

# 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

13    Forecasts issued by the Australian Bureau of Meteorology are based on  
14    model data that is edited by human forecasters. Two common edits to the  
15    wind fields aim to improve how boundary layer mixing processes and the  
16    land-sea breeze are treated in the forecast. In this study we compare the di-  
17    urnally varying component of the edited wind forecast, with those of station  
18    observations and unedited model guidance datasets, to assess the additional  
19    accuracy provided by the edits. We consider coastal locations across Aus-  
20    tralia over June, July and August 2018, assessing performance at three differ-  
21    ent spatial scales, and on both a daily and seasonal basis. The results show  
22    that the edited forecast generally only produces lower daily errors than the  
23    unedited model guidance products at the coarsest spatial scale we considered  
24    (1000 - 2000 km), but can achieve lower seasonal biases over all three spatial  
25    scales. However, the edited forecast only reduces errors at particular times  
26    and locations, and rarely produces lower errors than all model guidance prod-  
27    ucts simultaneously. This suggests that forecaster skill lies mostly in making  
28    the choice of model guidance, rather than in making edits. To better diagnose  
29    the causes of errors in the diurnal wind cycles, we fit a modified ellipse to the  
30    climatological diurnal cycle hodographs. Performance varies with location  
31    for multiple reasons, including biases in the directions sea-breezes approach  
32    coastlines, amplitude and shape biases in the hodographs, and disagreement  
33    as to whether sea-breeze or boundary layer mixing processes contribute most  
34    to the diurnal cycle.

## 35 1. Introduction

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

46 Australian forecasters generally make two types of edits to the surface wind fields on a routine  
47 daily basis. The first is to edit the surface winds after sunrise at locations where the forecaster be-  
48 lieves the model guidance is providing a poor representation of boundary layer mixing processes.  
49 Boundary layer mixing occurs as the land surface heats up, producing an unstable boundary layer  
50 which transports momentum downward to the surface layer; before this mixing occurs winds are  
51 typically both weaker and ageostrophically oriented due to surface friction (Lee 2018), and so  
52 mixing can affect both the speed and direction of the surface winds. How do the boundary layer  
53 mixing tools in GFE currently work? While I was in Darwin you picked a height  $z$  and a percent-  
54 age  $p$ , and the tool essentially formed an average of the surface winds and winds at  $x$  weighted by  
55  $p$ .

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

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

61 Forecasters, and the weather services that employ them, have good reasons for ensuring the  
62 diurnally varying component of their wind forecasts are as accurate as possible. Dai and Deser  
63 (1999) fitted the first two harmonics to seasonal averages of wind speed at different times of day  
64 and showed that over land surfaces the average amplitude of the wind speed diurnal cycle varied  
65 from 1.2 to 2.1 kn, (knots are used throughout this paper because this is the unit forecasters work  
66 with, and the unit that is used in Jive) and that the fitted harmonics accounted for 50 to 70% of  
67 the daily variability. Table 1 provides the mean wind speeds for the eight Australian capital city  
68 airport stations shown in Fig. 1, over December, January, February 2017/18 and June, July and  
69 August 2018, suggesting that the amplitude of the mean diurnal cycles are approximately 10 to  
70 34% of the mean wind speeds across Australia.

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

78 To our knowledge, no published work has assessed the diurnal component of human edited  
79 forecasts, although some previous studies have assessed the performance of different operational  
80 models at specific locations. Svensson et al. (2011) examined thirty different operational model  
81 simulations, including models from most major forecasting centres utilising most commonly used

82 boundary layer parametrisation schemes, and compared their performance with a large eddy sim-  
83 ulation (LES), and observations at Kansas, USA, during October 1999. They found that both the  
84 models and LES failed to capture the sudden  $\approx 6$  kn jump in wind speeds shortly after sunrise, and  
85 underestimated morning low level turbulence and wind speeds. Other studies have assessed near-  
86 surface wind forecasts, verifying the total wind speeds, not just the diurnal component. Pinson  
87 and Hagedorn (2012) studied the 10 m wind speeds resolved by the European Centre for Medium  
88 Range Weather Forecasting (ECMWF) operational model ensemble across western Europe over  
89 December, January, February 2008/09. They found that the worst performing regions were coastal  
90 and mountainous areas, and attributed this poor performance to the small scale processes, e.g. sea  
91 and mountain breezes, that are under-resolved at ECMWF’s coarse 50km spatial resolution.

92 Any attempt to validate model data against observations must confront the *representation prob-*  
93 *lem* (e.g. Zaron and Egbert 2006). Because models cannot resolve physical processes occurring  
94 at sub-grid scales, a value predicted by an operational model for a given grid-cell must be inter-  
95 preted as a prediction of the filtered, or Reynolds averaged value over that grid-cell. Therefore,  
96 comparing model data with observational data can be an unfair test of model performance, and for  
97 this reason model forecasts are often verified against reanalysis hind-casts that use the same model  
98 (e.g. Lynch et al. 2014).

99 However, the way the representation problem applies to the verification of forecasts issued to the  
100 public is more nuanced. In this case, a forecast issued by a national weather service is attempting  
101 to represent either reality itself, or the filtered version of reality *that is of interest to the end users*.  
102 Thus, Pinson and Hagedorn (2012) disregarded the representation problem entirely, arguing that  
103 the end user is not interested in spatiotemporal scales of models, only the “best forecast” at the  
104 time and place of their choice. However, different users will have different ideas about what  
105 the “best” forecast entails. Some users may desire a forecast that minimises error between the

106 forecast and observations, others a forecast that most accurately reproduces the observed wind  
107 speed distribution, regardless of temporal details. An operational forecast should therefore be  
108 assessed according to the representation needs of the end user.

109 Note that BoM verifies its wind forecasts by comparing hourly forecast data with station obser-  
110 vations that are first averaged 10 minutes either side of the hour: implicit in this practice is the  
111 assumption that end users do not care about wind turbulence at temporal scales less than 20 min-  
112 utes. Furthermore, the fact that the BoMs forecast is formed from model datasets with different  
113 resolutions, and the choice of model guidance can change even over the course of a single day,  
114 (e.g. Fig. 3 b), means it is difficult to determine precisely what BoM forecasts intend to repre-  
115 sent, and hence addressing the representation problem difficult. Note that clicking locations on  
116 the MetEye map seems to bring up the forecast for the nearest station or population centre. Does  
117 this imply that the Official forecast is intending to represent spatial scales at least as fine as the  
118 distance between stations? Or is it intending to represent averaged parameters at the ACCESS-R  
119 or ACCESS-C resolutions? Note also that grids are only provided every 3 hours in MetEye; do  
120 these grids represent the hourly values from the Official forecast, provided just at these three hour  
121 intervals, or a three hourly average?

122 Related to the representation problem is the question of how mesoscale models, which run at  
123 spatial resolutions of between 1 and 10 km, should be verified. BoM now regularly runs mesoscale  
124 models over some Australian capital cities as part of its daily forecasting routine, and the edits  
125 performed by human forecasters are also often mesoscale in nature. Note that mesoscale models  
126 resolve topography and its effects on the atmosphere in more detail, and explicitly simulate most  
127 convective and boundary layer processes. In this sense they are more realistic than coarser scale  
128 models, although they can actually perform worse than coarse models on standard verification  
129 scores whenever there are timing or location differences between features in the models and in

130 observations. Mass et al. (2002) found that in the northwestern United States, mean square errors  
131 in forecasts produced using mesoscale models decreased with resolution down to 10 to 15 km,  
132 whereas in the eastern United States where the topography is much flatter, this threshold was  
133 considerably larger, at 20 to 40 km.

134 Mass et al. (2002) therefore argued that existing verification approaches needed reform, suggest-  
135 ing that verification could instead be performed on spatially or temporally averaged parameters,  
136 an approach now known as *upscale* (Ebert 2008). Alternatively, Mass et al. (2002) argued that  
137 “feature based” identification metrics be developed, which reward models for realistically simulat-  
138 ing atmospheric features, even if the timing or location of these features is slightly incorrect. Rife  
139 and Davis (2005) developed such a method for the verification of surface winds, defining a wind  
140 “object” as a wind change of at least one standard deviation occurring within a 12 hour interval,  
141 then assessing whether a mesoscale model could replicate the “objects” present in observations.

142 The present study has two goals. First, to describe a method for comparing the diurnal cycles  
143 of human edited wind forecasts to those of unedited model guidance forecasts, in order to assess  
144 where and when human edits produce an increase in accuracy, and to do so in a way that re-  
145 spects the representation and mesoscale verification challenges discussed above. Second, to apply  
146 this methodology across Australian coastal locations to better understand the performance of both  
147 boundary layer mixing and land-sea breeze forecaster edits. The remainder of this paper is organ-  
148 ised as follows. Section 2 describes the methodology and datasets to which it is applied, section 3  
149 provides results, and sections 4 and 5 provide a discussion and a conclusion, respectively.

## 150 **2. Data and Methods**

151 This study compares both human edited and non human edited Australian Bureau of Meteorol-  
152 ogy wind forecasts with automatic weather station (AWS) data across Australia. The comparison

153 is performed by first isolating the diurnal signals of each dataset, then comparing these signals  
154 on an hour-by-hour basis. Much of the analysis is conducted through the BoM's *Jive* verification  
155 platform (Provide either a URL link to an Official Jive page if one exists by the time of publica-  
156 tion, or a link to copy of Jive I've been using on my laptop, which I could host on my personal  
157 GitHub page.), which provides an archive of forecast, model guidance and observational data, and  
158 a software library for calculating basic statistics.

#### 159 *a. Data*

160 Four datasets are considered in this study; the Official BoM wind forecast data that is issued to  
161 the public, model data from ECMWF, (is this the mean of the ECMWF operational ensemble?)  
162 model data from the Australian Community Climate and Earth System Simulator (ACCESS), (I  
163 haven't considered the BoM's Operational Consensus Forecast in this study, because it was not  
164 used in the Official forecast for winds while I was in the NT. If it is used frequently in other  
165 states, the study can be reinterpreted as a verification of both forecaster edits and OCF against  
166 unedited, un-post-processed model guidance.) and observational data from automatic weather  
167 stations (AWS) across Australia. (With the Bureau's permission, I would like to make the datasets  
168 used in this paper open access, and host them on the NCI catalogue page.) The Official and  
169 ECMWF data are at a ?, ? degree spatial resolution respectively. What are the resolutions of these  
170 datasets as they're used in Jive? I know ACCESS uses nested grids, so what is the resolution of  
171 the ACCESS dataset when used in GFE and Jive? Are the outer grids interpolated to the resolution  
172 of the inner grids, or are the inner grids upscaled? Official, ACCESS and AWS data exists at each  
173 UTC hour, but ECMWF data only exists at a three hour resolution. Why is this? What are the  
174 actual time-steps of the models? To be consistent with the other data sets, ECMWF is therefore  
175 linearly interpolated to an hourly resolution: note that this is also what happens when forecasters



176 load ECMWF wind data into the GFE, and the linearly-interpolated ECMWF data is therefore  
177 an appropriate model guidance dataset to compare with Official. To facilitate comparison with  
178 observations, Official, ACCESS and ECMWF data is (tri-linearly?) interpolated in all three spatial  
179 dimensions to the locations of the weather stations. AWS wind data is recorded every minute at  
180 each station, and the hourly AWS datasets used in this study are formed by taking 10 minute  
181 averages either side of each UTC hour. (My memory is that Jive uses a ten minute average either  
182 side, but need to confirm this as it could be five minutes either side.) Stations are quality controlled  
183 by... I've excluded the list of known problematic wind stations in the Jive documentation, but I  
184 can't remember what the other quality control methods were.

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

## 192 *b. Assessing Diurnal Cycles*

193 Forecasters typically edit model guidance data to account for under-resolved sea-breezes and  
194 boundary layer mixing processes. Instead of attempting to assess each type of edit individually,  
195 we study the overall diurnal signal by subtracting a twenty hour centred running mean *background*  
196 *wind* from each zonal and meridional hourly wind data point. This provides a collection of zonal  
197 and meridional wind *perturbation* datasets.

One measure of the accuracy of the Official, ACCESS and ECMWF diurnal cycles is to compare the Euclidean distances of the perturbations at each hour with the corresponding AWS perturbations. For example, to assess whether the Official forecast perturbations,  $\mathbf{u}_O$ , or ACCESS perturbations,  $\mathbf{u}_A$ , best match the AWS observations,  $\mathbf{u}_{AWS}$ , we calculate the *Wind Perturbation Index* (WPI), defined by

$$\text{WPI}_{OA} = |\mathbf{u}_{AWS} - \mathbf{u}_A| - |\mathbf{u}_{AWS} - \mathbf{u}_O|. \quad (1)$$

The analogously defined quantities  $\text{WPI}_{OE}$  and  $\text{WPI}_{EA}$  can then be used to provide a comparison of the Official and ECMWF perturbations, and of the ACCESS and ECMWF perturbations, respectively. We can then 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  $\overline{\text{WPI}}$ .

