ORIGINAL ARTICLE

Journal Section

Land-Sea Breeze Forecast Verification

Ewan Short[1] | **Ben Price**[2] | **Derryn Griffiths**[3] | **Alexei Hider**[3]

[1]ARC Centre of Excellence for Climate Extremes, School of Earth Sciences, University of Melbourne, Parkville, VIC, 3010, Australia

[2]Bureau of Meteorology, Casuarina, NT, 0810, Australia

[3]Bureau of Meteorology, Melbourne, VIC, 3208, Australia

Correspondence
Ewan Short, ARC Centre of Excellence for Climate Extremes, School of Earth Sciences, University of Melbourne, Parkville, VIC, 3010, Australia
Email: ewan.short@unimelb.edu.au

Funding information
ARC Centre of Excellence for Climate System Science

This study presents a methodology for comparing the performance of Australian Bureau of Meteorology forecasts of the land-sea breeze with unedited model guidance products, such as those of the European Center for Medium-Range Weather Forecasting (ECMWF) and the Australian Community Climate and Earth System Simulation (ACCESS). The methodology is applied to the 8 Australian capital city airports. The results indicate that at some airports, human intervention to model guidance products adds value to land-sea breeze forecasts, whereas at other airports it does not.

KEYWORDS
land-sea breeze, forecast verification, Australia, Airports

1 | INTRODUCTION

Modern weather forecasts are produced by models in conjunction with human forecasters. For instance, a forecaster working for the Australian Bureau constructs a seven day forecast by first loading model data into the Graphical Forecast Editor (GFE) software package, then manually editing this model data as they see fit. 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 underesolved at synoptic scale model resolutions, or to address known biases of the models being used.

It is therefore important to assess not only the overall accuracy of weather forecasts, but also the contribution human forecaster edits make to this accuracy. If effective, but routine, editing procedures can be identified they can be automated, freeing forecasters up to focus on other tasks. One common edit involves changing the surface wind fields near coastlines to try to represent sea-breezes more realistically. Forecasters invest time in making sea-breeze edits because accurate predictions of near-surface winds are highly valued by a number of users, such as the aviation and energy (Smith et al., 2009) industries. Accurate sea-breeze forecasts are also valuable to environmental monitoring authorities, as these winds provide ventilation to coastal urban areas.

Assessing the accuracy of a weather forecast is a task far more nuanced than it might first appear. For instance, attempting to assess the accuracy of a precipitation forecast by comparing the rainfall amounts measured at an individual weather station to the closest grid point of a model prediction will often give poor results. Although the synoptic drivers of convection are usually well predicted, excatly where convective cells form, and where the most rain falls, is highly unpredictable. As such, it is often appropriate to use “fuzzy" verification metrics which measure the agreement between prediction and observation in a more indirect way. For instance, one approach known as “upscaling" is to first average forecast and observational data over a given spatial domain before calculating verification scores. Ebert (2008) provided a review of current “fuzzy verification" methodologies, and a framework for how they can be used to determine the spatial scales at which a given forecast has predictive skill.

Relatively few forecast verification studies have focused on near-surface winds, and the ones that have generally only considered wind speeds. Pinson and Hagedorn (2012) performed a verification study of the ECMWF 10 m wind speeds 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 underesolved 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.

Lynch et al. (2014) also performed a verification study of ECMWF 10 m wind speed data, with the goal of assessing skill at lead times of between 14 to 20 days. They compared ECMWF 32-day forecast model wind speeds with gridded ERA-Interim wind speeds between 2008-12, with both datasets analysed at a six hour temporal resolution. Before conducting the comparison, the wind speed data were transformed into wind-speed “anomaly" data by first calculating the mean wind speed at 0000, 0600, 1200 and 1800 UTC for each calendar day from the entire ERA-Interim record, and from a 20 year ECMWF 32-day model hindcast, then subtracting these means from the ERA-Interim and ECMWF 32-day model data respectively. Wind speed anomaly data was used so that stable seasonal and diurnal cycles did not contribute to verification scores. At the 14-20 day timescale around western Europe, the greatest skill was found in the boreal winter (austral summer) months of December, January and February.

Pinson and Hagedorn (2012) and Lynch et al. (2014) restricted their verification studies to wind speeds, but wind directions are also crucial to diagnosing whether land sea breezes - and the diurnal wind cycle more generally - are being forecast correctly. Furthermore, no previous published work has proposed a verification methodology to assess the accuracy of the diurnal wind cycle in forecasts, or of the contributions made to this accuracy by human forecaster edits of model output. Finally, no previously published work has considered the performance of ACCESS near surface winds, which together with ECMWF, are the model guidance products most widely used by Australian forecasters. Thus, the present study has two goals. First, to describe a methodology for comparing human edited forecasts of the land-sea breeze to unedited model guidance forecasts, in order to assess where and when human edits are producing 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 in detail, section 3 provides results, and sections 4 and 5 provide a discussion and a conclusion, respectively.

