Verifying Operational Forecasts of Land-Sea Breeze and Boundary Layer Mixing Processes

Ewan Short*

School of Earth Sciences, and ARC Centre of Excellence for Climate Extremes, The University of Melbourne, Melbourne, Victoria, Australia.

Ben ?. Price

Bureau of Meteorology, Casuarina, Northern Territory, Australia

Derryn ?. Griffiths and Alexei ?. Hider

Bureau of Meteorology, Melbourne, Victoria, Australia

**Corresponding author address:* School of Earth Sciences, The University of Melbourne, Melbourne, Victoria, Australia.

E-mail: shorte1@student.unimelb.edu.au

ABSTRACT

This paper presents a method for verifying the diurnally varying component of the wind forecasts issued by the Australian Bureau of Meteorology. These wind forecasts are based on model data that is then edited by human forecasters. The model datasets most commonly used by Australian forecasters for winds are those of the European Center for Medium-Range Weather Forecasting (ECMWF) and the Australian Community Climate and Earth System Simulator (ACCESS). The methodology is applied to the coastal weather stations across Australia over June, July and August 2018, at three different spatial scales, on both a daily and seasonal basis. The results indicate that while the Official forecast outperforms unedited ACCESS and ECMWF at certain locations and times of day, it rarely outperforms both at once. The causes of the differences in the performance of each dataset vary by location, but can include biases in the direction at which the sea-breeze approaches the coast, amplitude biases in the diurnal cycle, and disagreement as the whether sea-breeze or boundary layer mixing processes contribute most to the diurnal cycle. Furthermore, when winds are compared at small spatial scales on a daily basis, ECMWF outperforms Official and ACCESS simply because its coarser resolution creates less internal variability than Official or ACCESS. These results have implications for both forecasting practice and verification methodology.

1. Introduction

Modern weather forecasts are typically produced by models in conjunction with human forecasters. Forecasters working for the Australian Bureau of Meteorology (BoM) construct a seven day forecast by loading model data into a software package called the Graphical Forecast Editor (GFE), then editing this model data using tools within GFE. Is this also how things work at the U.S National Weather Service and U.K. Met Office? Forecasters can choose which model to base their forecast on, and refer to this as a choice of *model guidance*. Edits are typically made to account for processes that are under-resolved at synoptic scale model resolutions, or to correct known biases of the models being used. The resulting gridded forecast datasets are then provided to the public through the BoM's online MetEye data browser (Bureau of Meteorology 2019); the gridded forecast datasets are also translated into text and icon forecasts algorithmically.

Australian forecasters generally make two types of edits to the surface wind fields on a routine daily basis. The first is to edit the surface winds after sunrise at locations where the forecaster believes the model guidance is providing a poor representation of boundary layer mixing processes. Bounday layer mixing occurs as the land surface heats up, producing an unstable boundary layer which transports momentum downward to the surface layer, where winds are both weaker and ageostrophically oriented due to surface friction (Lee 2018). The forecaster may edit both speed and direction on the basis of climatological knowledge, theory or recent upper level wind soundings from nearby stations. How do the boundary layer mixing tools in GFE currently work? While I was in Darwin you picked a height z and a percentage p, and the tool essentially formed an average of the surface winds and winds at x weighted by p.

The second type of edit involves changing the afternoon and evening surface winds around those coastlines where the forecaster believes the model guidance is resolving the sea-breeze poorly.

How do the sea-breeze tools in GFE currently work? While I was in Darwin you traced out the relevant coastline graphically, chose a wind speed and a time, and GFE would add in winds perpendicular to the traced coastline at this speed, and smoothly blend them in spatially and temporally.

Forecasters and national weather services have good reasons for ensuring the diurnally varying component of their wind forecasts are as accurate as possible. Dai and Deser (1999) fitted the first two harmonics to seasonal averages of wind speed at different times of day and showed that over land surfaces the average amplitude of the wind speed diurnal cycle varied from 1.2 to 2.1 kn, (knots are used throughout this paper because this is the unit forecasters work with, and the unit that is used in Jive) and that the fitted harmonics accounted for 50 to 70% of the daily variability. Table 1 shows the mean wind speeds for the Australian capital city airport station shown in Fig. 1, over December, January, February 2017/18 and June, July and August 2018, suggesting that the amplitude of the mean diurnal cycles are approximately 10 to 34% of the mean wind speeds across Australia.

Beyond their contribution to the overall wind field, diurnal wind cycles are important in and of themselves to the ventilation of pollution, with sea-breezes transporting clean maritime air inland, where it helps flush polluted air out of the boundary layer (Miller et al. 2003). The Victorian Latrobe Valley provides an important Austrailan example of this effect (Physick and Abbs 1992). Furthermore, diurnal wind cycles affect the function of wind turbines (Englberger and Dörnbrack 2018) and the design of wind farms (Abkar et al. 2016), as daily patterns of boundary layer stability affect turbine wake turbulence, and the losses in wind power that result.

To our knowledge, no published work has assessed the diurnal component of human edited forecasts, although some previous studies have assessed the performance of different operational models at specific locations. Svensson et al. (2011) examined thirty different operational model simulations, including models from most major forecasting centres and utilising most commonly used boundary layer parametrisation schemes, and compared their perfomance with a large eddy simulation (LES), and observations at Kansas, USA during October 1999. They found that both the models and LES failed to capture the sudden ≈ 6 kn jump in wind speeds shortly after sunrise, and underestimated morning low level turbulence and wind speeds.

Other studies have assessed near-surface wind forecasts, verifying the total wind speeds, not just the diurnal component. Pinson and Hagedorn (2012) performed a verification study of the 10 m wind speeds resolved by the European Centre for Medium Range Weather Forecasting (ECMWF) operational model ensemble across western Europe over December, January, February 2008/09. First, they interpolated ECMWF model data onto the locations of weather stations across Europe, then they compared the interpolated model data at these stations with the station observations themselves. They found that the worst performing regions were coastal and mountainous areas, and attributed this poor performance to the small scale processes, e.g. sea and mountain breezes, that are under-resolved at ECMWF's coarse 50km spatial resolution. They noted that future work could better identify the effect of diurnal cycles on verification statistics by considering forecasts at different times of day.

Thus, the present study has two goals. First, to describe a method for comparing the diurnal cycles of human edited wind forecasts to those of unedited model guidance forecasts, in order to assess where and when human edits produce an increase in accuracy. Second, to apply this methodology across Australia. The remainder of this paper is organised as follows. Section 2 describes the methodology and datasets to which it is applied, section 3 provides results, and sections 4 and 5 provide a discussion and a conclusion, respectively.

2. Data and Methods

This study compares both human edited and non human edited Australian Bureau of Meteorology wind forecasts with automatic weather station (AWS) data across Australia. The comparison is performed by first isolating the diurnal signals of each dataset, then comparing these signals on an hour-by-hour basis. Much of the analysis is conducted through the BoM's *Jive* verification platform (URL link), which provides an archive of forecast, model guidance and observational data, and a software library for calculating basic statistics.

a. Data

Four datasets are considered in this study; the Official BoM wind forecast data that is issued to the public, model data from ECMWF (is this the mean of the ECMWF operational ensemble?), model data from the Australian Community Climate and Earth System Simulator (ACCESS), and observational data from automatic weather stations (AWS) across Australia. The Official, ECMWF and ACCESS data are at a ?, ? degree spatial resolution respectively. What are the resolutions of these datasets as they're used in Jive? Official, ACCESS and AWS data exists at each UTC hour, but ECMWF data exists at a three hour resolution Why is this? What are the actual time-steps of the models? To be consistent with the other data sets, ECMWF is therefore linearly interpolated to an hourly resolution: note that this is also what happens when forecasters load ECMWF wind data into the GFE. Furthermore, to facilitate comparison with observations, Official, ACCESS and ECMWF data is (linearly?) interpolated in all three spatial dimensions to the locations of the weather stations. Possibly worth noting here that this is the same "station-centric" approach advocated by Pinson and Hagedorn (2012).

Both ACCESS and ECMWF use parametrisation schemes to simulate sub-grid scale boundary layer mixing and turbulence. ACCESS uses the schemes of Lock et al. (2000) and Louis (1979) for unstable and stable boundary layers respectively (Bureau of Meteorology 2010). ECMWF uses similar schemes that they develop in-house (European Center for Medium Range Weather Forecasting 2018). Data covers the austral winter months of June, July and August 2018; this short time period was chosen to reduce the effect of changing seasonal and climatic conditions, changing forecasting practice and staff, and of developments to the ACCESS and ECMWF models.

Figure 1 shows the stations considered in this study. Analyses are conducted at three spatial scales, with stations grouped accordingly. The finest scale is that of the individual station. This study focuses on the 8 capital city airport stations, marked by stars in Fig. 1, as their high operational significance means that they are typically the most accurate and most well maintained. The next spatial scale is formed by taking the 10 stations closest to each capital city airport station, with some flexibility allowed to ensure stations are roughly parallel to the nearest coastline. These station groups are referred to as the *airport station groups*. The coarsest spatial scale is formed by taking all stations within 150 km of the nearest coastline, and grouping these by state. This is done because Australian forecasts are currently produced on a state by state basis at forecasting centres based in each state capital, with each forecasting centre utilising different staff, different model guidance preferences, and different editing practices. Indeed, the Official gridded forecast typically shows slight discontinuities across state boundaries (Bureau of Meteorology 2019). Note that the Western Australian coastline is subdivided into three pieces, and stations along the Gulf of Carpentaria, north Queensland Peninsula, and Tasmanian coastlines are neglected, in order to ensure each station group corresponds to an approximately linear segment of coastline. These eight station groups are referred to as the *coastal station groups*.

b. Assessing Diurnal Cycles

Although close to coastlines we expect the land-sea breeze to be the dominant diurnal wind process, the overall diurnal signal may also include boundary layer mixing processes, mountain-valley breezes, atmospheric tides, and urban heat island circulations. Forecasters typically edit model output to account for under-resolved sea-breezes and boundary layer mixing processes. Instead of attempting to assess each type of edit individually, we study the overall diurnal signal by subtracting a twenty hour centred running mean *background wind* from each zonal and meridional hourly wind data point. This provides a collection of zonal and meridional wind *perturbation* datasets.

The accuracy of the Official, ACCESS and ECMWF diurnal cycles is then quantified by comparing the Euclidean distances of the perturbations at each hour with the corresponding AWS perturbations. For example, to assess whether the Official forecast perturbations $\boldsymbol{u}_{\mathrm{O}}$ or ACCESS perturbations $\boldsymbol{u}_{\mathrm{A}}$ best match the AWS observations $\boldsymbol{u}_{\mathrm{AWS}}$, we calculate the *Wind Perturbation Index* (WPI), defined by

$$\mathrm{WPI}_{\mathrm{OA}} = |\boldsymbol{u}_{\mathrm{AWS}} - \boldsymbol{u}_{\mathrm{A}}| - |\boldsymbol{u}_{\mathrm{AWS}} - \boldsymbol{u}_{\mathrm{O}}|. \tag{1}$$

At a given time, the Official forecast wind perturbation is closer to the AWS perturbation than that of ACCESS if and only if $\mathrm{WPI}_{\mathrm{OA}} > 0$. The analogously defined quantities $\mathrm{WPI}_{\mathrm{OE}}$ and $\mathrm{WPI}_{\mathrm{EA}}$ can then be used to provide a comparison of the Official and ECMWF perturbations, and of the ACCESS and ECMWF perturbations, respectively. To asses which dataset provides, in general, the most accurate representation of the diurnal cycle over the study period of June, July and August 2018, we take means of the WPI on an hourly basis; i.e. all the 00:00 UTC WPI values are averaged, all the 01:00 UTC values are averaged, and so forth, and denote such an average by $\overline{\mathrm{WPI}}$.