$\overline{\text{WPI}}$  is the difference of two mean absolute errors. A  $\overline{\text{WPI}}_{OA}$  value of 0.5 kn at 00:00 UTC means that the Official 00:00 UTC perturbations are, on average, 0.5 kn closer to the observed perturbations than are those of ACCESS. The WPI compares just *one aspect* of the Official forecast with model guidance; it says nothing, for instance, about whether the variability of the Official forecast is closer to that of the AWS than the model guidance. As such, any statements about performance made throughout this paper refer solely to WPI, and no claim is being made that WPI is sufficient to completely characterise the accuracy, or value to the user, of how the surface diurnal wind cycle is represented in competing forecasts.

Note that sea-breeze and boundary layer mixing processes depend crucially on the background atmospheric conditions in which they occur. By comparing wind perturbations rather than the overall wind fields we are not claiming these background conditions are irrelevant. However, when a forecaster makes an edit of a wind forecast to better resolve these processes, they are

220 implicitly assuming that future background conditions will be close enough to either climatology,  
 221 or model predictions of background conditions themselves, to justify making the edit. Thus, it  
 222 makes perfect sense to compare forecast perturbations to observed perturbations, as long as errors  
 223 are interpreted as the consequence not only of how the forecaster or model resolves the diurnal  
 224 cycle, but of how errors in the background state contribute to errors in the perturbations. To  
 225 minimise the significance of background state errors, this study focuses exclusively on lead-day  
 226 one forecasts.

227 Given the large degree of turbulence and unpredictable variability in both the AWS, Official,  
 228 and model datasets, care must be taken to ensure we do not pre-emptively conclude Official has  
 229 outperformed the model guidance when  $\overline{\text{WPI}} > 0$  purely by chance. The method for estimating  
 230 confidence in  $\overline{\text{WPI}}$  is based on a method proposed by Griffiths et al. (2017) as a general frame-  
 231 work for BoM verification metrics. Note first that WPI is defined so as to minimise the temporal  
 232 autocorrelations within each dataset, and to avoid having to consider correlations between the  
 233 zonal and meridional components within and between datasets. Time series formed from the WPI  
 234 values at a particular time, say 00:00 UTC, across the three month time period, can therefore be  
 235 idealised as an independent random sample of a random variable  $W$ . The sampling distribution for  
 236 each  $\overline{\text{WPI}}$  can be modelled by a Student's  $t$ -distribution, and from this we calculate the probability  
 237 that  $W$  is positive, denoted  $\Pr(W > 0)$ . Although temporal autocorrelations of WPI, i.e. corre-  
 238 lations between WPI values at a particular hour from one day to the next, are in practice small  
 239 or non-existent thanks to how WPI is defined, they are still accounted for by reducing the “ef-  
 240 fective” sample size to  $n(1 - \rho_1) / (1 + \rho_1)$ , where  $n$  is the actual sample size and  $\rho_1$  is the lag-1  
 241 autocorrelation (Zwiers and von Storch 1995; Wilks 2011). Note that in the standard language of  
 242 statistical hypothesis testing, we would reject the null hypothesis that  $W = 0$  at significance level  
 243  $\alpha$  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  $\Pr(W < 0) = 1 - \Pr(W > 0)$ . We say Official outperforms model guidance with “high confidence” if  $\Pr(W > 0) \geq 95\%$ , or that model guidance outperforms Official with “high confidence” if  $\Pr(W > 0) \leq 5\%$ , with high confidence implicit whenever it is not explicitly mentioned. Much of this explanation is probably unnecessary, but would like to get feedback before I trim it.

To investigate the consequences of the representation and mesoscale verification challenges discussed in section 1, we apply *upscaling* (Ebert 2008), where forecast and observational data are first averaged to coarser spatiotemporal scales before being compared. In this study we consider three spatial and two temporal scales. The finest spatial 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 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 forecasting 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, so as to best resolve the land-sea breeze signal after spatial averaging

267 (e.g. Vincent and Lane 2016). These eight station groups are referred to as the *coastal station*  
268 *groups*.

269 We also consider both daily and seasonal time scales. For daily time scales, we either consider  
270 just the individual airport stations, or modify the definition of WPI in equation (1) so that each  
271 perturbation dataset is first spatially averaged over either the airport or coastal station groups.  
272 Confidence scores are calculated for the airport and coastal station groups in the same way as for  
273 the single airport stations, treating the spatially averaged data as a single time series. This provides  
274 a conservative way to deal with spatial correlation between the stations in each group (Griffiths  
275 et al. 2017).

276 For the seasonal scale comparison we define the *Climatological Wind Perturbation Index*  
277 (CWPI) by

$$\text{CWPI}_{\text{OA}} = |\overline{\mathbf{u}}_{\text{AWS}} - \overline{\mathbf{u}}_{\text{O}}| - |\overline{\mathbf{u}}_{\text{AWS}} - \overline{\mathbf{u}}_{\text{A}}|, \quad (2)$$

278 where the over-bars denote temporal averages of the perturbations at a particular hour, across the  
279 three month time period. These temporally averaged perturbations represent the climatological  
280 diurnal wind cycle over the three month study period for each dataset.  $\text{CWPI}_{\text{OE}}$  and  $\text{CWPI}_{\text{EA}}$  are  
281 defined analogously. The three spatial scales are considered in the same way as for WPI, with the  
282 spatial average taken before the temporal average. Uncertainty in the CWPI is estimated through  
283 bootstrapping (Efron 1979). This is done by performing resampling with replacement on the  
284 underlying perturbation datasets, and calculating the CWPI multiple times using these resampled  
285 datasets. This provides a distribution of CWPI values, which analogously to with WPI, we treat as  
286 a sample from a random variable  $C$ , and use this to estimate  $\Pr(C > 0)$ .

287 Although the WPI and CWPI provide quantitative information on the accuracy of the diurnal cycle  
288 at different times of day, they do not provide much information on the structure of the diurnal wind  
289 cycles of each dataset. Gille et al. (2005) obtained summary statistics on the observed structure of

the climatological diurnal wind cycles across the globe by using linear regression to calculate the coefficients  $u_i, v_i$   $i = 0, 1, 2$ , for the fits

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

$$v = v_0 + v_1 \sin(\omega t) + v_2 \cos(\omega t), \quad (4)$$

where  $\omega$  is the angular frequency of the earth and  $t$  is the local solar time in seconds. If a temporal hodograph of this fit is considered, i.e. if  $(u, v)$  are plotted in an  $x, y$  plane for different values of  $t$ , the tips of the vectors trace out an ellipse. Gille et al. (2005) used this to calculate descriptive quantities, like the eccentricity of the ellipse, and the angle the semi-major axis of the ellipse makes with the horizontal, directly from the coefficients  $u_1, u_2, v_1$  and  $v_2$ . Gille et al. (2005) applied this fit to scatterometer data, which after temporal averaging resulted in just four zonal and meridional values per location, and as such the fit performed very well.

However, equations (3) and (4) do not provide a good fit for hourly wind data, primarily because they assume a twelve hour symmetry in the evolution of the diurnal cycle. In practice, asymmetries between daytime heating and nighttime cooling (e.g. Svensson et al. 2011) result in surface wind perturbations accelerating rapidly just after sunrise, but remaining comparatively stagnant at night (e.g. Fig. 9). Thus, we instead fit the equations

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

$$v = v_0 + v_1 \sin(\alpha(\psi, t)) + v_2 \cos(\alpha(\psi, t)), \quad (6)$$

to the climatological perturbations, with  $\alpha$  the function from  $[0, 24) \times [0, 2\pi) \rightarrow [0, 2\pi)$  given by

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

with  $t$  the time in units of hours UTC, and  $\psi$  providing the time when the wind perturbations vary least with time. For each climatological diurnal wind cycle, we solve for the seven parameters  $u_0,$

307  $u_1, u_2, v_0, v_1, v_2$  and  $\psi$  using nonlinear regression, performed using the `least_squares` function  
308 from the `scipy.optimize` python module (SciPy 2019).

309 Note Gille et al. (2005) fit equations (3) and (4) to the temporally averaged wind fields, so that  
310  $(u_0, v_0)$  can be interpreted as the mean wind over the study's time period, and the remaining terms  
311 providing the climatological diurnal perturbations. In this study we fit equations (5) and (6) to the  
312 climatological perturbations, with  $(u_0, v_0)$  now necessary to offset the asymmetry introduced by  
313  $\alpha$ , i.e. to ensure the time integral of the fitted perturbation values is approximately zero. Following  
314 Gille et al. (2005), the ellipse's orientation, i.e. the angle the semi-major axis of the ellipse makes  
315 with lines of latitude, as well as the ellipse's eccentricity are calculated algebraically, but the  
316 perturbation speed maximum, and the time at which this maximum achieved, are instead obtained  
317 numerically.

### 318 **3. Results**

319 In this section, the methods described in section 2 are applied to Australian forecast and station  
320 data over the months of June, July and August (austral winter) 2018. First, differences are assessed  
321 on a daily basis using the Wind Perturbation Index (WPI) at three different spatial scales. Second,  
322 overall seasonal biases during this time period are assessed using the Climatological Wind Per-  
323 turbation Index (CWPI), and by comparing structural indices derived from ellipses fitted to the  
324 climatological wind perturbations.

#### 325 *a. Daily Comparison*

326 Figure 2 provides the WPI values and confidence scores for the coastal station groups for  $\overline{\text{WPI}}_{\text{OA}}$ ,  
327  $\overline{\text{WPI}}_{\text{OE}}$  and  $\overline{\text{WPI}}_{\text{EA}}$ , which represent the the Official versus ACCESS, Official versus ECMWF, and  
328 ECMWF versus ACCESS comparisons, respectively. The results indicate that for the majority

of station groups and hours, both the unedited ACCESS and ECMWF models outperform the Official forecast. The lowest  $\overline{\text{WPI}}$  values occur at the NT station group at 23:00 and 00:00 UTC for both  $\overline{\text{WPI}}_{\text{OA}}$  and  $\overline{\text{WPI}}_{\text{OE}}$ . Although Official outperforms at least one of ACCESS or ECMWF at multiple times and station groups, the only group and time where it outperforms both is 05:00 UTC over the South WA station group, although the  $\overline{\text{WPI}}$  values are comparatively low. ECMWF generally outperforms ACCESS from 10:00 - 14:00 UTC, with the South WA station group being the main exception.

Figures 3 and 4 provide case studies of the NT and South WA station groups, respectively. Figure 3 a) provides a time series of WPI for the NT station group at 23:00 UTC. The time series shows significant temporal variability, with WPI frequently exceeding  $-2$  kn. Figures 3 b) and c) show hodographs of the winds and wind perturbations, respectively, for the AWS observations, Official forecast, and ACCESS and ECMWF model datasets, at each hour UTC on the 3<sup>rd</sup> of July, which provides an interesting example.

Figure 3 b) 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 the other datasets. This rapid change is clearer in the perturbation hodograph shown in Fig. 3 c). At this time of year the prevailing winds throughout the NT are east-southeasterly, and 22:00 UTC corresponds to  $\approx 08:30$  local solar time (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 5 a) shows the first ten values from wind soundings at Darwin Airport at 12:00 UTC on July 3<sup>rd</sup> and 00:00 UTC on July 4<sup>th</sup>. 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 southeasterly component of the winds too much. Similar issues create low WPI scores on the 8<sup>th</sup> of June and 9<sup>th</sup> and 10<sup>th</sup> of July.

Figure 4 a) provides a time series of WPI for the South WA station group at 05:00 UTC. As with the NT station group there is significant temporal variability, with WPI frequently exceeding 1 kn. Figures 4 b) and c) provides hodographs of the winds and wind perturbations, respectively, on the 9<sup>th</sup> of June, which is an interesting example. The perturbation hodograph shows both ECMWF and ACCESS under-predicting 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.

Figure 5 shows wind sounds throughout the first two km of the atmosphere between 12:00 UTC on the 8<sup>th</sup> June and 12:00 UTC on the 9<sup>th</sup> June; the soundings were taken at Perth Airport, which is the nearest station to the South WA station group to provide wind soundings. The 8<sup>th</sup> June 12:00 UTC sounding shows surface northerlies of  $\approx$  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 9<sup>th</sup> 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. 4 c), 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. One explanation for this discrepancy is that the Official forecast has been edited based on the June 8<sup>th</sup>

377 12:00 UTC sounding, with the winds above the surface changing direction in the subsequent 12  
378 hours. A similar explanation can be given for the high WPI scores on the 3<sup>rd</sup> of August, although  
379 in this case the Official forecast slightly improves both the magnitude and direction of the 05:00  
380 UTC wind perturbations.

