Bias Correction of Tropical Cyclone Parameters in the ECMWF Ensemble Prediction System in Australia

SAIMA AIJAZ, JEFFREY D. KEPERT, AND HUA YE

Australian Bureau of Meteorology, Melbourne, Australia

ZHENDONG HUANG

The University of Melbourne, Melbourne, Australia

ALISTER HAWKSFORD

Australian Bureau of Meteorology, Melbourne, Australia

(Manuscript received 28 October 2018, in final form 23 August 2019)

ABSTRACT

Global ensemble prediction systems have considerable ability to predict tropical cyclone (TC) formation and subsequent evolution. However, because of their relatively coarse resolution, their predictions of intensity and structure are biased. The biases arise mainly from underestimated intensities and enlarged radii, in particular the radius of maximum winds. This paper describes a method to reduce this limitation by bias correcting TCs in the ECMWF Ensemble Prediction System (ECMWF-EPS) for a region northwest of Australia. A bias-corrected TC system will provide more accurate forecasts of TC-generated wind and waves to the oil and gas industry, which operates a large number of offshore facilities in the region. It will also enable improvements in response decisions for weather sensitive operations that affect downtime and safety risks. The bias-correction technique uses a multivariate linear regression method to bias correct storm intensity and structure. Special strategies are used to maintain ensemble spread after bias correction and to predict the radius of maximum winds using a climatological relationship based on wind intensity and storm latitude. The system was trained on the Australian best track TC data and the ECMWF-EPS TC data from two cyclone seasons. The system inserts corrected vortices into the original surface wind and pressure fields, which are then used to estimate wind exceedance probabilities, and to drive a wave model. The bias-corrected system has shown an overall skill improvement over the uncorrected ECMWF-EPS for all TC intensity and structure parameters with the most significant gains for the maximum wind speed prediction. The system has been operational at the Australian Bureau of Meteorology since November 2016.

1. Introduction

The northwest shelf off Western Australia is a highly active region for tropical cyclone (TC) genesis and experiences more TCs per year than elsewhere in the Australian region. The region is characterized by substantial offshore oil and gas reserves, terrestrial mining resources, and associated communities. Resource industries in the region are highly vulnerable to TC hazards and rely on accurate weather and marine forecasts for critical decision-making that ensure safe and economical operations. This exposure is further exacerbated for remote offshore facilities as responders

can take several days to prepare safely for a TC impact.

The available forecast service capability is driven by numerical weather prediction (NWP) model guidance from the Australian Bureau of Meteorology (BoM) and from reputable meteorological centers around the world. Key models used in the BoM's forecast process include the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (ECMWF-EPS) (Molteni et al. 1994; ECMWF 2012, 2016), the Australian Community Climate and Earth-System Simulator (ACCESS) (Puri et al. 2013), and an Australian implementation of WAVEWATCH III known as AUSWAVE (Durrant and Greenslade 2011). It is well-known that all NWP systems contain

Corresponding author: Saima Aijaz, saima.aijaz@bom.gov.au

DOI: 10.1175/MWR-D-18-0377.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

systematic biases (Atger 2003). A bias-correction system for the northwest (NW) Australian forecasts would reduce such biases and would provide more accurate forecasts of TC related phenomena to the offshore oil and gas industry. Unlike other bias-correction systems that correct systematic errors in model fields, this scheme corrects for systematic errors in specific weather systems, namely the TCs.

In recent years, ensemble prediction systems (EPS) have been shown to improve the predictive skills over deterministic models (Buizza et al. 2008; Buizza and Leutbecher 2015; Richardson 2000) and are considered more valuable than single forecasts. This is mainly because they provide users with more information about future weather scenarios. Moreover, the ensembles give more consistent successive forecasts. For example, consecutive ensemble-mean forecasts issued 24 h apart and valid at the same time are generally found to jump less than corresponding single forecasts from the high-resolution forecast or the ensemble control forecast (Buizza and Richardson 2017). By using the whole ensemble, the unpredictable features are averaged out, and the predictable features (the signals) can be extracted. The latest verification metrics of ECMWF models (Haiden et al. 2018) show that the mean absolute error of the cyclone intensity and mean absolute error of cyclone motion speed are virtually the same for the 2018 version of the ECMWF-EPS and the ECMWF high-resolution deterministic models.

However similar to all NWP, the ensemble forecasts are biased (Toth et al. 2003) and underdispersive (Hamill and Colucci 1997). Therefore, postprocessing of the NWP forecasts to remove bias has become routine (Wilks 2006) to enhance the forecast accuracy. Recent techniques have also focused on bias-correcting ensemble forecasts, to provide reliable probability forecasts (Cui et al. 2012; Gneiting et al. 2005), but global NWP ensembles may not be applicable for all weather phenomena. In particular, TCs are usually too weak and have enlarged eyes in global NWP forecasts primarily because they have to be run at relatively coarse resolution (grid resolution of $\sim 10\text{--}20\text{ km}$), which results in reduced TC intensities. Hence their use is limited in TC-intensity forecasting. To sufficiently resolve the inner core of a TC well enough to capture the processes that affect intensity requires a resolution some 10 times finer. For TCs, bias depends on the weather system, rather than the specific geographic location where the TCs occur, and therefore a novel approach to bias correction is necessary.

This paper describes the development, testing and verification of a bias-correction system (referred hereafter as the EPS-BC) for TC forecasts in NW Australia.

It has been applied to the ECMWF-EPS, which we will refer to as EPS. However, it would be suitable for use on TC forecasts in any NWP EPS system.

The paper is organized as follows. The methodology is given in section 2, and a description of the training datasets, methods of data extraction from the EPS, and the development of the EPS-BC is presented in section 3. The EPS and the best track data are reviewed in section 4. Section 5 outlines the details of the bias-correction methods and application of the statistical models. Verification metrics and results are discussed in section 6. We explain the vortex insertion techniques in section 7. The final bias-corrected product is described in section 8 and the conclusions are summarized in section 9.

2. Methodology

Tropical cyclones can be characterized with a reasonable degree of accuracy by measuring relatively few parameters, such as their location, intensity, size of eye, and size of outer circulation. The wind and pressure field can then be reconstructed from these parameters. The bias-correction system corrects both the cyclone intensities (wind speed and central pressure) and the wind radii [radius of gale-force (34 kt) ($1\text{ kt} \approx 0.5144\text{ m s}^{-1}$) winds, R_{34} , and radius of maximum winds, $R_{\max}$]. Knowledge of TC intensity alone is not sufficient to determine the TC wind field structure as the structure is strongly related to the TC size, in particular to the maximum radial extent of R_{34} winds (Knaff et al. 2014).

The key components in the development of the EPS-BC system are as follows:

- detection of TCs in the EPS, and extraction of relevant TC parameters or predictors listed in Table 1;
- application of EPS-BC for correction of systematic bias in the TC properties (methods for handling absent predictors are discussed in section 5c);
- insertion of a new vortex, constructed to have the bias-corrected properties, replacing the existing vortex in the model fields; and
- calculation of wind exceedance probabilities; and use of the new model fields as forcing for an ensemble wave model (Zieger et al. 2018).

A schematic illustration of the process is shown in Fig. 1.

3. Training datasets

Our training dataset is comprised of the EPS historical forecasts and hindcasts and the observations or estimates of truth represented by the “Australian best track” data.

TABLE 1. Candidate predictors in the EPS.

Candidate predictors	
1	Longitude of the storm center (Lon)
2	Latitude of the storm center (Lat)
3	Mean symmetric radius from the storm center to 34-kt winds (R34)
4	Mean symmetric radius from the storm center to 30-kt winds (R30)
5	Mean symmetric radius from the storm center to 25-kt winds (R25)
6	Mean radius of the maximum wind speed ($R_{\max}$)
7	Storm intensity (i.e., maximum azimuthal wind speed, $V_{\max}$)
8	Minimum central pressure (CP)
9	Integrated kinetic energy at radii 500 km (IKE_{500})
10	Integrated kinetic energy at radii 400 km (IKE_{400})
11	Integrated kinetic energy at radii 300 km (IKE_{300})

a. ECMWF-EPS (EPS)

The EPS outputs 51-member forecast files for forecast lead times up to 15n days at a horizontal resolution (atmospheric model, Cycle 40r1) of 32 km for the dataset used in this study. The model resolution for the current EPS, Cycle 45r1 (ECMWF 2016) is 18 km (0.2°) up to forecast time of fifteen days. The ensemble members are generated by running the model 51 times, each of which is started from a slightly perturbed initial state (Molteni et al. 1994). Random perturbations to the model physics are also applied. The initial perturbations are defined using singular vectors (Buizza and Palmer 1995) of a linear approximation of the model and perturbed data assimilation.