Given the large degree of turbulence and unpredictable variability in both the AWS, Official, and model datasets, care must be taken to ensure we do not pre-emptively conclude Official has outperformed the model guidance when $\overline{\text{WPI}} > 0$ purely by chance. The method for estimating confidence in $\overline{\text{WPI}}$ is based on a method used by Griffiths et al. (2017) for precipitation. Note that WPI is defined so as to minimise the temporal autocorrelations within each dataset, and to avoid having to consider correlations between the zonal and meridional components within and between datasets. Time series formed from the WPI values at a particular time, say 00:00 UTC, across the three month time period, can therefore be idealised as an independent random sample of a random variable W. The sampling distribution for each $\overline{\text{WPI}}$ can then be modelled by a Student's t-distribution, and from this we can calculate the probability that W is positive, denoted $\Pr(W > 0)$. Although temporal autocorrelations of WPI, i.e. correlations between WPI values at a particular hour from one day to the next, are in practice small or non-existent thanks to how WPI is defined, they are still accounted for by reducing the "effective" sample size to $n(1-\rho_1)/(1+\rho_1)$, where n is the actual sample size and ρ_1 is the lag-1 autocorrelation (Zwiers and von Storch 1995; Wilks 2011). Note that in the standard language of statistical hypothesis testing, we would reject the null hypothesis that $W = 0$ at significance level α if $\Pr(W > 0) > 1 - \frac{\alpha}{2}$ or $\Pr(W < 0) > 1 - \frac{\alpha}{2}$. However, in this study we are interested in both whether $W > 0$ or whether $W < 0$, so prefer to simply state the value of $\Pr(W > 0)$, referring to this as a *confidence score*, and noting $\Pr(W < 0) = 1 - \Pr(W > 0)$. Much of this discussion is probably unnecessary, but would like to get feedback before I trim it.

To assess how well the diurnal perturbations of an overall region are predicted, for instance those of the Victorian coastal station group (see Fig. **??**), the perturbations across each station group are averaged before WPI values calculated. The temporal means and sampling distributions of the WPI are then calculated as before, with each value of WPI calculated from the spatially averaged perturbations treated as a single observation. This provides a conservative method for dealing with spatial correlation in the perturbations.

The advantage of the WPI method is it's clarity and simplicity: we are essentially just comparing the magnitudes of vector differences, then applying a two sided t-test to determine whether one dataset's perturbations are consistently closer to observations than another's. One factor that complicates interpretation of statistics of WPI, is that the near surface winds observed in AWS data are consistently noisier than those of the Official, ECMWF and ACCESS forecasts. This is likely due to unresolved subgrid scale turbulence in the Official, ECMWF and ACCESS model datasets. It would be unreasonable to expect forecasters to be able to predict this essentially random additional observed variability, and so a direct comparison of observed and modelled diurnal cycles is overly stringent.

Note that subtracting background winds may raise concerns, because perturbations obviously depend on background winds. However, the forecaster does not have knowledge of the observations when they make the diurnal process edits. They are implicitly assuming that the true mean state will be close enough to the predicted mean state - however this prediction is produced - to justify making diurnal edits on the basis of the predicted mean state.

To reduce the significance of unpredictable noise, we also compare temporal averages of the perturbations for each dataset. These comparisons have less operational significance: people generally care how well the actual weather forecast performed, not whether the average of a predicted quantity matched the average of an observed quantity. However, comparisons of averages arguably better represent what we can realistically expect from human forecaster edits, and from weather forecasts overall, particularly in regards to small scale processes like sea-breezes. Furthermore, when temporal averages of perturbations are considered, the diurnal signal becomes dramatically clearer, and structual differences become much easier to diagnose.

To quantify how closely the temporally averaged Official forecast perturbations match those of the AWS observations, we calculate $|\overline{\boldsymbol{u}}_{\mathrm{AWS}} - \overline{\boldsymbol{u}}_{\mathrm{O}}|$ for each hour. To assess the performance of the Official temporally averaged perturbations against those ACCESS, we define the *Climatological Wind Perturbation Index* (CWPI)

$$\mathrm{CWPI}_{\mathrm{OA}} \equiv |\overline{\boldsymbol{u}}_{\mathrm{AWS}} - \overline{\boldsymbol{u}}_{\mathrm{O}}| - |\overline{\boldsymbol{u}}_{\mathrm{AWS}} - \overline{\boldsymbol{u}}_{\mathrm{A}}|\,. \tag{2}$$

As with the WPI, the CWPI can also be used to provide a comparison of the Official and ECMWF datasets, or a comparison of the two model guidance datasets ACCESS and ECMWF. Uncertainty in the CWPI is estimated through bootstrapping (Efron 1979). This is done by performing resampling with replacement on the underlying perturbation datasets, and calculating the CWPI multiple times using these resampled datasets. This provides a distribution of CWPI values, from which the probability that CWPI > 0 can be calculated. Similarly to with the WPI, performance over a particular region can be assessed by first averaging perturbation values over multiple stations before the CWPI is calculated.

Although the WPI and CWPI provide quantitive information on the accuracy of the diurnal cycle at different times of day, they do not provide much information about the structure of the diurnal wind cycles of each dataset, or provide insight into the reason one dataset is outperforming another. Gille et al. (2005) obtained summary statistics on the observed structure of temporally averaged diurnal wind cycles across the globe by using linear regression to calculate the coefficients u_i, v_i $i = 0, 1, 2$, for the elliptical fit

$$u = u_0 + u_1 \cos(\omega t) + u_2 \sin(\omega t), \tag{3}$$

$$v = v_0 + v_1 \sin(\omega t) + v_2 \sin(\omega t), \tag{4}$$

where ω is the angular frequency of the earth and t is the local solar time in seconds. Descriptive quantities - like the angle the semimajor axis of the ellipse makes with the horizontal - were then calculated directly from the coefficients u_1, u_2, v_1 and v_2.

Gille et al. (2005) applied this fit to satellite scatterometer wind observations, which after temporal averaging provided only four temporal datapoints at each $0.25° \times 0.25°$ spatial grid cell. As such, their fit was very good, explaining over 90% of the wind variability in each spatial gridcell. However, the choice of ellipse parametrisation in equations 5 and 6 assumes that datapoints lie on the ellipse at equal intervals of time t. When observational or model data with an hourly or smaller timestep is considered, this assumption becomes too stringent, as heating asymmetries imply that wind perturbations evolve much more rapidly during the day than at night (see Fig. XX). Note I'm also basing this point on knowledge of the land vs sea breeze, and knowledge of heating vs cooling asymmetries (Brown et al. 2017, e.g.).

Thus, we model the climatological diurnal cycles with the equations

$$u = u_0 + u_1 \cos(\alpha(\psi,t)) + u_2 \sin(\alpha(\psi,t)), \tag{5}$$

$$v = v_0 + v_1 \sin(\alpha(\psi,t)) + v_2 \sin(\alpha(\psi,t)), \tag{6}$$

with α the function from $[0,24) \times [0,2\pi) \to [0,2\pi)$ given by

$$\alpha(\psi,t) \equiv \pi \left[\sin\left(\pi \frac{(t-\psi) \bmod 24}{24} - \frac{\pi}{2} \right) + 1 \right], \tag{7}$$

where t is time in units of hours UTC, and ψ gives to the time when the wind perturbations vary least with time. Need to confirm whether least or most! For each climatological diurnal wind cycle, we solve for the seven parameters u_0, u_1, u_2, v_0, v_1, v_2 and ψ using nonlinear regression.

Descriptive quantities can then be calculated from these parameters. The value of α at which the winds align with the semimajor axis, α_M, satisfies

$$\alpha_M = \frac{1}{2}\arctan\left(\frac{2(u_1u_2+v_1v_2)}{u_1^2+v_1^2-u_2^2-v_2^2}\right) \bmod \pi, \tag{8}$$

The time at which the perturbations align with the major axis t_M can then be calculated by inverting equation (7), fixing ψ to the value obtained from the nonlinear regression. The lengths of the semimajor and semiminor axes, and the angle the semimajor axis makes with lines of latitude ϕ, can then be calculated from α_M using the same expressions as Gille et al. (2005).

3. Results

In this section, the methods described in section **??** are applied to Australian forecast and station data over the months of June, July and August (austral winter) 2018. First, error is assessed on a daily basis using the Wind Perturbation Index (WPI) at three different spatial scales. Second, overall seasonal biases during this time period are assessed using the Climatological Wind Perturbation Index CWPI, and by comparing quantities derived from ellipses fitted to the climatological wind perturbations. Unless otherwise stated, values throughout this section are provided to two significant figures.

a. Daily Comparison

Figure 2 provides the mean wind perturbation index values $\overline{\text{wpi}}$ and confidence scores $P\left(\overline{\text{WPI}} > 0\right)$ for the coastal station groups for $\overline{\text{wpi}}_{\text{OA}}$, $\overline{\text{wpi}}_{\text{OE}}$ and $\overline{\text{wpi}}_{\text{EA}}$, which represent the the Official versus ACCESS, Official versus ECMWF, and ECMWF versus ACCESS comparisons, respectively. Values of $\overline{\text{wpi}}_{\text{OA}}$ and $\overline{\text{wpi}}_{\text{OE}}$ are negative for the majority of station groups and hours, and often both $P\left(\overline{\text{WPI}}_{\text{OA}} > 0\right) < 5\%$ and $P\left(\overline{\text{WPI}}_{\text{OE}} > 0\right) < 5\%$. This implies that at this level of spatial aggregation, there is often high confidence that both the unedited ACCESS and ECMWF models outperform the Official forecast. The lowest $\overline{\text{wpi}}$ values of -0.9 kn occur for the NT station group at 23:00 and 00:00 UTC for both $\overline{\text{wpi}}_{\text{OA}}$ and $\overline{\text{wpi}}_{\text{OE}}$, with $\overline{\text{wpi}}_{\text{EA}} = 0$ kn. Comparatively low values also occur at 08:00 UTC with $\overline{\text{wpi}}_{\text{OA}} = \overline{\text{wpi}}_{\text{OE}} = -0.6$ kn, but $\overline{\text{wpi}}_{\text{EA}} = 0$ kn. This suggests the Official forecast may be performing particularly poorly over the NT station group.

Although Official outperforms at least one of ACCESS or ECMWF with high confidence at a few dozen times and station groups, there is only one group and time where it outperforms both. At 05:00 UTC over the South WA station group, $\overline{\text{wpi}}_{\text{OA}} = 0.2$ kn and $\overline{\text{wpi}}_{\text{OE}} = 0.1$ kn, both with confidence scores $\geq 95\%$, although the actual $\overline{\text{wpi}}$ values are comparatively small. Note that ECMWF generally outperforms ACCESS from 10:00 - 14:00 UTC, with the South WA station group being the main exception.

Using the NT and South WA station groups as case studies, Figures 3 a) and b) provide time series of wpi_{OA} and wpi_{OE} for, a), the NT station group at 23:00 UTC, and b), the South WA station group at 05:00 UTC. The wpi_{OA} and wpi_{OE} values for the NT station group show significant temporal variability over the three month period, exceeding -2 kn on at least 10 days each, and occasionally becoming positive. The wpi values for the South WA station at 05:00 UTC also show significant temporal variability, with wpi_{OA} and wpi_{OE} each exceeding 1 kn on at least 9 seperate days, despite $\overline{\text{wpi}}_{\text{OA}}$ and $\overline{\text{wpi}}_{\text{OE}}$ being small.

Fig. 3 a) shows that there are four days where wpi_{OA} and wpi_{OE} are both less than -2 kn: the 8$^{\text{th}}$ of June and the 3$^{\text{rd}}$, 9$^{\text{th}}$ and 10$^{\text{th}}$ of July. Figures 3 c) and d) show hodographs of the winds and wind perturbations, respectively, at each hour UTC for the AWS observations, Official forecast, and ACCESS and ECMWF model datasets on the 3$^{\text{rd}}$ of July, which provides an interesting example. Figure 3 e) shows that the Official wind forecast on this day was likely based on edited ACCESS from 00:00 to 06:00 UTC, then edited ECMWF from 07:00 to 13:00 UTC, then unedited ACCESS from 15:00 to 21:00 UTC. The final two hours of the forecast show the Official winds acquiring a stronger east-northeasterly component than either the AWS observations, ACCESS, or ECMWF; this rapid, exaggerated change is even clearer in the perturbation hodograph shown in Fig. 3 f). Note that at this time of year the prevailing winds throughout the NT are east-southeasterly, and 22:00 UTC corresponds to $\approx$ 08:30 LST in this region, so the rapid departure of the Official forecast from ACCESS at this time likely represents an edit made by a forecaster to capture boundary layer mixing processes. Figure 4 a) shows the first ten values from wind soundings at Darwin Airport - the nearest station to issue vertical wind soundings - at 12:00 UTC on July 3$^{\text{rd}}$ and 00:00 UTC on July 4$^{\text{th}}$. In both instances the winds are indeed east-southeasterly, and so the rapidly changing wind perturbations at 22:00 UTC in the Official forecast likely reflect a boundary layer mixing edit that has been applied either too early, or has strengthened the south-easterly component of the winds too much. The 8$^{\text{th}}$ of June and 9$^{\text{th}}$ and 10$^{\text{th}}$ of July examples are all similar in this respect.