381 To provide a contrast to the coastal station group results, Fig. 6 presents the  $\overline{\text{WPI}}$  values and  
382 confidence scores for  $\overline{\text{WPI}}_{\text{OE}}$ , which represents the Official versus ECMWF comparison, for the  
383 airport stations, and airport station groups. The results for the airport stations are noisier than the  
384 results for the coastal station groups in Figs. 2 c) and d), although they share some similarities.  
385 Official outperforms ECMWF at 01:00 and 02:00 UTC at both the Darwin airport station and  
386 the NT station group, although ECMWF outperforms Official between 08:00 and 14:00 UTC  
387 at Darwin and Brisbane airports, and the corresponding NT and QLD station groups, with the  
388 exception of the QLD station group at 12:00 UTC. ECMWF also outperforms Official at Hobart  
389 airport at almost all hours of the day, and at Adelaide and Canberra airports from 11:00 to 14:00  
390 UTC.

391 For the remaining stations and times, Official only outperforms ECMWF at the Perth airport  
392 station at 06:00 UTC and the Melbourne airport station at 01:00 UTC, although in both cases WPI  
393 values are comparatively small in magnitude. Furthermore, in both cases there is no clear pattern  
394 to the  $\overline{\text{WPI}}_{\text{OE}}$  values over the rest of the day. Note that the *multiplicity problem* (Wilks 2011, p.  
395 178) requires care be taken before giving meaning to these two examples: i.e., given that we are  
396 calculating twenty four confidence scores for eight stations, then if WPI were uncorrelated across  
397 each station and hour we would expect to find  $0.05 \times 24 \times 8 \approx 10$  instances where  $P(W_{\text{OE}} > 0) \geq$   
398 95%, even if  $W_{\text{OE}}$  were in fact equal to zero.

399 For the airport station groups, ECMWF outperforms Official for the majority of station groups  
400 and times. The main exception is the Darwin airport station group, where Official outperforms

ECMWF at 02:00 UTC, 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  $\overline{\text{WPI}}_{\text{OA}}$  Official versus ACCESS comparisons (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 7 presents the  $\overline{\text{WPI}}$  values and confidence scores for  $\overline{\text{WPI}}_{\text{EA}}$ , which represents the ECMWF versus ACCESS comparison, for the airport stations, and airport station groups. As with the Official versus ECMWF comparison in Fig. 6, 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. **Might be interesting to note that ACCESS-C+ does not run over Darwin or Canberra, possibly explaining the better performance of ACCESS there.**

Naively, 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. However, these results are not surprising in light of the mesoscale verification challenges discussed in section 1. The AWS data resolves motion with time scales as low as 10 minutes, and at arbitrarily small spatial scales: it therefore includes more unpredictable turbulence than either model dataset. Furthermore, because ACCESS runs at higher spatiotemporal resolutions than ECMWF, it includes additional scales of motion, and therefore adds additional variability to the wind fields. Unless the additional variability in ACCESS is perfectly correlated with observations, the average of  $|\mathbf{u}_{\text{AWS}} - \mathbf{u}_{\text{A}}|$  will therefore increase, unless this additional variability is compensated for by a reduction in bias, i.e.  $|\overline{\mathbf{u}}_{\text{AWS}} - \overline{\mathbf{u}}_{\text{A}}|$  decreases. These ideas are discussed in greater detail in section 4. Note finally that the results for the Official versus ECMWF

comparison in Fig. 6 largely mirror those of the ECMWF versus ACCESS comparison in Fig. 7, suggesting that similar arguments apply to Official, as it is based on both ACCESS, ECMWF as well as forecaster edits, which contribute additional variability.

#### *b. Seasonal Comparison*

Figure 8 provides the Climatological Wind Perturbation Index (CWPI) values and confidence scores for the coastal station groups for  $CWPI_{OA}$ ,  $CWPI_{OE}$  and  $CWPI_{EA}$ , which represent 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 reasonable confidence. However, both ACCESS and ECMWF outperform Official at 23:00 and 00:00 UTC, consistent with the  $\overline{WPI}$  results of 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 9 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 very slight 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 correcting for this somewhat. Both Official and ECMWF slightly exaggerate the magnitude of the climatological sea-breeze, which peaks around 09:00 UTC, with ACCESS performing well in this respect.

At the South WA station group from 01:00 to 05:00 UTC, Official outperforms ECMWF with confidence scores of at least 88%. Figure 9 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 noticeably 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, Official outperforms both ACCESS and ECMWF with confidence scores of at least 95% and 75%, respectively. Figure 9 c) shows that these times correspond to “dimples” in the perturbation hodographs that are 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 9 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 02:00 to 05:00 UTC and 09:00 to 12:00 UTC, Official outperforms both ACCESS and ECMWF, although confidence scores do not exceed 88% and 65% respectively. Figure 9 d) shows that although the Official forecast captures the amplitude of the perturbations from 01:00 to 05:00 UTC almost perfectly, its diurnal cycle is out of phase with that of the AWS during this period, explaining why Official only slightly outperforms ACCESS in the results of Figures 8 a) and b).

For contrast, Fig. 10 presents the CWPI values and confidence scores for  $CWPI_{OE}$ , which represents the Official versus ECMWF comparison, for the airport stations, and airport station groups. These results show much greater similarity with the Official versus ECMWF comparisons at the coastal station groups shown in Figs. 8 c) and d), than do the analogous  $\overline{WPI}$  results in Fig. 6 and Figs. 2 c) and d). This likely because the temporal averaging has reduced the additional unpredictable variability in Official, revealing biases in Official and ECMWF that are partly shared across the three spatial scales. This point is discussed further in section 4. The analogous CWPI comparisons with ACCESS (not shown) are more ambiguous, although are generally more

472 favourable for Official than those for  $\overline{WPI}$ . For example Official outperforms both ACCESS and  
473 ECMWF at Darwin Airport from 02:00 to 03:00 and 15:00 to 17:00 UTC with at least 90% confi-  
474 dence. However, Official performs less well compared to ACCESS over the airport station groups,  
475 with CWPI values close to zero for most times and station groups, but ACCESS now strongly  
476 outperforming Official over the Darwin Airport station group.

477 Note that the hodographs in Fig. 9 are roughly elliptical in shape, suggesting that descriptive  
478 quantities can be estimated by fitting equations (5) and (6) to the zonal and meridional climatolog-  
479 ical perturbations, then calculating these quantities from the fit, as described in section 2. Figure  
480 11 provides the  $R^2$  values for the fits of the zonal and meridional perturbations to equations (5) and  
481 (6), respectively. The fit performs best at the coastal station group spatial scale, with  $R^2$  generally  
482 above 95%. It also performs well at the airport station and airport station group scales, with a few  
483 exceptions, including the ACCESS and Official meridional perturbations at the Canberra airport  
484 station group, and the ECMWF zonal perturbations at Melbourne airport.

485 The ellipse fits are used to derive four descriptive quantities: the maximum perturbation speed,  
486 the eccentricity of the fitted ellipse, the angle the fitted ellipse's semi-major axis makes with lines  
487 of latitude, and the time maximum perturbation speed occurs. Figure 12 provides these four quanti-  
488 ties for each dataset and location across the three spatial scales. A variety of structural differences  
489 are apparent at a number of locations and scales. For example, Fig. 12 a) shows that at Bris-  
490 bane airport, the maximum AWS perturbation is at least 1 kn greater than Official, ACCESS and  
491 ECMWF, and Fig. 12 c) shows that the orientation of the AWS fitted ellipse is at least 20 degrees  
492 anti-clockwise from the other datasets. Figures 13 a) and b) show hodographs of the Brisbane air-  
493 port perturbation climatology and ellipse fit, respectively. Although the ellipse fits suppress some  
494 of the asymmetric details, they capture the amplitudes and orientations of the real climatological  
495 diurnal cycles well. In this case the results show that the average AWS sea-breeze approaches

496 from the northeast, whereas the forecast and model sea-breezes approach more from the east-  
 497 northeast. To check whether this just represents a direction bias of the Brisbane Airport station,  
 498 Fig. 12 shows the climatological perturbations at the nearby Spitfire Channel station (see Fig. 1  
 499 for the location of this station, and other stations referred to in this section). While the amplitude  
 500 bias is smaller at Spitfire Channel than Brisbane Airport, the directional bias is at least as high.  
 501 A similar directional bias is evident at the nearby Inner Beacon station (not shown), although the  
 502 bias is smaller than at Spitfire Channel and Brisbane Airport. Thus, the directional bias in Official,  
 503 ACCESS and ECMWF at these stations is likely genuine, and not just a consequence of biased  
 504 AWS observations. Figure 1 shows there are two small islands to the east of Brisbane airport; the  
 505 more northwesterly orientation of the Brisbane Airport sea-breeze suggests these islands may be  
 506 redirecting winds between the east coast of Brisbane and the west coasts of these islands, and that  
 507 this local effect is not being captured in Official, ACCESS or ECMWF.

508 Another example is the Hobart Airport station. Figure 12 c) shows that the ellipse fits for the  
 509 AWS perturbations are oriented 31, 35 and 62 degrees anti-clockwise from the ECMWF, Official  
 510 and ACCESS ellipse fits, respectively. Figures 11 a) and b) show that the ellipse fit for the AWS  
 511 perturbations at Hobart airport only achieve  $R^2$  values of 59% and 68% for the  $u$  and  $v$  components,  
 512 respectively. However, figures 13 d) and e) show that the fit still captures orientations accurately,  
 513 although it underestimates the maximum AWS perturbation. Figure 13 f) shows the climatological  
 514 perturbations at the Hobart (city) station, which also show a large difference in orientation between  
 515 ACCESS and AWS. Given the timing of the westerly perturbations in ACCESS, and the fact  
 516 that the prevailing winds around Tasmania are westerly, these results suggest that ACCESS is  
 517 exaggerating the boundary layer mixing processes involved in the diurnal cycle around Hobart.

518 The South WA station group also provides an interesting example. Here the ACCESS and  
 519 Official ellipse fits are oriented at least 49 degrees anti-clockwise from those of AWS and ECMWF,

520 and the ECMWF perturbations peak between 1.2 and 2.5 hours after the other datasets. These  
521 differences occur because eccentricity values are low for this station group, and Figure 9 b) shows  
522 that the westerly perturbations associated with boundary layer mixing are weaker for ECMWF  
523 that the other datasets. A similar issue affects the VIC station group, explaining why the AWS  
524 ellipse fit is oriented at least 49 degrees anti-clockwise from those of the other datasets.

525 The Darwin Airport, Darwin Airport station group, and NT station group provide further exam-  
526 ples. In these cases there are timing differences between the perturbation maximums of up to 8.2  
527 hours. Figure 14 shows that these differences occur because for some datasets, the later north to  
528 northwesterly sea-breeze perturbations dominate the diurnal wind cycle, but for other datasets the  
529 earlier easterly to southeasterly boundary layer mixing effects dominate.

#### 530 **4. Discussion**

531 The results of section 3 may have implications for forecasting practice. If the goal of land-sea  
532 breeze and boundary layer mixing edits is to reduce absolute errors in the following day's forecast  
533 of the surface wind fields, then a necessary (but not sufficient) condition for this to occur is for  
534 these edits to at least reduce the absolute errors in the diurnal component of the surface wind fields.  
535 However, the WPI comparisons in Figs. 2 and 6 suggest that this is only possible when absolute  
536 error is calculated at coarse spatial scales. If the Official forecast is based, at least partly, on an  
537 edited high resolution model guidance dataset like ACCESS, then due to the mesoscale verification  
538 issues discussed in section 1, the larger absolute errors associated with a higher resolution model  
539 completely mask the effect of the edits, with a lower resolution unedited model like ECMWF  
540 scoring better overall. While the CWPI results in Figs. 8 and 6 suggest that forecaster edits can  
541 improve the accuracy of diurnal wind cycles in a climatological sense, it is not clear if these  
542 improvements have operational significance.



