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 3

The paper covers an important aspect of wind meteorology: atmospheric stability. It also provides a thorough analysis of experimental data to identify trends and limitations of the different methods used to characterize stability.

The introduction provides readers with an overview of research efforts to characterize the impact of stability on turbines and wind farms. It highlights the lack of general understanding of the impact on power and wear and indicates that characterizing stability is an expensive process.

The methods section details the data sampling, equations of the indicators used (L, TKE, and ϵ), how the dataset was cleaned, and which averaging windows were used. The methods section also introduces a study on the impact of the Reynolds decomposition time (1 to 60 minutes) on heat flux and friction velocity.

The conclusion helps wind farm designers determine where to locate met masts and understand the risks of taking 10-meter wind speed measurements and extrapolating them to derive an atmospheric stability assessment, which is likely inaccurate. The authors suggest using LES in complex terrain to better identify locations where it is important to place a measurement tower and to use Louvain group theory to identify these locations, as well as to identify similarities using Python code.

We thank the reviewer for their thoughtful appraisal.

Overall, the paper is well-structured and presented. However, it would have been helpful to include more justification for the analysis of the results. For example: "the plot indicates that ... because (explanation)" I provide a concrete example of this in my comment on L297.

We thank the reviewer for an opportunity to be more complete in our discussion of our results. We provide more complete justifications throughout and address your specific comment where it is outlined in the end.

The abstract could also be improved. It is unclear where the authors provide general information and where they present the paper's findings. Additionally, the analysis of the sampling window for heat flux and friction velocity, which occupies an important portion of the paper, is not mentioned.

We thank the reviewer for their feedback on the abstract. We adapt our more general comment about the sampling window to be more direct. We also shift the sentence structure to more clearly separate specific and general claims.

"Characterizing atmospheric stability becomes challenging in heterogeneous complex terrain. We use data from 47 meteorological towers associated with the Perdigão field campaign to recommend data processing approaches and to assess the limitations of shorter or fewer towers. We quantify atmospheric stability according to the Obukhov Length, the turbulence kinetic energy, and the turbulence dissipation rate using two decomposition periods, including consistent 10 minute periods to match convention in the wind energy community and consistent 30 minute periods to match convention in the atmospheric science community. We also demonstrate a methodology that can indicate the necessary number and location of towers to characterize atmospheric stability. We find that the 10 minute Reynolds decomposition window underestimates turbulence patterns. Additionally, 10 m measurements do not provide reliable 100 m hub-height stability predictions. Holistically, this work addresses challenges in relying on sparse surface measurements."

Below are some general comments and questions that I recommend addressing in the paper.

- Can you define or explain the tilt correction methodology of the sonic anemometer?

We thank the reviewer for pointing out this oversight. The section starting on line 154 now includes a reference and discussion of the planar fit method used for this analysis:

"The sonic anemometer data were then tilt-corrected using the planar fit method following the approach of Wilczak et al. (2001)."

- What about cloud cover? Is Perdigao rarely covered by clouds, or were cloudy days filtered out?

We thank the reviewer for this question. Clouds were identified on several days throughout the field campaign and were logged on a daily basis in the supplement to Fernando et al. (2019). The revised analysis now filters out these days and the motivation for this screening is discussed, starting on line 174:

"[D]ata on days with clouds or precipitation were screened out of the analysis. The Perdigão field log (Supplement to Fernando et al., 2019) continues a daily breakdown of the synoptic conditions according to eight European standard patterns defined in Santos et al. (2016). This field log, along with determining the relevant synoptic

regime for a day, also notes relevant weather patterns, such as cloud presence and precipitation. Based on this analysis, we screen out any days that include low or medium clouds as well as those with fog or precipitation. In situations where these weather patterns only affect targeted hours of the day, we remove the entire day to avoid biasing diurnal patterns. The net result of this screening is that 22 of the 45 days are removed (Appendix A).”

