# Reviewer 1:

Thank you etc.

*1. From the explanation on "Edits", I hardly understand what kind of "edit" is this to improve the model performance! The author completely missed to explain clearly, what are Edits? Edits is only choosing the base model forecast or there are many more rules and regulations to guide the daily forecasts. How does Edits differ from Data assimilation?*

To construct the Bureau’s official forecast dataset, Australian forecasters use a two-step process.

1. Choose a model dataset to base the official forecast dataset on. This is referred to as a choice of *model guidance*.
2. Manually *edit* the model guidance dataset using the Graphic Forecast Editor (GFE) software package.

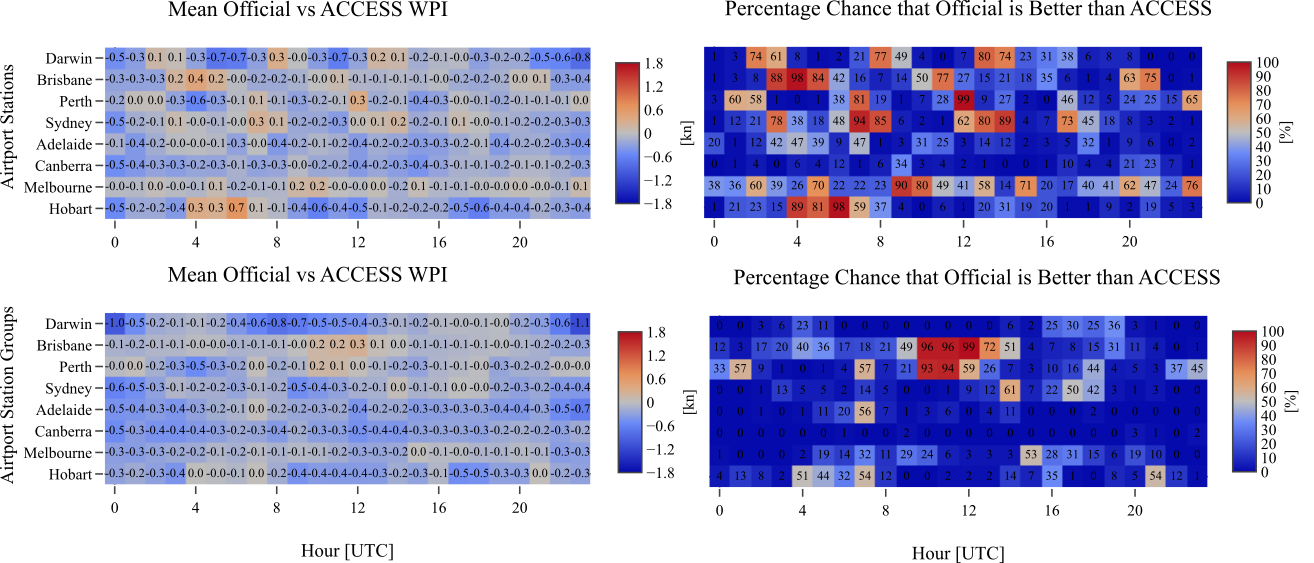
I view the “editing” step as distinct from the “choosing model guidance” step. Different forecasting centres across Australia have different practices for precisely how the forecasters choose model guidance and make edits, I don’t believe there are strict rule and regulations to guide this, just the managerial structures that exist within forecasting centres.

Data assimilation is a technique for ingesting observations into models during model initialisation. Operational forecasters in Australia do not run models themselves, and they therefore have limited input into decisions made by modellers, such as method of data assimilation, choice of parametrisation schemes and so forth.

I discussed these issues in the abstract, and throughout lines 31-41, and lines 51-66 of the original manuscript. I have edited lines 9-11 of the abstract to better distinguish between “choice of model guidance” and “edits”, also rewording the final sentence of the abstract to stay under the word limit. I have also re-written sections of the introduction for clarity, and restructured the order of paragraphs to ensure I finish discussing edits before moving on. I have also added the sentence “Forecasters themselves are rarely directly involved in model setup or calibration, instead models are run and post-processed by other teams either within the BoM or internationally” to the second paragraph of the new introduction, to try and distinguish the types of decisions forecasters make, from those of modellers, such as how to perform data assimilation.