543 To investigate these ideas further, consider first just the zonal components of the AWS and  
 544 Official wind perturbations, denoted by  $u_{\text{AWS}}$  and  $u_{\text{O}}$  respectively. Considering just the values at  
 545 a particular hour UTC, at a particular station, over the entire June, July, August time period, the  
 546 mean square error  $\text{mse}(u_{\text{AWS}}, u_{\text{O}}) = \overline{(u_{\text{AWS}} - u_{\text{O}})^2}$  can be decomposed

$$\text{mse}(u_{\text{AWS}}, u_{\text{O}}) = \underbrace{\text{var}(u_{\text{AWS}}) + \text{var}(u_{\text{O}}) - 2 \cdot \text{covar}(u_{\text{AWS}}, u_{\text{O}})}_{\text{var}(u_{\text{AWS}} - u_{\text{O}})} + \underbrace{(\bar{u}_{\text{AWS}} - \bar{u}_{\text{O}})^2}_{\text{squared bias}} \quad (8)$$

547 where var, covar and over-bars denote the sample variance, covariance and mean respectively.  
 548 The first three terms are the total variance of  $u_{\text{AWS}} - u_{\text{O}}$ , whereas the last term is the square of the  
 549 bias between  $u_{\text{AWS}}$  and  $u_{\text{O}}$ . Note that the mean square error  $\text{mse}(u_{\text{AWS}}, u_{\text{O}})$  is closely related to  
 550  $\overline{\text{WPI}}$ , which is the difference between the mean absolute error of Official and AWS, and a model  
 551 guidance dataset and AWS. Similarly, the CWPI is closely related to the squared bias component  
 552  $(\bar{u}_{\text{AWS}} - \bar{u}_{\text{O}})^2$  of the mean square error. Equation (8) can also be applied to wind perturbations that  
 553 have first been spatially averaged over a station group, and to  $\text{mse}(u_{\text{AWS}}, u_{\text{E}})$  and  $\text{mse}(u_{\text{AWS}}, u_{\text{A}})$ ,  
 554 where  $u_{\text{E}}$  and  $u_{\text{A}}$  are the ECMWF and ACCESS zonal perturbations, respectively.

555 Figure 15 shows each term in the mean square error decomposition of equation 8 for both  
 556  $\text{mse}(u_{\text{AWS}}, u_{\text{O}})$  and  $\text{mse}(u_{\text{AWS}}, u_{\text{E}})$ , for Darwin Airport, the Darwin station group, and the NT  
 557 station group. This region provides an interesting case study because Fig. 6 shows that Offi-  
 558 cial has some skill at both Darwin Airport and over Darwin Airport station groups, in contrast  
 559 to most other locations. At Darwin Airport,  $\text{mse}(u_{\text{AWS}}, u_{\text{O}})$  exceeds  $\text{mse}(u_{\text{AWS}}, u_{\text{E}})$  from 04:00  
 560 to 16:00 UTC due to higher total variance, whereas outside of these times  $\text{mse}(u_{\text{AWS}}, u_{\text{E}})$  ex-  
 561 ceeds  $\text{mse}(u_{\text{AWS}}, u_{\text{O}})$  due to larger bias. The higher total variance of  $u_{\text{AWS}} - u_{\text{O}}$  occurs because  
 562  $\text{var}(u_{\text{O}}) > \text{var}(u_{\text{E}})$ . This additional variability is mostly random from 04:00 to 14:00 UTC, i.e.  $u_{\text{O}}$   
 563 is not sufficiently correlated with  $u_{\text{AWS}}$  at these times for the additional variability of  $u_{\text{O}}$  to pro-  
 564 duce a reduction in mean square error. Thus, while the bias between Official and AWS is lower,

565 or about the same, as that between ECMWF and AWS, the higher random variability of Official  
566 results in higher mean square error for most of the day. Figure 16 shows similar conclusions can be  
567 drawn for the meridional perturbations at Darwin Airport, although in this case  $\text{var}(u_O) > \text{var}(u_E)$   
568 for the entire day. Most of the difference between the WPI and CWPI scores for the Official ver-  
569 sus ECMWF comparison at Darwin Airport in Figures 6 and 10, respectively, can be explained  
570 through the different mean square error and bias terms for the zonal perturbations alone.

571 Figure 14 a) shows that ECMWF’s climatological perturbations at Darwin Airport underestimate  
572 the easterly perturbations from 00:00 to 03:00 UTC, which are presumably associated with bound-  
573 ary layer mixing processes. Official does a better job of resolving these easterly perturbations, but  
574 is generally outperformed by ECMWF in resolving the northerly sea-breeze perturbations. Similar  
575 points can be made for the Darwin and NT coastal station groups. While spatial averaging reduces  
576 a portion of the unpredictable variability in Official, Official also often has larger meridional biases  
577 at these scales compared to ECMWF. Figures 14 and 12 show that these biases can be explained  
578 in terms of amplitude and orientation differences between Official, ECMWF and AWS. Figures  
579 analogous to Figs. 15 and 16, but for other locations around Australia, show similar results, but  
580 generally without large biases in the Official forecast at the coarser scales like those present in the  
581 meridional perturbations over the Darwin Airport station group and NT coastal station group.

582 These examples illustrate the idea that the additional unpredictable variability introduced by a  
583 higher resolution edited forecast needs to be “paid for” by a reduction in bias, otherwise the net  
584 result will just be an increase in error. One obvious way to reduce the influence of unpredictable  
585 variability at higher resolutions is to move to an ensemble forecasting approach. However, compu-  
586 tational resources typically require a choice between either a single model of higher resolution, or  
587 an ensemble of models at lower resolution, and there is a long and spirited debate in the literature  
588 about the relative merits of each (Brooks and Doswell III 1993). If ensemble forecasting is not

possible, careful thought must be given to precisely what scales of motion the Official forecast is intended to represent. If the end user doesn't care about the "realism" of the forecast and simply wants the lowest errors, and if daily errors are larger with a higher resolution, edited forecast, than with a coarse model guidance product, there may be an argument for smoothing or filtering the higher resolution forecast before it is provided to the end user, assuming of course the higher resolution forecast actually reduces biases.

Careful thought should therefore be given to precisely what the Official forecast intends to represent. It is unclear if the Official forecasts wind fields should be regarded as predictions of the actual winds at a specific location, or a predicted Reynold's average, and if so, at what scale. If we assume the BoM's Official wind forecast intends to represent variability at hourly timescales, and horizontal scales less than 50 kms, then sea-breeze and boundary layer mixing edits appear to have little effect, because the intended reduction in error is washed out by the unpredictable turbulent variability at these scales, and lower errors can be achieved simply by using a coarser resolution unedited model forecast. However, some users may be more interested in whether the variability of the forecast wind field matches those of observations, than in whether the forecast minimises absolute error, and a higher resolution edited forecast will likely perform better than a coarse resolution model in this regard.

Ambiguity in what the Official forecast represents, and the resulting verification challenge, points to an important role for human forecasters: post-processing and editing model data so that the forecast consistently represents what is of interest to individual users. One barrier to this is that the Official forecast is provided to the national public as a whole, which includes diverse users with different representation needs. Part of the solution may therefore be the development of a secondary economy, either within national weather services like the BoM, or in the private sector, where human forecasters ensure their forecast products consistently represent the "filtered

613 version of reality” of interest to the end user. This would then help address the representation  
614 problem as it applies to the verification of operational forecasts, as the individual user’s represen-  
615 tation needs could then be taken as the intended representation of the forecast, and appropriate  
616 verification metrics chosen accordingly.

## 617 **5. Conclusion**

618 In this paper we have presented a method for verifying the diurnal component of wind forecasts  
619 issued to the public, with the intended application being the assessment of the edits Australian  
620 forecasters make to model guidance datasets in order to better resolve land-sea breeze and bound-  
621 ary layer mixing processes. We considered two temporal scales, and three spatial scales, but the  
622 method is immediately generalisable to other scales.

623 When the method is applied to Australian forecast data, the results indicate that when the Of-  
624 ficial, edited forecast, is assessed on a daily basis, it only produces lower absolute errors in the  
625 diurnal wind cycle at very coarse spatial scales of at least 500 km. Even at these scales, the im-  
626 provements are isolated to particular times of day, and only apply at some locations. Furthermore,  
627 while the Official forecast can outperform the two most commonly used model guidance products  
628 ACCESS and ECMWF, in the sense of absolute error, it rarely outperforms both simultaneously,  
629 suggesting that forecaster skill lies more in making the choice of model guidance than in making  
630 edits. When the Official forecast is assessed on a seasonal basis, i.e. the average or climatological  
631 diurnal wind cycle is assessed, the Official forecast performs better than when assessed on a daily  
632 basis, particularly at the single station spatial scale. However, its performance is not overwhelm-  
633 ing, as it struggles to unambiguously produce lower absolute error than ACCESS.

634 An alternative approach to forecast verification is to assess the realism of “structural” features of  
635 the atmospheric, and following Gille et al. (2005), we do this in an objective way by fitting ellipses

636 to hodographs of the climatological diurnal wind cycles. In the Australian context this approach  
637 reveals structural biases in the Official forecast, including directional biases in the approach of the  
638 sea-breeze at Brisbane airport, eccentricity biases along the coast of NSW, and amplitude biases  
639 along the southwest coast of WA.

640 Future research could extend this study in multiple directions. An important goal would be to  
641 identify precisely the spatial scale at which the Official forecast can produce lower absolute error  
642 on a daily basis than a coarse model like ECMWF: for Australia over the time period considered,  
643 our study shows that this occurs somewhere between our airport station group scale (50 - 200  
644 km), and coastal station group scale (1000 - 2000 km). Another interesting question is whether the  
645 diurnal component of the Official forecast can outperform a climatological diurnal cycle calculated  
646 from previous observations.

647 In summary, we have shown that forecaster edits can reduce errors in the diurnal cycle of sur-  
648 face winds, but only at very large spatial scales, or in a climatological sense. This scale sensitivity  
649 suggests careful thought needs to be given to how the representation problem applies to the ver-  
650 ification of operational forecasts, although the consistent representation of a particular “filtered  
651 version of reality” is difficult for national forecasting centres due to the diverse representation  
652 needs of forecast users, and in the Australian context, due to the hybrid nature of the Official  
653 forecast.

## 654 **References**

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

- 658 Brooks, H. E., and C. A. Doswell III, 1993: New technology and numerical weather prediction  
659 — a wasted opportunity? *Weather*, **48** (6), 173–177, doi:10.1002/j.1477-8696.1993.tb05877.x,  
660 URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/j.1477-8696.1993.tb05877.x>, <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/j.1477-8696.1993.tb05877.x>.  
661
- 662 Bureau of Meteorology, 2010: Operational implementation of the ACCESS numerical weather  
663 prediction systems. Tech. rep., Bureau of Meteorology, Melbourne, Victoria. [Available online  
664 at <http://www.bom.gov.au/australia/charts/bulletins/apob83.pdf>].
- 665 Bureau of Meteorology, 2019: Meteye. Bureau of Meteorology, [Available online at <http://www.bom.gov.au/australia/meteye/>].  
666
- 667 Dai, A., and C. Deser, 1999: Diurnal and semidiurnal variations in global surface wind and diver-  
668 gence fields. *Journal of Geophysical Research*, **104**, 31 109–31 125.
- 669 Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed  
670 framework. *Meteor. Appl.*, **15** (1), 51–64, doi:10.1002/met.25, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.25>, <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/met.25>.  
671  
672
- 673 Efron, B., 1979: Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7** (1),  
674 1–26, doi:10.1214/aos/1176344552.
- 675 Englberger, A., and A. Dörnbrack, 2018: Impact of the diurnal cycle of the atmospheric  
676 boundary layer on wind-turbine wakes: A numerical modelling study. *Boundary-Layer Me-*  
677 *teorology*, **166** (3), 423–448, doi:10.1007/s10546-017-0309-3, URL <https://doi.org/10.1007/s10546-017-0309-3>.  
678

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>].

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), [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>.

Lynch, K. J., D. J. Brayshaw, and A. Charlton-Perez, 2014: Verification of european sub-seasonal wind speed forecasts. *Monthly Weather Review*, **142** (8), 2978–2990, doi:10.1175/

MWR-D-13-00341.1, URL <https://doi.org/10.1175/MWR-D-13-00341.1>, <https://doi.org/10.1175/MWR-D-13-00341.1>.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bulletin of the American Meteorological Society*, **83** (3), 407–430, doi:10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2, URL [https://doi.org/10.1175/1520-0477\(2002\)083<0407:DIHRPM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2), [https://doi.org/10.1175/1520-0477\(2002\)083<0407:DIHRPM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2).

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), [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>.

Rife, D. L., and C. A. Davis, 2005: Verification of temporal variations in mesoscale numerical wind forecasts. *Monthly Weather Review*, **133** (11), 3368–3381, doi:10.1175/MWR3052.1, URL <https://doi.org/10.1175/MWR3052.1>, <https://doi.org/10.1175/MWR3052.1>.



SciPy, 2019: Optimization and root finding (scipy.optimize). SciPy, [Available online at <https://docs.scipy.org/doc/scipy/reference/optimize.html>].

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 gabl's 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>.

Vincent, C. L., and T. P. Lane, 2016: Evolution of the diurnal precipitation cycle with the passage of a madden–julian oscillation event through the maritime continent. *Monthly Weather Review*, **144** (5), 1983–2005, doi:10.1175/MWR-D-15-0326.1, URL <https://doi.org/10.1175/MWR-D-15-0326.1>, <https://doi.org/10.1175/MWR-D-15-0326.1>.