We used the ECMWF tracking method (Van der Grijn et al. 2005; Vitart et al. 1997) to identify the TC location and obtain the intensity and structure from the EPS gridded datasets. We verified our tracker by comparing against the Thorpex Interactive Grand Global Ensemble tracks and further validated against the ECMWF tracks.

We extracted a subdomain of EPS data in the north-west Australian region extending from 0° to 40°S and 90° to 145°E (Fig. 2) with a grid resolution of $0.25^\circ \times 0.25^\circ$ and 1 vertical level (at 10 m above sea level) at 6-hourly intervals for forecast time scales of up to 240 h (10 days). The size of the training set is a balance between a longer training set with a potential for greater confidence in the bias-correction coefficients, and a shorter training set that would produce more up-to-date bias-correction coefficients and encompass fewer model upgrades. We assessed the suitability of using time frames of two TC seasons (2013–14 and 2014–15) and three TC seasons (2012–13, 2013–14, and 2014–15) of EPS data and found that using the two-seasons (2-yr) dataset leads to smaller errors in bias-corrected R34 and $R_{\max}$ than the 3-yr data. We attributed this improvement to the enhancements in

FIG. 1. Schematic of the bias-correction system. Note that the intensity, eye size, and outer size of the cyclones have been modified in the EPS-BC. The bias-corrected symmetrical vortex is inserted within the EPS asymmetrical wind field.

the ECMWF model physics, data assimilation, and the ensemble configuration of the model done after November 2013 (Bauer and Richardson 2014; Buizza and Richardson 2017). The use of an overly long period is disadvantageous since the EPS model is typically updated every one or two years. We therefore used the 2-yr dataset for bias-correction analyses. To ensure that meaningful statistics are achieved, we ascertained that at least 20 storms exist in the selected dataset and that each storm is of at least 3–4 days duration. The statistical model coefficients defining the EPS-BC need to be recomputed at least every two years to incorporate recent storms and account for updates to the EPS.

We further evaluated the use of additional 20-yr ensemble hindcast data (ECMWF 2017) comprised of 5 ensemble members (currently upgraded to 11 members) output twice per week for bias correction. Our analyses indicated that the errors in the hindcast variables were significantly larger when compared with that of the EPS. Moreover, the hindcast was also not able to capture high intensity storms because of the lower resolution of the initial conditions and different ensemble characteristics. We therefore excluded the hindcast dataset from being part of the training data as it would potentially deteriorate the skill of the EPS-BC.

We used the R34 winds in the northwest, northeast, southeast, and southwest quadrants to determine the effects of asymmetry in the TC structure in our bias-correction analyses. We found that the errors in the corrected asymmetric radii are significantly higher in comparison with the errors in the symmetric radii and are not always available. Therefore, we corrected only the symmetric radii.

The key predictants from EPS used in the statistical analyses are listed in Table 1. Model development and evaluation datasets are discussed in section 6.

b. TC best track

We selected the Australian Bureau of Meteorology best track (BT) (Bureau of Meteorology 2011, 2014)

FIG. 2. Domain extent for the EPS-BC model.

data as the “observed truth” for verification and statistical analysis of the bias-correction models. This dataset is prepared using all available data and observations including the advanced scatterometer ASCAT and wind speed estimates from the Special Sensor Microwave Imager satellites. Extensive reanalysis is undertaken to estimate key TC parameters.

The BT database consists of more than fifty TC parameters from recorded tropical cyclone tracks over the region south of the equator between 90° and 160°E. The time resolution for the parameters is at least 6-hourly. Additional times are added for significant TCs particularly when close to the coast and within radar range.

Unless there are surface observations, the central pressure in the BT is derived from the maximum wind estimate using the wind–pressure relationship from Courtney and Knaff (2009), although historically a variety of other wind–pressure relationships have been used. For TCs over land this derivation is known to produce higher pressures than actual, therefore the estimates are then based on land-based observations when available.

In addition to uncertainties arising from the observation errors, there is uncertainty in the estimates of intensity and track location from the reanalysis in the BT. The uncertainty in track position has been found to be greater for weak storms than for intense storms (Torn and Snyder 2012; Landsea and Franklin 2013). As a check of the reliability of the BT dataset, we compared the R34 statistics of our BT data to those presented in the recent climatological analysis of Chan and Chan (2015) and found satisfactory agreement.

In general, $R_{\max}$ is poorly observed in nature. The documentation of the BT database (Bureau of Meteorology 2011) lists a variety of methods used to estimate $R_{\max}$. These include aircraft and surface observations, radar,

TABLE 2. Standard deviation and range of key BT parameters used in the statistical analyses.

Parameter	Standard deviation	Range minimum, maximum
CP (hPa)	20.6	930, 1009
$V_{\max}$ (m s^{-1})	9.7	5.0, 56.6
R34 (km)	59.3	64, 310
$R_{\max}$ (km)	31.3	15, 185

satellite, estimates using the outer closed isobar, and other methods. In reality, surface or aircraft observations are rarely available in the Australian region. The $R_{\max}$ in the BT database is determined mostly from satellite and radar data. Detailed estimates of uncertainties in the BT parameters based on TC observations were not available. Therefore to provide an indication of the variability in the BT data, we have computed the standard deviation, minimum and maximum of the BT parameters (Table 2).

4. Data review

We reviewed the extracted TC data from the EPS and the TC BT datasets to identify any irregularities or anomalies that may have existed in the datasets or could have resulted from the data extraction procedures, for example detection of low pressure systems by the TC tracker, that could be monsoonal or extratropical.

In TC forecasting analysis, the reduced number of ensemble members predicting TC genesis at longer lead times leads to reduced data at these lead times. Therefore, to increase the data sample and to provide adequate time continuity in the statistical model, we grouped the EPS TC data into periods of 48 h centered on the lead time except for lead times of 24 and 240 h. For example, the data for 48-h lead time are comprised of data from forecast times of 24, 30, 36, 42, 48, 54, 60, 66, and 72 h. For the 24-h lead time, we excluded the 0-h lead time. The data for 240-h lead time was grouped into a 24-h period centered on 228 h. This grouping of various lead times resulted in a several-fold increase in the sample data when compared with the method using single lead times. It is recognized that the expected autocorrelation among the nine samples for each storm likely leads to the effective sample size being closer to the sample size without enlargement. However, increasing the samples by this procedure is still a net positive to forecast skill relative to not enlarging the sample.

The scatterplots of EPS versus BT variables for individual ensemble members are presented for a lead time of 24 h (grouped as described above) in Fig. 3. Similar correlations between the EPS and BT variables for forecast lead times ranging from 24 to 240 h were also calculated.

FIG. 3. Scatterplots for individual forecast members at forecast lead time of 24 h (before bias correction). Linear regression lines are shown in red while $x = y$ line is represented in aqua. All EPS parameters are plotted against the corresponding BT except R25, R30, and IKE₅₀₀ are plotted against BT R34.

The statistical relationships show that at 24-h lead time, the EPS R34 correlates with the BT R34 with an r^2 of 0.61. Using R30 as a surrogate for R34 yields slightly better performance with an r^2 of 0.66. The BT-CP versus EPS-CP and the BT- $V_{\max}$ versus EPS- $V_{\max}$ have an r^2 of 0.65. The correlation of the longitude and latitude are nearly perfect with $r^2 > 0.95$ for both at lead time of 24 h. This indicates that the EPS model performs well in predicting the position of the cyclone center.

