

## Response to reviewer1

1. I do find it difficult to understand that the metric is being designed based on physical considerations, rather than on the practical considerations resulting from the grid-spacing of the model. Perhaps you could rename this the ‘Multi-Scale Variability Index’, and avoid this problem of claiming that it separates the mesoscale and synoptic scales? The relationship of this to the mesoscale and synoptic scale variability could still be discussed, but this would also make the methodology quite flexible for others to apply in other scenarios where it is of interest to separate the variability from two time-scales.

**Reply:** We added a paragraph to the manuscript to explain better our approach. Although such a definition always relies on a combination of physical and practical considerations we prefer not to use the term multiscale spatial variability index, since this term does not explicitly refer to the scale it aims to capture.

*One limitation is that this metric is derived using a model run of 1.5 km grid spacing, which is a slightly higher resolution than most wind atlases (e.g. DOWA Wijnant et al. (2019) and NEWA Petersen et al. (2014); Hahmann et al. (2020); Dörenkämper et al. (2020)), so the smallest signals that can be captured are at the effective resolution of 10×10 grid points (Kapper et al., 2010) (~ 15 km). As size for the small window this effective resolution is chosen, thereby capturing most of the mesoscale variability, although the smallest mesoscale signals are averaged out by this or any other product based on state-of-the-art mesoscale model simulations. The synoptic scale starts at ~ 100 kilometres (Oblack, 2020), but when taking this size for the large window, frontal systems were averaged out. Therefore, we took approximately half of this as dimensions for the large window to fully capture the synoptic background velocity field including the frontal systems. A size of 30 grid points (~ 45 km) is chosen for the large window size, which is in area nine times larger than the small window. With these window sizes, the MSVI turned out to identify mesoscale systems like gravity waves, convective systems and land-sea breezes (see Section 3.3).*

2. Line 220-231: I appreciate the added attempt to explain the window sizes. There is a grammatical problem in the added text: “we average out thunderstorms and smaller mesoscale systems, whereas cloud clusters and fronts are not”. I suggest changing this to “thunderstorms and smaller mesoscale systems are averaged out, whereas cloud clusters are not”.

**Reply:** This has been changed in the manuscript.

*Using the large window thunderstorms and smaller mesoscale systems are averaged out, whereas cloud clusters and fronts are not.*

3. Line 64: “The offshore part of sea breezes can have an influence on the power output of a wind farm as it, in general, opposes the synoptic wind flow (Steele et al., 2015).” -> I don’t think there is anything inherent about the offshore part of sea breezes opposing the synoptic wind flow. This would be sites-specific, and depend on the orientation of the coastline and the prevailing background wind. Also, are the authors talking about the offshore part of the Seabreeze, or the landbreeze?

**Reply:** This has been adjusted in the manuscript.

*The offshore part of pure sea breezes can have an influence on the power output of a wind farm as it, in general, opposes or reinforces the synoptic wind flow (Steele et al., 2015).*

4. Line 67: What aspect of MCSs is not understood? I agree this is a complex topic, but it would be helpful to understand the relevant of this sentence to this study.

**Reply:** This has been clarified in the manuscript.

*A class of convective systems of which the conditions for formation, evolution and stability are not completely understood yet are the mesoscale convective systems (MCS) (Houze Jr, 2004).*

5. Line 280: ...”meaning that parametrising the effects of the wind farm with Fitch wind introduce a larger error than not taking it into account”. If the Fitch parametrisation is going to be mentioned here, then it needs to be introduced and referenced properly. I also don’t understand the connection between the first and second parts of this sentence.

**Reply:** In response to a comment from the other reviewer this statement has been cut from the manuscript.

6. Line 296: I don’t believe the Van der Hoven (1957) spectrum is an appropriate reference for this paper. This work was based on an extremely limited dataset of only a few hours for the higher frequencies, and the higher end of the spectrum was measured during a tropical cyclone, which is not relevant to this study.

**Reply:** We refer to a more modern measurement of the atmospheric spectrum.

*For the lower frequencies the periodogram seems to be leveling off, which is indicative of the synoptic weather peak (Larsén et al., 2016).*

## Response to reviewer 2

1. L77-78: NREL's WIND Toolkit is available at 5 min resolution, so adjust the sentence to say something like "which is not available from wind atlases in our region of interest"

**Reply:** This has been changed in the manuscript.

*Our analysis has been performed using 10-minute wind speed data, which is not available from wind atlases in our region of interest.*

2. While I still believe that the lidar measurements should be excluded from this study, I will not press this larger issue further, and I appreciate that the authors have added more caveats. I have also now seen the editor's initial concerns, and I can see additional added value in doing lidar analysis in this context. That being said, I do request some small tweaks regarding the words used around the lidar analysis:
  - a. L149: Does availability rise from 12 to 36%, or normalized availability rise from 12% to 36%? Fig A3 suggests normalized availability.

**Reply:** In the manuscript this has been changed to the latter.

*Therefore a sub period of the measurement campaign was used, one in which the Anholt wind farm was still under construction and the normalised availability of operation wind farms increased from 12% to 36% (see fig. A3 for the availability of the Anholt wind farm).*

- b. L154: I'm being pedantic, but it's only "appropriate" to compare time averages and spatial averages if we know that the data is ergodic, and this data probably isn't strictly ergodic. Could you soften this statement to say "This is complementary" or something along those lines?

**Reply:** We have softened this statement in the manuscript.

*Note that the COSMO wind speed values are a grid cell average. This spatial averaging, similar to the temporal averaging of the lidar data, should diminish the impact of wind gusts on the analysis.*

- c. L264-265: I'm not convinced of this statement. If the winds were weak during this period, then Fitch would have handled the low availability just fine. Please cut the "During this... into account." Statement.

**Reply:** This statement has been cut from the manuscript.