We also include a discussion of shading starting on line 352 with a new figure, Fig. 3:

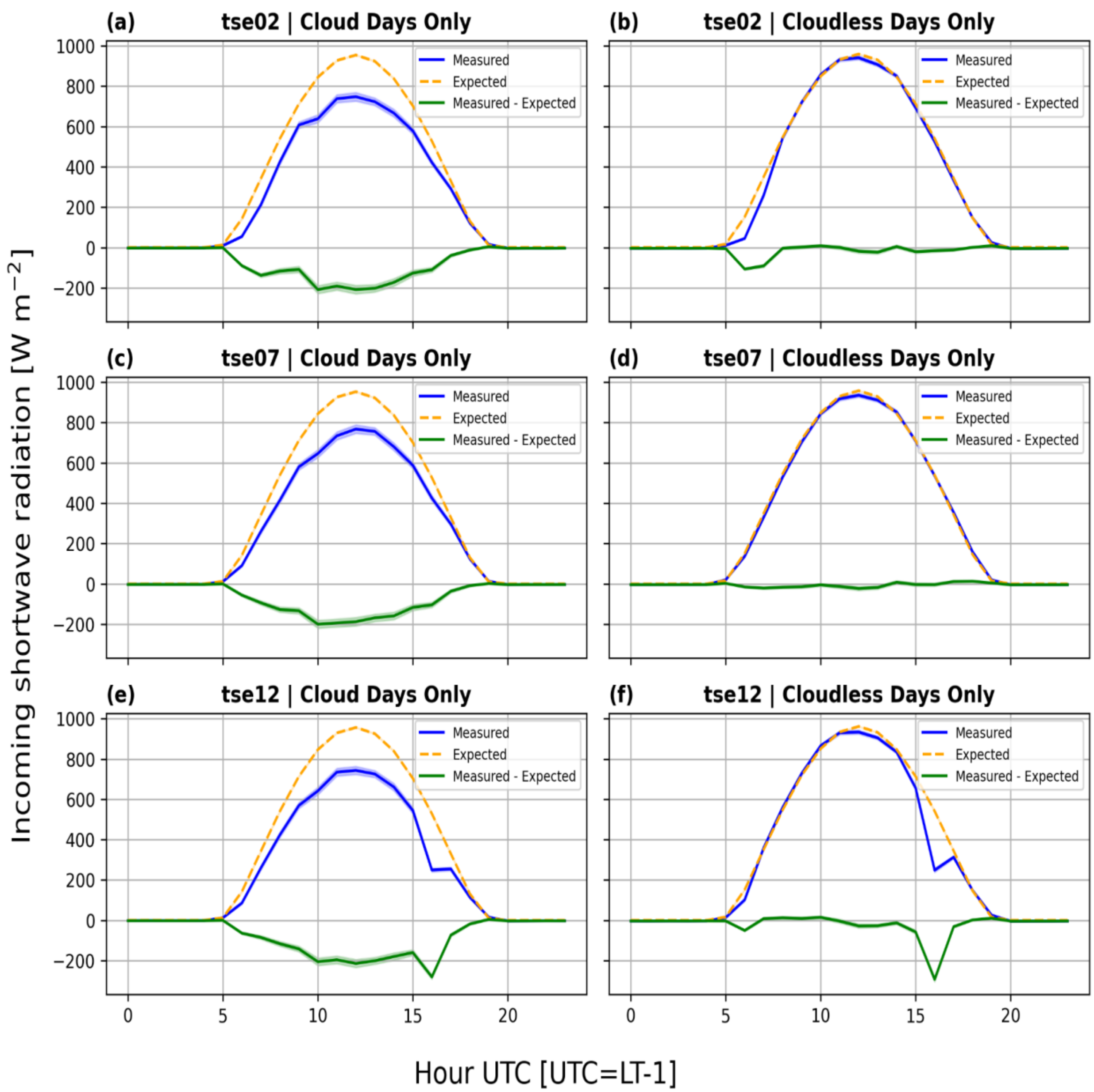

Figure 3. Comparison between expected and measured incoming shortwave radiation adjacent to each of the three 100 m towers. The line represents the mean value and

the band represents the standard error. (a) tse02 (SW ridge) cloud days only; (b) tse07 (valley) cloud days only; (c) tse12 (NE ridge) cloud days only; (d) tse02 (SW ridge) cloudless days only; (e) tse07 (valley) cloudless days only; (f) tse12 cloudless days only.