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

Zaron, E. D., and G. D. Egbert, 2006: Estimating open-ocean barotropic tidal dissipation: The hawaiian ridge. *Journal of Physical Oceanography*, **36** (6), 1019–1035, doi:10.1175/JPO2878.1, URL <https://doi.org/10.1175/JPO2878.1>, <https://doi.org/10.1175/JPO2878.1>.

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), [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).

740 **LIST OF TABLES**

741	<b>Table 1.</b>	Average 10 m wind speeds for austral winter (June, July August) 2018, and	
742		austral summer (December, January, February) 2017/18 across the eight Aus-	
743		tralian capital city airport weather stations. . . . .	35

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

747	<b>Fig. 1.</b>	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. . . . .	37
748			
749	<b>Fig. 2.</b>	Heatmaps of $\overline{WPI}$ values, a), c), e), and confidence scores, b), d), f), for each coastal station group and hour of the day: Official versus ACCESS, a) and b), Official versus ECMWF, c) and d), ECMWF versus ACCESS, e) and f). Positive $\overline{WPI}$ values indicate that the former dataset in each pair is on average $\overline{WPI}$ kn closer to observations than the latter dataset is. Confidence scores provide the probability the population or “true” value of $\overline{WPI}$ is greater than zero. . . . .	38
750			
751			
752			
753			
754			
755	<b>Fig. 3.</b>	Time series, a), of $\overline{WPI}_{OA}$ and $\overline{WPI}_{OE}$ for the NT station group at 23:00 UTC, and b), hodographs showing hourly changes in winds, and c), wind perturbations, at the NT station group on the 3 <sup>rd</sup> of July 2018. . . . .	39
756			
757			
758	<b>Fig. 4.</b>	As in Fig. 3, but for, a), the South WA station group at 05:00 UTC, and b) and c), the winds and wind perturbations over the South WA station group on the 9 <sup>th</sup> June 2018. . . . .	40
759			
760	<b>Fig. 5.</b>	Wind soundings at, a), Darwin Airport, and b), Perth Airport. . . . .	41
761	<b>Fig. 6.</b>	The $\overline{WPI}_{OE}$ values, a) and c), and confidence scores, b) and d), for the airport stations, a) and b), and airport station groups, c) and d). . . . .	42
762			
763	<b>Fig. 7.</b>	As in Fig. 6, but for the $\overline{WPI}_{EA}$ values and confidence scores. . . . .	43
764	<b>Fig. 8.</b>	As in Fig. 2, but for the CWPI values and confidence scores. . . . .	44
765	<b>Fig. 9.</b>	Average wind perturbations over June, July and August 2018 for the, a), North WA, b) South WA, c) NSW and d), SA coastal station groups. . . . .	45
766			
767	<b>Fig. 10.</b>	As in Fig. 6, but for the CWPI values and confidence scores. . . . .	46
768	<b>Fig. 11.</b>	$R^2$ values as percentages for the fit of equation (5) to the zonal perturbations, a), c) and e), and equation (6) to the meridional perturbations, b), d) and f), for the airport stations, a) and b), airport station groups, c) and d), and coastal station groups, e) and f). . . . .	47
769			
770			
771	<b>Fig. 12.</b>	Metrics derived from fitting the elliptical equations (5) and (6) to the climatological perturbations: maximum perturbation speed, a), e) and i), eccentricity, b), f) and j), orientation, c), g) and k), and time of maximum perturbation, d), h) and l), for the airport stations, a) to d), airport station groups, e) to h), and coastal station groups, i) to l). . . . .	48
772			
773			
774			
775	<b>Fig. 13.</b>	Hodographs of the climatological perturbations at Brisbane and Hobart airports, a) and d), and the associated ellipse fits, b) and e). For comparison, c) and f) provide hodographs of the climatological perturbations at Spitfire Channel and Hobart (city), respectively. . . . .	49
776			
777			
778	<b>Fig. 14.</b>	Ellipse fits. <b>Could also include the ellipses, but this makes the figure very large.</b> . . . .	50
779	<b>Fig. 15.</b>	Actual perturbation standard deviation values. Note that official performs the worst at this scale! . . . . .	51
780			
781	<b>Fig. 16.</b>	Actual perturbation standard deviation values. Note that official performs the worst at this scale! . . . . .	52
782			

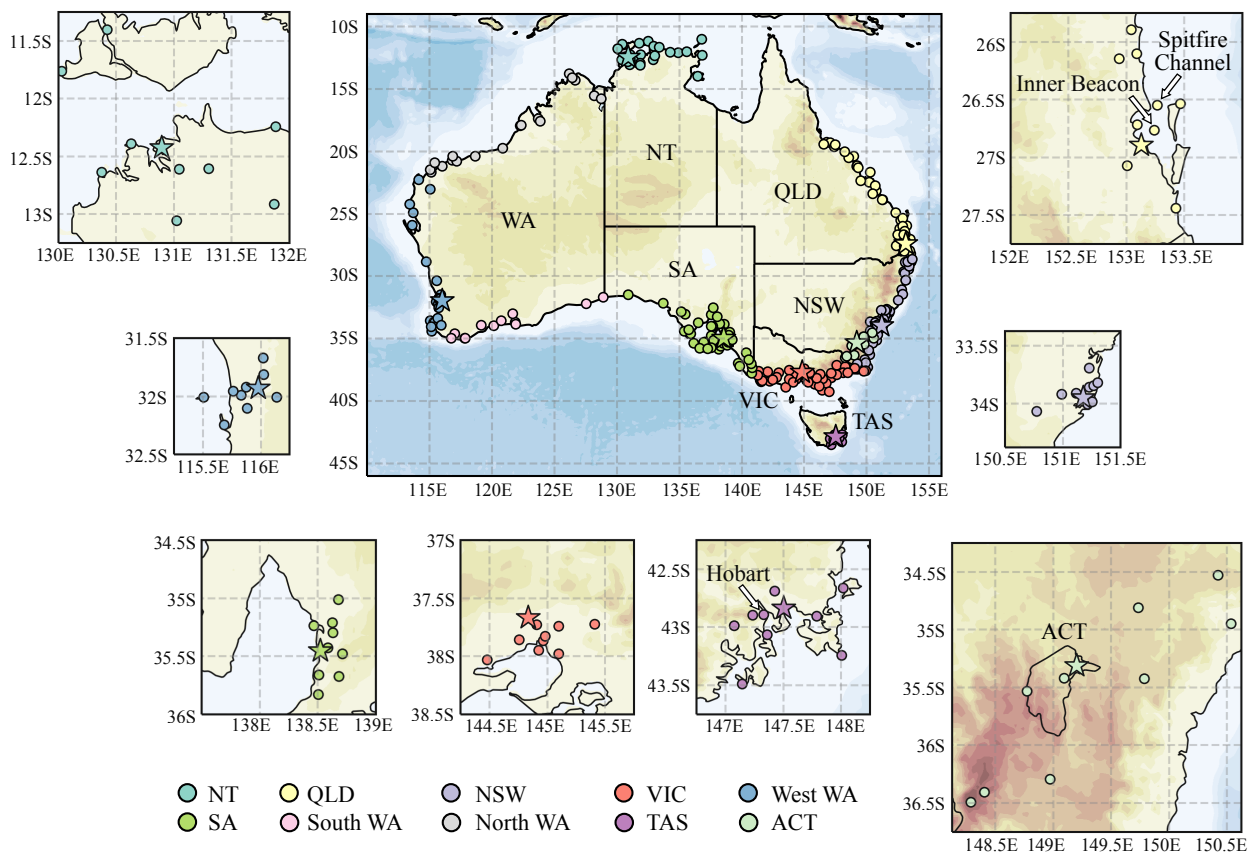


FIG. 1. Locations of the automatic weather stations used in this study. Stars indicate capital city airport stations. Height and depth shading intervals every 200 and 1000 m, respectively.

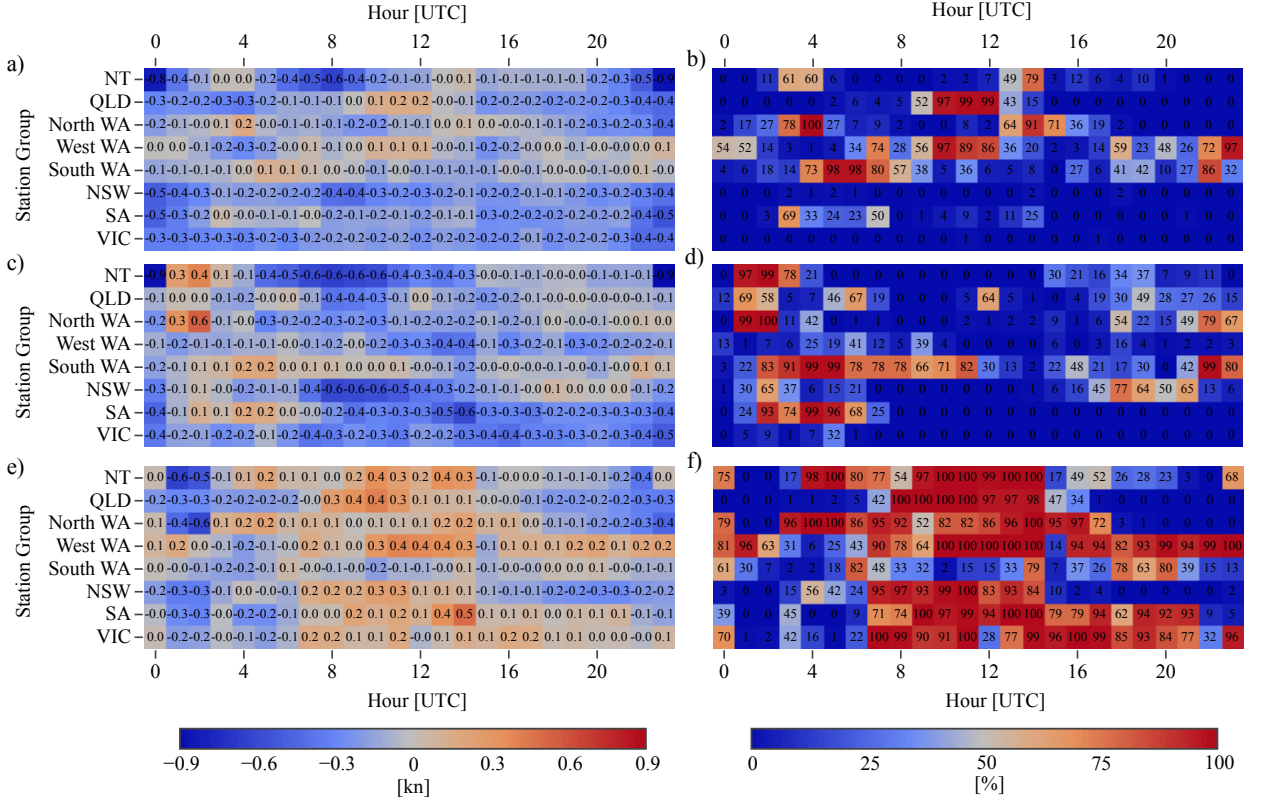


FIG. 2. Heatmaps of  $\overline{\text{WPI}}$  values, a), c), e), and confidence scores, b), d), f), for each coastal station group and hour of the day: Official versus ACCESS, a) and b), Official versus ECMWF, c) and d), ECMWF versus ACCESS, e) and f). Positive  $\overline{\text{WPI}}$  values indicate that the former dataset in each pair is on average  $\overline{\text{WPI}}$  kn closer to observations than the latter dataset is. Confidence scores provide the probability the population or “true” value of  $\overline{\text{WPI}}$  is greater than zero.