The $R_{\max}$ variable has the lowest correlation with BT, an r^2 of 0.19, RMSE of 82 km, and an average bias of 63 km at 24-h lead time. The relatively poor performance of $R_{\max}$ was expected, as most global models have little ability to predict it and very high resolutions (e.g., a few kilometers) are needed to avoid systematic biases.

In addition to the lack of model skill in predicting $R_{\max}$, it is possible that substantial errors may exist in the BT estimates. Please see [section 3b](#) for a brief description of methods of estimation of $R_{\max}$ in the BT database. [Figure 4](#) shows the reduction in correlation coefficient with increasing forecast lead times for R34, IKE₄₀₀, $V_{\max}$, and CP. There is a gradual decrease in r^2 from about 0.6 at 24 h to less than 0.2 at 240 h (10 days) for all variables except for R34 that shows a very rapid decline in r^2 in the first 72 h and then saturates to near zero after 96 h.

5. Bias-correction statistical models

We investigated a number of statistical models including the simple linear regression (SLR) model,

FIG. 4. Correlation coefficient (r^2) as a function of lead time. Blue line indicates BT R34 vs EPS R34 (km), green line represents BT R34 vs IKE₅₀₀ (km), aqua line represents the BT $V_{\max}$ vs EPS $V_{\max}$ (m s^{-1}), and black line represents BT CP vs EPS CP (hPa).

multivariate linear regression (MVR), principal component analysis (PCA), and quantile mapping statistical models for bias correction of the EPS data.

The two linear regression models and the PCA performed significantly better than the quantile mapping in achieving smaller RMSE of the predictions. Therefore, we will discuss only the first three methods in this paper.

a. Ensemble mean and ensemble spread

The ensemble mean is often considered as a single forecast when analyzing ensemble forecasts (Whitaker and Loughé 1998). It has been shown that the mean of an ensemble of forecasts is generally known to have a smaller error than the mean error of any of the individual members of the ensemble (Murphy 1988). Like previous studies, our analysis also revealed that there was considerable improvement in the RMSE in the regression of EPS against the BT both before and after the application of bias correction when using the ensemble mean of each candidate predictor. The largest reduction among all predictors was determined for R34 where the RMSE decreased from 44 to 38 km for a lead time of 24 h (using the ensemble mean in the EPS-BC).

The ensemble spread should be representative of the uncertainty in the mean to produce a reliable ensemble (Johnson and Bowler 2009). We found that the ensemble spread (represented by the standard deviation) collapses at longer lead times if bias correction is applied to individual ensemble members instead of the ensemble mean. This is a strong indication that the spread does not provide a useful indication of the forecast skill in this situation. Normally we would expect the forecast spread to increase with lead time, because the forecast accuracy decreases. The loss of spread in our analysis was a direct consequence of correcting individual members, and we

therefore developed a strategy (section 5c) to preserve the ensemble spread for the EPS-BC.

b. Simple linear regression

A simple linear model (SLR) is a linear regression model that relates two variables. The ensemble mean of the EPS TC variable, is regarded as the predictor, explanatory, or independent variable while the BT TC variable is regarded as the response, outcome, or dependent variable as shown in Eq. (1) below:

$$\text{BT} = a_0 + a_1 \overline{\text{EPS}}, \quad (1)$$

where a_0 (intercept) and a_1 (slope) are the regression coefficients. We regress on the ensemble mean rather than individual ensemble members for reasons explained below.

The regression coefficients for all TC variables at various lead times are computed using (1), then the bias correction is applied to the ensemble mean as

$$\overline{\text{EPS}}_{\text{bc}} = a_0 + a_1 \overline{\text{EPS}}, \quad (2)$$

where $\overline{\text{EPS}}_{\text{bc}}$ is the bias-corrected ensemble mean, and $\overline{\text{EPS}}$ is the ensemble mean of the uncorrected EPS data.

c. Preserving ensemble spread

After computing the bias correction using the ensemble mean, we estimated each individual EPS-BC ensemble member using the following strategy that preserves the ensemble spread that is its displacement from the mean.

We illustrate the technique using the SLR, but it is also applicable to the other regression methods. For the SLR, storm parameters are predicted using Eq. (1). As the forecast lead time increases, the model skill decreases and so the correlation between model and the verifying data decreases. Thus, the slope coefficient a_1 in Eq. (1) approaches 0 and the intercept a_0 in Eq. (1) approaches the climatological mean of the particular variable. Thus, at long lead times as the EPS skill tends toward zero, all the ensemble members are adjusted to be fairly close to the climatological mean leading to insufficient spread.

If we measure the ensemble spread by its variance, then the spread of the adjusted ensemble is

$$\text{Var}(\text{EPS}_{\text{bc}}) = \text{Var}(a_0 + a_1 \text{EPS}_i) = a_1^2 \text{Var}(\text{EPS}). \quad (3)$$

The bias-correction procedure rescales the variance of the uncorrected ensemble by a factor a_1^2 , which tends to zero at long lead times. While this behavior of the bias-corrected forecast tending toward climatology at long lead times is correct for a single deterministic forecast, it

TABLE 3. Final predictors for each of dependent variables of CP, R34, $V_{\max}$, and $R_{\max}$.

Predicted variables	CP	R34	$V_{\max}$	$R_{\max}$ ^a
Final predictors	$V_{\max}$ CP	$V_{\max}$ CP	$V_{\max}$ CP	$V_{\max}$ (BT) Lat (BT)
	IKE ₅₀₀	IKE ₅₀₀	IKE ₅₀₀	—
	IKE ₄₀₀	IKE ₄₀₀	IKE ₄₀₀	—
	IKE ₃₀₀	IKE ₃₀₀	IKE ₃₀₀	—

^a $R_{\max}$ was not subject to BIC.

gives a false impression of the likely range of possibilities when applied to an ensemble.

We resolve this issue by applying the bias correction to the ensemble mean using Eq. (2) and then computing each individual bias-corrected member to preserve its displacement from the mean using

$$\text{EPS}_{i\text{bc}} - \overline{\text{EPS}}_{\text{bc}} = \text{EPS}_i - \overline{\text{EPS}}, \quad (4)$$

where $\text{EPS}_{i\text{bc}}$ is the bias-corrected i th individual ensemble member, and EPS_i is the i th individual uncorrected EPS ensemble member.

From Eq. (4), the variance of the bias-corrected ensemble is

$$\begin{aligned} \text{Var}(\text{EPS}_{\text{bc}}) &= E[(\text{EPS}_{\text{bc}} - \overline{\text{EPS}}_{\text{bc}})^2] \\ &= E[(\text{EPS}_i - \overline{\text{EPS}})^2] = \text{Var}(\text{EPS}). \end{aligned} \quad (5)$$

We can see that the bias correction now preserves the ensemble spread.

While offering significant advantages, there are some practical issues with this method. The first is that it is necessary to calculate the mean parameters for each TC present in the ensemble. This requires some method of determining whether two vortices in different ensemble members are likely to represent the same storm. This is straightforward for storms that exist at the initial time since they will be close together, but more difficult for storms that form later during the forecast period, since not all ensemble members will capture the genesis at the same time. For such cases, we group the tracks that are deemed to represent the same storm into “clusters.” We define the “track” as the path of a storm in an ensemble member, and a “cluster” as a group of tracks from

FIG. 5. Comparison of RMSE (cross validated) before bias correction from EPS (aqua) and after bias correction from the EPS-BC using: SLR (orange), MVR (pink), and PCA-MV (pink dashed line). Note that RMSE for EPS $R_{\max}$ is on a different (right-hand scale) than RMSE for EPS-BC $R_{\max}$. The error bars indicate standard deviation for each EPS parameter.

TABLE 4. Comparison of RMSE (cross validated) before bias correction (EPS) and after bias correction from the EPS-BC using SLR, MVR, and the PCA-MV; n is the number of sample points. Note that data are presented only for selected lead times of 0, 24, 120, and 240 h.

