We thank both the editors and reviewers for considering this manuscript. Below, reviewer comments are in black and **our responses are in blue bold.**

Reviewer 1

This paper addresses an important and timely topic: characterizing atmospheric stability in complex terrain using Perdigao tower data. The authors evaluate stability metrics with different Reynolds decomposition intervals, and quantify the predictive skill of low level vs hub height stability. They apply clustering methods to recommend a necessary number of met masts which is good but rather academic because usually the problem is the opposite: how many more met masts than 1 are needed to properly sample the site conditions (and where numerical estimates of wind field covariance are helpful - see the other general comment).

The analysis is carefully executed, the dataset is of very high quality, and the topic is relevant to both atmospheric science and wind energy.

We thank the reviewer for their time and careful consideration of our work.

However, the paper in its current form is incomplete. Perdigao has also been extensively studied with mesoscale and large-eddy simulation (LES) modeling, yet the study relies exclusively on observational/statistical analyses, while ignoring the context and assistance that numerical models can provide. Without at least a discussion — and preferably some demonstration — **of how models complement observations**, the results risk being overly narrow and less generalizable, especially when the conclusions are drawn from data from a single site.

We thank the reviewer for this thoughtful perspective. We also welcome the opportunity to demonstrate how our proposed Louvain methodology can be extended to model data. As such, we have introduced another analysis of the Louvain community detection algorithm that relies on the publicly-available LES analysis of the Perdigão valley performed in Robey and Lundquist (2024). We hope that this established case study, as well as its corresponding analysis, demonstrate one potential application of our proposed methodology as well as motivate future research. This work can be found in the new Appendix C:

Here we demonstrate a proof-of-concept implementation of the Louvain methodology applied to large-eddy simulation (LES) data. This Louvain implementation considers near-surface static stability from the LES of Robey and Lundquist (2024) (Table C1), based on Wise et al. (2021). These data represent a roughly 3 hr period early in the morning. Although simulation output is available at 1 second resolution, here we sample every 30 min, consistent with other analyses in this work. We also spatially coarsened the 100 m output to the effective model resolution of 500 m (Skamarock,

2004) and then subset to the identified Perdigão tower locations. Static stability was then calculated from these spatially- and temporally-coarsened data as:

$$\frac{\partial \theta}{\partial z} = \frac{\theta_2 - \theta_1}{z_2 - z_1}$$

such that θ represents a potential temperature (K), z represents a model height above ground (m) and the two model levels taken are those closest to the surface (10 m and 2 m). The Louvain algorithm was then applied to these static stability calculations, resulting in seven partitions (Fig. C1). Overall, these tower groupings reflect broad terrain boundaries, with opportunities for mixing. Orange is predominately restricted to the upper valley. Dark blue exists mostly on ridges and the outer transects. The relatively small gray partition exists in the slope between the NE ridge and the valley. Light blue reflects most of the SW ridge, with occasional representation in the valley. The salmon, brown, and green transects are less geographically informed. Of course, because this analysis is based on 3 hours of LES during stably stratified conditions rather than the full measurement campaign period, the partitions are different from those presented in the main paper.

Table A1. Model parameters used in LES Louvain analysis, based on Robey and Lundquist (2024).

Model Parameter	Value
Start time	2017-06-14 03:24:00
End time	2017-06-14 05:16:25
Effective horizontal resolution	500 m
Horizontal grid resolution	100 m
Vertical grid resolution	160 vertical levels
Vertical levels	2 m and 10 m
Land cover	CORINE Land Cover 2006 with mixed shrubland–grassland roughness length updated to 0.5 m
Terrain	1 arcsec terrain from the Shuttle Radar Topography Mission (SRTM)
Turbulence subgrid-scale model	non-linear backscatter and anistropy (NBA2)