"Before proceeding with stability analysis, we ensured that influences like topographic shading of our measurements did not directly impact parts of our analysis. In particular, we ensured that our designated "stable" and "unstable" hours did not experience inequitable solar exposure between towers. Importantly, these designations also translated to our CFAW-CM convergence analysis, which informed the Reynolds decomposition window.

"All three 100 m towers likely experience shading during some part of the day. The incoming shortwave solar radiation near all towers shows a typical diurnal cycle for a mid-latitude land-based site (Fig. 3). This diurnal cycle is defined by low (i.e. zero) incoming shortwave radiation in the late (18-24 UTC) and early morning (00-04 UTC) hours with an increase throughout the morning (05-11 UTC), a midday peak (12-14 UTC), and a late afternoon drop (15-17 UTC) (Fig. 3). Within this diurnal cycle, differences in solar exposure throughout the day emerge. These differences in solar exposure contribute to differences in residuals. While negligible positive residuals may reflect measurement uncertainties, negative residuals are assumed to reflect shading. Both the morning (05-07 UTC) and early evening (15-17 UTC) experience shading (Fig. 3). Cloudy days (Fig. 3a,c,e) naturally exhibit more shading than cloudless days (Fig. 3b,d,f) near each of the three 100 m towers. This diurnal difference in shading is especially pronounced in the valley, where negligible shading is suggested during cloudless days (Fig. 3d). Non-cloud-based shading due to diurnal differences in solar exposure also influences the two 100 m ridge towers. While shading in the valley is only imposed by clouds (Fig. 3c,d), the two ridges show stronger terrain-induced shading (Fig. 3a,b,e,f). Further, while the SW ridge experiences minimal terrain-induced shading in the early evening (Fig. 3b), the NE ridge shows the strongest terrain-induced shading during both the early morning and early evening (Fig. 3f). Local canopy effects may then contribute–but not-define–non-cloud shading. The NE ridge shows more diurnal shading than the SW ridge and also has a taller canopy cover. However, the valley has canopy cover but experiences no non-cloud shading. Further, the SW ridge, with a negligible canopy cover, does experience non-cloud shading. Thus, shading from the larger ridge topography may also be contributing to the experience diurnal shading. Regardless of the exact shading source, these site-based differences in shading suggest that sites do not have equivalent diurnal exposure to the incoming solar radiation. Our CFAW-CM analysis, described below, accounts for these differences in solar exposure by restricting our analysis to cloudless days and our stable and unstable periods to 00-02 UTC and 12-14 UTC, respectively."

-- Along those lines, what is the sun exposure near the ridges and in the valley? Is it full sun between 12 p.m. and 4 p.m., or is it partly shadowed by the terrain?

We thank the reviewer for this question. To address the reviewer's curiosity, we performed an additional analysis and determined that terrain-induced shading may affect both the SW ridge and NE ridge after 15 UTC. As such, we adapt our unstable definition to include 12-14 UTC. The analysis is summarized starting on line 181:

"Radiometer measurements informed the second, terrain-based, shading screening. These measurements–located near the 100 m towers at tse02, tse07, and tse12–were collected at either 20 m or 30 m and reported at 5 min intervals. These (cloudless, as defined above) incoming shortwave radiation measurements were diurnally-averaged and compared to the amount of incoming shortwave radiation expected by the Ineichen-Perez clear sky model (Ineichen and Perez, 2002), as defined in the python package pvlib (Anderson et al., 2023). The residual (measured-expected) was then assumed to trace hours where terrain shading might be assumed to occur. Instead of discarding data associated with shading, by identifying inequitable patterns in solar exposure between the towers, we were able to tailor our definitions of "stable" and "unstable" to only focus on hours that were not affected by these differences. Shown later, these stable and unstable hour definitions also inform our determination of an appropriate Reynolds decomposition window for two of three stability metrics. Taken together, these two shading evaluations–albeit strict–are crucial to minimize spatial and temporal biases between towers and support a more equitable evaluation."