	Lead time (h)	EPS (uncorrected)	EPS-BC			n from MVR
			SLR	MVR	PCA-MV	
CP (hPa)	0	13.2	10.1	12.1	9.3	128
	24 (1 day)	14.5	11.7	11.0	10.3	1122
	120 (5 days)	17.7	14.2	12.8	11.9	1285
	240 (10 days)	18.1	14.8	12.9	12.2	706
R34 (km)	0	43.5	33.0	30.3	37.4	67
	24 (1 day)	36.5	30.5	28.1	31.9	561
	120 (5 days)	46.0	45.1	36.4	36.4	635
	240 (10 days)	62.1	50.0	43.0	44.2	347
$V_{\max}$ (m s^{-1})	0	12.6	7.0	7.1	7.1	128
	24 (1 day)	13.3	7.9	7.8	7.3	1122
	120 (5 days)	15.9	9.5	9.1	8.6	1285
	240 (10 days)	16.6	10.0	9.0	8.7	706
$R_{\max}$ (km)	0	74.1	12.9	13.0	12.4	1222
	24 (1 day)	101.9	14.4	13.7	14.1	1222
	120 (5 days)	201.1	14.6	14.4	13.7	1222
	240 (10 days)	255.2	14.4	13.8	14.4	1222

different members that are diagnosed to represent the same storm.

The method of grouping tracks into clusters consists of looping through all ensemble members for each forecast time. If the start of a track for a certain member is found, then we check whether it is “close” to a cluster. The “closeness” is determined when the distance from the closest center in the cluster is less than 7° or the mean distance from the members of the cluster is less than 10° . If it is close to a cluster then the corresponding track is assigned to that cluster. If the point is too far away from the all the existing clusters, then a new cluster is created.

The second issue relates to missing data, particularly the wind radii. The EPS model is sometimes unable to predict R34 because of the inherent biases in the system that result in underestimates of the maximum wind intensities. This problem is especially severe at longer lead times. If a TC is above gale force in only half of the ensemble members, then only those members will possess an R34 and so only R34 from those members can be corrected using the above method. Bias correction of the intensity of the ensemble will very likely mean that substantially more members are now at least of gale-force intensity. However, our only means of providing an R34 for the weaker members would be a separate bias-correction step using different predictors, removing our ability to systematically deal with the ensemble spread. This issue arises with R34 (and with wind radii for winds greater than 34 kt), since the tracking algorithm always returns values for all other TC parameters. For very weak storms, R30 and R25 may also be missing in the EPS. The BT does not include R30 and R25,

therefore we excluded these wind radii as predictors in our analyses.

To overcome the issue of missing R34 values, we can use IKE as a predictor for R34 as IKE is calculated for all storms using

$$R34_{\text{ibc}} = a_0 + a_1 \text{IKE}_i, \quad (6)$$

where IKE can be either IKE_{500} or IKE_{400} or IKE_{300} . This method will still experience the problem of loss of ensemble spread at long lead times if individual members are corrected:

$$\text{Var}(R34_{\text{bc}}) = a_1^2 \text{Var}(\text{IKE}), \quad (7)$$

because at longer lead times, a_1 will tend to zero as the correlation between the ensemble system and the BT diminishes.

We can extend the technique in Eqs. (2) and (3) to predict the mean:

$$\overline{R34}_{\text{bc}} = a_0 + a_1 - \overline{\text{IKE}} \quad (8)$$

and

$$R34_{\text{ibc}} = \overline{R34}_{\text{bc}} + a(\text{IKE}_i - \overline{\text{IKE}}). \quad (9)$$

That is we use the spread in the IKE to provide the spread in R34. This method has the following property:

$$\text{Var}(R34_{\text{bc}}) = a^2 \text{Var}(\text{IKE}). \quad (10)$$

FIG. 6. Spread–skill relationships of BT and EPS. The red line indicates a linear fit. Each dot represents a different lead time. EPS R30, EPS R25, and EPS IKE₅₀₀ are plotted against BT R34 as BT does not contain R30, R25, or IKE₅₀₀. There are no data shown for IKE₅₀₀ because meaningful skill (RMSE) cannot be computed between an energy parameter (IKE₅₀₀) and a length parameter (BT R34).

The coefficient a is determined from a climatological regression of the ensemble R34 perturbations onto the ensemble IKE perturbations.

d. Multivariate linear regression

The multivariate or multiple linear regression (MVR) analysis relates one dependent variable to other independent variables or predictors (Glahn and Lowry 1972). The equation can be written as

$$\overline{\text{EPS}}_{\text{bc}} = a_0 + a_1 \overline{\text{EPS}}_1 + a_2 \overline{\text{EPS}}_2 + \cdots + a_n \overline{\text{EPS}}_n, \quad (11)$$

where $\overline{\text{EPS}}_{\text{bc}}$ is the dependent or the estimated EPS-BC mean variable, $\overline{\text{EPS}}_1$, $\overline{\text{EPS}}_2$, etc. represent the ensemble mean of the candidate variables or predictors, a_0 is the regression constant, and a_1 , a_2 , etc. are the regression coefficients.

We applied the MVR model to perform bias correction for CP, R34, V_{max} , and R_{max} . We have shown earlier using simple linear regressions that the IKE variables

are skillful predictors of R34. Also, R34 is not always predicted by EPS. Considering these issues, we excluded R34 (and all other wind radii) as potential predictors in the MVR and the Principal Component Analysis (section 5e). We further excluded R_{max} as a predictor as we have shown that the EPS model possesses poor skill in predicting R_{max} (section 4a). We use EPS ensemble means of CP, V_{max} , IKE₅₀₀, IKE₄₀₀, IKE₃₀₀ as candidate predictors.

The MVR model has the advantage of selecting a subset of predictors from a large number of candidate predictors. It integrates the effect of different candidate predictors on a single dependent parameter. As it is generally not known a priori which or how many predictors to include for the estimation of a dependent variable (Glahn and Lowry 1972), we applied a technique called the Bayesian information criterion (BIC) to screen and select the candidate variables.

The BIC balances the residual sum of squared error and the model complexity. If a model is applied on a particular training dataset, the BIC score gives an estimate

FIG. 7. Spread-skill relationships of BT and EPS-BC using SLR. The red line indicates a linear fit. Each dot represents a different lead time. EPS R30, EPS R25, and EPS IKE₅₀₀ are each regressed against BT R34 as BT does not contain these parameters.

of the model performance on a new training dataset. BIC attempts to manage the risk of overfitting by introducing a penalty term, which grows with the number of parameters. This allows filtering out unnecessarily complicated models. The model with the lowest BIC is selected over a whole set of other candidate predictors.

Using the MVR, we generated a matrix of candidate predictors for each forecast lead time for each of CP, R34, and $V_{\max}$. Then we extracted those candidate predictors for each variable that occurred in at least 6 of the 11 forecast lead times. The final predictors are listed in Table 3.

The CP and $V_{\max}$ were estimated by applying the bias correction using Eq. (11) where EPS_{bc} is the bias-corrected ensemble mean of the CP or $V_{\max}$, and EPS_1 , EPS_2 , etc. are the final predictors from Table 3. Each individual EPS member was estimated using Eq. (3) or using Eqs. (8) and (9) in case of R34.

As we have excluded the EPS $R_{\max}$ as a potential predictor for calculating R34, $V_{\max}$, CP, and $R_{\max}$ itself, we now estimate $R_{\max}$ by using the long-term historical BT $V_{\max}$ and BT latitude data spanning over a 17-yr period (2000–16). We first transform $R_{\max}$ into its logarithmic form as the logarithm tends to make the skewed distribution of $R_{\max}$ more Gaussian and can provide a better fit. Using the logarithm of one of the variables makes the effective relationship nonlinear, while still preserving the linear model (Benoit 2011). The logarithm further ensures that only positive values of $R_{\max}$ are computed.