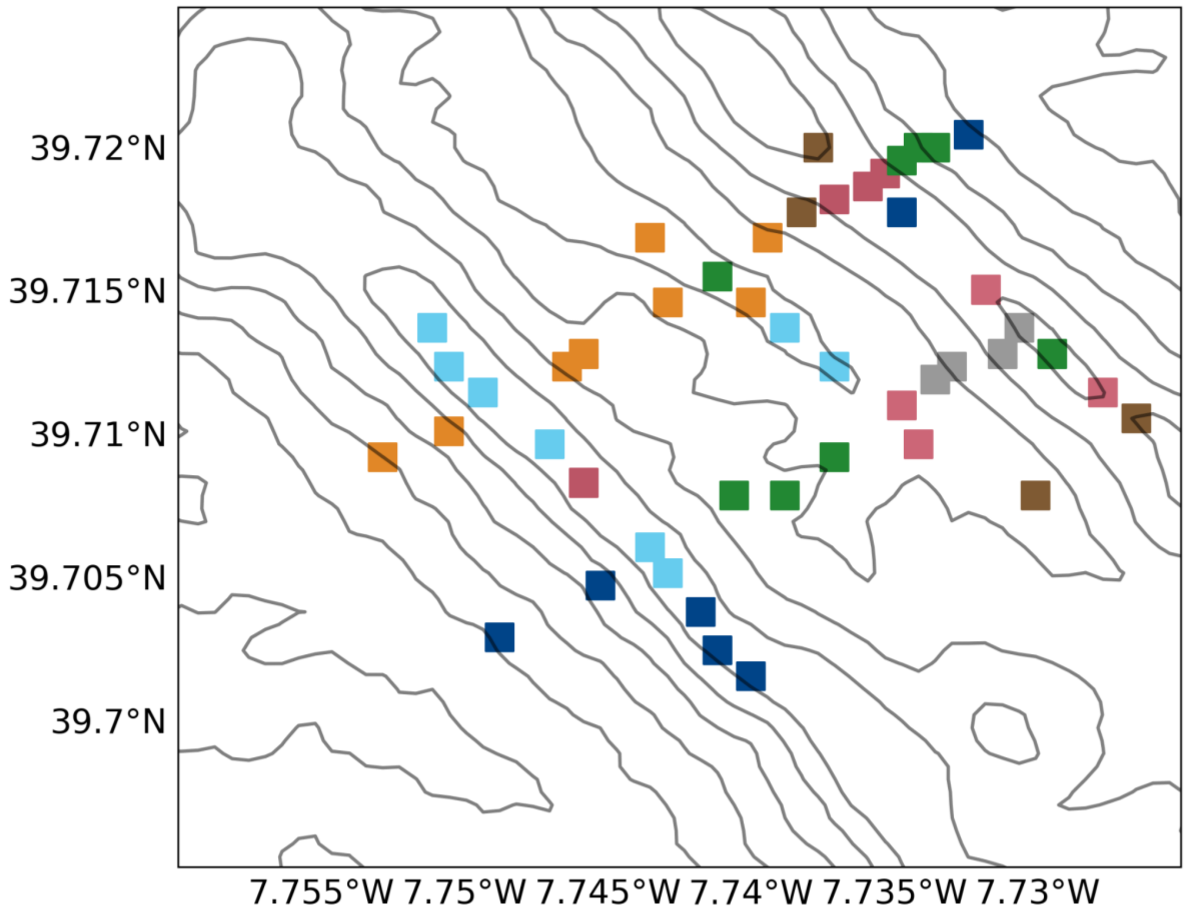

Figure A1. Static stability partitions determined by the Louvain community detection algorithm with LES output from Robey and Lundquist (2024) at the locations of the station towers. Model parameters are defined in Table A1.

"This implementation is designed to serve several stakeholders and future research directions. Because this analysis builds on publicly available LES output from Robey and Lundquist (2024), it provides a reproducible benchmark for future Louvain or LES-based studies. The short duration and dense observations create opportunities to explore similarity metrics beyond the temporal correlation weight used here. We also introduce static stability as an alternate metric that depends only on temperature measurements, recognizing that many stakeholders, including wind farm developers, rely on sparse instrumentation. To further mimic these practical constraints, we adopt a 30-minute temporal resolution.

"This case study therefore provides both a benchmark and a flexible framework that can be extended to other datasets, parameterizations, and applications."

I therefore recommend major revision to address the points above, and a few specific comments below.

Specific comments

L32: Discussing various papers about the effects of stability before at least defining it broadly. Some of these papers even analyze data from complex terrain.