The results of this additional analysis are presented in a new Fig. 3 and discussed starting on line 352:

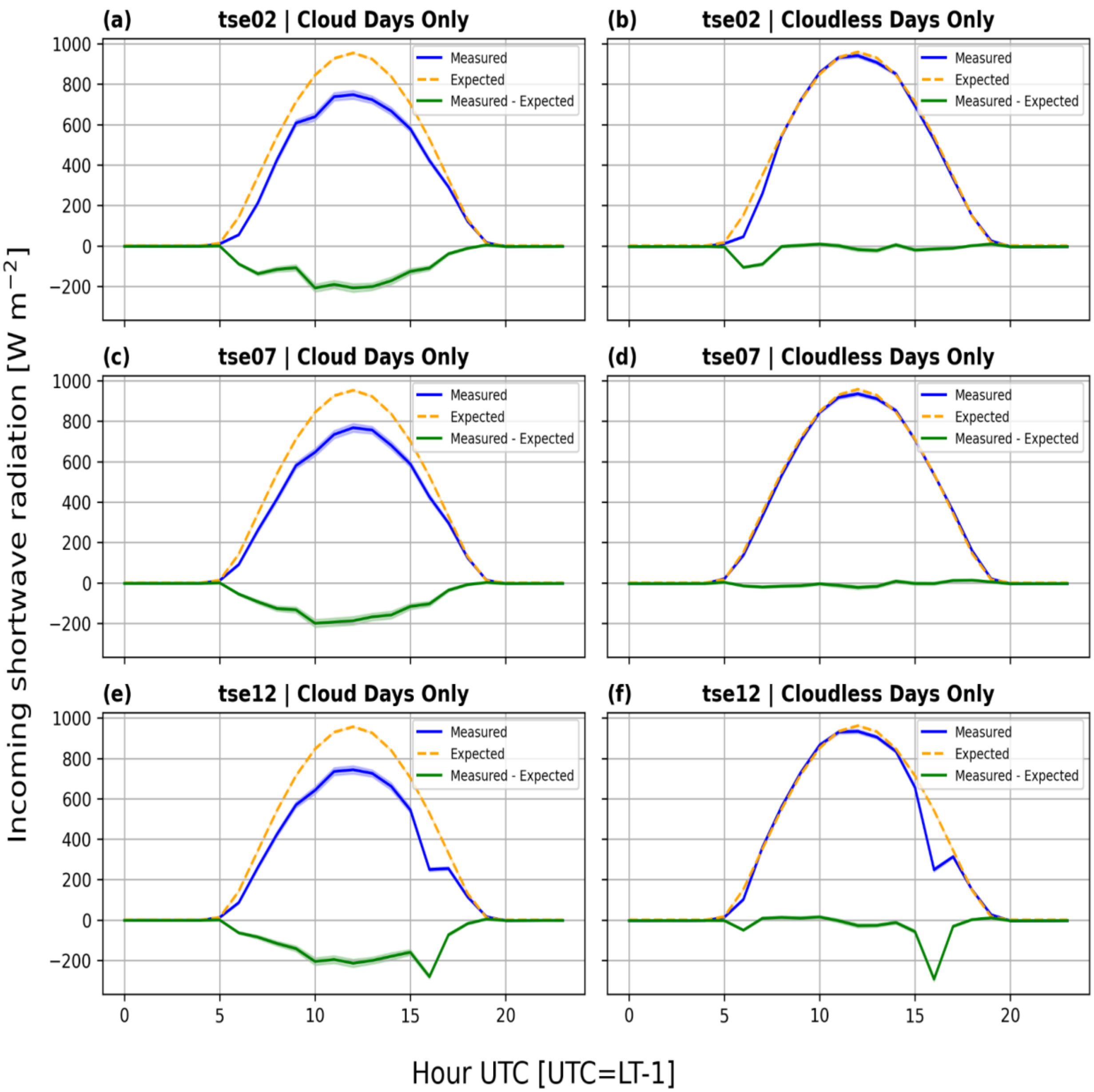

Figure 3. Comparison between expected and measured incoming shortwave radiation adjacent to each of the three 100 m towers. The line represents the mean value and the band represents the standard error. (a) tse02 (SW ridge) cloud days only; (b) tse07 (valley) cloud days only; (c) tse12 (NE ridge) cloud days only; (d) tse02 (SW ridge) cloudless days only; (e) tse07 (valley) cloudless days only; (f) tse12 cloudless days only.

"Before proceeding with stability analysis, we ensured that influences like topographic shading of our measurements did not directly impact parts of our analysis. In particular, we ensured that our designated "stable" and "unstable" hours did not experience inequitable solar exposure between towers. Importantly, these

designations also translated to our CFAW-CM convergence analysis, which informed the Reynolds decomposition window.

"All three 100 m towers likely experience shading during some part of the day. The incoming shortwave solar radiation near all towers shows a typical diurnal cycle for a mid-latitude land-based site (Fig. 3). This diurnal cycle is defined by low (i.e. zero) incoming shortwave radiation in the late (18-24 UTC) and early morning (00-04 UTC) hours with an increase throughout the morning (05-11 UTC), a midday peak (12-14 UTC), and a late afternoon drop (15-17 UTC) (Fig. 3). Within this diurnal cycle, differences in solar exposure throughout the day emerge. These differences in solar exposure contribute to differences in residuals. While negligible positive residuals may reflect measurement uncertainties, negative residuals are assumed to reflect shading. Both the morning (05-07 UTC) and early evening (15-17 UTC) experience shading (Fig. 3). Cloudy