*2. Similar to Fig 6, how does Official vs ACCESS perform for Airport station and City Stations?*

Figure 1 (below) shows results analogous to Fig. 6 of the manuscript, but for the Official versus ACCESS comparison. The results are similar to Fig. 6 in that there are only a few times and locations where Official unambiguously outperforms ACCESS. I noted this in the original manuscript on lines 299-304. I am happy to include the below figure in a revised manuscript, and elaborate on this discussion if the reviewer/editor desires. I consider it unnecessary to include these figures, as I think the more interesting discussion is why HRES, which unlike ACCESS is provided in three hourly time steps and not calibrated to Australian conditions, is still able to produce a “better” diurnal cycle (in the sense of the metrics defined in the paper) than both the official forecast dataset and other model datasets like ACCESS.

**

**Figure 1:** test

*Official Vs HRES shows very low confidence over most of the locations.*

Remember that the confidence score provides the probability that the population or “true” mean difference of absolute errors is greater than zero. A score near 0% implies *high* confidence that this mean is less than zero, whereas a score near 100% implies *high* confidence it is greater than zero. A score between, for instance, 5-95%, implies low confidence either way. This is described in the caption to Fig. 2 and on lines 177 to 186. Out of context the term “confidence” may be confusing, but read in the context of the caption and the description in the text I believe it to be appropriate.

*DAE (Difference in Absolute error) is a very simple skill score and been obsolete. What about the skill scores of forecasts based on probabilistic forecasts, e.g Ranked Probability Skill Score (RPSS)?! It is observed that most of the time probabilistic skills are superior than dynamical skill.*

Unfortunately the Bureau doesn’t issue probabilistic wind forecasts, so metrics like the Ranked Probability Skill Score cannot be used. However, I agree that probabilistic forecasts are desirable. I also agree that the DAE is a very simple metric, but for me thats it’s appeal, as it has a simple and immediate interpretation: “how much closer (in a euclidean distance sense) is vector\_1 to an observed vector, than vector\_2 is to the same observed vector?” There are innumerable ways I could have made the metric more complex, for instance by applying the Huber loss function to both terms in equation (1) before taking a difference, i.e., or to the overall expression, or to both the consituent terms and the overall expression.

Such options grow extremely rapidly, and given that I hadn’t seen anything else in the literature focused on diurnal wind signals, I decided to start simple, and leave it to others to experiment with more complex variations of the metric.

*3. HRES and ACCESS carry lesser error and higher confidence (Fig 7) and compared to Edits. A method which is inferior in the forecast, in general been rejected. This research work shows the forecasts skills are very random in nature. Thus, with the facts shown in the illustrations, it is a bit hard to trust in these forecasts.*

Don’t really understand what’s being asked of me here. Add schematic to show how method works.

*4. Knowing the "edits" are constrained to and much depended on the expertise of the human, who performs the edits. I think, Machine learning and AI will be superior options to choose a base model! However, the error of model bias and random variability in the forecast by edits will retain questionable?*

Again, not sure what’s being asked of me here. Wait till I’ve added OCF discussion.

# Reviewer 2:

*In Figs. 2, 6-8, 10, it might be very helpful to convert the abscissa to Local Time, rather than UTC. Additionally, the groupings along the ordinate should be arranged from geographic locations in the east near the top, those in central Australia in the middle, and those in western Australia at the bottom. This follows the natural march of the sun across the sky during the day.*

*In its present form, the figures require the reader to mentally translate UTC to local time for each of the geographical groupings, particularly for those readers who reside outside of Australia. And the choice or order of geographic groupings is seemingly random, with Brisbane (far east) plotted 1 row above Perth (far west). This must be contributing to the "noisiness" in the plots, which the author frequently notes throughout the text.*