2 | DATA AND METHODS

This study compares both edited and non-edited Australian Bureau of Meteorology forecast data with automatic weather station (AWS) data across Australia. The comparison is performed by first isolating the diurnal signals of each

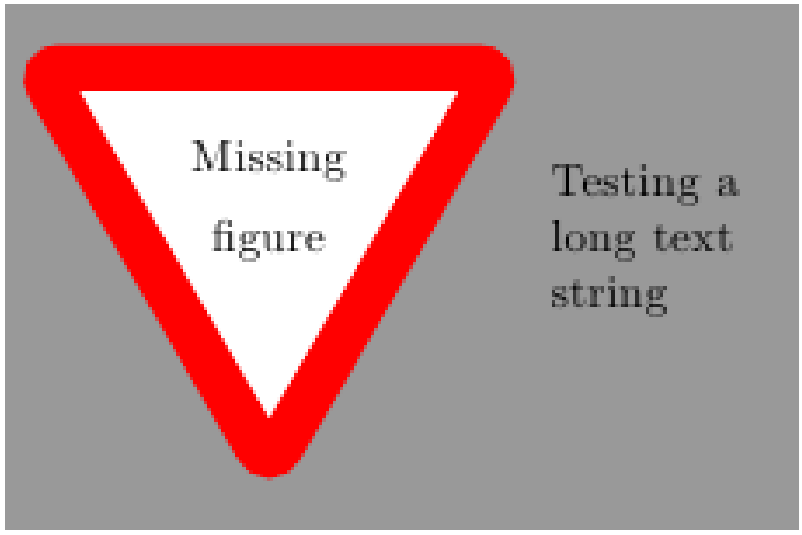

FIGURE 1 Locations of the automatic weather stations used in this study.

dataset, then comparing these signals on an hour-by-hour basis.

2.1 | Data

Four datasets are considered in this study; they are the Australian Bureau of Meteorology's Official wind forecast data, model data from the European Center for Medium Range Weather Forecasting (ECMWF), model data from the Australian Community Climate and Earth System Simulator (ACCESS), and observational data from automatic weather stations. The Official, ECMWF and ACCESS data are at a XX, XX 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. ECMWF data exists at a three hour resolution. To be consistent with the other data sets, ECMWF is therefore linearly interpolated to an hourly resolution: this is also what happens in practice when forecasters load ECMWF wind data into the GFE. Two time periods are considered, the austral summer months (December, January, February) of 2017/18, and the austral winter months (June, July, August) of 2018.

Only station data from the seven Australian capital city airport automatic weather stations are considered; Official, ECMWF and ACCESS data is (linearly?) interpolated to the coordinates of the airport weather stations. Capital city airports have been chosen as the focus of this study for a number of reasons. Automatic weather stations located at airports tend to provide the most accurate wind data, and wind forcasts at airports are important to the aviation industry. Moreover, the capital city airports are all reasonably close to coastlines, resulting in a clear diurnal signal. Finally, these airports are also all close to their respective capital cities, which are high priority regions for accurate forecasting. The datasets are hosted on the Bureau's Jive database, but are not currently generally available, although

the long term plan is for this to change. Can I extract and host the data I need myself? Can I obtain copies of the relevant Jive Functions so that I can post complete code online?

As described above, the Australian Bureau of Meteorology's official wind forecast is constructed out of model data, which is then edited by human forecasters using the Graphical Forecast Editor (GFE) software package. Australian forecasters typically construct wind forecasts out of model data either from the European Center for Medium Range Weather Forecasting (ECMWF), or the Australian Community Climate and Earth System Simulator (ACCESS). Testing whether the official forecast data conforms more closely to the AWS observations than ECMWF or ACCESS therefore provides a way to assess the extra accuracy gained by forecaster edits.

2.2 | Assessing Diurnal Cycles

Although close to coastlines the land-sea breeze is generally the dominant diurnal wind process, the overall diurnal signal may also include mountain-valley breezes, boundary layer mixing processes, atmospheric tides, and urban heat island circulations. Forecasters typically edit model output to account for *both* unresolved sea-breezes *and* unresolved boundary layer mixing; attempting to focus solely on sea-breezes without examining the entire diurnal cycle therefore risks erroneous conclusions, with the effects of one category of edit mistaken for another. In general it is hard to seperate boundary layer mixing edits from sea-breeze edits in the diurnal cycle composites, so this point maybe needs to be reworked. Or could simply comment on this in the discussion.