Existing models of $R_{\max}$ are generally based on $V_{\max}$ or CP (Quiring et al. 2011), or R34 and larger wind radii (Takagi and Wu 2016; Kossin et al. 2007). $R_{\max}$ has an inverse relationship with $V_{\max}$ and has also been shown to be positively correlated with the absolute value of the latitude (Kossin et al. 2007; Willoughby and Rahn 2004). Previous studies have found that the correlation

FIG. 8. Spread–skill relationships of BT and EPS-BC using MVR. The red line indicates a linear fit. Each dot represents a different lead time.

between $R_{\max}$ and $V_{\max}$ varies significantly in different ocean basins. While some studies have found a moderate inverse correlation of $R_{\max}$ with $V_{\max}$ in the Atlantic and Caribbean (Quiring et al. 2011), Takagi and Wu (2016) found only a weak correlation (r^2 of 0.112) in the western North Pacific. We found that BT $R_{\max}$ correlated with BT $V_{\max}$ with an r^2 of 0.3 in the NW Australian region of the Indian Ocean. Although the correlation of BT $R_{\max}$ with BT latitude is weak (r^2 of 0.02) for our study region, we include the latitude as a predictor of $R_{\max}$ in addition to $V_{\max}$ because of variations of the local Coriolis force (Kossin et al. 2007), and the dependency of wind field structure on latitude (Kossin et al. 2007) for wind radii greater than 50 km. There is also a known tendency of TCs to get larger toward the poles, which is a further indication of the dependence of wind radii on latitude.

We apply the MVR to compute the regression coefficients from

$$\text{Log } R_{\max} = a_0 + a_1 V_{\max_{\text{BT}}} + a_2 \text{Lat}_{\text{BT}}, \quad (12)$$

where $V_{\max_{\text{BT}}}$ and Lat_{BT} are taken from the long-term BT data.

The bias correction is then applied as follows:

$$\text{Log } R_{\max_{\text{ibc}}} = a_0 + a_1 V_{\max_{\text{ibc}}} + a_2 \text{Lat}_i, \quad (13)$$

where $R_{\max_{\text{ibc}}}$ is the bias-corrected $R_{\max}$, $V_{\max_{\text{ibc}}}$ is the bias-corrected EPS-BC $V_{\max}$ from Eqs. (13) and (3), and Lat_i is the EPS latitude.

We excluded all values of BT $R_{\max}$ greater than 150 km from the regression because we found that including outliers that exceed 150 km results in a poor linear fit and subsequently large errors in the estimation of $R_{\max}$. Also, $R_{\max}$ of 150 km represents a value greater than 3σ from the sample mean, which is a commonly used convention for identifying outliers (Kazmier 2003; Grafarend 2006; Seo 2006).

e. Principal component analysis-multivariate regression

Principal component analysis (PCA) is a mathematical procedure that converts several possibly correlated variables into a (smaller) number of uncorrelated variables called principal components. A main advantage of PCA is that it interprets the response variable using independent new predictors, thus eliminating the correlation structures (which occur among the predictors in the EPS

FIG. 9. Rank histograms before bias correction from EPS. Note that there are no data for IKE_{500} as IKE_{500} is not present in the BT.

data) of the observed variables, and hence leads to a more robust estimate of coefficients in the regression that is followed (Wilks 2011).

The bias-correction methodology consists of applying the PCA to the EPS data to obtain a new set of corresponding components, then undertaking a multivariate regression on the system bias using the components. We refer to the procedure as PCA-MV. It can be described by an $m \times n$ observation matrix and an $n \times 1$ response vector, where m is the number of predictors and n is the sample size. The predictors used in the PCA-MV are the EPS ensemble means of CP, $V_{\max}$, $R_{\max}$ (logarithm), IKE_{500} , IKE_{400} , IKE_{300} , latitude, and longitude. We compute scores from the predictors and then use these scores to train the MVR. The final models are selected based on the significance levels of the regression coefficients.

6. Verification

We assess the model skill of predicting TC parameters after bias correction by using the standard

root-mean-square-error (RMSE), spread-skill relationships, rank histograms, frequency histograms, box-and-whisker plots, cumulative probability distributions, and Q-Q plots of the TC parameters. We evaluate the model performance of TC parameters but do not verify wind probabilities since there are insufficient data for wind observations in the study area.

a. RMSE

A comparison of RMSE resulting from all three bias-correction modeling techniques is presented in Fig. 5 and Table 4. All three statistical models have been found to successfully reduce the RMSE for each of CP, R34, $V_{\max}$, and $R_{\max}$ relative to the corresponding uncorrected EPS TC parameters.

All statistics presented here have been cross validated. That is, regression coefficients are calculated for “ n ” number of training subsets, where n is the number of datapoints using the leave-one-out cross-validation method. This method consists of leaving one data point out and treating the remaining datapoints as training

FIG. 10. As in Fig. 9, but rank histograms represent EPS-BC using the SLR. IKE_{500} has been used as a predictor for R34.

data. The procedure is repeated for all datapoints, and the scores are averaged. This ensures that the performance estimate is stable and less sensitive to the partitioning of the data into various samples. Cross validation is used primarily to assess the predictive capability of a model on a dataset that has not been previously used for “training,” thereby preventing “overfitting” of data.

The improvement in RMSE for R34 when EPS R34 is used as predictor ranges from 3 km at lead time of 24 h to 16 km at lead time of 240 h for the SLR. A more substantial gain was obtained when IKE_{500} was used as a predictor for R34. IKE has the advantage of being available in storms that are too weak to possess significant wind radii of winds while R34, for example, can only be calculated when the storm intensity is at least 34 kt. Regressing on IKE_{500} (or IKE_{400} or IKE_{300}) provides a technique that is more accurate, and more widely applicable.

Although the application of the linear model leads to a substantial reduction in RMSE in the range 60–240 km for R_{max} relative to the EPS R_{max} , we found

that the EPS R_{max} is not well correlated with the BT R_{max} , so this regression is achieving little more than replacing the EPS R_{max} with the climatological mean. In any event, modest deficiencies in the R_{max} will lead to only local errors in wind probabilities but it does influence the wave response.

The MVR and the PCA-MV perform better than the SLR. The improvements in RMSE of the MVR and PCA-MV over the SLR are greater for CP and R34 than for V_{max} or R_{max} .

Following detailed comparison of the three statistical models, we selected MVR for bias correction of the EPS for operational forecasting. While the performance of the MVR and the PCA-MV was found to be very similar, the MVR is less complex and less computationally intensive than the PCA-MV, and therefore more efficient to implement than the PCA-MV.

b. Spread–skill relationships

Following Whitaker and Loughe (1998), we define skill as the root-mean-square (RMS) distance between

FIG. 11. As in Fig. 9, but rank histograms represent EPS-BC using the MVR.

the BT and the ensemble mean. The standard deviation or the average RMS distance between the individual ensemble members and the ensemble mean represents the spread.

The scatterplots of spread and skill of the EPS (Fig. 6) and the bias-corrected EPS-BC using the simple linear (Fig. 7) and MVR methods (Fig. 8) show that the skill for all parameters in EPS-BC from both the models has significantly improved relative to the EPS (Fig. 6).

However, the spread is small and remains unchanged between EPS and EPS-BC as we have preserved the spread for EPS-BC (section 5c). For example, the spread in the $V_{\max}$ (speed) varies from 1.5 to 3 m s^{-1} for both EPS and EPS-BC. The EPS data are underspread. There are two likely causes for this significant level of underspread. First, the grid resolution of the global model is too coarse to generate high intensity storms, so random errors in cyclone intensity are much more likely to be on the low side, rather than equally distributed on both sides of the true value. Second, the model perturbations are designed to represent spread at all scales within the EPS, and are verified against global data, not cyclone-specific data.

c. Rank histograms

Rank histograms are common means to verify ensemble forecasts from ensemble prediction systems. They measure how often the observation falls between each pair in an ordered set of ensemble values. The rank histogram thus assesses the distribution characteristics of an ensemble relative to the distribution of the corresponding observations, specifically to determine how well the average spread of the ensemble compares to the spread of the observations (World Meteorological Organization 2013). Rank histograms can also indicate unconditional biases in the sample of forecasts.