We thank the reviewer for this comment. We agree that beginning more broadly helps contextualize the scope. To avoid describing the effects of something that is not yet defined, we propose beginning by contextualizing atmospheric stability with respect to its connection to atmospheric boundary layer as a whole:

"The atmospheric boundary layer (ABL) directly interacts with Earth's surface, and the surface/atmosphere interactions that occur in this layer dictate transport of heat, momentum, sediment, and moisture into the larger atmosphere (Stull, 1988). Understanding these exchange processes has implications for fields like renewable energy (Pérez Albornoz et al., 2022) and agriculture (Tang et al., 2022). These exchange processes are also highly dynamic and sensitive to atmospheric stability (Garratt, 1994)."

L152: The Obukhov length is not proportional to the height above the surface. It is the height above the surface.

Thank you. We adapt this line, now line 210, to properly read:

"The Obukhov Length, *L*, is the height above the surface…"

L162: Please define Tv. Is it even meaningful to use theta-v and Tv in the context when theta is then anyway assumed to be constant?

We appreciate the reviewer's thoughtful observation regarding the differences in temperature characterization. We agree that moisture and pressure corrections to air temperatures could allow for fine adjustments to the analysis. We also believe that the most appropriate way to respond to the available measurements is to name the experimental limitation as opposed to either misrepresenting the theory or introducing proxies that further propagate uncertainties. To improve clarity, this line, now line 220, reads:

"where T_v is the virtual temperature (K) approximated as the air temperature; p_0 is a reference pressure (Pa); p is the pressure at a given height (Pa); and $\frac{R}{C_p} \approx 0.286$.

Herein, θ_v was approximated as the air temperature due to a lack of localized

pressure and moisture measurements. Further, because not all 47 towers had temperature measurements available at 10 m, the virtual potential temperature for the horizontal homogeneity analysis (described later) was assumed to be a uniform 300 K."

L250: Please clarify if the linear regression is calculated at specific time-stamps? If yes, please discuss how the vertical information propagation may adversely affect this metrics.

We thank the reviewer for this comment. The linear regression incorporates all 30 min (or 10 min) timesteps. Thus, while decoupling occurs for 8-11% of the (cloudless) field campaign, the vertical information propagation over the 100-m tower is also significantly shorter than the Reynolds averaging window. We acknowledge the concern on line 311:

"The vertical information propagation over the 100-m tower is also assumed to be much shorter than the Reynolds decomposition window."

L274, Figure 2: The case (e) tse09 (valley) stable exhibits the opposite behavior from every other case. This would merit some discussion.

We thank the reviewer for this observation. We agree that the stable conditions in the valley are notable and worthy of discussion. Discussion of this behavior, in part, has informed the development of a new section of the manuscript, 3.1.3 Vertical Profiles and a new figure, Fig. 6, all of which highlight the unique behavior of the stable profiles in the valley.