This is an excellent suggestion. For context, my initial choice of UTC and layout were motivated by a few factors. First, during my time at the Darwin forecasting centre in July 2018, I noticed that forecasters seemed to think and work in terms of UTC, and so figured my results might be more intelligible to them if I used UTC in place of my preferred local solar time (LST). Second, because the coastal station groups are averages over many degrees of longitude, reporting results in LST would require choosing an “average LST” value for each group. Finally, to plot results on a single LST abscissa, different LST values for each location would have to be interpolated, or approximately matched with a single set of LST values: interpolating risks additional, and inconsistent smoothing of data at different locations, and a “nearest-neigbour” type matching would result in a loss of temporal precision. As to my choice of row layout, this was done based on approximate latitude of each group, as one of my initial questions with this work was to see whether latitude played a role in verification outcomes, given the significant effect of latitude on land-sea breeze dynamics (e.g. Rotunno 1983).

I entirely agree that given how the results turned out, it makes much more sense to order rows by longitude, and this has been done. Furthermore I agree that providing an indication of LST is vital, but for the reasons given above I am uncomfortable translating all results onto a common, approximate, LST abscissa. As a compromise, I have instead opted to display both UTC and LST on the relevant plots, with Perth LST displayed along the top line, and Brisbane LST displayed along the bottom line.

*I am motivated by the desire for the reader to able to instinctively determine the points on those plots where sunrise and sunset occur, and for the rows along the ordinate to be logically arranged by time zone, following the march of the sun. Doing so will make these plots far more interpretable, and far more accessible. Also, I have the sense that by re-ordering the abscissa to local time is likely to reduce the "noisiness" in the plots.*

An excellent and insightful comment. The reformulated plots are indeed more coherent and easier to interpret. Etc.

# Reviewer 3:

*I would like to thank the author for submitting his manuscript for review. The manuscript contains some interesting concepts, but at this stage, I do not believe it is ready for publication in Weather and Forecasting. My reasons for this recommendation are outlined below.*

*a. Does the paper fit within the stated scope of the journal?*

*Yes*

*b. Does the paper identify a gap in scientific knowledge and add new knowledge to the overall body of scientific understanding?*

*The paper does identify the importance of verifying human edited forecasts.*

*c. Is the paper free of errors in logic?*

*Yes.*

*d. Do the conclusions follow from the evidence?*

*I have some concerns about the methodology. Until these concerns are addressed, I do not know how much weight to put on the results from which the conclusions are drawn.*

*e. Are alternative explanations explored as appropriate?*

*Yes.*

*f. Are biases, limitations and assumptions clearly stated, and uncertainty quantified?*

*I am not sure the author is aware of the limitations in his methodology. For example, the spatial domains he defines are of widely varying sizes, and I think this may well make comparison of the results from different locations difficult.*

*g. Is the methodology explained in sufficient detail so that the paper's scientific conclusions could be tested by others?*

*I think the methodology probably could be reproduced by others. But the author could have made his descriptions of the method much easier for readers to understand.*

*h. Is previous work and current understanding cited and represented correctly?*

*Mostly. But the author either ignores or is unaware of the current forecaster practise of using a gridded consensus forecast as the starting point for manual edits.*

*i. Is information conveyed clearly enough to be understood by the typical reader.*

*I think this is a major weakness in the paper. The method and results could be conveyed much more clearly. This paper is difficult to read.*

*j. Are all the figures and tables necessary, appropriate, legible and annotated (as appropriate)?*

*No. I have provided a minimum set of improvements for the figures in the comments below.*

*Major comments*

*1. The author correctly identifies the procedures that Bureau of Meteorology forecasters use to prepare their official forecasts. First gridded guidance is loaded into the Graphical Forecast Editor (GFE), and then the forecaster makes manual edits to the guidance. It is correct that the forecasters have access to several NWP models for initial guidance, including the ACCESS-R and ECMWF models. However, additionally, there is a consensus forecast available called Gridded OCF (Operational Consensus Forecasts). This is the guidance which forecasters are now expected to use as their first guess in most situations. The Gridded OCF system is described in several "Operations Bulletins" available at:*