Sea-breezes are therefore analysed by examining the overall diurnal signal in each dataset, with the assumption that close to coastlines the land-sea breeze is the dominant diurnal process. The diurnal signal is identified 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. Note that thinking of land-sea breezes in terms of perturbations from a background wind may require a conceptual shift from the usual operational definitions. A forecaster would likely define a sea-breeze to be a reversal in wind direction from a primarily offshore flow during the night and morning, to an onshore flow in the afternoon and evening. However, even if the wind is offshore the entire day, sea-breeze *perturbations* are generally still detectable as a weakening of the offshore flow throughout the afternoon and evening.

Once the wind perturbation datasets have been constructed, the accuracy of the Official, ACCESS and ECMWF diurnal cycles are quantified by first calculating the Euclidean distances of the perturbations at each hour from the correspoding AWS perturbations. For instance, to quantify how closely the Official forecast perturbations match the AWS observations, we calculate the Euclidean distances $|\boldsymbol{u}_{\mathrm{AWS}} - \boldsymbol{u}_{\mathrm{O}}|$ at each time step. The accuracy with which the Official and ACCESS datasets resolve the diurnal cycle can then be compared by defining the *Wind Perturbation Index* (WPI)

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

At a given time, the Official forecast wind perturbation is closer to the AWS perturbation than that of ACCESS if and only if $\mathrm{WPI} > 0$. Similarly, the WPI can used to provide a comparison of the Official and ECMWF datasets, or a comparison of the two model guidance datasets ACCESS and ECMWF.

To asses which dataset provides, in general, the most accurate representation of the diurnal cycle, we 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. The sampling distributions of these means can then be modelled as Student's t-distributions, and from this we can calculate the probability that $\overline{\mathrm{WPI}} > 0$ at each hour, where the bar denotes a temporal average.

Temporal autocorrelations of WPI, i.e. correlations between WPI values at a particular hour from one day to the next, are accounted for using the standard method of 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), although in practice temporal autocorrelations of WPI are either non-existant or very small. To assess how well the diurnal perturbations of an overall region are predicted, for instance those of the Victorian coastal station group (see Fig. 1), 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 consistenty 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.

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)

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

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 $\mathrm{CWPI} > 0$ 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 (**?**, e.g.).

Thus, we use non-linear regression to perform the fit

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

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

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

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

where t is time in units of hours UTC, and ψ is the hour that the slowest evolution of the diurnal cycle occurs. The value of α at which the winds align with the semimajor axis, α_M, satisfies

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

The time at which the perturbations align with the major axis t_M then satisfies

$$t_M = \frac{24}{\pi} \left[\arcsin\left(\frac{\alpha_M}{\pi} - 1 \right) + \frac{\pi}{2} \right] + \psi. \tag{9}$$

The lengths of the semimajor and semiminor axes a and b, 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 2 are applied to Australian forecast and station data over the months of June, July and August (austral winter) 2018. First, mean errors are assessed using the Wind Perturbation Index (WPI) at three different spatial scales. Second, overall biases during this time period are assessed using the Climatological

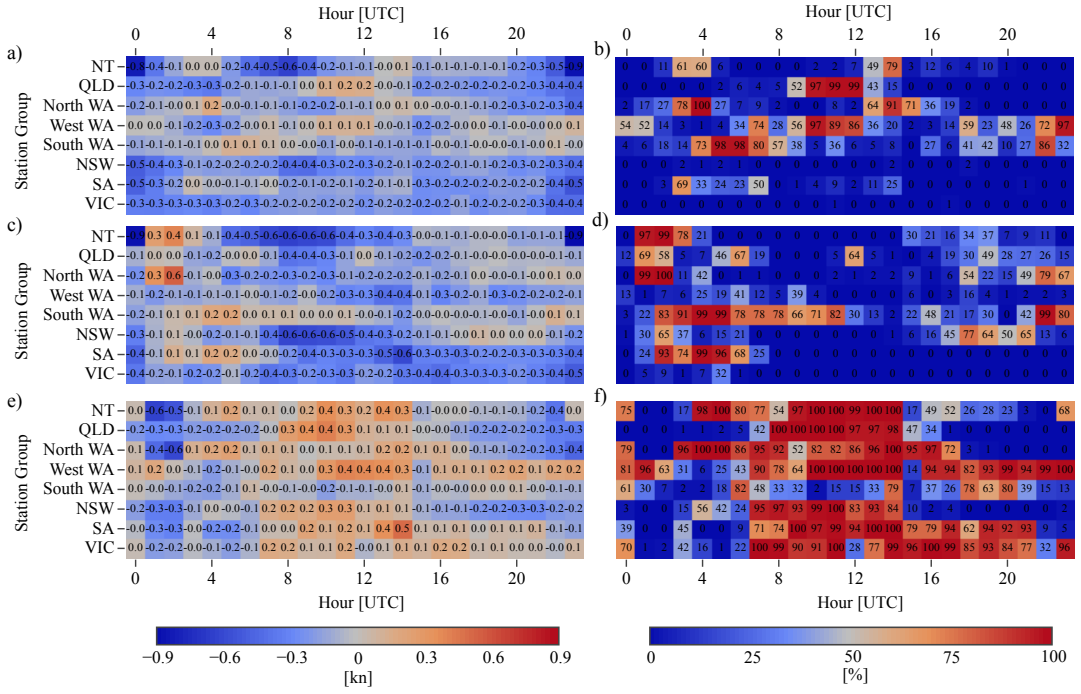

FIGURE 2 Official vs Access and Official vs ECMWF WPI values, standard deviations and confidence. Use just the coastal station group data - other data not a "fair" comparison.