Considering now the South WA station group, Fig. 3 b) shows that wpi_{OA} and wpi_{OE} both exceed 1 kn on the 9$^{\text{th}}$ of June and the 3$^{\text{rd}}$ of August. Figures 3 c) and d) show hodographs of the winds and wind perturbations, respectively, at each hour UTC for the AWS observations, Official forecast, and ACCESS and ECMWF model datasets on the 9$^{\text{th}}$ of June, which is the more interesting example. The perturbation hodograph shows both ECMWF and ACCESS underpredicting the amplitude of the diurnal wind cycle on this day. In each dataset the 05:00 UTC perturbations are westerly to northwesterly, and given the orientation of the South WA coastline (see Fig. 1) and the fact that 05:00 UTC corresponds to $\approx$ 13:00 local solar time (LST) in this region, the perturbations likely indicate boundary layer mixing processes, rather than the land-sea breeze. Furthermore, the AWS perturbations rapidly become northwesterly between 01:00 and 02:00 UTC, $\approx$ 09:00 - 10:00 LST, which would be about three hours after the sun has risen, consistent with a boundary layer mixing mechanism.

Figure 4 provides hodographs of wind with height throughout the first two km of the atmosphere between 12:00 UTC on the 8th June and 12:00 UTC on the 9th June; the soundings were taken at Perth Airport, which is the nearest station to the South WA station group to provide wind soundings. The 8th June 12:00 UTC hodograph shows surface northerlies of ≈ 6 kn, becoming west to northwesterlies of over 20 kn 2.4 km above the surface. A forecaster basing a model edit of the following days winds on this sounding would therefore gradually strengthen the westerly component of the surface winds in the hours after sunrise. However, the subsequent sounding at 00:00 UTC on the 9th of June shows that the winds acquire a strong northerly component of 30 kn in the first 500 m of the atmosphere, with the final sounding indicating a strong northwesterly wind at 725 m persisting until 12:00 UTC. In Fig. 3 d), the Official perturbations from 04:00 to 07:00 UTC show stronger westerly perturbations than either ACCESS or ECMWF, improving the amplitude of Official's diurnal wind cycle. However, the AWS perturbations are more northerly than those of Official, and so the Official forecast winds have been strengthened in a slightly incorrect direction. An explanation for this discrepancy is that the Official forecast for the southwest region of WA has been edited based on the June 8th 12:00 UTC Perth Airport sounding, with the winds above the surface changing direction in the subsequent 12 hours. Note that the 3rd of August example is similar, although in this case the Official forecast slightly improves both the magnitude and direction of the 05:00 UTC wind perturbations.

Figure 5 presents the $\overline{\text{wpi}}$ values and confidence scores for the Official versus ECMWF comparisons, i.e. $\overline{\text{wpi}}_{\text{OE}}$ and $P\left(\overline{\text{WPI}}_{\text{OE}} > 0\right)$, for the airport stations, and airport station groups. The results for the airport stations are noisier than the analogous results for the coastal station groups in Figures 2 c) and d), although they do share some similarities. Official outperforms ECMWF at 01:00 and 02:00 UTC at both the Darwin airport station and the NT station group, although ECMWF outperforms Official between 08:00 and 14:00 UTC at Darwin and Brisbane airports, and the corresponding NT and QLD station groups, with the exception of the QLD station group at 12:00 UTC where $\overline{\mathrm{wpi}}_{\mathrm{OE}} = 0$. ECMWF also outperforms Official at Hobart airport at almost all hours of the day, and at Adelaide and Canberra airports from 11:00 to 14:00 UTC.

For the remaining stations and times, only the Perth airport station at 06:00 UTC and the Melbourne airport station at 01:00 UTC exhibit $\overline{\mathrm{wpi}}_{\mathrm{OE}} > 0$ with $P\left(\overline{\mathrm{WPI}}_{\mathrm{OE}} > 0\right) \geq 95\%$. However, in both cases $\overline{\mathrm{wpi}}_{\mathrm{OE}} = 0.3$, which is small compared to the maximum value of 1.0 which occurs at the Darwin airport station at 02:00 UTC. Furthermore, in both cases there is no clear pattern to the $\overline{\mathrm{wpi}}_{\mathrm{OE}}$ values over the rest of the day. Given the random appearance of the $\overline{\mathrm{wpi}}_{\mathrm{OE}}$ values, the *multiplicity problem* (Wilks 2011, p. 178) requires care be taken before giving meaning to these two examples: i.e., given that we are calculating twenty four confidence scores for eight stations, then assuming WPI were uncorrelated across each station and hour we would expect to find $0.05 \times 24 \times 8 \approx 10$ instances where $P\left(\overline{\mathrm{WPI}}_{\mathrm{OE}} > 0\right) \geq 95\%$, even if $\overline{\mathrm{WPI}}_{\mathrm{OE}}$ was in fact equal to zero. Comment on performance versus ACCESS.

For the airport station groups, ECMWF outperforms Official for the majority of station groups and times. The main exception is the Darwin airport station group, where Official outperforms ECMWF at 02:00 UTC, and there is ambiguity as to whether Official or ECMWF performs better at 01:00, 03:00 and 04:00 UTC, and from 15:00 to 22:00 UTC. In the analogous comparisons of Official and ACCESS (not shown), the airport station results are similarly noisy, although the airport station group results are slightly more favourable to Official, with Official outperforming ACCESS from 10:00 to 12:00 UTC at the Brisbane station group, and fewer occasions overall where ACCESS outperforms Official than ECMWF does.

Figure 5 shows the $\overline{\mathrm{wpi}}$ values and confidence scores for the ECMWF versus ACCESS comparisons, i.e. $\overline{\mathrm{wpi}}_{\mathrm{EA}}$ and $P\left(\overline{\mathrm{WPI}}_{\mathrm{EA}} > 0\right)$, for the airport stations, and airport station groups. As with the Official versus ECMWF comparison in Fig. 5, the results for the airport stations are noisy, but more often than not show that ECMWF outperforms ACCESS. The results for the airport station group show ECMWF usually outperforms ACCESS, the main exceptions being the Darwin and Canberra airport station groups.

At face value, the fact that ECMWF generally outperforms ACCESS at these scales is surprising, as ACCESS runs at a higher spatiotemporal resolution than ECMWF, and is calibrated for Australian conditions, so one would expect ACCESS would better resolve small scale processes like the land-sea breeze and boundary layer mixing processes. However, these results are unsurprising if one considers the scales at which predictable atmospheric motion occurs, and the scales being resolved by AWS, ACCESS and ECMWF. The AWS data resolves motion with time scales as low as 10 minutes, and arbitrarily small spatial scales: it therefore includes highly unpredictable eddy turbulence. This explains why the results for the airport stations are noisier than for the airport station groups or coastal station groups. Furthermore, because ACCESS runs at a higher resolution than ECMWF, it includes additional scales of motion, and therefore adds additional variability to the wind fields. Unless this additional variability in ACCESS is perfectly correlated with observations, the average of $|\boldsymbol{u}_{\text{AWS}} - \boldsymbol{u}_{\text{A}}|$ will therefore increase, unless this additional variability is compensated for by a reduction in bias, i.e.$|\overline{\boldsymbol{u}}_{\text{AWS}} - \overline{\boldsymbol{u}}_{\text{A}}|$ decreases. These ideas are discussed in greater detail in section 4. Note finally that the results for the Official versus ECMWF comparison in Fig. 5 largely mirror those of the ECMWF versus ACCESS comparison in Fig. 5, e.g. for the Darwin airport station and station group, Official outperforms ECMWF at the same times that ACCESS does, suggesting that either the Official forecast at these spatial scales is largely based on ACCESS, or that ECMWF is highly biased at these scales and times.

b. Seasonal Comparison

Figure 5 provides the climatological wind perturbation index values, cwpi, and confidence scores, $P(\mathrm{CWPI} > 0)$, for the coastal station groups for $\mathrm{cwpi}_{\mathrm{OA}}$, $\mathrm{cwpi}_{\mathrm{OE}}$ and $\mathrm{cwpi}_{\mathrm{EA}}$, which represent the the Official versus ACCESS, Official versus ECMWF, and ECMWF versus ACCESS comparisons, respectively. At the NT station group Official outperforms both ACCESS and ECMWF at 03:00 UTC with $\mathrm{cwpi}_{\mathrm{OA}} = \mathrm{cwpi}_{\mathrm{OE}} = 0.4$, $P(\mathrm{cwpi}_{\mathrm{OA}} > 0) = 94\%$ and $P(\mathrm{cwpi}_{\mathrm{OE}} > 0) = 93\%$. However, both ACCESS and ECMWF outperform Official at 23:00 and 00:00 UTC, consistent with the $\overline{\mathrm{wpi}}$ results in Fig. 2. The NT station group results are discussed in more detail in section 4.

At the North WA station group at 01:00, 03:00 and 04:00, Official outperforms ACCESS with confidence scores of 77, 78 and 90%, respectively; Official also outperforms ECMWF at 01:00 and 02:00 UTC with confidence scores above 99%. Figure 6 a) shows that ECMWF's poor performance at 01:00 and 02:00 UTC is simply due to its linear interpolation at these times, whereas Official's outperformance of ACCESS at 01:00, 03:00 and 04:00 is due to ACCESS's climatological diurnal cycle being slightly out of phase with that of the AWS observations, and the Official forecast appearing to correct for this somewhat. Both Official and ECMWF slightly exaggerate the magnitude of the climatological sea-breeze with ACCESS doing a good job in this regard.

At the South WA station group from 01:00 to 05:00 UTC, $\mathrm{cwpi}_{\mathrm{OE}}$ is positive with confidence scores of at least 88%, although $\mathrm{cwpi}_{\mathrm{OA}}$ is negative or zero at these times. Figure 6 b) shows that ECMWF underestimates the westerly perturbations at these times, with these perturbations likely associated with boundary layer mixing processes, as discussed in section a. Each of Official, ACCESS and ECMWF underestimate the amplitude of the diurnal cycle between 02:00 and 10:00 UTC, including both the westerly perturbations and the southerly sea-breeze perturbations.

At the NSW station group from 17:00 to 19:00 UTC, cwpi_{OA} and cwpi_{OE} are at least 0.4 and 0.1 kn, respectively, with confidence scores of at least 95% and 75%, respectively. Figure 6 c) shows that these times correspond to a strange "dimple" in perturbation hodograph that is present in all four datasets. The Official hodograph closely resembles that of ACCESS, except for this dimple, which has been exaggerated relative to ACCESS. Don't know what is going on here. Figure 6 c) also shows that although ECMWF exaggerates the amplitude of the easterly sea-breeze perturbations, it captures the narrower shape of the AWS hodograph better than Official or ACCESS.

At the SA station group from 01:00 to 05:00 UTC and 09:00 to 11:00 UTC both cwpi_{OA} and cwpi_{OE} are positive, with maximum values of 0.4 and 0.1 kn, although confidence scores do not exceed 88% and 65% respectively. Figure 6 shows that the Official forecast captures the amplitude of the perturbations from 01:00 to 05:00 UTC almost perfectly, matching the amplitude of the AWS perturbations better than both ACCESS and ECMWF. However, the Official diurnal cycle is slightly out of phase with the AWS cycle during this period, explaining why Official only slightly outperforms ACCESS in the results of Figures 5 a) and b).