*http://www.bom.gov.au/australia/charts/bulletins/nmoc\_bulletin.shtml*

*In particular, Bulletin 91 may be of interest to the author:*

*http://www.bom.gov.au/australia/charts/bulletins/apob91.pdf*

*For the period of the author's study (mid 2018), forecasters may have used ACCESS-R and ECMWF guidance, but they will also have utilised the Gridded OCF guidance. Unfortunately, it is not clear if there are records of which guidance the forecasters used for any particular official forecast. Given that forecasters now mainly rely on Gridded OCF guidance, it is disappointing that the author didn't include this data in his study.*

*2. The difference of absolute errors metric (pages 7 and 8)*

*This metric is interesting. However, I have some concerns about its applicability to a vector wind field. In meteorology, wind fields commonly contain sharp discontinuities at fronts. It is not uncommon for the wind direction to change by very large amounts in a period of minutes.*

*The author attempts to identify diurnal cycles by subtracting a twenty four hour centered running mean from the observed or modeled wind. But consider this hypothetical situation: the first twelve hours of winds are northerlies at a constant speed, and the next twelve hours are southerlies with the same speed. The twenty four hour vector mean will be zero, and the perturbations at each hour will be quite large. I question whether in this situation the perturbation is of use for identifying diurnal cycles.*

*The above considerations aside, the difference of absolute errors metric may still be of value. The author may find that it would be possible to produce a shorter paper focusing solely on the applicability of this metric for weather forecast verification.*

*3. Coarser spatial scales (city station groups and coastal station groups)*

*The author looks at a number of spatial scales. The city station groups comprise the ten stations closest to each capital city. The coastal station groups comprise all stations within 150 km of the nearest coastline. I have the following concerns:*

*- The area of the city spatial groups varies quite a lot. The largest city spatial group is three or four times the area of the smallest group. Is comparison of statistics from these differing sized groups valid?*

*- It is debatable if Canberra can be considered a coastal area. It is too far from the sea to be subject to sea breezes.*

*I believe the above concerns need to be adequately addressed before the results can be published.*

*Minor comments and typos*

*There are numerous minor concerns with the paper. Some are listed below:*

*1. Line 98 and elsewhere: I believe the ACCESS model the author is using is ACCESS-R. There are other configurations (a global ACCESS-G and high resolution ACCESS-city models as well, so it is important to specify which model configuration is being used.*

*2. Lines 110-111: Although the ACCESS time steps may be in the order of 5 minutes, forecasters only get to see hourly data. It may pay to note this point.*

*3. Line 128: I think upscaled should be downscaled*

*4. Line 131-132, last sentence: A reference describing the standard approach the BoM takes would be helpful.*

*5. Pages 7 and 8: A diagram may help readers more easily understand how the perturbations are calculated.*

*6. Line 147 and elsewhere: u appears to refer to the wind vector. But in meteorology, the standard usage of u is for the west-east component of the wind vector. Perhaps a different symbol could be used to avoid confusion?*

*7. Line 151: "means" → "arithmetic means".*

*8. Page 10 and elsewhere: A map with place names would be helpful. Remember that the audience of Weather and Forecasting is international, so international readers will be less familiar with Australian geography.*

*9. Multiple places in the results section: Rather than mention UTC after all dates and times, at the beginning of the section you could mention that all dates and times are in UTC.*

*10. Figure 1:*

*1. No scale for the height colours*

*2. Inadequate place name labelling*

*3. State name abbreviations (WA, ACT etc) need defining.*

*11. Figure 2: The right hand side panels are difficult to read. The black numbers are hidden by the dark blue colouring.*

*12. Various figures and other places in the text: Wind speed units of knots are used. Although knots are commonly used in aviation (where the ICAO abbreviation is kt, i.e. different from the ISO abbreviation of kn), perhaps these units should be converted to m/s?*

*13. Figures 6-8 and elsewhere: Same dark colour problem as in Figure 2.*

*14. Figure 9 (and elsewhere): Where panels in a plot refer to different places, rather than labelling the panels a, b, c and d and then mentioning what these refer to in the caption, put a heading on each panel (e.g. Northern Territory, South Western Australia and so on).*

*15. Figure 11: The colour scale here is not very helpful; most of the boxes are the same red colour.*