Wind Perturbation Index CWPI. Finally, ellipse based indices are applied and discussed.

FIgure **??** provides example winds and wind perturbations at the Darwin Airport station on 19/06/2018. *Discuss structural differences between observations, Official, ECMWF and ACCESS.*

Figure 9 presents the WPI values, and confidence scores for the Official versus ACCESS and Official versus ECMWF and ECMWF versus ACCESS comparisons over the eight airport stations. *Discuss how although the WPI values are small, standard deviations large, so errors can actually be large on some days. Don't really need to show std values everywhere if you're going to show a time series example.* Note that the four stations/times where official outperforms ACCESS can also be explained by the fact ECMWF outperforms ACCESS at these times - i.e. these results could be obtained if ECMWF was used for the Official forecast.

Understanding results? Figure 10 shows time series plots and hodograph for 1st of July at Darwin airport at 10 UTC (compare with WPI for Off vs ECM, value of -1.7 and v good confidence.) Freak result - likely data entry error. Can see how extreme sea breeze been blended in.

Figure **??** is analogous to figure 9, but presents the results for the coastal station groups. Here, WPI values are first averaged over station groups, before time series statistics are calculated. Note that almost everywhere Official outperforms ACCESS with high confidence, ECMWF also outperforms ACCESS with high confidence.