The rank histograms are generated by repeatedly tallying the rank of the observations relative to values from an ensemble sorted from lowest to highest. This can lead to overpopulation of the extreme ranks in rank histograms from ensemble forecasts because of errors in the observation (Hamill 2001). Underspread of the forecast ensemble may be another reason for the higher frequency in the extreme ranks, so in seeking to diagnose whether an ensemble is underspread, it is necessary to account for observation error. To account for the observation error, we followed Hamill (2001), and

FIG. 12. Histograms showing the (left) R_{34} and (right) $R_{\max}$ distributions for the BT (blue), EPS (aqua), and EPS-BC (pink). All ensemble members are included. The bar representing EPS-BC is centered on the value of the bin.

generated rank histograms by ranking the observations from the BT relative to the sorted ensemble, after first adding a random observational noise to each ensemble member. The observations in the BT are described in section 3b. The noise is added using random Gaussian method with mean = 0 and standard deviation of 5 m s^{-1} for $V_{\max}$, 10 hPa for CP, 25% for $R_{\max}$, and 20 km for R_{34} . We chose these values empirically based on discussions with researchers having a sound knowledge of BT TC parameters. Rank histograms are shown below for 24-h lead time for EPS (Fig. 9), EPS-BC using SLR (Fig. 10), and the EPS-BC using the MVR (Fig. 11).

Figures 9–11 show that there is a tendency for most of the TC variables to have higher frequency in the extreme ranks which is consistent with our earlier finding that the ensemble is underspread. After the bias correction, the middle ranks become more populated indicating a reduced bias. The exceptions are the position (i.e., longitudes and latitudes), which represent quite reliable spread-skill. The $R_{\max}$ shows large biases in EPS (Fig. 9) with ranks greater than 20 barely populated, while the bias-correction systems, EPS-BC (Figs. 10

and 11) are improved, but still indicate underspread. The CP shows the most gains after bias correction in both the simple linear and the MVR models where the rank histograms (Figs. 10 and 11) are fairly uniform while still showing high frequency in the lowest rank. This means that the CP is occasionally overestimated.

The $R_{\max}$ and speed histograms approach a U-shape where the high frequencies are found at the extreme ensemble ranks. Comparing to the spread-skill diagrams, this most likely indicates that the ensemble is underspread, although an element of residual bias may also contribute.

d. Histograms of climatological frequency

We compared the frequency distribution of the TC parameters in the BT to the EPS and EPS-BC data. A clear demonstration of the improvement in the model skill after bias correction is presented in the histograms in Figs. 12 and 13. These show the frequency distribution of the BT data along with the EPS and EPS-BC for R_{34} , $R_{\max}$, CP, and $V_{\max}$ for several lead times. For $R_{\max}$, there are significant gains after bias correction (Fig. 12)

FIG. 13. Histograms showing the (left) CP and (right) $V_{\max}$ distributions for the BT (blue), EPS (aqua), and EPS-BC (pink) using MVR. All ensemble members are included. The bar representing EPS-BC is centered on the value of the bin.

when the EPS-BC can predict most $R_{\max}$ below 100 km in agreement with the BT, which the EPS was unable to estimate. However, EPS-BC does not perform as well in predicting larger $R_{\max}$, which may occur during weak storms. It is evident from Fig. 13 that the EPS severely underestimates the $V_{\max}$ with most predictions falling within the 10 m s^{-1} bin and some values in the 20 m s^{-1} bin. After bias correction, the model is able to predict the larger speeds exceeding $30\text{--}40 \text{ m s}^{-1}$, which are in better agreement with the BT data. The CP distribution (Fig. 13) has similar characteristics to that of the $V_{\max}$ where the EPS is unable to predict low CP and severely overestimates high CP. Again, the bias correction substantially improves this situation.

It is clear that the EPS-BC remains underrepresented in the tails of these distributions, even though it is significantly better than the EPS. This is probably partly due to the known tendency for statistical prediction schemes to be reluctant to predict record or near-record values, and partly due to the tendency for the ensemble to be underspread.

e. Box-and-whisker plots

The box-and-whisker plots in Fig. 14 present the climatological distributions of BT data in comparison with the EPS and EPS-BC using MVR. The plots clearly show that the distributions for all four predicted TC parameters (CP, R_{34} , $V_{\max}$, and $R_{\max}$) after bias correction (pink) are in better agreement with the BT (blue) than before the bias correction (aqua). The cyclone intensity (represented by an overestimated EPS CP and an underestimated EPS $V_{\max}$) was severely underestimated before the bias correction and shows a significant improvement after the bias correction. The EPS $R_{\max}$ correlated poorly with the BT before the bias correction, which is consistent with our analyses presented in preceding sections (sections 6b, 6c, and 6d). The EPS-BC leads to improved distributions of all four parameters.

f. Cumulative distribution functions

Cumulative distribution function (CDF) plots are a useful tool for assessing similarities between distributions

FIG. 14. Box-and-whisker plots showing distributions of the BT (blue box and whiskers, and equivalent gray shading for reference), EPS (aqua), and EPS-BC (pink) using the MVR. The top of the box represents the upper quartile (Q3) while the bottom of the box represents the lower quartile (Q1). The median is shown by a horizontal line across the box. The whiskers extend from the edges of box to show the range of the data. The position of the whiskers is set to $1.5 \times (Q3 - Q1)$ from the edges of the box. Outlier points (represented by plus signs) are those that extend beyond the end of the whiskers.

of different datasets. Figure 15 displays the CDF of the climatological BT data with the CDF of the forecast EPS and the EPS-BC from the MVR at a lead time of 24 h. The CDFs of all four EPS-BC TC parameters show an increased similarity to the BT CDF. For example, the EPS CP shows occurrences of less than 1% below 980 hPa while the BT shows more than 20% occurrence for the same CP (Fig. 15). The EPS-BC does reasonably well in estimating the occurrences of CP of around 980 hPa, however, it still predicts a lower probability of occurrence of CP less than 980 hPa, and estimates significantly more occurrences that are higher than 980 hPa relative to the BT. In any case, the CDF for EPS-BC CP is much closer to the BT CDF than the EPS versus the BT CDF. Similarly, for $V_{\max}$, the EPS-BC CDF has more similarities to the BT than the EPS has with the BT CDF. R34 shows improvements in EPS-BC but

these are not as dramatic as those for the other TC parameters.

g. Q-Q plots

The quantile–quantile (Q–Q) plot is used to assess differences in distributions of two datasets and determine whether they have similar distributions. It displays the quantiles of the first dataset against the quantiles of the second dataset which are determined to be equivalent if the points fall approximately along a straight 45° reference line. The greater the departure from this reference line, the greater the evidence to conclude that the two datasets have come from populations with different distributions.

Q–Q plots were generated for the BT against EPS, and BT against EPS-BC using the MVR. It is evident from Fig. 16 that the bias correction leads to the distributions moving closer to the reference line thus indicating

FIG. 15. Cumulative distribution functions at a lead time of 24 h for the BT (blue), EPS (aqua), and EPS-BC (pink) using the MVR.

that the differences between EPS and the BT have been greatly reduced after the application of the bias correction.

7. Vortex insertion

Once the TC parameters are estimated and verified, bias-corrected wind fields are generated by inserting a synthetic bias-corrected vortex in the EPS wind fields. The corrected wind fields were then used to force the wave model. As the EPS can predict asymmetries reasonably well due to the storm motion, and due to the response to the surrounding environment including vertical wind shear, and the bias correction is less successful for individual quadrant wind radii, therefore we preserve the modeled EPS asymmetry and only correct the symmetric component of R34. The method consists of generating a symmetric vortex using the bias-corrected TC structure and replacing the wind field in the original EPS. The symmetric vortex consists of the modified Rankine vortex (Holland et al. 2010) given by

$$v = \begin{cases} V_{\max} \left(\frac{r}{R_{\max}} \right), & r < R_{\max} \\ V_{\max} \left(\frac{R_{\max}}{r} \right)^{\alpha}, & r \geq R_{\max} \end{cases}, \quad (14)$$

where v is the wind speed at distance r from the center, $R_{\max}$ is the radius of maximum winds, $V_{\max}$ is the maximum wind speed, and α is a scaling parameter that can be easily calculated if R34 is known. This basic profile is smoothed near $R_{\max}$ as in Kepert (2013). The asymmetric part of the original wind field is then added to the symmetric vortex to form the final vortex wind field. Thus, the original vortex asymmetry is preserved. Finally, we incorporate a surface inflow angle using the parametric model of Zhang and Uhlhorn (2012) that is based on dropsonde data.