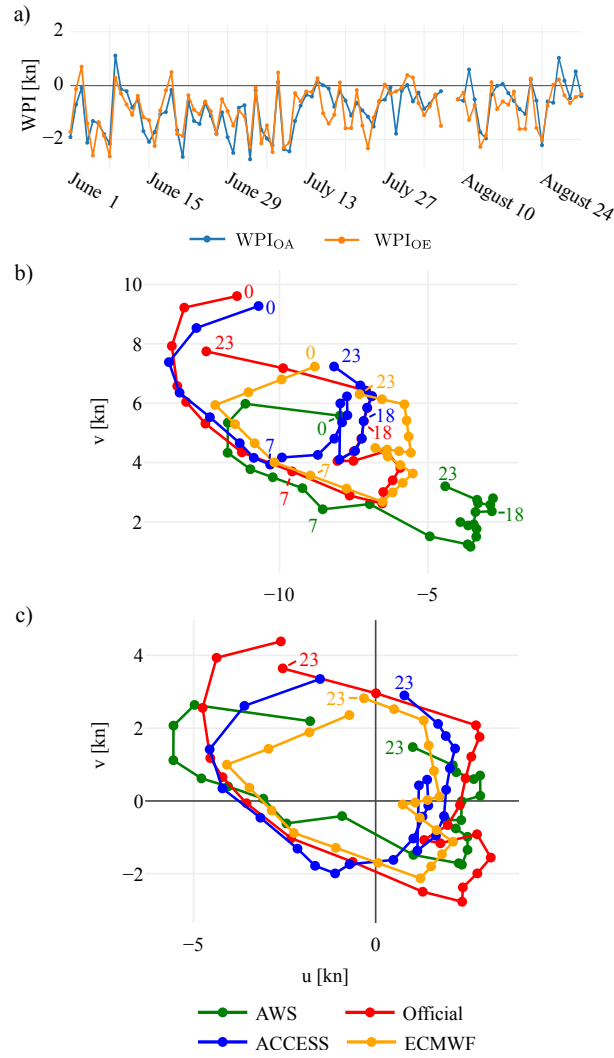


FIG. 3. Time series, a), of  $\overline{WPI}_{OA}$  and  $\overline{WPI}_{OE}$  for the NT station group at 23:00 UTC, and b), hodographs showing hourly changes in winds, and c), wind perturbations, at the NT station group on the 3<sup>rd</sup> of July 2018.

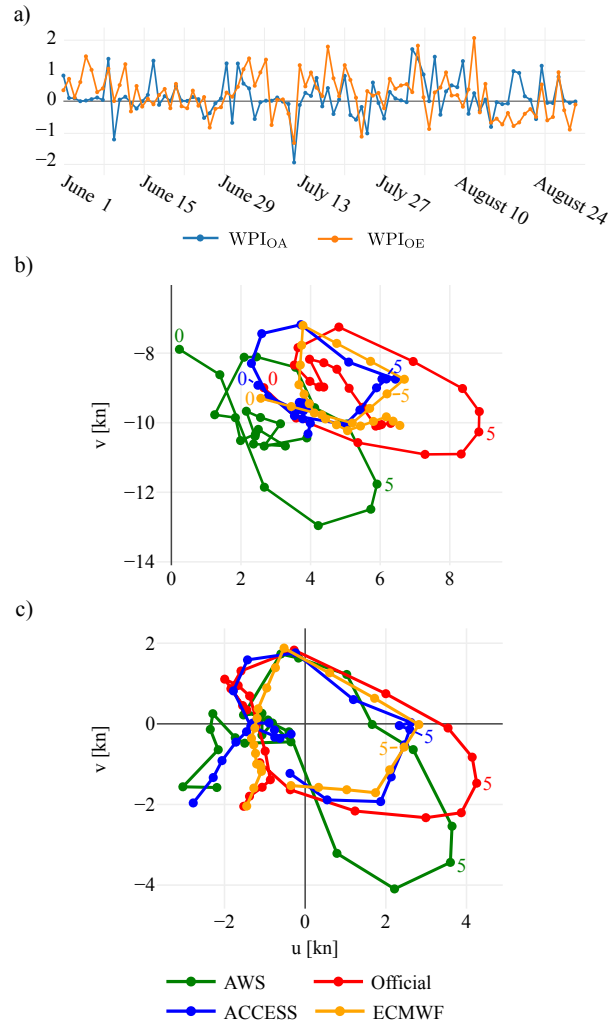


FIG. 4. As in Fig. 3, but for, a), the South WA station group at 05:00 UTC, and b) and c), the winds and wind perturbations over the South WA station group on the 9<sup>th</sup> June 2018.



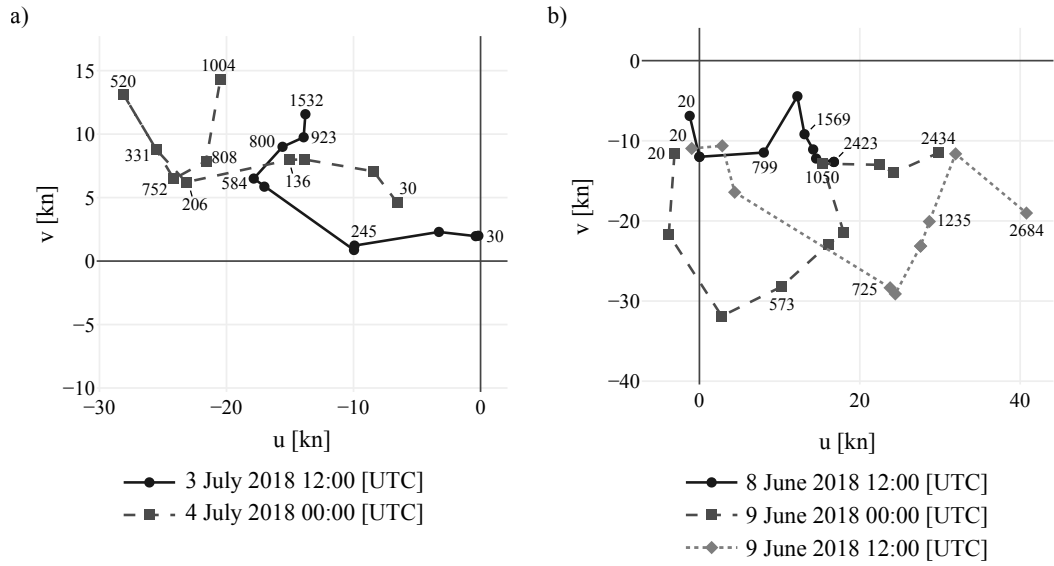


FIG. 5. Wind soundings at, a), Darwin Airport, and b), Perth Airport.

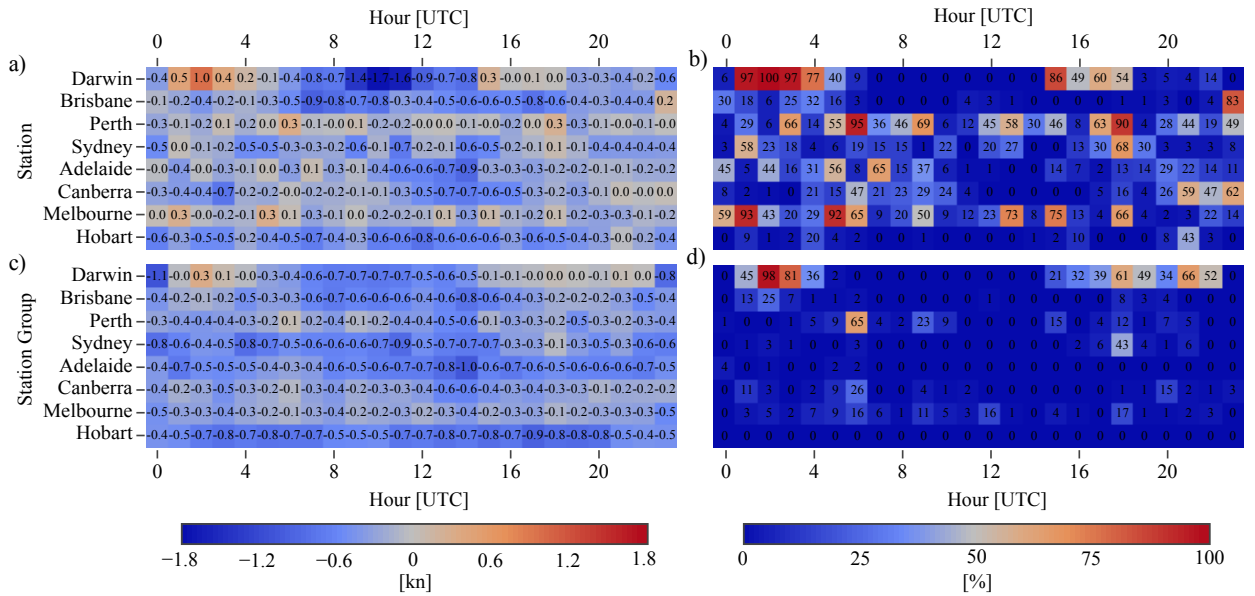


FIG. 6. The  $\overline{\text{WPI}}_{\text{OE}}$  values, a) and c), and confidence scores, b) and d), for the airport stations, a) and b), and airport station groups, c) and d).

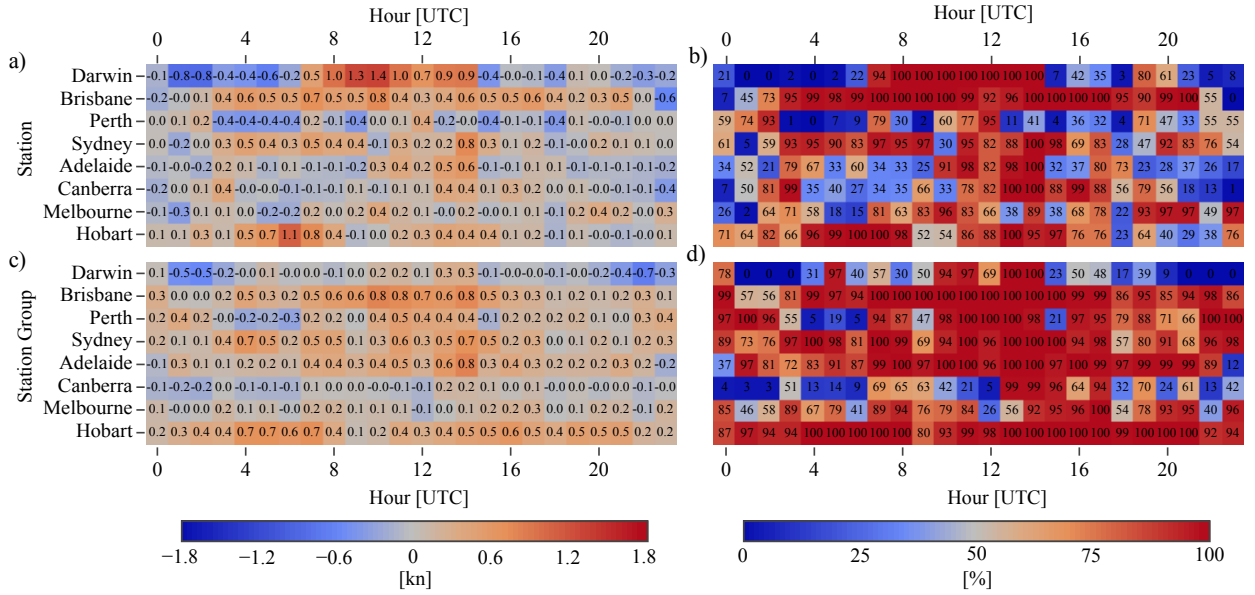


FIG. 7. As in Fig. 6, but for the  $\overline{\text{WPI}}_{\text{EA}}$  values and confidence scores.