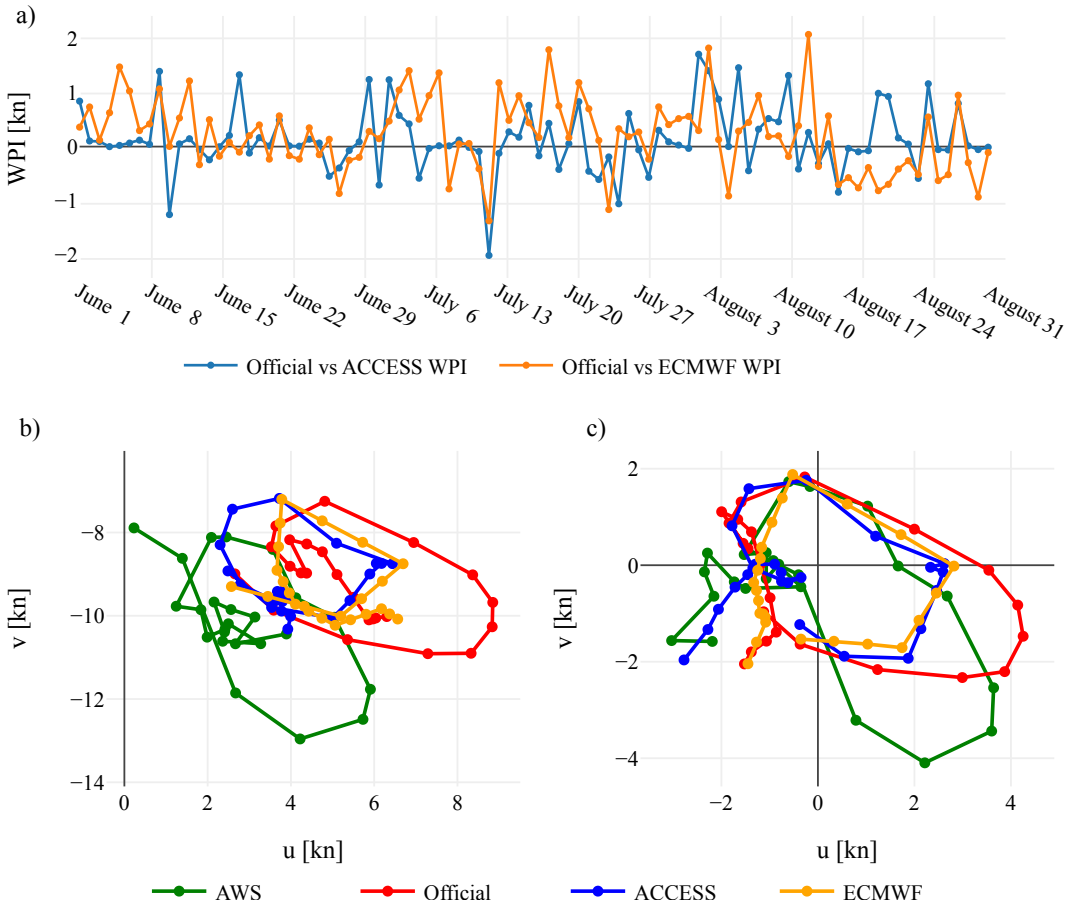

FIGURE 3 Case study 1? South West WA 5 UTC

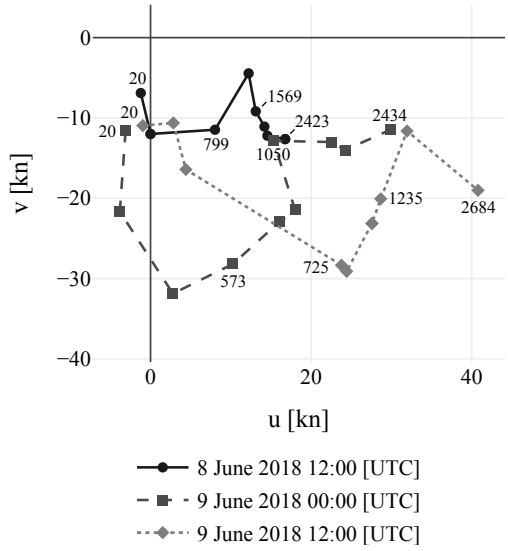

FIGURE 4 Perth Sounding

4 | DISCUSSION

The methods developed in this study can be readily extended to analyse *just* the sea-breezes satisfying the operational definition above. For instance, to study the sea-breezes at a station near a coastline with inward pointing normal vector $\widehat{\boldsymbol{n}}$, the wind perturbation datasets could be restricted to just those days where the corresponding raw wind vector $\boldsymbol{u}$ satisfies $\widehat{\boldsymbol{n}} \cdot \boldsymbol{u} > 0$ for at least one of the hours of that day.

How much time should forecasters spend on sea-breeze edits (if any)? What is the value of an improved diurnal cycle climatology? Improving the accuracy of forecast climatologies will have little value to the typical forecast user. Are there applications where a higher performing climatological forecast yields better outcomes, even if errors increase or even get worse?

Increasing the resolution of a forecast may reduce bias, but increase error.

Error, not bias, that generally matters for the forecast user. Standard methods for “improving" forecasts (adding parametrisations, increasing resolution) reduce bias, but actually increase errors!

Although they have similar definitions, $\overline{\text{WPI}}$ and CWPI measure different things. They do not converge as the length of the time period grows - they don’t even necessarily approach the same sign. As a simple example, suppose that for each day, the observed and Official wind perturbations are given by $\boldsymbol{p}_{\text{AWS}} = (5\cos\omega t, 5\sin\omega t)$ and $\boldsymbol{p}_{\text{O}} = (6\cos\omega t, 6\sin\omega t)$, respectively. Furthermore, suppose that the ACCESS perturbations alternate between $\boldsymbol{p}_{\text{A}} = (7\cos\omega t, 7\sin\omega t)$ and $\boldsymbol{p}_{\text{A}} = (3\cos\omega t, 3\sin\omega t)$ from one day to the next. Then for any contiguous period of n days, $\overline{\text{WPI}} = 2 - 1 = 1$, but CWPI ≈ -1, with the approximation becoming exact for even n. Moreover $\overline{\text{WPI}} = 1$ with a confidence of 1, and using the bootstrapping procesure described above, the confidence that CWPI $= -1$ approaches 1 as $n \to \infty$. This example shows that while the WPI and CWPI are sensitive both to random error and consistent biases between the different datasets, the CWPI becomes increasingly less sensitive to random error as the length of the time period being considered grows. Thus while the WPI arguably provides a more meaningful operational metric, as it measures the accuracy of actual forecast data, it may favour a more biased dataset over a less