While the cwpi values and confidence scores of Fig. 5 provide detailed information on which dataset's climatological diurnal cycle best matches those of the AWS observations, cwpi on it's own reveals little about the structure of the diurnal cycle, and provides little insight into forecast accuracy could be improved. Note that the hodographs in Fig. 6 are roughly elliptical in shape, suggesting that descriptive quantities can be estimated by fitting equations (5) and (6) to the zonal and meridional climatological perturbations, then calculating these quantities from the fit, as described in section 2.

Figure 8 provides the R^2 values for the fits of the zonal and meridional perturbations to equations (5) and (6), respectively. The fit performs best at the coastal station group spatial scale, with R^2 generally above 95%. It also performs well at the airport station and airport station group scales, with a few exceptions, including the ACCESS and Official meridional perturbations at the Canberra airport station group, and the ECMWF zonal perturbations at Melbourne airport.

The ellipse fits are used to derive four descriptive quantities: amplitude (half the length of the semi-major axis), eccentricity, orientation (the angle the semi-major axis makes with lines of latitude) and the time of the peak in the diurnal cycle (the time at which the perturbations align with the semi-major axis, ignoring translational coefficients). Figure 9 provides these four quantities for each dataset and location across the three spatial scales. A variety of structural differences are apparent at a number of locations and scales. For example, Fig. 9 a) shows that at Brisbane airport, the amplitude of the AWS diurnal cycle is at least 1 kn greater than Official, ACCESS and ECMWF, and Fig. 9 c) shows that the orientation of the AWS diurnal cycle hodograph is at least 20 degrees (anti-clockwise) from the other datasets. Figures 10 a) and b) show hodographs of the Brisbane airport perturbation climatology and ellipse fit, respectively. Although the ellipse fit suppresses some of the asymmetric details, it captures the amplitudes and orientations of the real climatological diurnal cycles well. In this case the results show that the average AWS sea-breeze approaches from the northeast, whereas the forecast and model sea-breezes approach more from the east-northeast. To check whether this just represents a direction bias of the Brisbane Airport station, Fig. 9 shows the climatological perturbations at the nearby Spitfire Channel station (see Fig. **??** for the location of this station, and other stations referred to in this section). While the amplitude bias is smaller at Spitfire Channel than Brisbane Airport, the directional bias is at least as high; a similar directional bias is evident at the nearby Inner Beacon station, although the bias is smaller than at Spitfire Channel and Brisbane Airport. Thus, the directional bias in Official, ACCESS and ECMWF at these stations is likely genuine, and not just a consequence of biased AWS observations. Figure 1 x) shows there are two small islands to the east of Brisbane airport; the more northwesterly orientation of the Brisbane Airport sea-breeze suggests these islands may be channelling winds between the east coast of Brisbane and the west coasts of these islands, and that this local effect is not being captured in Official, ACCESS or ECMWF.

Another example is the Hobart Airport station. Figure 9 c) shows that the ellipse fits for the AWS perturbations are oriented 31, 35 and 62 degrees anti-clockwise from the ECMWF, Official and ACCESS ellipse fits, respectively. Figures 8 a) and b) show that the ellipse fit for the AWS perturbations at Hobart airport only achieve R^2 values of 59% and 68% for the u and v components, respectively, although figures 10 d) and e) show that the fit still captures the orientation accurately; the deficiency is more with the amplitude of the AWS diurnal cycle. Figure 8 c) shows the climatological perturbations at the Hobart (city) station, which also show a large difference in orientation between ACCESS and AWS. Given the timing of the westerly perturbations in ACCESS, and the fact that the prevailing winds around Tasmania are Hobart, these results suggest that ACCESS is exaggerating the boundary layer mixing processes involved in the diurnal cycle, whereas ECMWF better captures the southerly sea-breeze component of the cycle.

The South WA station group also provides an interesting example. Here the ACCESS and Official ellipse fits are oriented at least 49 degrees anti-clockwise from those of AWS and ECMWF, and the time of the peak in the diurnal cycles of ACCESS and Official is at least 4.3 hours earlier than AWS and ECMWF. This occurs because eccentricity values are low for this station group, and Figure 6 b) shows that the westerly perturbations associated with boundary layer mixing are slightly faster than the corresponding southerly sea-breeze perturbations, which peak later, for both ACCESS and Official, but slightly slower for ECMWF and Official. A similar issue affects the VIC station group, explaining why the AWS ellipse fit is oriented at least 49 degrees anti-clockwise from those of the other datasets.

Finally, figure 9 suggests that at the Darwin Airport, Darwin Airport station group, and NT station group, the AWS wind perturbations align with the semi-major axis after those of the other datasets, and in the case of the NT station group alignment occurs at least 2.3 hours later; furthermore, the amplitude of the Official ellipse fit is in each case higher than those of the other datasets. “Alignment” is probably the wrong word here. Figure 11 shows that these biases are indeed evident in the perturbation climatologies themselves, with the exception of the Darwin Airport amplitude bias, where the asymmetric hodograph shapes lead to the ellipse fit underestimating the amplitude of the AWS diurnal cycle Needs to be clarified to better distinguish between “ellipse” amplitude and diurnal cycle amplitude. Furthermore, should we interpret the NT station group results as genuine evidence of a timing bias?

4. Discussion

The two most important results of section 3 to explain are, first, why equations 5 and 6 provide such a good fit to the climatological perturbations and, second, why there are such substantial changes in the performance of the Official forecast at the different spatial and temporal scales.

The idea that diurnal wind cycles can be described by ellipses originated with Haurwitz (1947). Haurwitz obtained exact solutions for u and v resembling equations (3) and (4) for the simple model

$$\frac{du}{dt} - fv + ku = F_x - F(t) \tag{9}$$

$$\frac{dv}{dt} + fu + kv = F_y \tag{10}$$

where u, v are taken in a coordinate system where the u axis is normal to the coast, f is the Coriolis parameter, k is a linear friction coefficient, (F_x, F_y) represents a constant synoptic scale pressure gradient force, ω is the angular frequency of earth’s rotation, and

$$F(t) = \frac{A}{\pi} + \frac{A}{2}\cos(\omega t) \tag{11}$$

is the pressure gradient force normal to the coastline induced by the diurnally varying air temperature contrast over the land and sea surfaces. Kusuda and Alpert (1983) extended this model slightly, but the fundamentals are the same. The limitations of this model are discussed extensively by Haurwitz (1947) and Kusuda and Alpert (1983), but the most important given the results of section 3 involve the choice of $F(t)$, which does not sufficiently capture the asymmetries in daytime heating and nighttime cooling, and the fact that the model has no vertical dimension, and therefore cannot capture the boundary layer mixing processes that in many cases play a significant role in the diurnal wind cycle. This might work better in the introduction or methods sections - there's not actually very much to discuss!

I tried modifying the pressure perturbation terms $\frac{A}{\pi}+\frac{A}{2}\cos(\omega t)$ so that the new ellipse fit of equations (5) and (6) become solutions to equations (9) and (10), but with no luck. For example, simply changing to $F(t)$ to $\frac{A}{\pi}+\frac{A}{2}\cos(\alpha(\psi,t))$ doesn't work, nor does expanding this expression as a Fourier series and solving each term individually.

The second result of section 3 that requires explanation are the differences in the performance of the Official forecast at the different spatial and temporal scales. Consider first just the zonal components of the AWS and Official wind perturbations, denoted by u_{AWS} and u_{O} respectively. Considering just the values at a particular hour UTC, at a particular station, over the entire June, July, August time period, the mean square error $\text{mse}(u_{\text{AWS}},u_{\text{O}})=\overline{(u_{\text{AWS}}-u_{\text{O}})^2}$ can be decomposed

$$\text{mse}(u_{\text{AWS}},u_{\text{O}})=\underbrace{\text{var}(u_{\text{AWS}})+\text{var}(u_{\text{O}})-2\cdot\text{covar}(u_{\text{AWS}},u_{\text{O}})}_{\text{var}(u_{\text{AWS}}-u_{\text{O}})}+\underbrace{(\bar{u}_{\text{AWS}}-\bar{u}_{\text{O}})^2}_{\text{bias}^2} \qquad (12)$$

where var, covar and over-bars denote the sample variance, covariance and mean respectively. The first three terms are the total variance of $u_{\text{AWS}}-u_{\text{O}}$, whereas the last term is the square of the bias between u_{AWS} and u_{O}. This decomposition can also be applied to wind perturbations that have first been spatially averaged over a station group, and to mse$(u_{\mathrm{AWS}}, u_{\mathrm{E}})$ and mse$(u_{\mathrm{AWS}}, u_{\mathrm{A}})$, where u_{E} and u_{A} are the ECMWF and ACCESS zonal perturbations, respectively.

Figure 12 shows each term in the mean square error decomposition of equation 12 for both mse$(u_{\mathrm{AWS}}, u_{\mathrm{O}})$ and mse$(u_{\mathrm{AWS}}, u_{\mathrm{E}})$, for Darwin Airport, the Darwin station group, and the NT station group. At Darwin Airport, mse$(u_{\mathrm{AWS}}, u_{\mathrm{O}})$ exceeds mse$(u_{\mathrm{AWS}}, u_{\mathrm{E}})$ from 04:00 to 16:00 UTC due to higher total variance, whereas outside of these times mse$(u_{\mathrm{AWS}}, u_{\mathrm{E}})$ exceeds mse$(u_{\mathrm{AWS}}, u_{\mathrm{O}})$ due to larger bias. The higher total variance of $u_{\mathrm{AWS}} - u_{\mathrm{O}}$ occurs because var$(u_{\mathrm{O}}) >$ var(u_{E}), with this additional variability mostly random from 04:00 to 14:00 UTC, i.e. u_{O} is not sufficiently correlated with u_{AWS} at these times for the additional variability of u_{O} to produce a reduction in mean square error. Thus, while the bias between Official and AWS is lower, or about the same, as that between ECMWF and AWS, the higher random variability of Official results in higher mean square error for most of the day. Figure 13 shows similar conclusions can be drawn for the meridional perturbations at Darwin Airport, although in this case var$(u_{\mathrm{O}}) >$ var(u_{E}) for the entire day. Most of the difference between the WPI and CWPI scores for the Official versus ECMWF comparison at Darwin Airport in Figures 5 and 7, respectively, can be explained through the different mean square error and bias terms for the zonal perturbations alone. Figure 11 a) shows that ECMWF's climatological perturbations underestimate the easterly perturbations from 00:00 to 03:00 UTC, which are presumably associated with boundary layer mixing processes. Official does a better job of resolving these easterly perturbations, but is generally outperformed by ECMWF in resolving the northerly sea-breeze perturbations. Similar points can be made for the Darwin and NT coastal station groups. While spatial averaging reduces a portion of the unpredictable variability in Official, Official also often has larger meridional biases at these scales compared to ECMWF. Figures 11 and 9 show that these biases can be explained in terms of amplitude and orientation differences between Official, ECMWF and AWS.

These examples illustrate the idea that the additional unpredictable variability introduced by a higher resolution edited forecast needs to be "paid for" by a reduction in bias, otherwise the net result will just be an increase in error. However, although a high resolution edited forecast may have higher mean squared error compared with observations than an unedited low resolution model, the former may capture variability more realistically, and hence better represent the possibility of extremes, even if the timing of these extremes is unpredictable; which of the two constitutes a better forecast therefore depends entirely on the application. For instance, in engineering applications, the possibility of wind extremes of a certain magnitude may be most important, regardless of when they occur, whereas in aviation or sailing it may be more important to minimise the mean square error. This is obviously speculation as I know little about either of these applications. I hope there are more appropriate examples. The fact that high and low resolution model guidance products are used at different times, and on different days, implies that the Official forecast is inconsistent in which measures of accuracy it intends to maximise, and more thought therefore needs to be given to this issue.

5. Conclusion

We have

References

Abkar, M., A. Sharifi, and F. Porté-Agel, 2016: Wake flow in a wind farm during a diurnal cycle. *Journal of Turbulence*, **17 (4)**, 420–441, doi:10.1080/14685248.2015.1127379, URL https://doi.org/10.1080/14685248.2015.1127379, https://doi.org/10.1080/14685248.2015.1127379.