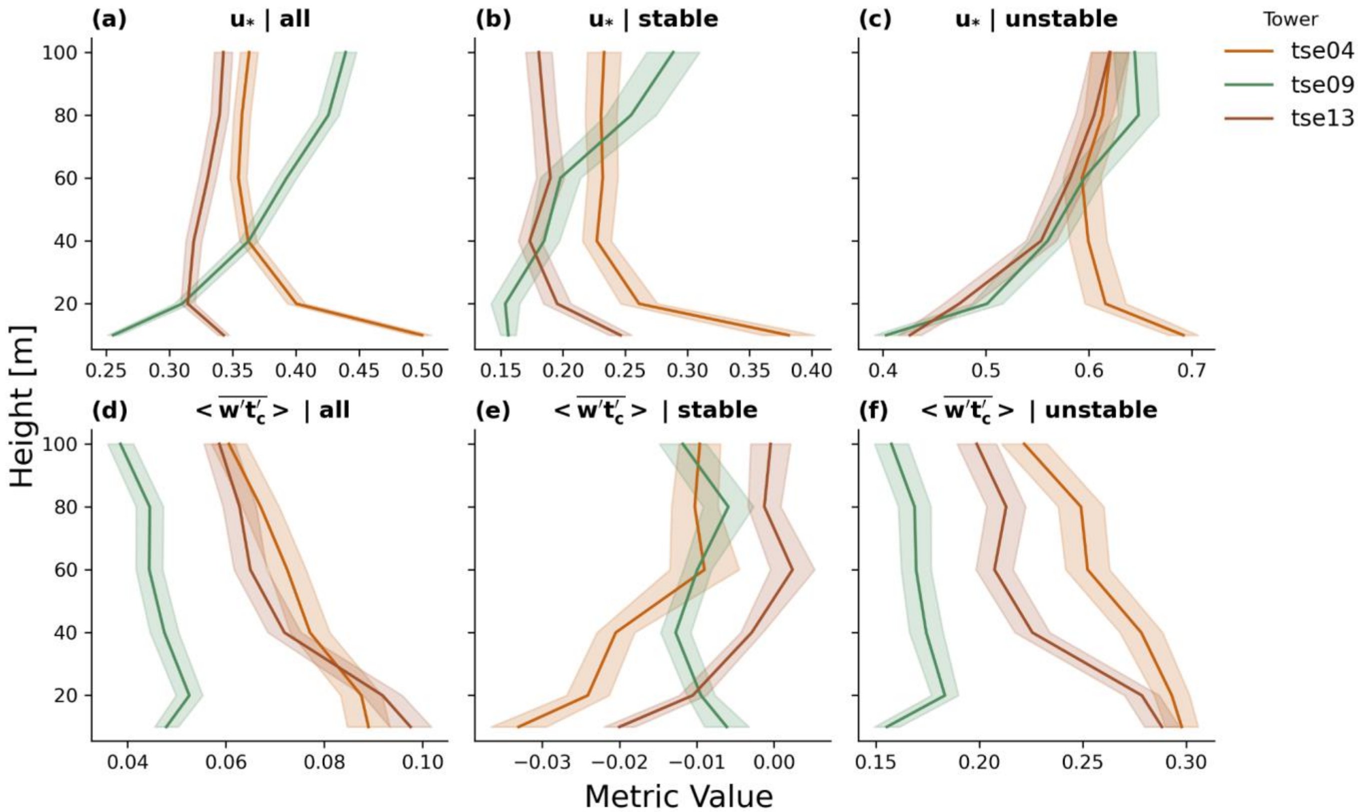

Figure 6. Perdigão field campaign vertical profiles with 30 min averaging at the three 100 m towers. The solid line indicates the mean value and the band represents the standard error. (a) friction velocity all (00-23 UTC); (b) friction velocity stable (00-02 UTC); (c) friction velocity unstable (12-14 UTC); (d) heat flux all (00-23 UTC); (e) heat flux stable (00-02 UTC); (f) heat flux unstable (12-14 UTC).

The dedicated discussion of the stable heat flux profiles in the valley begins on line 418:

"Stable heat flux profiles also reflect profoundly distinct shapes between the ridges and the valley (Fig. 6e). Stable heat flux profiles along the two ridges are more stable than in the valley at the surface, but as the height increases, the stable heat fluxes along the two ridges become increasingly weaker (Fig. 6e). In contrast, the stable heat flux profile in the valley (Fig. 6e) becomes more stable from the surface to 40 m before stabilizing above. Thus, while the valley is well-mixed during stable conditions, the ridges are not (Fig. 6e). This unique heat flux behavior in the valley during stable conditions is also corroborated by the heat flux CFAW-CM analysis, where stable heat fluxes in the valley (Fig. 4e) are much tighter and smaller than those along the two ridges (Fig. 4b,h). These ridge/valley differences in stable heat fluxes may signal an upside down boundary layer in the valley (Fig. 6e)."

We also complement this discussion with another, detailed discussion of an upside down boundary layer in the valley later on, starting on line 503:

“The TKE in the valley (Fig. 14c, d) shows stratified layers during nighttime hours, with smaller values at the surface and larger values with increasing height.