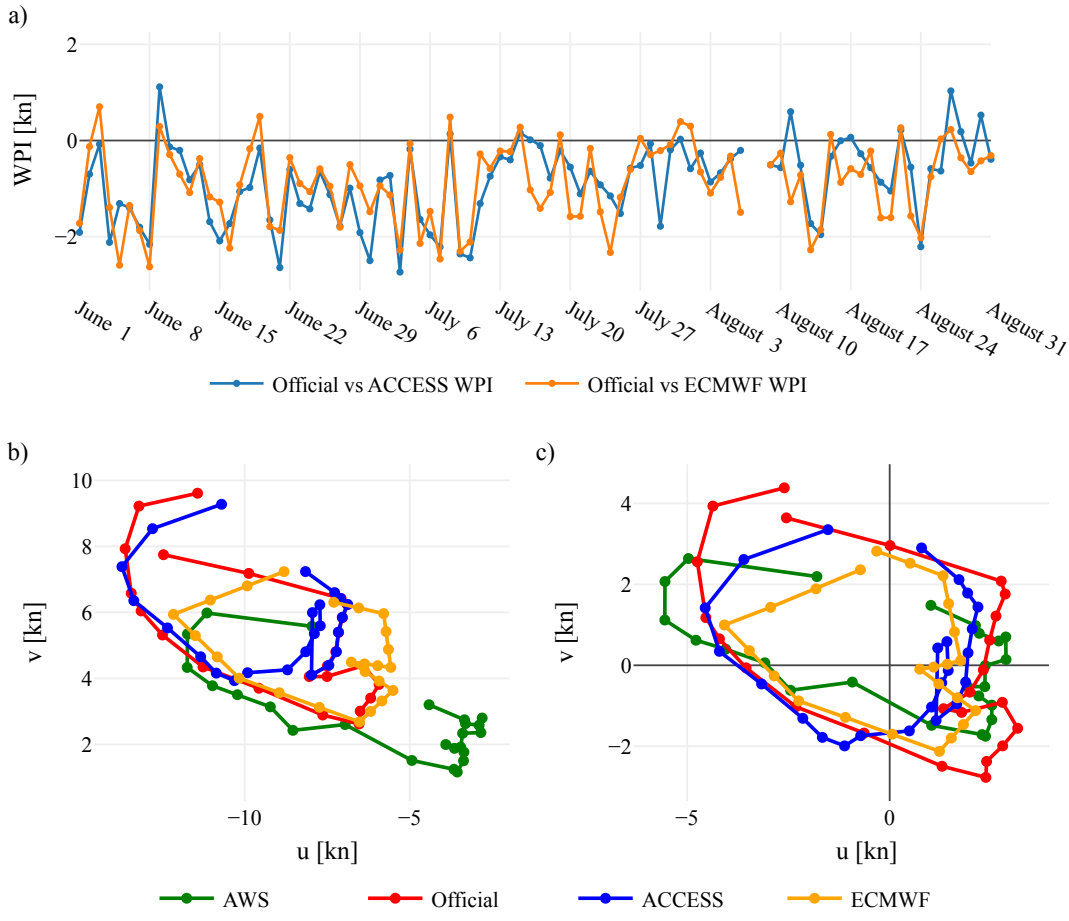

FIGURE 5 Case study 2? Darwin July 3 23 UTC

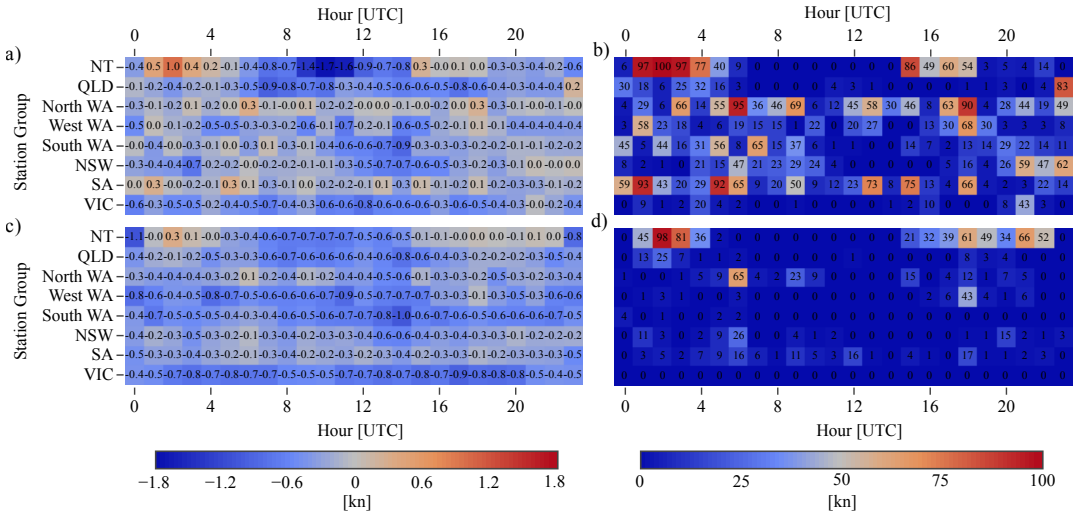

FIGURE 6 Actual perturbation standard deviation values. Note that official performs the worst at this scale!

