatology. Whiteman et al. (1997) distinguish a warm season of April-September from a cold season of October-March at this site. The quarters April-June and July-September together partition the warm season, and October-December and January-March together partition the cold season, so no stratum mixes the two regimes. Stratifying at quarter resolution rather than at season resolution gives finer control on seasonal composition while preserving exactly the climatological boundary the referee has in mind.

What was missing was the explanation, not the alignment, and we have added it to Sect. 2.4. We also note the purpose of the balancing, which is to prevent differences in seasonal composition between the pre-, during-, and post-construction samples from masquerading as farm effects; this requires only that the strata be applied identically to all three periods, which they are.

We are grateful to the referee for the question, since the choice was defensible but entirely unexplained in the submitted manuscript.

RC2.7 Figure 5 — bootstrapping, interpretation of the CIs, and data spread

Referee comment.

> *How would the results shown in Fig. 5 look like if the bootstrapping was not applied? Also, how exactly should the 90% MBB CIs be interpreted? The variability of the data must be much higher than the narrow CI values shown, especially in panel c). What would a plot look like that shows "error bands" defined by the standard deviation of the data?*

Response.

The referee has identified a real presentational problem, and the diagnosis is correct. The bands are 90% confidence intervals for the period mean, estimated using the moving-block bootstrap. They are narrow because the sample is large, and they carry no information about the spread of the underlying data, which

is indeed far greater. Nothing in the caption prevents a reader from taking them for data spread, and the figure is misleading to that extent.

We have done both of the things the referee asks. The caption and text now explicitly state that the bootstrap bands represent intervals around the mean, not the spread of the data, and Fig. 5 has been redrawn to show that spread as well, plotting the raw mean with one standard deviation of the hourly wind speeds alongside the bootstrap interval for the mean. Both quantities are therefore visible in the same panel, and the caption distinguishes them, so the misreading the referee identifies is no longer possible.

On the effect of the bootstrapping itself: it is used to balance seasonal composition and up-samples the three periods by 3 %, 23 %, and 12 %, respectively, so it cannot move the period means far. We have reported those figures alongside the interpretation so that the reader can judge their influence directly.

RC2.8 Detection methods with multiple maxima; length of the case B/C discussion; Figs. 8 and 9

Referee comment.

> *The discussion of the 2 traditional LLJ detection methods is interesting but I have questions about how exactly these methods handle e.g. profiles with more than one maximum etc. and I think the discussion of the results could be shortened. While it is interesting to see the profiles in Fig. 8 for the 4 different scenarios, most of the discussion that focuses on the cases when only once of the methods worked could be significantly shortened. In Fig. 8, I would again suggest adding "error bands" that show the standard deviation of the profiles at each height. Fig. 9 is quite misleading as the sample sizes for cases A, B, and C are so different. Case B/C include only ~5/2 % of the data and it's not clear to me what we can learn from comparing the distributions for these two cases during the 3 time periods (Pre, During, and Post). I would suggest focusing only on case A and show bar graphs comparing the 3 time periods for these cases. This would allow the reader to get a much better sense if the construction of wind farms had any effect. As I pointed out at the beginning, I would also find it valuable to focus on the nocturnal periods.*

Response.

Several distinct points. We accept all of them, and we have acted on most.

Multiple maxima. The manuscript does not describe this behavior, and it should. Both criteria operate on the absolute maximum of the wind speed profile within the analysis window: a profile with more than one local maximum is assigned a single nose, at the global maximum, and secondary maxima are not detected separately. We have added this statement to Sect. 3.

Length. The discussion of cases B and C has been cut substantially: the case-by-case commentary on Figs. 8 and 9 is reduced to a single paragraph each, and mentions of cases B and C in the main text fall from 33 to 14.

Figure 9. We agree the figure invites comparisons its sample sizes will not support. Rather than replot, we have restricted the main-text discussion to case A, stated the case B and C sample sizes prominently in the caption so no reader can mistake their weight, and presented the case B and C distributions explicitly as illustrative of profile morphology rather than as a basis for period comparison. (The panels of Fig. 9 are the three construction periods; the cases are distinguished by color within each panel.)