days (Fig. 3a,c,e) naturally exhibit more shading than cloudless days (Fig. 3b,d,f) near each of the three 100 m towers. This diurnal difference in shading is especially pronounced in the valley, where negligible shading is suggested during cloudless days (Fig. 3d). Non-cloud-based shading due to diurnal differences in solar exposure also influences the two 100 m ridge towers. While shading in the valley is only imposed by clouds (Fig. 3c,d), the two ridges show stronger terrain-induced shading (Fig. 3a,b,e,f). Further, while the SW ridge experiences minimal terrain-induced shading in the early evening (Fig. 3b), the NE ridge shows the strongest terrain-induced shading during both the early morning and early evening (Fig. 3f). Local canopy effects may then contribute–but not-define–non-cloud shading. The NE ridge shows more diurnal shading than the SW ridge and also has a taller canopy cover. However, the valley has canopy cover but experiences no non-cloud shading. Further, the SW ridge, with a negligible canopy cover, does experience non-cloud shading. Thus, shading from the larger ridge topography may also be contributing to the experience diurnal shading. Regardless of the exact shading source, these site-based differences in shading suggest that sites do not have equivalent diurnal exposure to the incoming solar radiation. Our CFAW-CM analysis, described below, accounts for these differences in solar exposure by restricting our analysis to cloudless days and our stable and unstable periods to 00-02 UTC and 12-14 UTC, respectively."

Similarly, for the heat flux analysis, c**an the changes in sun exposure throughout the day be excluded?** For example, should we expect the results between 12 and 13 hours to be the same as those between 15 and 16 hours, or should we expect some variability in e.g. the sun radiation from the ground?