“One way to explain this apparent decoupling in the valley between the 10 m and 100 m measurements is to consider remote generation of turbulence that is then advected to the upper levels of the valley tower. This advection could occur horizontally, such that turbulence generated on the ridge is advected into the valley, where it is dissipated. This process could also occur vertically through the presence of an upside-down boundary layer (Parker and Raman, 1993; Mahrt, 1999). Warm air on the ridges may be advected over cold air pooled and trapped in the valley, leading to shear generation at the top of the cold pool (Mahrt, 1999). The explanation of an upside-down boundary layer in the valley is supported by a consistent TKE increase across intermediate heights between 10 m and 100 m within the valley (Fig. 14c, d), but not at the two ridge locations (Fig. 14a,b, e, f). This unique, well-mixed behavior during stable conditions in the valley is also reflected in the heat fluxes (Fig. 6e).”

L277: Not entirely clear how it is discernible from Fig. 2h that the stable ogive shifts to mesoscale fluctuations at 60 min. Do you mean that the curves which appear to have flattened, suddenly receive a kink?

We thank the reviewer for the opportunity to be more clear in our descriptions. The reviewer is correct that we are in fact commenting on the sudden shift/kink that occurs after the otherwise flattened behavior. We adapt this line, now line 399, to read:

“On the NE ridge, heat fluxes (Fig. 4h) are mostly constant through 30 min before increasing from 30-60 min.”

L307: "more diffuse" is not the best choice of words. It would be better said that the winds are less bidirectional, or aligned with the terrain.

We thank the reviewer for this comment. We alter this line (now line 448) to read “less bidirectional.”

L310, Figure 3: Please discuss why the ogives (are they really ogives, strictly speaking?) for u* are so different from those for the heat flux?

Thank you for both the motivation for discussing u* vs. heat flux as well as challenging our labelling of our method as an ogive method.

Regarding the first issue, we argue that the friction velocity ogives differ from those for heat flux because of the increased surface roughness from surface and canopy cover.

Regarding the second issue, whether these are strictly speaking ogives, we have carefully considered the differences between our analysis in the time domain and traditional ogive analysis in the frequency domain. While these approaches should result in the same conclusion, the fact that we did our analysis in the time domain does suggest that we should label the approach with a different name. Therefore, we have labelled our approach the "Cumulative Flux Averaging-Window Convergence Method (CFAW-CM)" and distinguished it from the ogive method in the text as below:

"An appropriate averaging window was determined using an ogive-type method, the Cumulative Flux Averaging-Window Convergence Method (CFAW-CM). While the ogive method (Desjardins et al., 1989; Oncley et al., 1996; Babic et al., 2012) identifies an appropriate Reynolds-averaging window through spectral integration of the flux cospectrum in the frequency domain, the convergence method determines the window in the time domain by identifying the averaging period beyond which additional low-frequency variability no longer contributes to the flux covariance. This assessment was performed to determine whether a uniform averaging period could be utilized for heat fluxes ($\overline{w'\theta'_{v}}$) and friction velocities ($\overline{u_{*}}$) across various heights, times of day, and tower locations.

"By evaluating the cumulative eddy flux contribution according to increasing averaging periods, an asymptote separates the local turbulent fluctuations from the larger-scale (mesoscale) fluctuations. For this dataset, heat flux and friction velocity values were calculated at each height for each tower location according to 1, 5, 10, 20, 30, and 60 min Reynolds decomposition windows.

L477: It is too optimistic to claim that this work improves the characterization of atmospheric stability. It only demonstrates the challenges, based on data from one site.

We thank the reviewer for this comment. We alter new line 626 to read "This work demonstrates the challenges in atmospheric stability characterization in complex terrain…"

L486: "... towers that extend to hub height ...". Should they not be extended to the rotor top?

Arguably, similar if not larger discrepancies will occur when the top of the boundary layer, which often is at about HH, intersects the rotor area.

We thank the reviewer for this comment. We also agree that the rotor top may present an additional transition in stability characterization. However, because measurements are not available at the rotor top for this analysis, we restrict our conclusions to a designated 100 m hub-height. As such, we propose to keep the following line, now line 634, to read:

"Thus, it may be necessary to invest in meteorological towers that extend to hub height to make effective hub-height predictions."

We also clarify this decision in the text starting on line 309:

"While many representative heights may be potentially considered in a wind farm resource assessment, the hub height is one such representative measurement. Further, the hub height, with typical values near 100 m, aligns with the highest available measurement from the meteorological towers during the Perdigão field campaign."