Standard-deviation bands. These have been added to Fig. 5, as described under RC2.7. We have not added them to Fig. 8; instead, we state the spread in words rather than imply a change we have not made.

Nocturnal focus. Adopted — see RC2.1, where we note that the evening and night stratification already exists in Sect. 4 and has been foregrounded and motivated from the physics.

RC2.9 Percentages in Figure 11

Referee comment.

> *Is it possible to show percentages in Fig. 11 rather than counts?*

Response.

Yes. Figure 11 shows the percentages now. Referee #1 also asks that the axes be swapped and questions the figure's purpose; both are addressed under RC1.22, and the figure is moved to Appendix B.

RC2.10 Table 3 — meaning of the values and definition of the quadrants

Referee comment.

> *It is not clear to me what the values shown in Table 3 mean. The 4 cases were defined in section 3 comparing the skill of the 2 traditional LLJ detection methods. How are the 4 quadrants defined for the cluster analysis?*

Response.

The quadrants are not redefined for the clustering. Table 3 cross-tabulates the profile clusters against the same four detection cases A–D from Sect. 3, each cell giving pre / during / post counts. Its purpose is to connect the data-driven clusters to the conventional definitions — and it is the basis on which the manuscript classifies clusters 1–5 as non-LLJ and 6–9 as LLJ clusters at L393–395.

The caption states what the cells contain but not what the table is for, and the surrounding text does not supply the purpose either, which is why it reads as an orphan. We have rewritten the caption to be self-contained on both counts and relocated the table next to the cluster classification it supports, where it does visible work.

RC2.11 Focus on the LLJ-like clusters; the two channels of farm effect; Figs. 13–15

Referee comment.

> *Fig. 12 clearly shows that only clusters 7-9, possibly cluster 6, are LLJ like profiles. Wind farm effects could show up in 2 ways: (i) a shift in the LLJ clusters that are most prominent during the 3 time periods (Pre, During, and Post) and (ii) differences in the profile shape near the ground as the clustering may focus more on the height and strength of the LLJ nose. Why does the rest of the analysis not focus on the clusters with LLJ signatures., what is the value of including all the other clusters? I find the pdf comparisons introduced after line 414 very confusing and the results shown in Figs. 13-15 difficult to interpret.*

Response.

This is a particularly constructive suggestion and adopting it improves the results section.

The referee's reading of Fig. 12 agrees with our own classification, which the manuscript states at L393–395: for the case-D populations in Table 3, clusters 1–5 are non-LLJ and 6–9 are LLJ, with 7–9 consistent at over 80%. The problem is that this appears within a paragraph about the clustering's internal consistency and is then never used to organize anything. We have promoted it to a labeled definition at the head of the results and carried it through, which we take to be what the referee is really asking for.

On the two channels: the decomposition into (i) a shift among LLJ clusters and (ii) a change in near-ground profile shape corresponds to hypotheses H2 and H3 of the new Sect. 1.3, and we have reported them separately rather than as one undifferentiated PDF difference. This states more clearly what the paper found, namely that the evidence for (ii) is stronger than for (i).

On why the non-LLJ clusters are retained: they carry the near-ground shear signal that constitutes channel (ii), so restricting to LLJ clusters would discard the evidence for the effect we actually observe. That reasoning is nowhere stated, and without it, the nine-cluster presentation looks indiscriminate. We have stated it and reported LLJ-cluster occurrence as the headline metric while retaining the full distribution as the basis for channel (ii).

On Figs. 13–15: we accept that they are hard to read, but we have concluded that relegating them would be the wrong remedy, since they carry the principal evidence for channel (ii) and the argument would then reference material the reader cannot see. We have instead made them legible. A new paragraph preceding them explains how they are constructed and how to read them: each figure is arranged in two columns and two rows, the rows being the evening-transition and established-night bins; the left-hand panels give per-cluster occurrence rates before (blue) and after (red) construction, the right-hand panels give the difference normalized by the strength of difference Δ so that positive bars sum to +1 and negative bars to -1, with markers showing the same differences computed within sub-periods as an indication of background drift, and reading a right-hand panel from left to right therefore shows which profile shapes became more or less common. The scalar Δ is quoted in the text as the summary, with the bars showing its composition.