We thank the reviewer for this question. To address the reviewer's curiosity, we performed an additional analysis and determined that terrain-induced shading may

affect both the SW ridge and NE ridge after 15 UTC. As such, we adapt our unstable definition to include 12-14 UTC. The analysis is summarized starting on line 181:

“Radiometer measurements informed the second, terrain-based, shading screening. These measurements–located near the 100 m towers at tse02, tse07, and tse12–were collected at either 20 m or 30 m and reported at 5 min intervals. These (cloudless, as defined above) incoming shortwave radiation measurements were diurnally-averaged and compared to the amount of incoming shortwave radiation expected by the Ineichen-Perez clear sky model (Ineichen and Perez, 2002), as defined in the python package pvlib (Anderson et al., 2023). The residual (measured-expected) was then assumed to trace hours where terrain shading might be assumed to occur. Instead of discarding data associated with shading, by identifying inequitable patterns in solar exposure between the towers, we were able to tailor our definitions of "stable" and "unstable" to only focus on hours that were not affected by these differences. Shown later, these stable and unstable hour definitions also inform our determination of an appropriate Reynolds decomposition window for two of three stability metrics. Taken together, these two shading evaluations–albeit strict–are crucial to minimize spatial and temporal biases between towers and support a more equitable evaluation.”

The results of this additional analysis are presented in a new Fig. 3 and discussed starting on line 352:

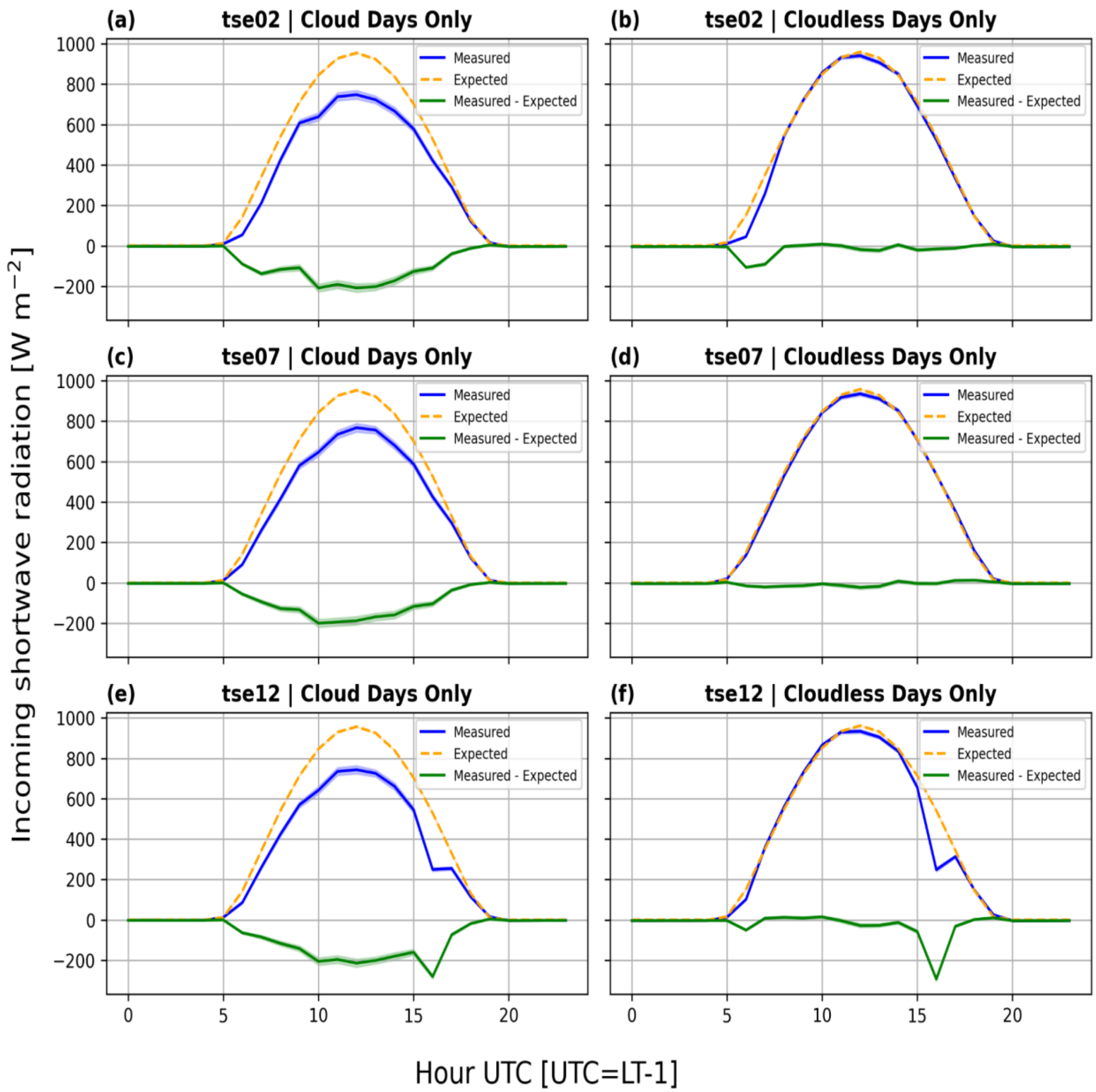

Figure 3. Comparison between expected and measured incoming shortwave radiation adjacent to each of the three 100 m towers. The line represents the mean value and the band represents the standard error. (a) tse02 (SW ridge) cloud days only; (b) tse07 (valley) cloud days only; (c) tse12 (NE ridge) cloud days only; (d) tse02 (SW ridge) cloudless days only; (e) tse07 (valley) cloudless days only; (f) tse12 cloudless days only.

"Before proceeding with stability analysis, we ensured that influences like topographic shading of our measurements did not directly impact parts of our analysis. In particular, we ensured that our designated "stable" and "unstable" hours did not experience inequitable solar exposure between towers. Importantly, these

designations also translated to our CFAW-CM convergence analysis, which informed the Reynolds decomposition window.

"All three 100 m towers likely experience shading during some part of the day. The incoming shortwave solar radiation near all towers shows a typical diurnal cycle for a mid-latitude land-based site (Fig. 3). This diurnal cycle is defined by low (i.e. zero) incoming shortwave radiation in the late (18-24 UTC) and early morning (00-04 UTC) hours with an increase throughout the morning (05-11 UTC), a midday peak (12-14 UTC), and a late afternoon drop (15-17 UTC) (Fig. 3). Within this diurnal cycle, differences in solar exposure throughout the day emerge. These differences in solar exposure contribute to differences in residuals. While negligible positive residuals may reflect measurement uncertainties, negative residuals are assumed to reflect shading. Both the morning (05-07 UTC) and early evening (15-17 UTC) experience shading (Fig. 3). Cloudy days (Fig. 3a,c,e) naturally exhibit more shading than cloudless days (Fig. 3b,d,f) near each of the three 100 m towers. This diurnal difference in shading is especially pronounced in the valley, where negligible shading is suggested during cloudless days (Fig. 3d). Non-cloud-based shading due to diurnal differences in solar exposure also influences the two 100 m ridge towers. While shading in the valley is only imposed by clouds (Fig. 3c,d), the two ridges show stronger terrain-induced shading (Fig. 3a,b,e,f). Further, while the SW ridge experiences minimal terrain-induced shading in the early evening (Fig. 3b), the NE ridge shows the strongest terrain-induced shading during both the early morning and early evening (Fig. 3f). Local canopy effects may then contribute–but not-define–non-cloud shading. The NE ridge shows more diurnal shading than the SW ridge and also has a taller canopy cover. However, the valley has canopy cover but experiences no non-cloud shading. Further, the SW ridge, with a negligible canopy cover, does experience non-cloud shading. Thus, shading from the larger ridge topography may also be contributing to the experience diurnal shading. Regardless of the exact shading source, these site-based differences in shading suggest that sites do not have equivalent diurnal exposure to the incoming solar radiation. Our CFAW-CM analysis, described below, accounts for these differences in solar exposure by restricting our analysis to cloudless days and our stable and unstable periods to 00-02 UTC and 12-14 UTC, respectively."