Brown, A. L., C. L. Vincent, T. P. Lane, E. Short, and H. Nguyen, 2017: Scatterometer estimates of the tropical sea-breeze circulation near Darwin, with comparison to regional models. *Quart.*

J. Roy. Meteor. Soc., doi:10.1002/qj.3131.

Bureau of Meteorology, 2010: Operational implementation of the ACCESS numerical weather prediction systems. Tech. rep., Bureau of Meteorology, Melbourne, Victoria. [Available online at http://www.bom.gov.au/australia/charts/bulletins/apob83.pdf].

Bureau of Meteorology, 2019: Meteye. Bureau of Meteorology, [http://www.bom.gov.au/australia/meteye/].

Dai, A., and C. Deser, 1999: Diurnal and semidiurnal variations in global surface wind and divergence fields. *Journal of Geophysical Research*, **104**, 31 109–31 125.

Efron, B., 1979: Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7 (1)**, 1–26, doi:10.1214/aos/1176344552.

Englberger, A., and A. Dörnbrack, 2018: Impact of the diurnal cycle of the atmospheric boundary layer on wind-turbine wakes: A numerical modelling study. *Boundary-Layer Meteorology*, **166 (3)**, 423–448, doi:10.1007/s10546-017-0309-3, URL https://doi.org/10.1007/s10546-017-0309-3.

European Center for Medium Range Weather Forecasting, 2018: *Part IV : Physical processes*. No. 4, IFS Documentation, European Center for Medium Range Weather Forecasting, [Available online at https://www.ecmwf.int/node/18714].

Gille, S. T., S. G. Llewellyn Smith, and N. M. Statom, 2005: Global observations of the land breeze. *Geophysical Research Letters*, **32 (5)**, doi:10.1029/2004GL022139, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004GL022139.

Griffiths, D., H. Jack, M. Foley, I. Ioannou, and M. Liu, 2017: Advice for automation of forecasts: a framework. Tech. rep., Bureau of Meteorology, Melbourne, Victoria. [Available online at http://www.bom.gov.au/research/publications/researchreports/BRR-021.pdf].

Haurwitz, B., 1947: Comments on the sea-breeze circulation. *Journal of Meteorology*, **4 (1)**, 1–8, doi:10.1175/1520-0469(1947)004⟨0001:COTSBC⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1947)004⟨0001:COTSBC⟩2.0.CO;2, https://doi.org/10.1175/1520-0469(1947)004⟨0001:COTSBC⟩2.0.CO;2.

Kusuda, M., and P. Alpert, 1983: Anti-clockwise rotation of the wind hodograph. part i: Theoretical study. *Journal of the Atmospheric Sciences*, **40 (2)**, 487–499, doi:10.1175/1520-0469(1983)040⟨0487:ACROTW⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1983)040⟨0487:ACROTW⟩2.0.CO;2.

Lee, X., 2018: *Fundamentals of boundary-layer meteorology.* Springer atmospheric sciences, Springer.

Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith, 2000: A new boundary layer mixing scheme. part i: Scheme description and single-column model tests. *Monthly Weather Review*, **128 (9)**, 3187–3199, doi:10.1175/1520-0493(2000)128⟨3187:ANBLMS⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0493(2000)128⟨3187:ANBLMS⟩2.0.CO;2, https://doi.org/10.1175/1520-0493(2000)128⟨3187:ANBLMS⟩2.0.CO;2.

Louis, J.-F., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteorology*, **17 (2)**, 187–202, doi:10.1007/BF00117978, URL https://doi.org/10.1007/BF00117978.

Miller, S. T. K., B. D. Keim, R. W. Talbot, and H. Mao, 2003: Sea breeze: Structure, forecasting, and impacts. *Reviews of Geophysics*, **41 (3)**, doi:10.1029/2003RG000124, URL https://doi.org/10.1029/2003RG000124.

Physick, W. L., and D. J. Abbs, 1992: Flow and plume dispersion in a coastal valley. *Journal of Applied Meteorology*, **31 (1)**, 64–73, doi:10.1175/1520-0450(1992)031⟨0064:FAPDIA⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0450(1992)031⟨0064:FAPDIA⟩2.0.CO;2, https://doi.org/10.1175/1520-0450(1992)031⟨0064:FAPDIA⟩2.0.CO;2.

Pinson, P., and R. Hagedorn, 2012: Verification of the ecmwf ensemble forecasts of wind speed against analyses and observations. *Meteor. Appl.*, **19 (4)**, 484–500, doi:10.1002/met.283, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.283, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/met.283.

Svensson, G., and Coauthors, 2011: Evaluation of the diurnal cycle in the atmospheric boundary layer over land as represented by a variety of single-column models: The second gabls experiment. *Boundary-Layer Meteorology*, **140 (2)**, 177–206, doi:10.1007/s10546-011-9611-7, URL https://doi.org/10.1007/s10546-011-9611-7.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences. [electronic resource].* International geophysics series: v. 100, Elsevier.

Zwiers, F. W., and H. von Storch, 1995: Taking serial correlation into account in tests of the mean. *Journal of Climate*, **8 (2)**, 336–351, doi:10.1175/1520-0442(1995)008⟨0336:TSCIAI⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0442(1995)008⟨0336:TSCIAI⟩2.0.CO;2, https://doi.org/10.1175/1520-0442(1995)008⟨0336:TSCIAI⟩2.0.CO;2.

LIST OF TABLES

Airport	Austral Summer	Austral Winter
Darwin	6.3 kn	6.2 kn
Brisbane	8.6 kn	7.0 kn
Perth	11.3 kn	7.9 kn
Sydney	12.2 kn	10.2 kn
Adelaide	9.5 kn	10.3 kn
Canberra	7.4 kn	7.9 kn
Melbourne	10.0 kn	12.1 kn
Hobart	10.0 kn	8.7 kn

TABLE 1. Average 10 m wind speeds for austral winter (June, July August) 2018, and austral summer (December, January, February) 2017/18 across the eight Australian capital city airport weather stations.

LIST OF FIGURES

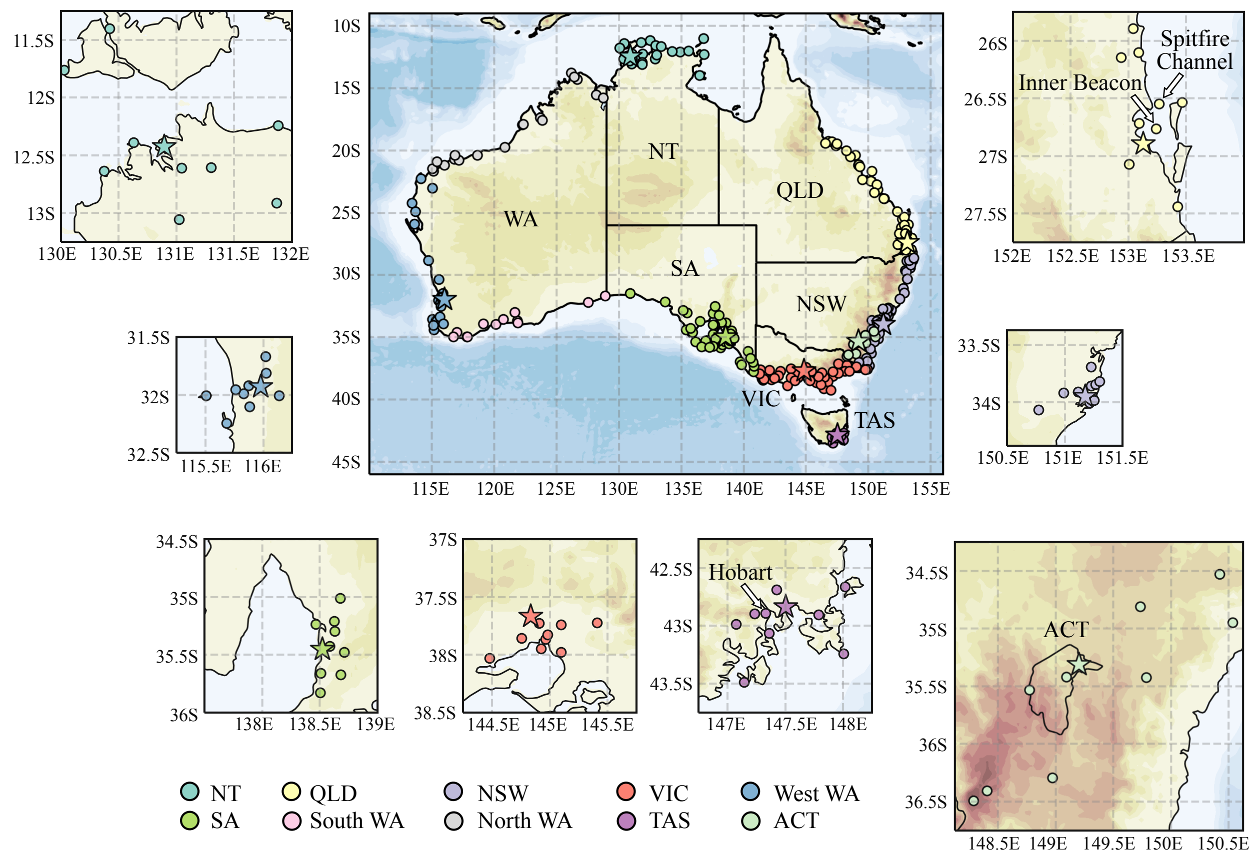


FIG. 1. Locations of the automatic weather stations used in this study. Stars indicate capital city airport stations. Height and depth shading intervals every 200 and 1000 m, respectively.

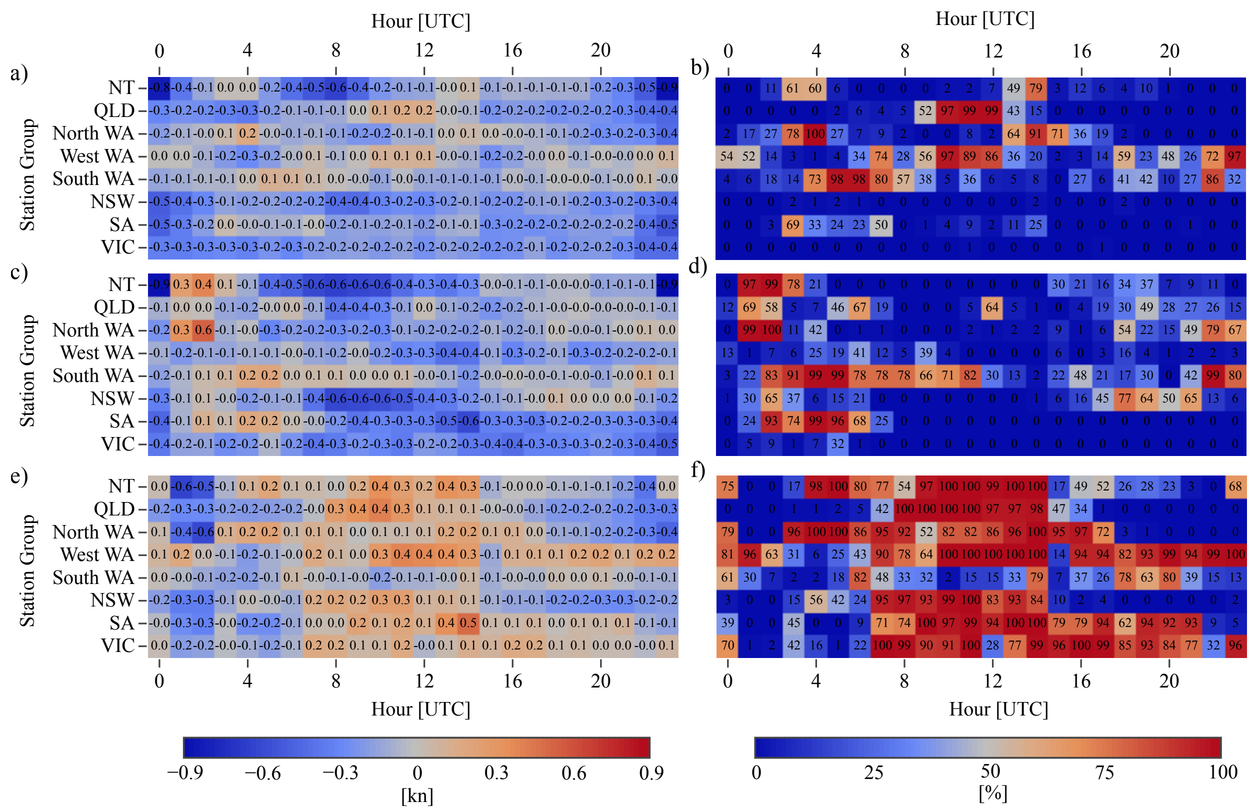


FIG. 2. Heatmaps of $\overline{\text{WPI}}$ values and confidence scores for each coastal station group and hour of the day: a) and b), Official versus ACCESS, c) and d) Official versus ECMWF, e) and f) ECMWF versus ACCESS. Positive $\overline{\text{WPI}}$ values mean that the former dataset in each pair is on average closer to observations than the latter dataset. Confidence scores provide the probability the population $\overline{\text{WPI}}$ is greater than zero. Values within the heatmaps are accurate to two significant figures.