Editor (Johan Arnqvist, 24 September 2026)

EC1 Editor's comment — Johan Arnqvist, 24 September 2026

Referee comment.

> *I have decided to close the discussion and allow for the authors to respond to the questions raised by the reviewers. Both reviewers raised major concerns relating to the lack of discussion of physical mechanisms for the LLJs and the accompanying interaction/modification of LLJs by wind turbine wakes as well as the uncertainty involved in the application of sparse resolution radar data to capture LLJ characteristics. It is naturally expected that the authors answer all questions raised by the reviewers, but I regard improvements in the above mentioned two points as crucial for a potential future acceptance of the work.*

Response.

We thank the Editor for identifying the two decisive issues so clearly, and we summarize our response to each here.

Crucial point 1 — physical mechanisms and wake modification of LLJs. The revised Introduction adds a subsection on LLJ formation in the Southern Great Plains, covering frictional decoupling and the inertial oscillation, diurnal thermal forcing over the sloping terrain, and their combination, followed by a subsection stating three falsifiable hypotheses for how a farm cluster could perturb them: delayed and weakened decoupling through wake-generated mixing (H1), reduced jet amplitude through momentum extraction (H2), and reduced near-ground shear through enhanced mixing and effective roughness (H3). Material already in the manuscript is then recast as evidence bearing on these hypotheses — the momentum and mixing arguments at L39–49, the inertial-oscillation reasoning behind our evening and night windows at L447–452, and the stronger night-time than evening effect at L620–626. We recognize that this material was placed too late to guide the reader and that its reorganization, not its addition, is what was required. Full details are in our responses to RC1.M4 and RC2.1.

Crucial point 2 — uncertainty from the radar's vertical sampling. We set out four independent arguments in a new subsection: that profiling radar is a standard and well-validated instrument for LLJ work and the only one offering a homogeneous record spanning the construction era; that the Great Plains jet nose lies predominantly between about 300 and 800 m at this site (Whiteman et al., 1997), which our window resolves with roughly 3 to 11 gates below the nose, while a radiosonde — six times finer in the vertical — drifts 2–4 km during the ascent and so samples a slanted path rather than a column, and is therefore not a reference standard for this measurement without careful matching; that our detection rates are of the same order as independent estimates at this site, offered as context rather than as validation; and that a sampling bias constant in time largely cancels in the pre/post contrast, the instrument, its siting and its processing being unchanged across the periods compared. The threshold sweep reported in Section. 3 supports this but does not establish it, and we present it as a robustness result for the detection method rather than as a direct test of resolution. We have therefore presented our absolute occurrence rates as conditional on the sampling, while showing that the pre/post contrast the paper concludes from is not sensitive to it.

We want to be straightforward about the limits of this. We have not performed the comparison against degraded high-resolution soundings that Referee #1 proposes, and we do not claim our arguments are equivalent to it. They establish that the conclusion is robust to the resolution, not that the absolute climatology is unbiased. We have stated that distinction in the Conclusions and identified the sounding

comparison as the natural next step, together with the specific form we would recommend, since a straightforward sonde-versus-radar comparison would conflate instrument and sampling effects.

Beyond these two points, the manuscript is restructured around a single research question and method-verification material is moved to appendices, so that the core of the paper reads continuously; and every referee comment is addressed individually in the pages that follow.

References added in revision

The following works are cited by the referees or required by our responses and are added to the bibliography.

Baas, P., Bosveld, F. C., Klein Baltink, H., and Holtslag, A. A. M.: A climatology of nocturnal low-level jets at Cabauw, J. Appl. Meteorol. Climatol., 48, 1627–1642, https://doi.org/10.1175/2009JAMC1965.1, 2009.

Kalverla, P. C., Steeneveld, G.-J., Ronda, R. J., and Holtslag, A. A. M.: An observational climatology of anomalous wind events at offshore meteomast IJmuiden (North Sea), J. Wind Eng. Ind. Aerodyn., 165, 86–99, 2017.