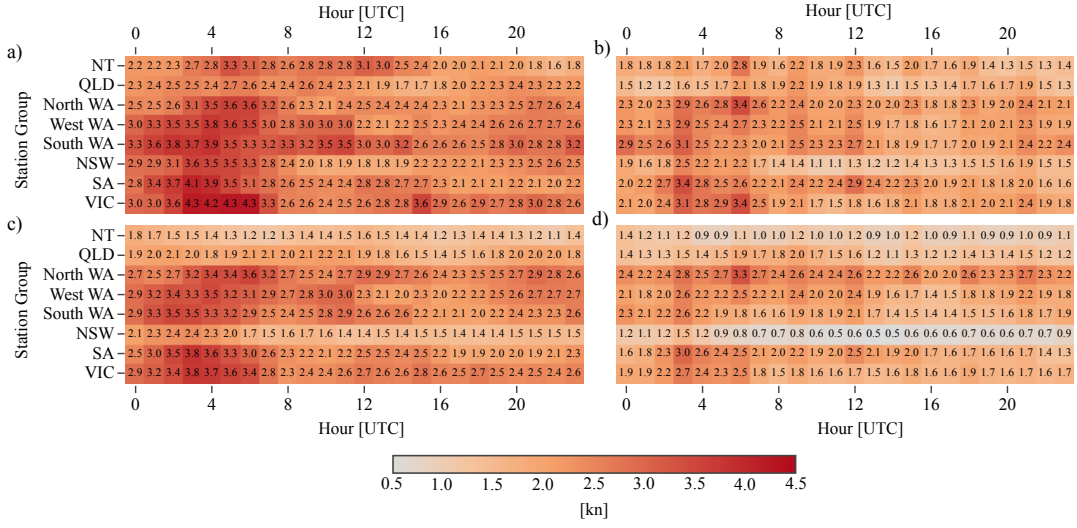

FIGURE 7 Actual perturbation standard deviation values. Note that official performs the worst at this scale!