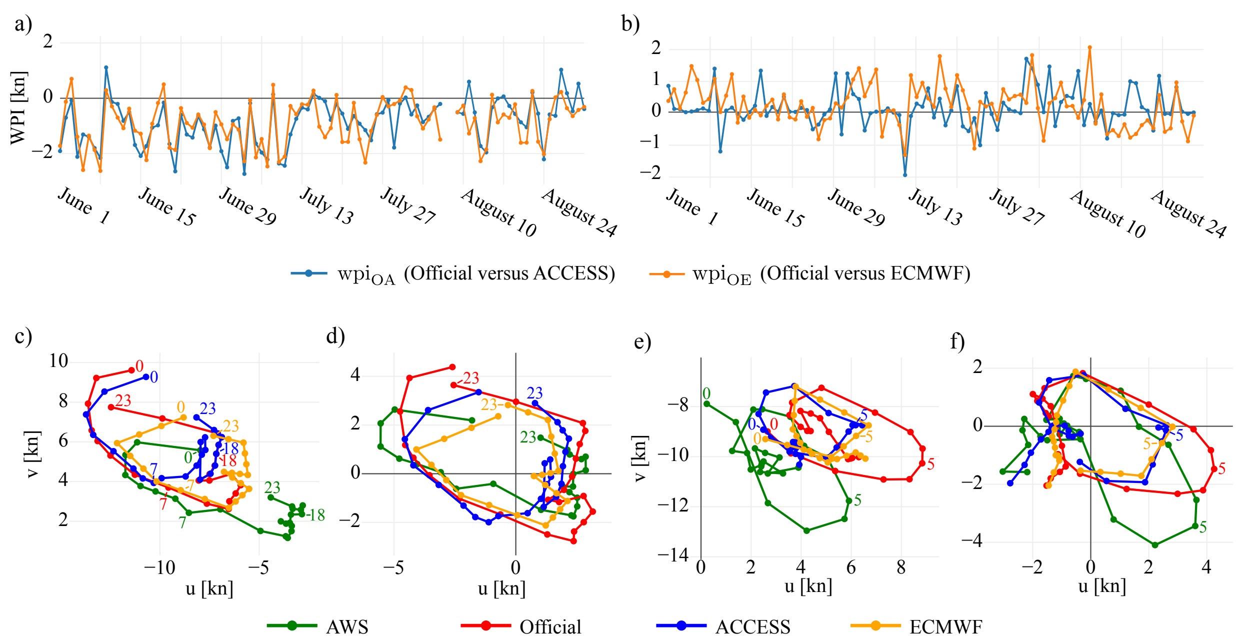


FIG. 3. Time series, a) and b), of $\overline{\text{wpi}}_{\text{OA}}$ and $\overline{\text{wpi}}_{\text{OE}}$ for, a), the NT station group at 23:00 UTC, and b), the south WA station group at 05:00 UTC. Hodographs, c) to f), showing change in winds, c) and e), and wind perturbations, d) and f), for the NT station group, c) and d), and south WA station group, e) and f).

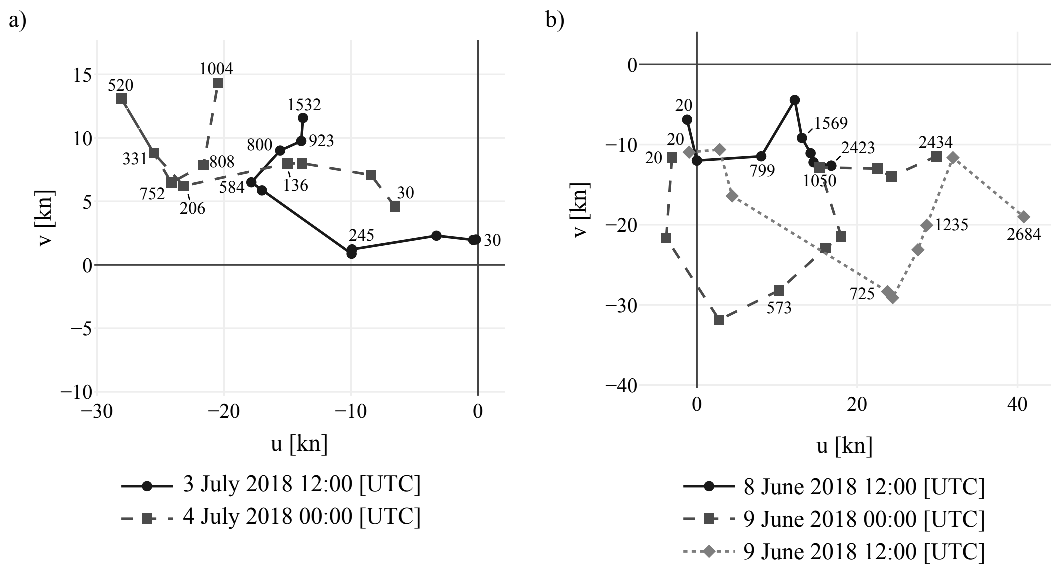


FIG. 4. Hodographs showing change in winds with height at, a), Darwin Airport, and b), Perth Airport.

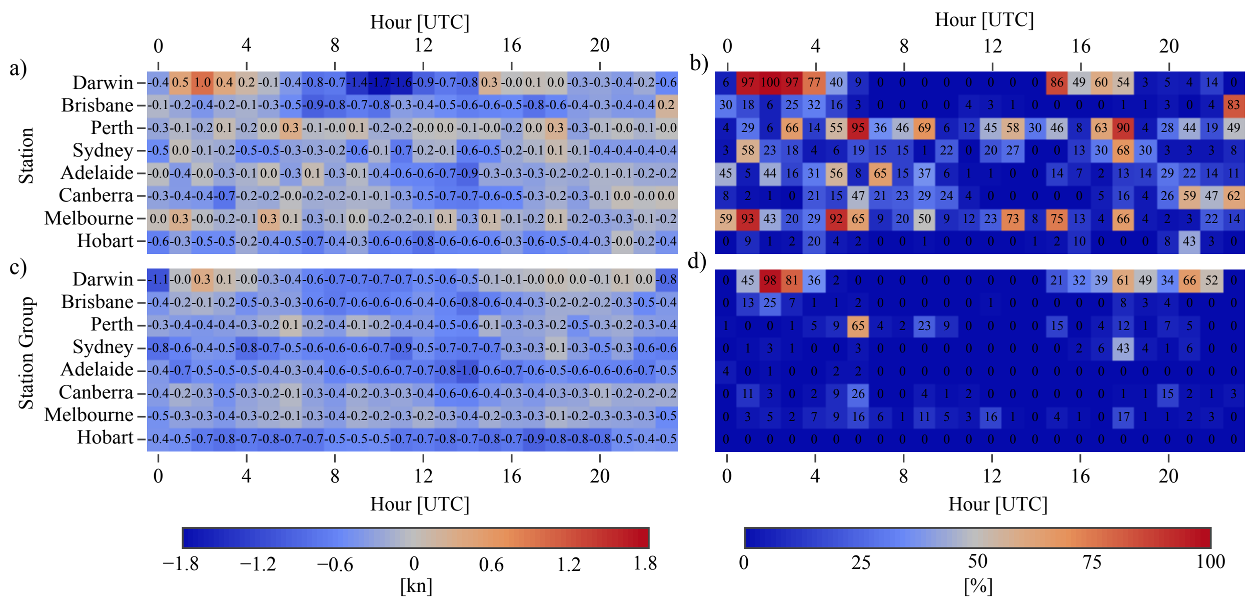
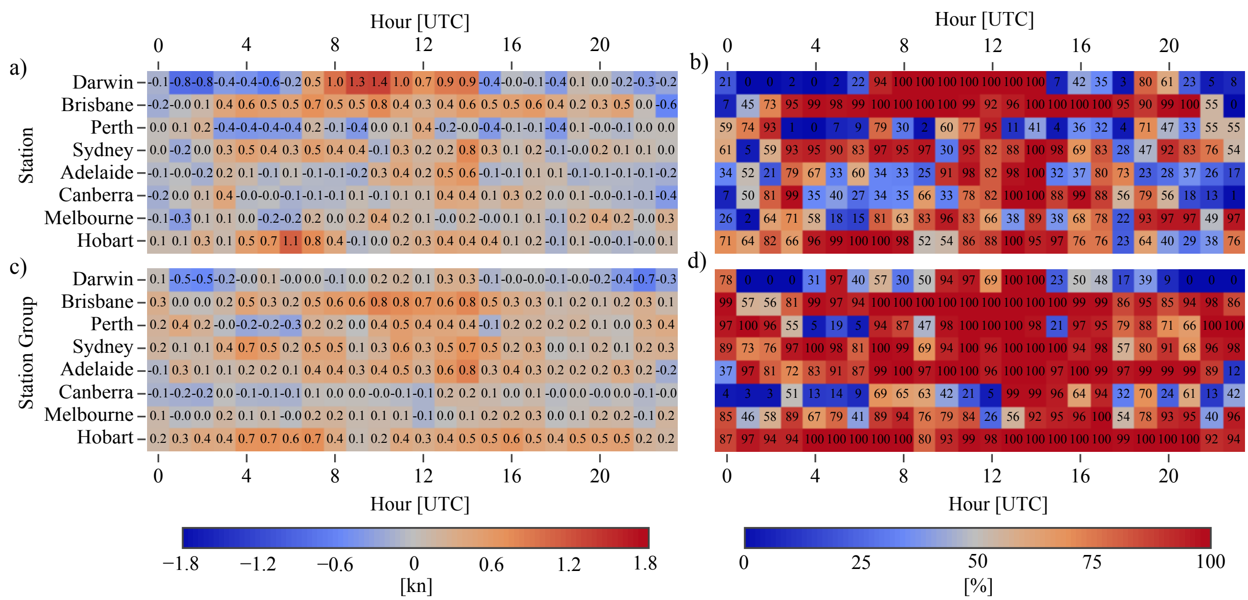
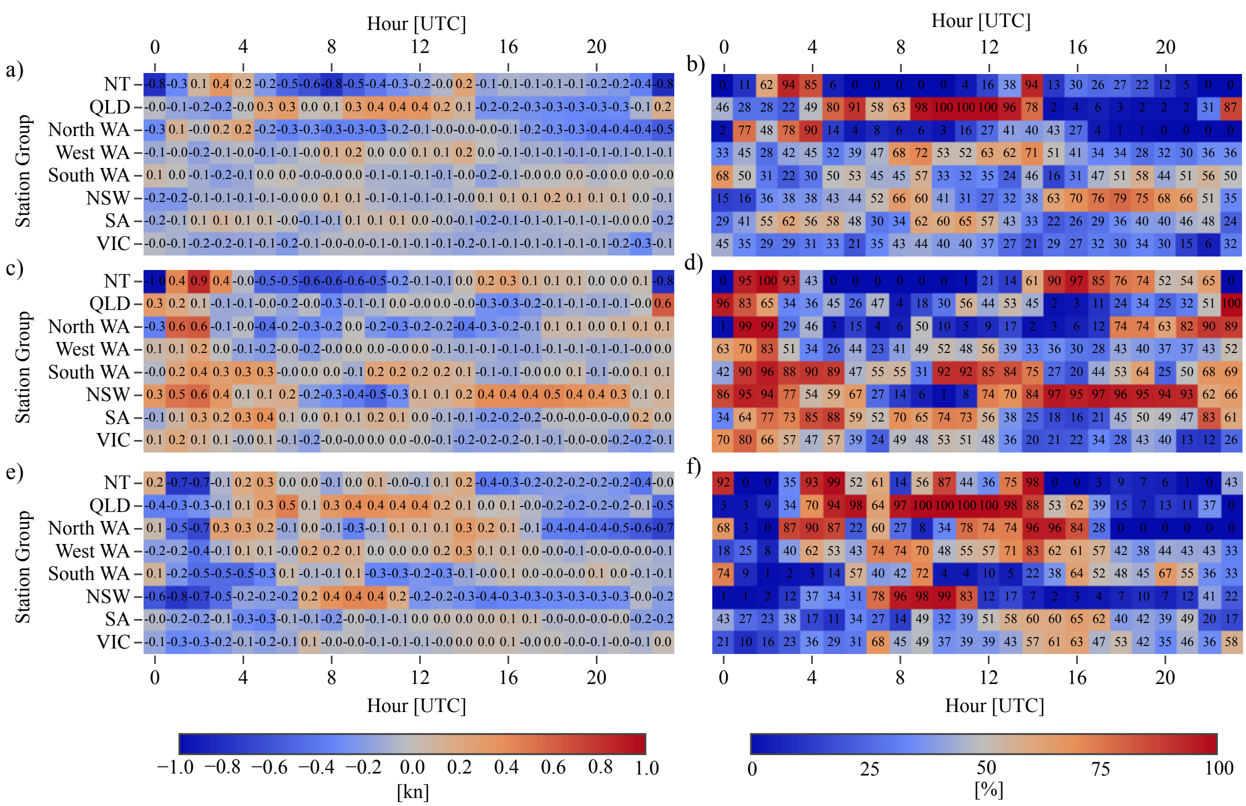