- The paper does not specify the timeframe of the data. Did you select a few days or years? Which exact dates do the data represent? During the filtering process, are significant time periods (days or weeks) flagged out? What impact does this have on, for example, the 60-minute sampling? Were dummy values inserted, or were full 60-minute blocks removed?

We thank the reviewer for this inquiry. Line 104 notes that we start with the 45-day period of 1 May - 15 June 2017. Based on the reviewer's comments above regarding clouds, we re-performed our analysis after screening out 22 of the 45 available days of the field campaign. This process is described starting on line 174:

“[D]ata on days with clouds or precipitation were screened out of the analysis. The Perdigão field log (Supplement to Fernando et al., 2019) presents a daily breakdown of the synoptic conditions according to eight European standard patterns defined in Santos et al. (2016). This field log, along with determining the relevant synoptic regime for a day, also notes relevant weather patterns, such as cloud presence and precipitation. Based on this analysis, we screen out any days that include low or medium clouds as well as those with fog or precipitation. In situations where these weather patterns only affect targeted hours of the day, we remove the entire day to avoid biasing diurnal patterns. The net result of this screening is that 22 of the 45 days are removed (Appendix A).”

- Regarding your use of the Louvain community detection algorithm, are any tolerances or parameters used to create the groupings? It would be useful for the reader to know what threshold was used to determine the borders between the groups.

We thank the reviewer for this close observation. The Louvain algorithm can be tuned to several parameters, including a resolution as well as a modularity gain threshold. We stick to the default parameters in our analysis and to address this confusion in the manuscript, we now mention on line 337:

“While the Louvain community detection algorithm is also sensitive to several tuning parameters, we make no adjustments to the default modularity gain threshold (1e-7) or resolution (1) and introduce no artificial seeds or establish restrictions on the number of optimization cycles.”

- You mention hub height as being important in characterizing stability. On what criteria was the hub height chosen? Can stability be characterized by parameters other than turbine height, or is this choice based on the assumption that we are mostly interested in knowing what the turbine "sees"? Please explain this in the paper.

We thank the reviewer for this inquiry. We base our criteria on the hub height of the turbine both because the hub-height is a standard measure as well as the fact that hub height aligns well with the top altitude available for the measurement towers. We 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, naturally aligns with the highest available measurement from the meteorological towers during the Perdigão field campaign.”

Also, I have two comments related to lines:

L62: It's unclear what is meant by "stretch into an ellipse". From where is the ellipse seen? (The wake is 3D).

Thank you, we adapt this line, now line 55, to read "stretch into a 3D ellipse."

L297: "unstable friction velocity ogives at all tower locations show asymptotic behavior for a 30 min averaging period and a shift to mesoscale fluctuations for a 60 min averaging period"

-> Please help the reader by explaining that this is seen by the increase in u* for the 60-min average period, or otherwise if needed.

We thank the reviewer for the opportunity to be more clear and complete in our discussion of the results, especially in our discussion of the appropriate Reynolds decomposition section.

We have explicitly gone through and restructured this section to be more clear by discussing these ogive-type results tower by tower. As such, we discuss the unstable friction velocities first in the valley on line 389:

"During unstable conditions [in the valley], heat fluxes (Fig. 4f) level off after 10 min and increase again between 30-60 min whereas friction velocities (Fig. 5f) stabilize at 5 min before increasing again after 30 min."

Then on line 395 for the two ridges:

On the SW ridge, unstable friction velocities (Fig. 5c)–especially above 10 m–stabilize by 10 min, whereas on the NE ridge they increase until about 20 min at heights above 20 m and remain constant from 1-30 min at heights at or below 20 m) (Fig. 5i).

Together, we hope that this makes our discussion of the unstable friction velocities more clear and complete.