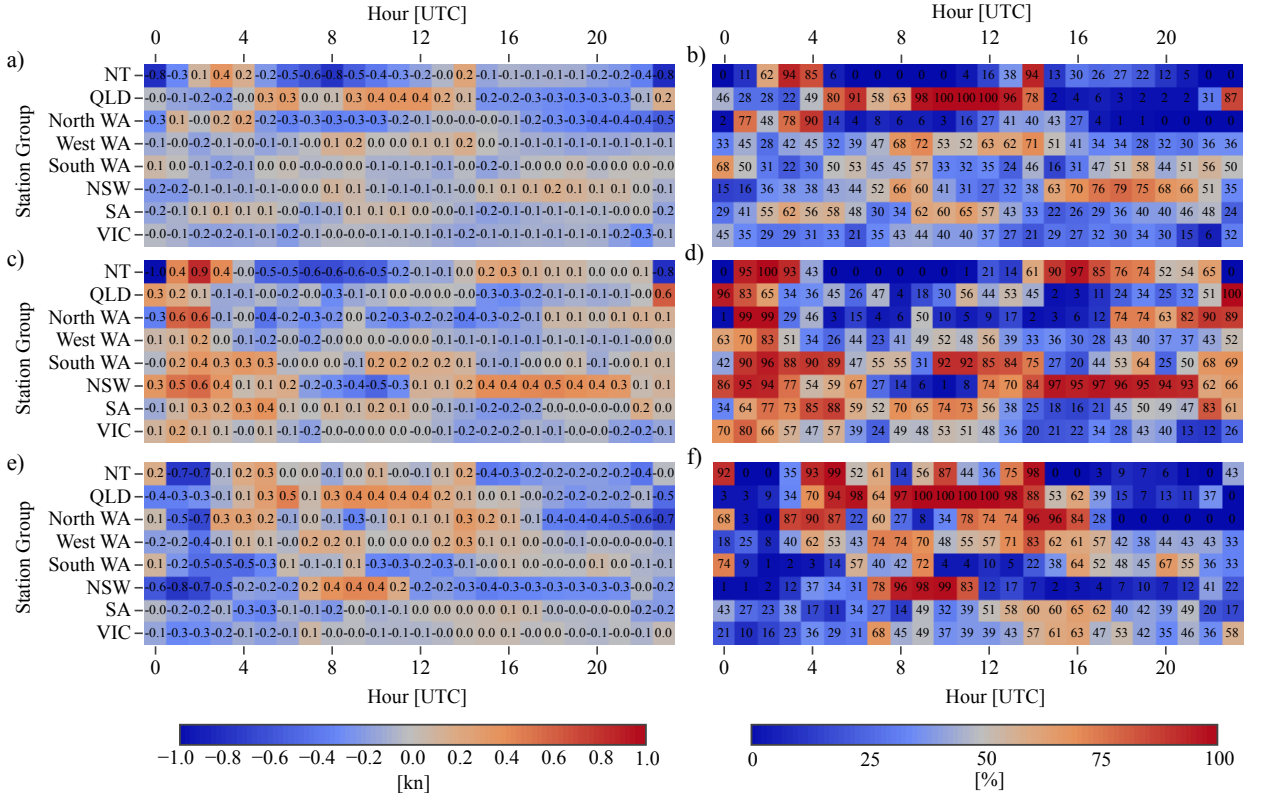
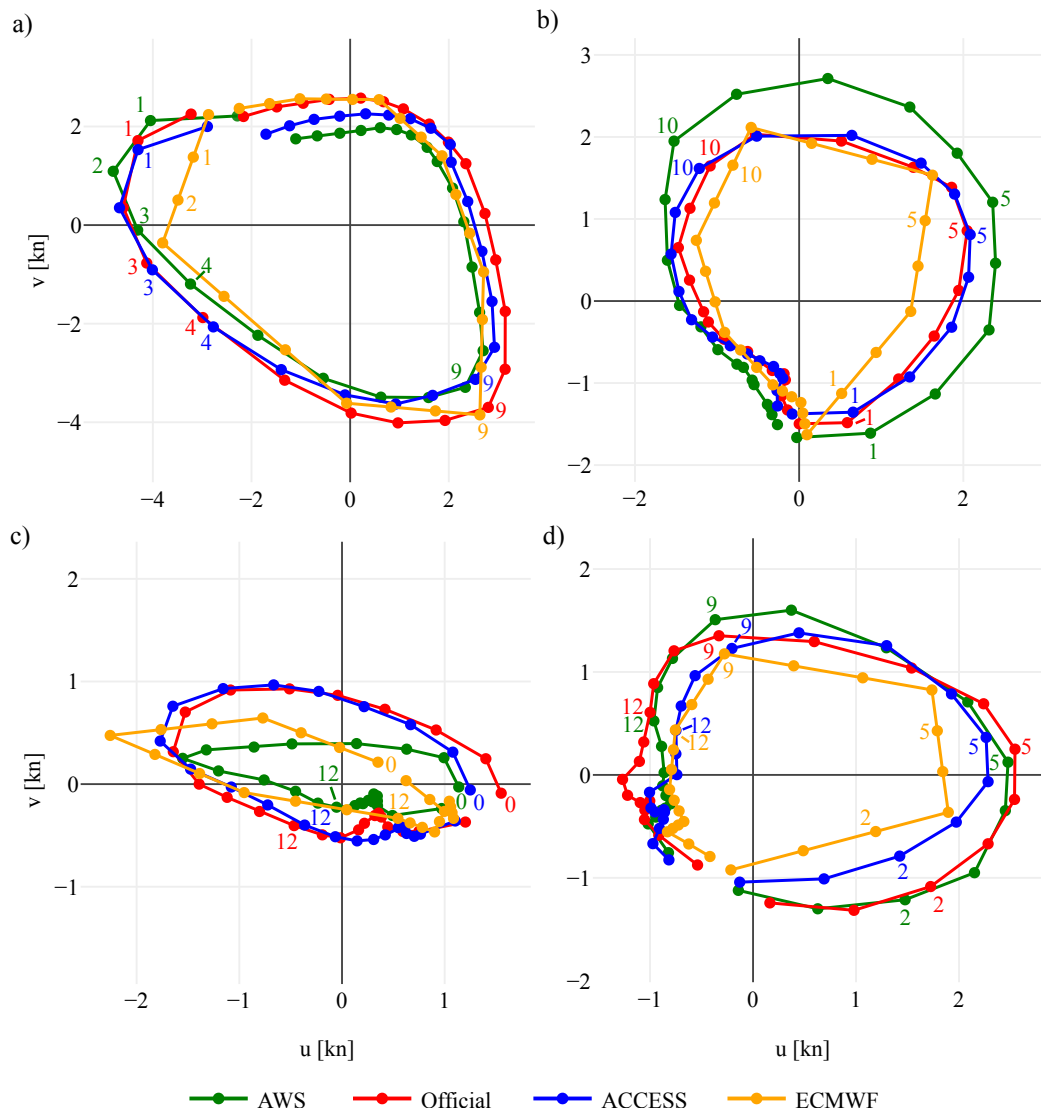


FIG. 8. As in Fig. 2, but for the CWPI values and confidence scores.



796 FIG. 9. Average wind perturbations over June, July and August 2018 for the, a), North WA, b) South WA, c)  
 797 NSW and d), SA coastal station groups.

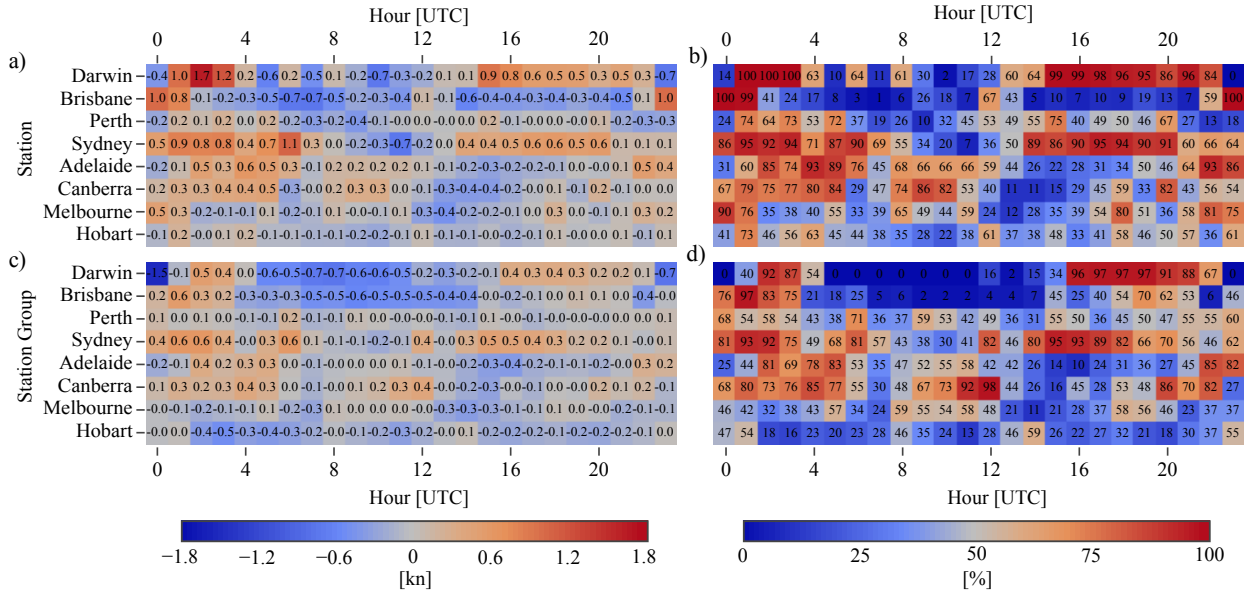


FIG. 10. As in Fig. 6, but for the CWPI values and confidence scores.

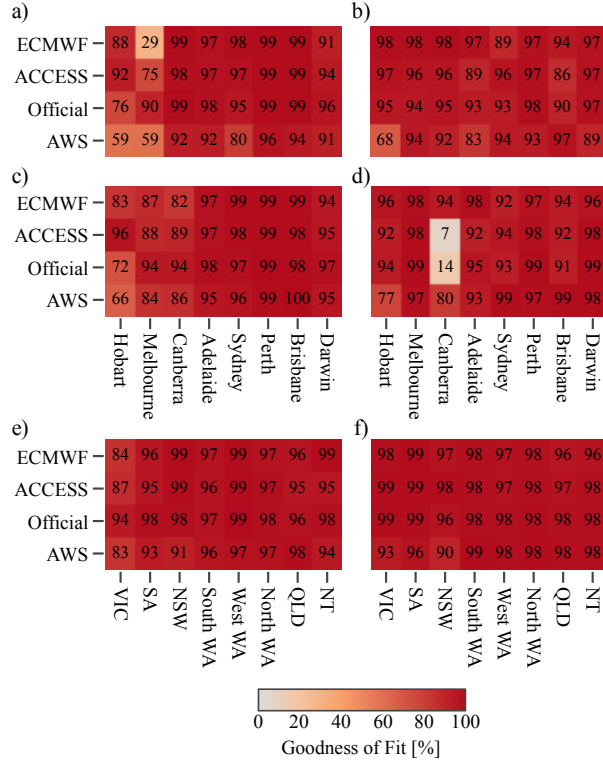


FIG. 11.  $R^2$  values as percentages for the fit of equation (5) to the zonal perturbations, a), c) and e), and equation (6) to the meridional perturbations, b), d) and f), for the airport stations, a) and b), airport station groups, c) and d), and coastal station groups, e) and f).

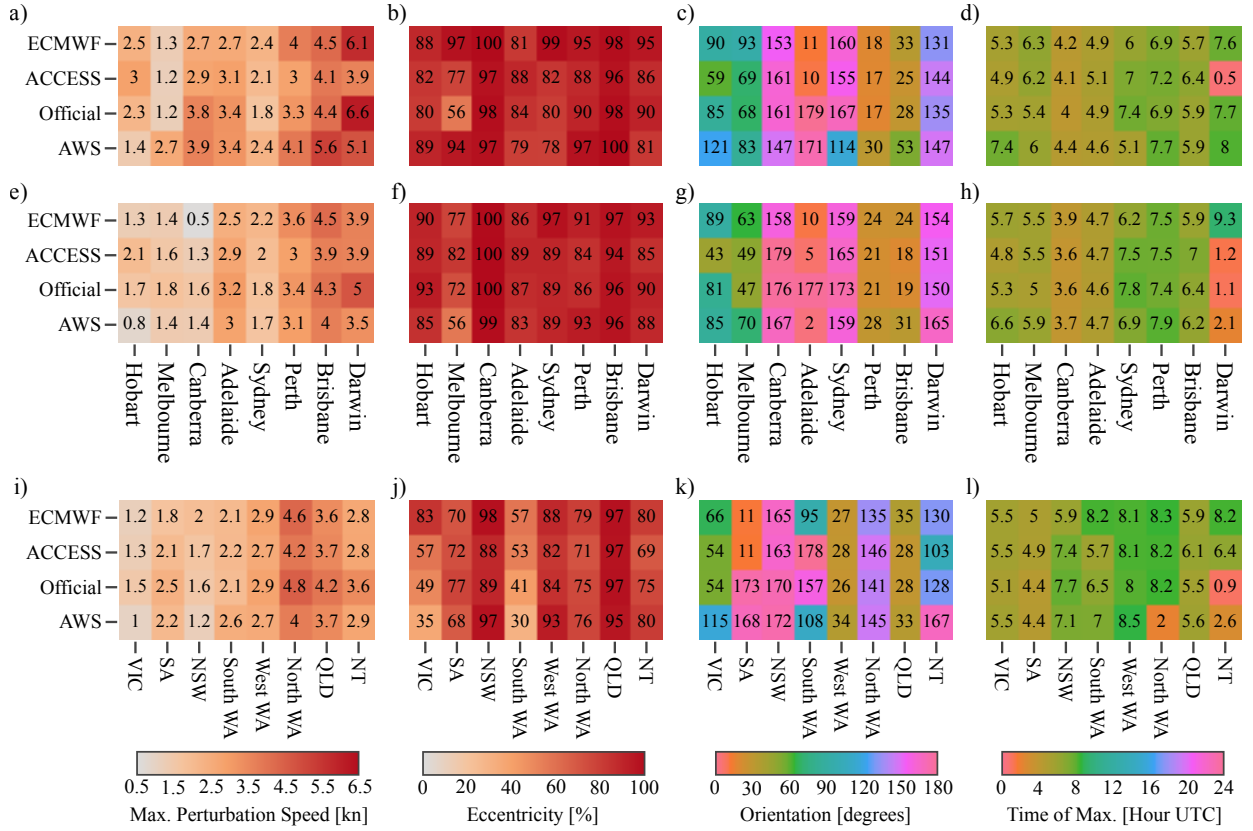


FIG. 12. Metrics derived from fitting the elliptical equations (5) and (6) to the climatological perturbations: maximum perturbation speed, a), e) and i), eccentricity, b), f) and j), orientation, c), g) and k), and time of maximum perturbation, d), h) and l), for the airport stations, a) to d), airport station groups, e) to h), and coastal station groups, i) to l).