FIG. 5. The $\overline{\mathrm{wpi}}_{\mathrm{OE}}$ (Official versus ECMWF comparison) values, a) and c), and confidence scores, b) and d), for the airport stations, a) and b), and airport station groups, c) and d), respectively.


a)
b)
c)
d)
Hour [UTC]
Station
Station Group
Darwin
Brisbane
Perth
Sydney
Adelaide
Canberra
Melbourne
Hobart
[kn]
[%]


a)
b)
c)
d)
e)
f)
Hour [UTC]
0
4
8
12
16
20
Station Group
NT
QLD
North WA
West WA
South WA
NSW
SA
VIC
−1.0 −0.8 −0.6 −0.4 −0.2 0.0 0.2 0.4 0.6 0.8 1.0
[kn]
0 25 50 75 100
[%]

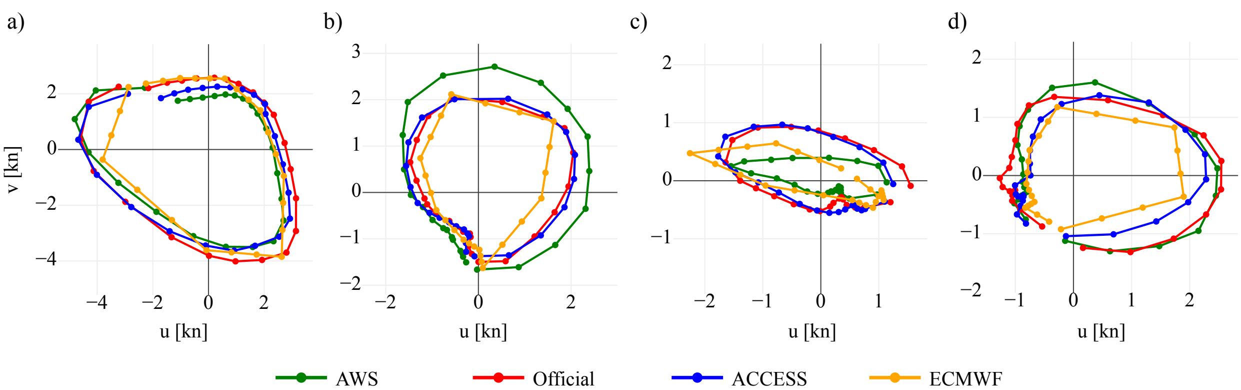


FIG. 6. Climatological hodographs.

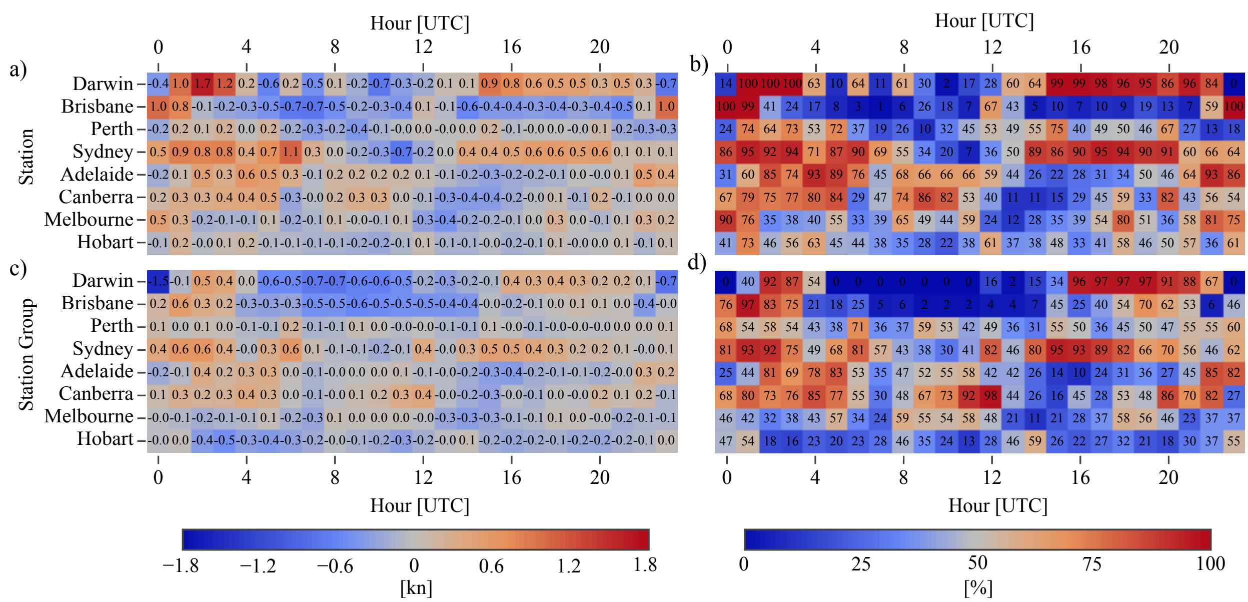
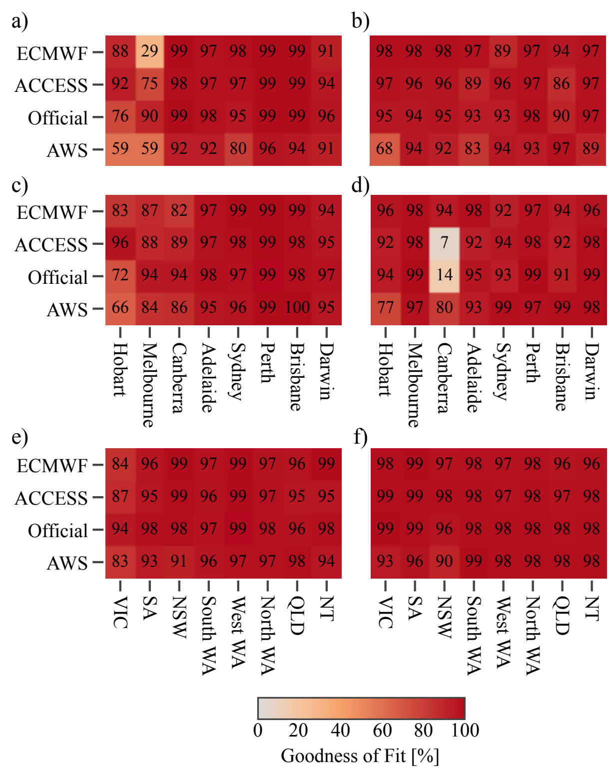


FIG. 7. As in Fig. 5, but for the cwpi values and confidence scores.



683 FIG. 8. Could also provide an analogous figure showing the use of the function α provides a significant improvement over the basic ellipse fit - or instead just quote some numbers? Or maybe these figures are entirely unnecessary?

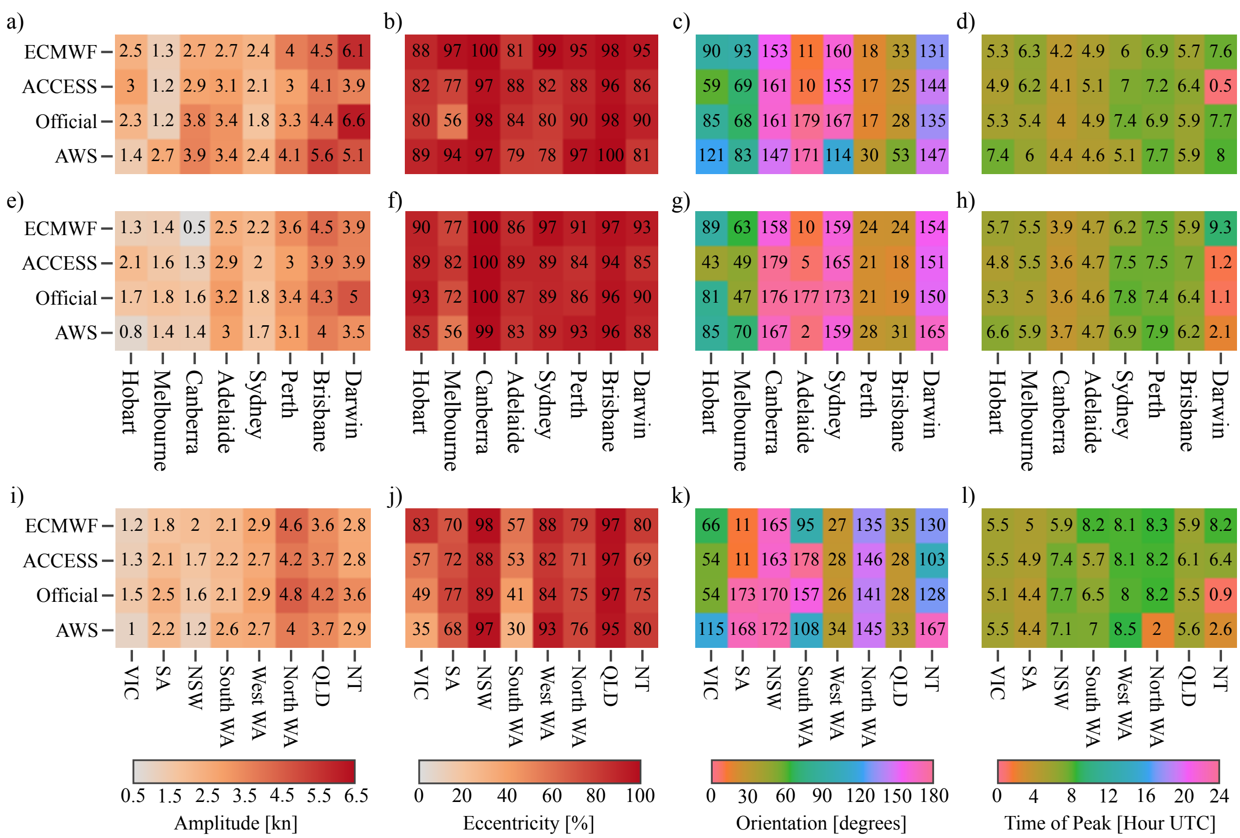


FIG. 9. Ellipse fits. If we were to include any analysis for alternative time periods (e.g. summer 2017/18 for contrast; or could do 18/19 if I were to go back to BoM to get the data) a copy of this figure could be a good choice. Could explain changes in diurnal cycle properties, e.g. amplitude, with seasonal changes to background winds, heating, etc. Note some issues with timing and amplitude values due to asymmetry - could instead just show eccentricity and orientation values?