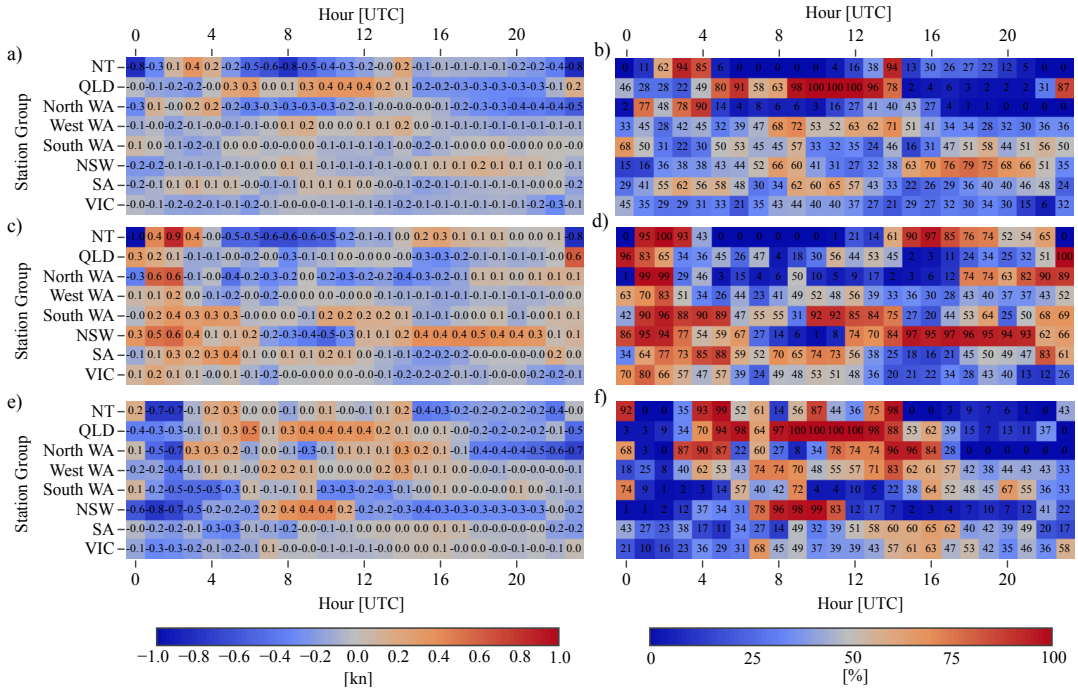

FIGURE 8 Climatological results at different scales

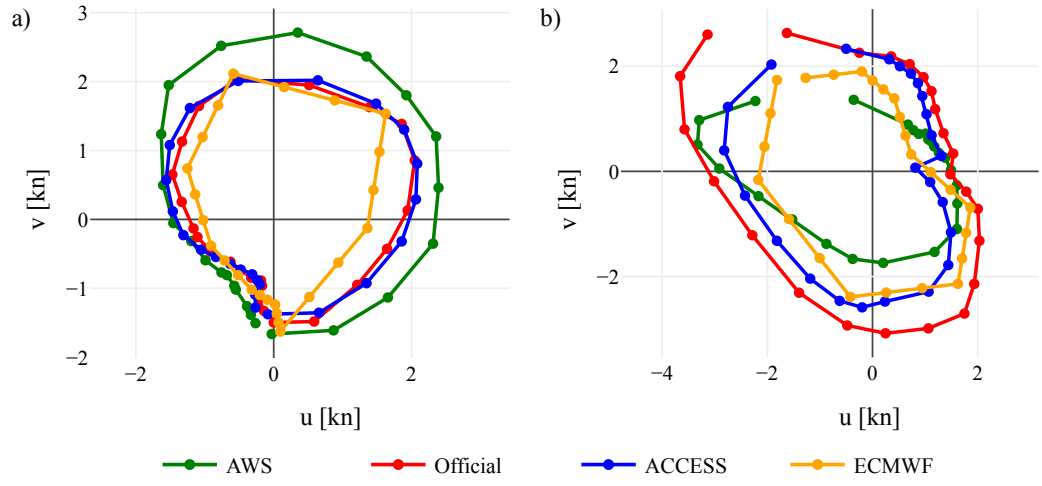

FIGURE 9 Climatological hodographs.

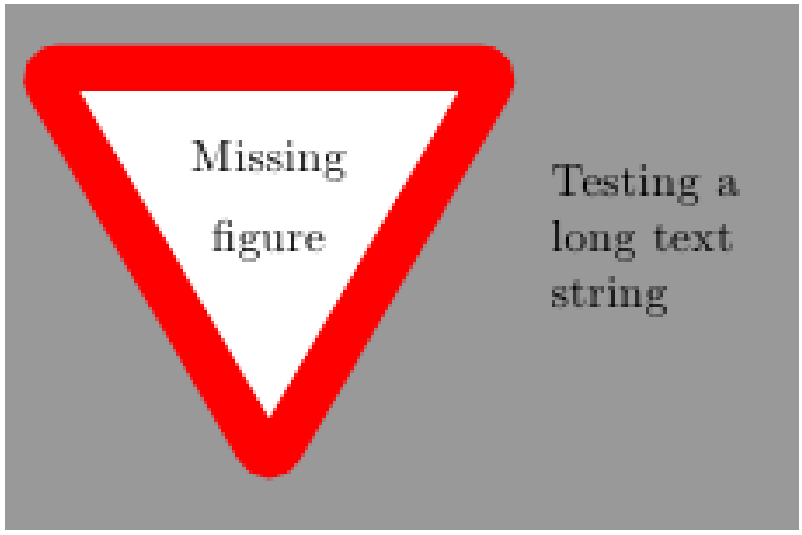

FIGURE 10 Ellipse fits.

biased one, just because the internal variability of that dataset is lower. One consequence of this is that model data at a lower spatiotemporal resolution may outperform in $\overline{\text{WPI}}$ model data of a higher resolution, purely because the internal variability is lower. In this way, the CWPI may actually provide more information about the performance of different forecasts.

Note that the Bureau has not yet moved to ensemble forecasting - and probabilistic forecasting methods therefore not appropriate.

5 | CONCLUSION

In this report, a methodology for comparing the performance of Bureau forecasts of diurnal wind processes to unedited model guidance products has been developed and applied to a case study of the Darwin airport. The key results may be summarised as follows.

1. During the dry season months of June, July and August 2017, the ECMWF sea-breeze is generally more accurate than that of the official forecast. However, during the wet season months of December, January and February 2017/18 this result is reversed, and the official forecast sea-breeze generally outperforms that of ECMWF.
2. In both seasons, boundary layer mixing processes are generally represented better in official forecasts than in ECMWF.
3. In the dry season, the climatological wind perturbations of the official forecast generally outperform those of ECMWF between 13:00 and 16:00 UTC. This is due to ECMWF not capturing the magnitude of the south-easterly

mean perturbations.

4. During the wet season, the climatological wind perturbations of the official forecast generally outperform those of ECMWF at 11:00 UTC. This is due to ECMWF underestimating the magnitude of the mean land-breeze perturbation.

There a number of ways that this work could be extended. The most pressing would probably be to investigate whether the results presented here change when a more operational definition of the sea breeze is used in place of the entirely perturbation based definition used here: this could be done using the method described in section 2.

references

Ebert, E. E. (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, **15**, 51–64. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.25.

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

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

Lynch, K. J., Brayshaw, D. J. and Charlton-Perez, A. (2014) Verification of european subseasonal wind speed forecasts. *Monthly Weather Review*, **142**, 2978–2990. URL: https://doi.org/10.1175/MWR-D-13-00341.1.

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

Smith, J. C., Thresher, R., Zavadil, R., DeMeo, E., Piwko, R., Ernst, B. and Ackermann, T. (2009) A mighty wind. *IEEE Power and Energy Magazine*, **7**, 41–51.

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

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