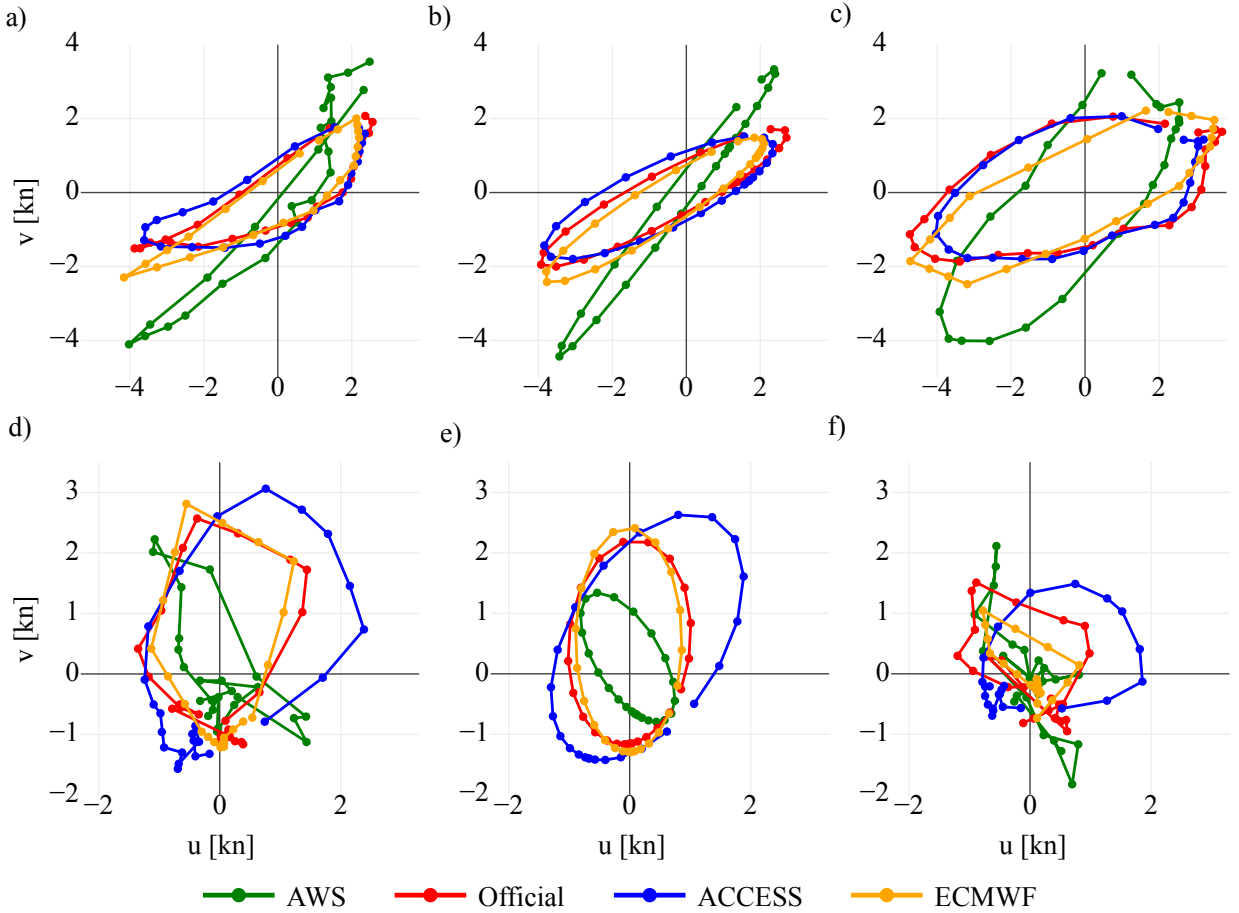


FIG. 13. Hodographs of the climatological perturbations at Brisbane and Hobart airports, a) and d), and the associated ellipse fits, b) and e). For comparison, c) and f) provide hodographs of the climatological perturbations at Spitfire Channel and Hobart (city), respectively.

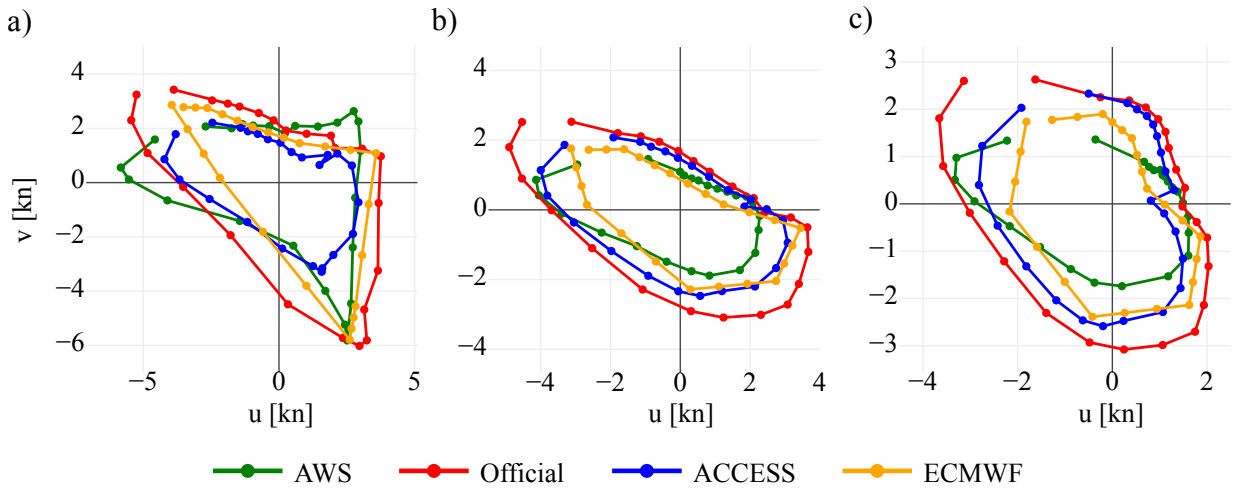


FIG. 14. Hodographs of the climatological perturbations at, a), Darwin Airport, b) the Darwin Airport station group, and c), the NT coastal station group.

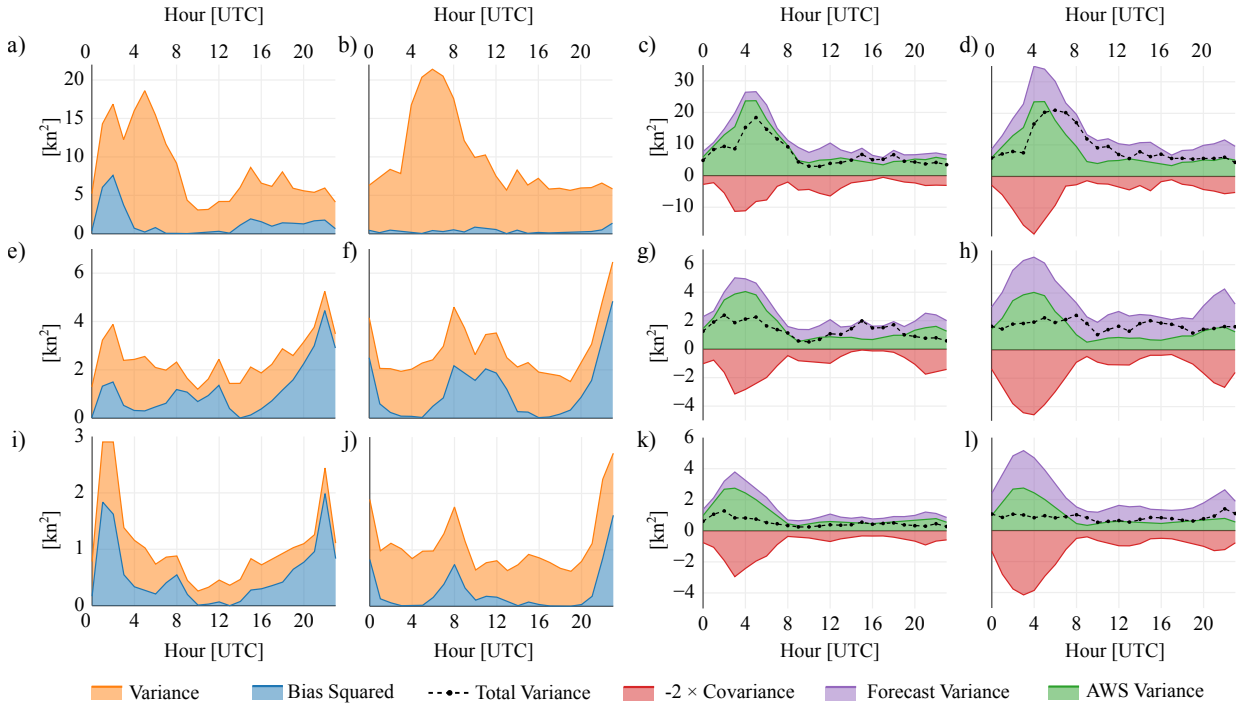


FIG. 15. Actual perturbation standard deviation values. Note that official performs the worst at this scale!

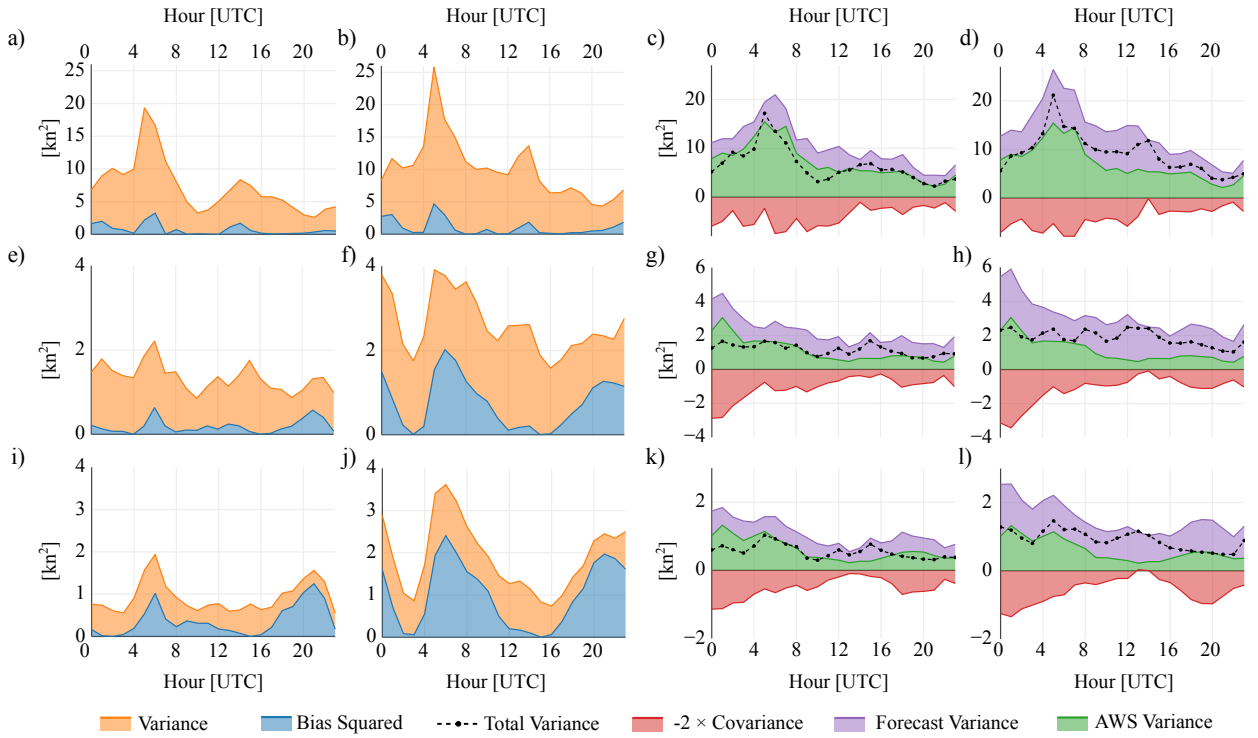


FIG. 16. Actual perturbation standard deviation values. Note that official performs the worst at this scale!