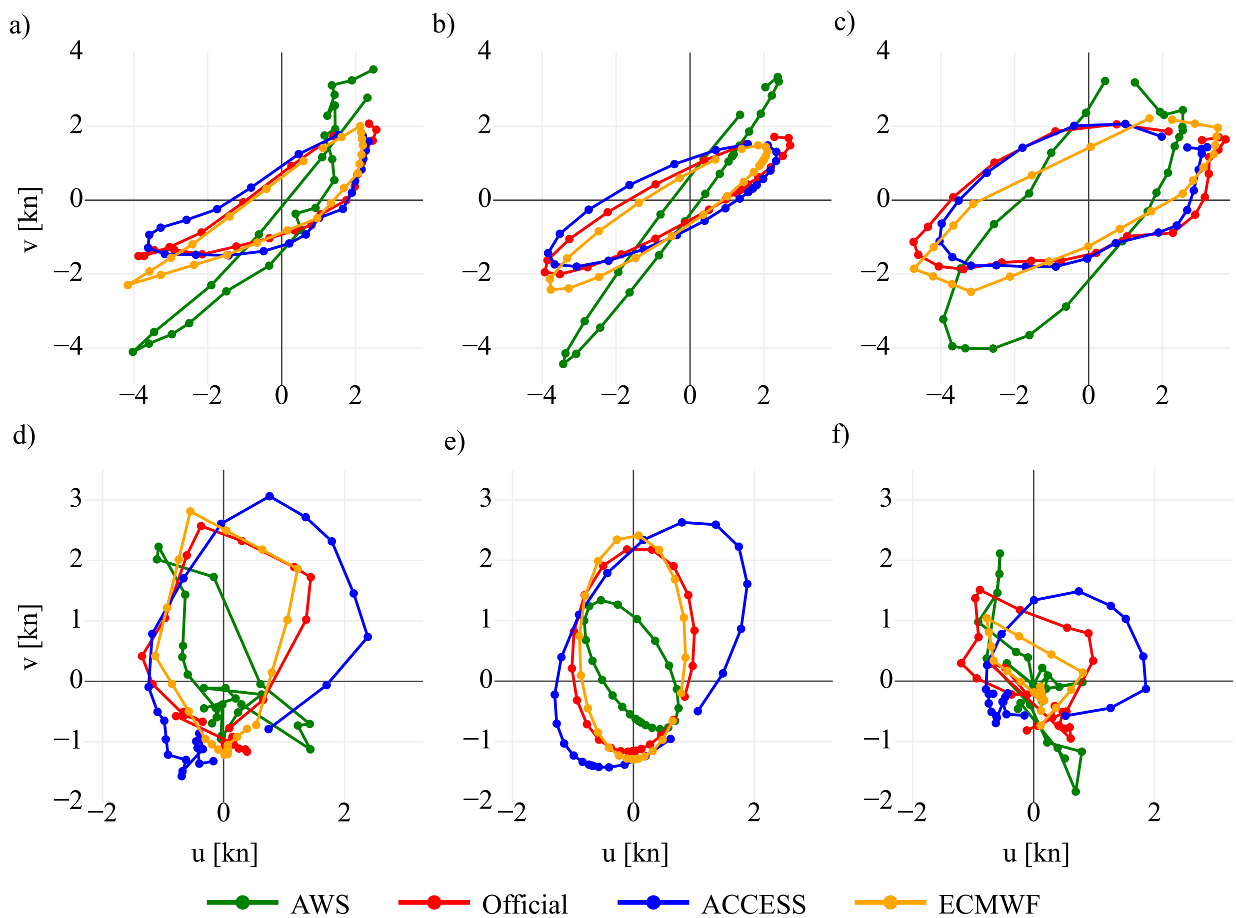


FIG. 10. Ellipse fits. Could instead just provide one example.

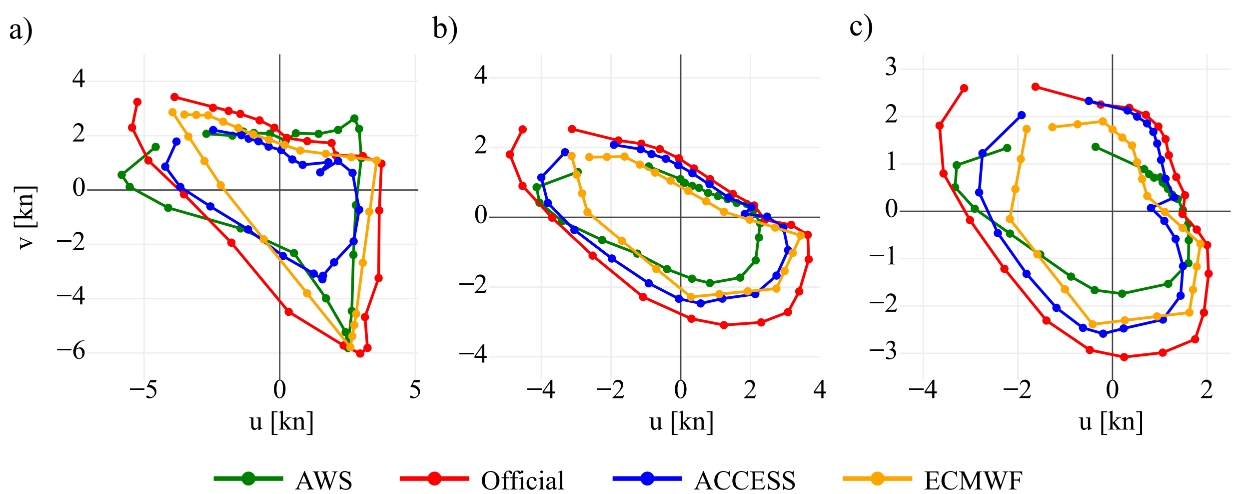


FIG. 11. Ellipse fits. Could also include the ellipses, but this makes the figure very large.

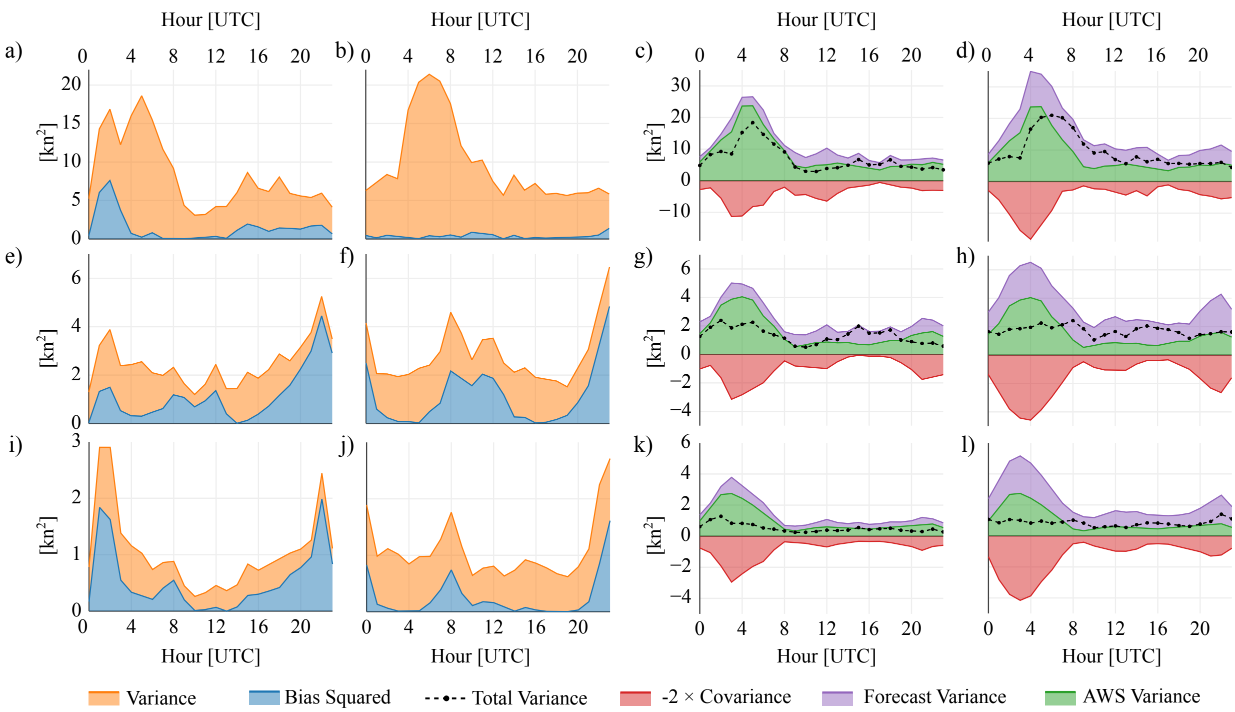


FIG. 12. Actual perturbation standard deviation values. Note that official performs the worst at this scale!

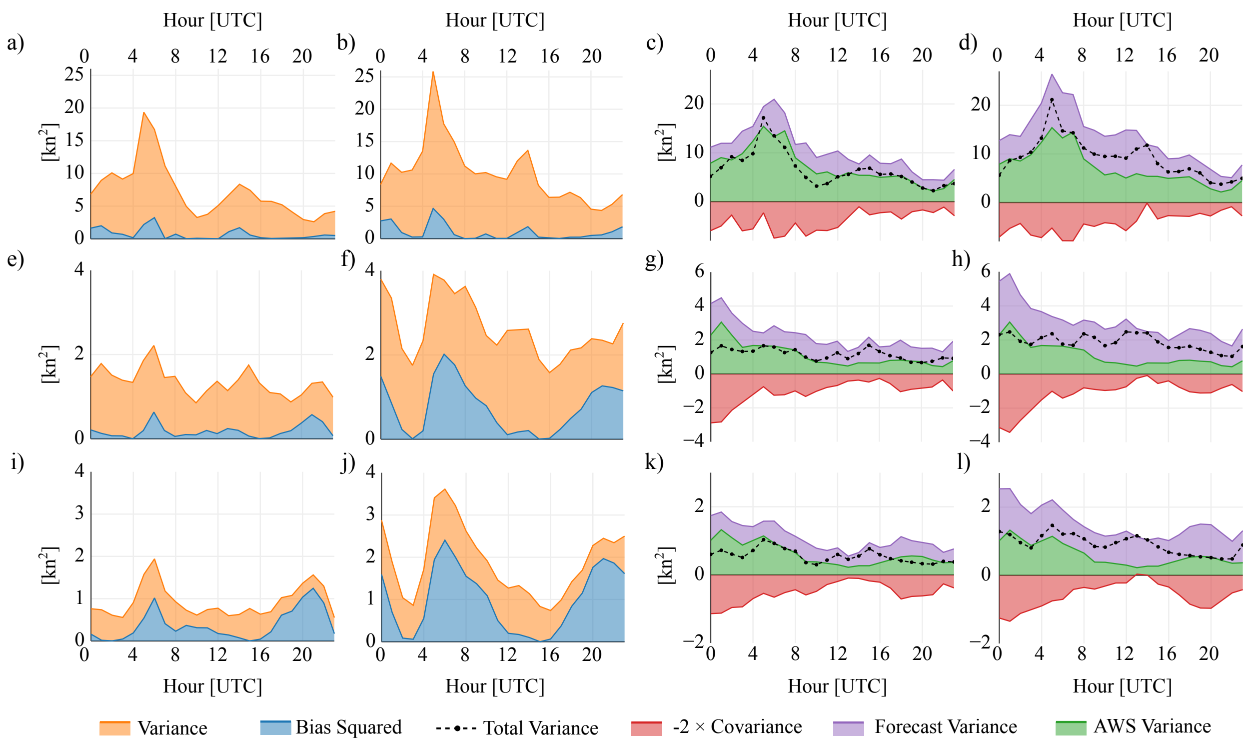


FIG. 13. Actual perturbation standard deviation values. Note that official performs the worst at this scale!