Figure 17 provides a demonstration of the insertion of a symmetric bias-corrected vortex within an asymmetric background EPS field, when run over a single ensemble member. A marked increase in intensity is apparent after the vortex insertion.

FIG. 16. Q–Q plots at a lead time of 24 h for the EPS (aqua) and EPS-BC (pink) using the MVR. The black-dashed line is the “reference line” indicating best similarities between the BT and EPS/EPS-BC.

8. Wind and wave risk maps

The bias-corrected system has been operational at the Australian Bureau of Meteorology since November 2016. The final product consists of forecast graphs at specific locations and maps of exceedance probabilities for various wind and wave thresholds produced every 12 h for lead times up to 10 days. The wind map shown in Fig. 18 was issued on 1800 WST (western standard time) 27 January 2018 and shows a probability of exceeding gale force winds of up to 60% along the NW Australian coast on 29 January (48-h lead time) and 30 January 2018 (72-h lead time).

Figures 19 and 20 provide a contrast between uncorrected EPS and EPS-BC forecasts at North Rankin (offshore platform at 19.584°S, 116.138°E) in the probability of exceeding gale, storm and hurricane force winds during TC Olwyn out to approximately 3 days. The uncorrected EPS forecast did not indicate any chance of experiencing storm or hurricane force winds at any lead time (Fig. 19) despite the observations peaking nearly at storm force (Fig. 20) and the cyclone intensity reaching almost 50 kt. If the standard industry operating

procedure were to act only on a likelihood of storm force winds, then no emergency response action would have been taken. However, if following the EPS-BC, the storm force warning would have been received 3 days in advance with forecasts consistently above 25% chance of exceedance from 3 days prior to the event (Fig. 20). Overall this leads to an increase in confidence of guidance as to the worst-case scenario up to 10 days ahead of time, enabling effective decision making. It also allows more accurate forecasting of the ocean response, in particular the wave field. We have not undertaken a systematic validation of wind probabilities as there are insufficient observations of wind speeds in the study region.

9. Summary and conclusions

We have developed and verified a multivariate statistical bias-correction scheme for tropical cyclone structure and intensity in the ECMWF-EPS. The scheme leads to more accurate forecasts of tropical cyclones, with better ensemble properties such as the

FIG. 17. (top) Pressure fields before bias-correction and after insertion of bias-corrected vortex, and (bottom) wind speed fields before bias correction and after insertion of bias-corrected vortex for TC Olwyn at 1200 UTC 12 Mar 2015. Note: the bias-corrected symmetrical vortex is inserted within the EPS asymmetrical wind field.

spread-skill relationship. The results are thus more suitable guidance in support of probabilistic forecasts than the uncorrected ensembles. They are also used to force an operational wave ensemble (Zieger et al. 2018). Key predictands of the scheme are the storm intensity, radius of maximum winds, and radius of gales. These are predicted from the raw storm parameters including intensity, latitude, and integrated kinetic energy.

Predictors were chosen to give good statistical performance, and to allow practical implementation. A variety of verification techniques including the standard RMSE, rank histograms, cumulative distribution functions, and Q-Q plots were employed to ensure confidence in the predicted variables.

To address the issue of the ensemble spread collapsing at longer lead times, we applied the bias correction to

FIG. 18. Wind risk map showing the probability of exceedance of gale force (34 kt), storm force (48 kt), and hurricane force (64 kt) winds issued at 1800 WST 27 Jan 2018. Probabilities are presented as percentage of model scenarios with winds exceeding specific threshold per day. Please note that the “raw” model output in the legend refers to the bias-corrected data as received by the forecasters. The forecasters do not further modify or reanalyze the data before issuing it to end users.

the ensemble mean and then computed each individual ensemble member to preserve the ensemble spread. In situations where the ensemble mean does not exist because the forecast data does not contain that parameter, for example R34, then our technique consists of applying ensemble perturbations to R34 using the IKE, which is always present in the forecast data.

We excluded $R_{\max}$ as a potential predictor in the bias-corrections system because the EPS indicated a poor skill in forecasting $R_{\max}$. To predict $R_{\max}$, we used a climatological relationship between $R_{\max}$ and the wind intensity and the storm latitude.

We trained the scheme on two years of operational forecast data, which we found to be the best balance between an adequately large training set with a consistent model setup. Performance was worse when 3 years of training data were used. We also tested long-term ensemble hindcast data as part of the training set but found it to be unsuitable.

We reconstructed the wind fields by applying a vortex insertion method that retains the modeled asymmetry and incorporates a surface wind inflow angle designed to match measurements from a large body of dropsonde data.

The bias-corrected system achieves an overall skill improvement over the uncorrected EMCWF-EPS for all TC intensity and structure parameters by 23%–46%. Although the bias-correction technique is applicable to strong and weak TCs, the technique is more beneficial in correcting intense TCs than in correcting weak TCs or systems undergoing extra-tropical transition.

The system has been operational at the Bureau of Meteorology since November 2016. It provides the resources industry with TC forecast guidance consisting of maps and graphs of exceedance probabilities of winds and waves for various wind and wave thresholds produced every 12 h for lead times up to 10 days during the TC season.

FIG. 19. Uncorrected EPS wind probability forecasts for TC Olwyn at North Rankin. Solid lines indicate the forecast chance of exceeding one of three risk thresholds: gale (green), storm (yellow), and hurricane (red) force winds. Dashed lines indicate the period in which the observations exceeded the same thresholds. Please note that probabilities for storm and hurricane winds is zero, hence yellow and red lines are not clearly visible.

FIG. 20. EPS-BC wind probability forecasts for TC Olwyn at North Rankin. Solid lines indicate the forecast chance of exceeding one of three risk thresholds: gale (green), storm (yellow), and hurricane (red) force winds. Dashed lines indicate the period in which the observations exceeded the same thresholds. Observed wind speeds are provided in the bottom panel in blue for reference.

Acknowledgments. The authors gratefully acknowledge funding and advice for this study from the Joint Industry Partnership Group–Shell, Woodside, Chevron, and INPEX. We thank Andrew Burton, Mike Bergin, and Stefan Zieger for their valuable comments and feedback. All data presented in this paper have been referenced in figures, tables, text, and references.

REFERENCES

- Atger, F., 2003: Spatial and interannual variability of the reliability of ensemble-based probabilistic forecasts: Consequences for calibration. *Mon. Wea. Rev.*, **131**, 1509–1523, [https://doi.org/10.1175/1520-0493\(2003\)131<1509:SAIVOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1509:SAIVOT>2.0.CO;2).
- Bauer, P., and D. S. Richardson, 2014: New model cycle 40r1. *ECMWF Newsletter*, No. 138, ECMWF, Reading, United Kingdom, p. 3, <https://www.ecmwf.int/en/elixir/14581-newsletter-no-138-winter-2013-14>.
- Benoit, K., 2011: Linear regression models with logarithmic transformations. Methodology Institute London School of Economics, 8 pp., <https://kenbenoit.net/assets/courses/ME104/logmodels2.pdf>.
- Buizza, R., and T. N. Palmer, 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434–1456, [https://doi.org/10.1175/1520-0469\(1995\)052<1434:TSVSOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2).
- , and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141**, 3366–3382, <https://doi.org/10.1002/qj.2619>.
- , and D. S. Richardson, 2017: 25 years of ensemble forecasting at ECMWF. *ECMWF Newsletter*, No. 153, ECMWF, Reading, United Kingdom, accessed 15 February 2019, <https://www.ecmwf.int/en/newsletter/153/meteorology/25-years-ensemble-forecasting-ecmwf>.
- , M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **134**, 2051–2066, <https://doi.org/10.1002/qj.346>.
- Bureau of Meteorology, 2011: Tropical cyclone database: Structure specification. Australian government, Bureau of Meteorology, 18 pp., http://www.bom.gov.au/cyclone/history/database/TC_Database_Structure_Oct2011.pdf.
- , 2014: A guide for best tracking tropical cyclones in the Australian region, version 2.9, Bureau of Meteorology Training Centre, Australian government, 25 pp.
- Chan, K. T. F., and J. C. L. Chan, 2015: Global climatology of tropical cyclone size as inferred from QuikSCAT data. *Int. J. Climatol.*, **35**, 4843–4848, <https://doi.org/10.1002/joc.4307>.
- Courtney, J., and J. A. Knaff, 2009: Adapting the Knaff and Zehr wind–pressure relationship for operational use in Tropical Cyclone Warning Centres. *Aust. Meteor. J.*, **58**, 167–179, <https://doi.org/10.22499/2.5803.002>.
- Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias correction for global ensemble forecast. *Wea. Forecasting*, **27**, 396–410, <https://doi.org/10.1175/WAF-D-11-00011.1>.
- Durrant, T. and D. Greenslade, 2011: Evaluation and implementation of AUSWAVE. Collaboration for Australian Weather and Climate Research, CAWCR Tech. Rep. 041, 62 pp., <https://pdfs.semanticscholar.org/df5d/19f0f3954d96091078a7a97f8b5a4d19499b.pdf>.
- ECMWF, 2012: The ECMWF Ensemble Prediction System: The rationale behind probabilistic weather forecasts. European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, 4 pp., <http://www.ecmwf.int/sites/default/files/elibrary/2012/14557-ecmwf-ensemble-prediction-system.pdf>.
- , 2016: New forecast model cycle brings highest-ever resolution. European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, accessed 15 February 2018, <http://www.ecmwf.int/en/about/media-centre/news/2016/new-forecast-model-cycle-brings-highest-ever-resolution>.
- , 2017: Re-forecast for medium and extended forecast range, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, accessed 1 May 2017, <http://www.ecmwf.int/en/forecasts/documentation-and-support/extended-range/re-forecast-medium-and-extended-forecast-range>.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, [https://doi.org/10.1175/1520-0450\(1972\)011<1203:TUOMOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2).
- Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, <https://doi.org/10.1175/MWR2904.1>.
- Grafarend, E. W., 2006: *Linear and Nonlinear Models: Fixed Effects, Random Effects, and Mixed Models*. Walter de Gruyter, 553 pp.
- Haiden, T., M. Janousek, J. Bidlot, R. Buizza, L. Ferranti, F. Prates, and F. Vitart, 2018: Evaluation of ECMWF forecasts, including the 2018 upgrade. ECMWF Tech. Memo. 831, ECMWF, Reading, United Kingdom, 54 pp., <https://www.ecmwf.int/en/elixir/18746-evaluation-ecmwf-forecasts-including-2018-upgrade>.
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, [https://doi.org/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2).
- , and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327, [https://doi.org/10.1175/1520-0493\(1997\)125<1312:VOERSR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2).
- Holland, G. J., J. I. Belanger, and A. Fritz, 2010: A revised model for radial profiles of hurricane winds. *Mon. Wea. Rev.*, **138**, 4393–4401, <https://doi.org/10.1175/2010MWR3317.1>.
- Johnson, C., and N. Bowler, 2009: On the reliability and calibration of ensemble forecasts. *Mon. Wea. Rev.*, **137**, 1717–1720, <https://doi.org/10.1175/2009MWR2715.1>.
- Kazmier, L., 2003: *Schaum's Outline of Business Statistics*. McGraw Hill Professional, 432 pp.
- Keper, J. D., 2013: How does the boundary layer contribute to eyewall replacement cycles in axisymmetric tropical cyclones? *J. Atmos. Sci.*, **70**, 2808–2830, <https://doi.org/10.1175/JAS-D-13-046.1>.
- Knaff, J. A., S. P. Longmore, and D. A. Molnar, 2014: An objective satellite-based tropical cyclone size climatology. *J. Climate*, **27**, 455–476, <https://doi.org/10.1175/JCLI-D-13-00096.1>.
- Kossin, J. P., J. A. Knaff, H. I. Berger, D. C. Herndon, T. A. Cram, C. S. Velden, R. J. Murnane, and J. D. Hawking, 2007: Estimating hurricane wind structure in the absence of aircraft reconnaissance. *Wea. Forecasting*, **22**, 89–101, <https://doi.org/10.1175/WAF985.1>.
- Landsea, C. W., and J. L. Franklin, 2013: How “good” are the best tracks?—Estimating uncertainty in the Atlantic Hurricane Database. *Mon. Wea. Rev.*, **141**, 3576–3592, <https://doi.org/10.1175/MWR-D-12-00254.1>.

- Molteni, F., R. Buizza, T.N. Palmer, and T. Petroliaigis, 1994: The ECMWF Ensemble Prediction System: Methodology and validation. ECMWF Tech. Memo. 202, ECMWF, Reading, United Kingdom, 56 pp., <https://www.ecmwf.int/en/elibrary/11189-ecmwf-ensemble-prediction-system-methodology-and-validation>.
- Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **114**, 463–493, <https://doi.org/10.1002/qj.49711448010>.
- Puri, K., and Coauthors, 2013: Implementation of the initial ACCESS numerical weather prediction system. *Aust. Meteor. Oceanogr.*, **63**, 265–284.
- Quiring, S., A. Schumacher, C. Labosier, and L. Zhu, 2011: Variations in mean annual tropical cyclone size in the Atlantic. *J. Geophys. Res.*, **116**, D09114, <https://doi.org/10.1029/2010JD015011>.
- Richardson, D. S., 2000: Skill and relative economic value of the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **126**, 649–668, <https://doi.org/10.1002/qj.49712656313>.
- Seo, S., 2006: A review and comparison of methods for detecting outliers in univariate data sets. M.S. thesis, Dept. of Biostatistics, University of Pittsburgh, 53 pp.
- Takagi, H., and W. Wu, 2016: Maximum wind radius estimated by the 50 kt radius: improvement of storm surge forecasting over the western North Pacific. *Nat. Hazards Earth Syst. Sci.*, **16**, 705–717, <https://doi.org/10.5194/nhess-16-705-2016>.
- Torn, R. D., and S. Snyder, 2012: Uncertainty of tropical cyclone best track information. *Wea. Forecasting*, **27**, 715–729, <https://doi.org/10.1175/WAF-D-11-00085.1>.
- Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley and Sons, 137–163.
- Van der Grijn, G., J.-E. Paulsen, F. Lalaurette, and M. Leutbecher, 2005: Early medium-range forecasts of tropical cyclones. ECMWF Newsletter, No. 102, ECMWF, Reading, United Kingdom, 7–14.
- Vitart, F., J. L. Anderson, and W. F. Stern, 1997: Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations. *J. Climate*, **10**, 745–760, [https://doi.org/10.1175/1520-0442\(1997\)010<0745:SOIVOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<0745:SOIVOT>2.0.CO;2).
- Whitaker, J. S., and A. F. Lough, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292–3302, [https://doi.org/10.1175/1520-0493\(1998\)126<3292:TRBESA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2).
- Wilks, D. S., 2006: Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteor. Appl.*, **13**, 243–256, <https://doi.org/10.1017/S1350482706002192>.
- , 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Willoughby, H. E., and M. E. Rahn, 2004: Parametric representation of the primary hurricane vortex. Part I: Observations and evaluation of the Holland (1980) model. *Mon. Wea. Rev.*, **132**, 3033–3048, <https://doi.org/10.1175/MWR2831.1>.
- World Meteorological Organization, 2013: Verification methods for tropical cyclone forecasts. WWRP/WGNE Joint Working Group on Forecast Verification Research, World Weather Research Program and Working Group on Numerical Experimentation, 84 pp.
- Zhang, J. A., and E. W. Uhlhorn, 2012: Hurricane sea surface inflow angle and an observation-based parametric model. *Mon. Wea. Rev.*, **140**, 3587–3605, <https://doi.org/10.1175/MWR-D-11-00339.1>.
- Zieger, S., D. Greenslade, and J. D. Kepert, 2018: Wave ensemble forecast system for tropical cyclones in the Australian region. *Ocean Dyn.*, **68**, 603–625, <https://doi.org/10.1007/s10236-018-1145-9>.