Prediction errors of tropical cyclones in the western north Pacific in the Met Office global forecast model

Prediction Errors of Tropical Cyclones in the Western North Pacific in the Met Office Global Forecast Model

K. I. HODGES AND N. P. KLINGAMAN

National Centre for Atmospheric Science, and Department of Meteorology, University of Reading, Reading, United Kingdom

(Manuscript received 11 January 2019, in final form 20 June 2019)

ABSTRACT

The prediction of tropical cyclones (TCs) in the western North Pacific (WNP) and the Philippines Area of Responsibility (PAR) has been explored in the Met Office (UKMO) global forecasting system over a 10-yr period at 0–7-day lead times. Both the high-resolution deterministic and lower-resolution ensemble systems have been considered. Location errors for verification against the observations are comparable for the deterministic, control, and ensemble mean forecasts; however, the ensemble spread indicates the ensemble is underdispersive. Intensity error metrics, for pressure and surface winds, show large biases relative to the observations, with the smallest biases for the deterministic system. For the intensity metrics the ensemble spread shows the ensemble is severely underdispersive primarily due to the large errors relative to the observations. Verification against the analyses shows similar results to verification against the observations for location. This is also the case for the intensities albeit with smaller errors and less underdispersion. The PAR region has larger intensity errors and biases and larger intensity ensemble spread compared with the broader WNP region. Forecast errors for location and intensity have reduced significantly with system upgrades over the period studied (2008–17) for the deterministic and ensemble systems. Intensity errors for the latest configuration of the deterministic system at day 4 are smaller than the initial errors of all the earlier configurations for both pressure and winds. The Madden–Julian oscillation (MJO) and boreal summer intraseasonal oscillation (BSISO) significantly affect the intensity forecast errors, but not the location errors. Intensity errors are lower at the initiation and for early lead times of the forecasts started in phases 6–7 and 7–8, when the MJO and BSISO are active in the WNP. These reduced errors appear to result mainly from the variations in intensity of the observed storms with MJO and BSISO phases, though the initial states of the forecasts are also affected. Over the studied period, the European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic and ensemble systems have lower errors and biases for both location and intensity than the UKMO forecast systems.

1. Introduction

The western North Pacific (WNP) region experiences the largest number of tropical cyclones (TCs) annually. The island nations and continental countries bordering this region are highly vulnerable to economic losses and loss of life from such storms. It is imperative to improve the prediction of TCs in the WNP to prevent loss of life and mitigate against economic damage from the associated storm surges, high winds, and intense precipitation. For example, Typhoon Haiyan in 2013, considered one of the most powerful typhoons to have made landfall, killed over 7000 people predominantly in the Philippines and caused estimated economic losses of $5.8 billion (U.S. dollars). The Philippines Area of Responsibility (PAR) experiences on average more than 20 TCs a year, of which 5–6 make landfall. Cinco et al. (2016) showed that although the frequency of TCs affecting the Philippines over the last 40 years has not changed significantly, the annual losses attributed to TCs have increased over the period. Other WNP countries are also severely impacted by TCs, including China, Vietnam, Japan, and the Korean Peninsula.

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WAF-D-19-0005.s1.

Corresponding author: K. I. Hodges, k.i.hodges@reading.ac.uk

DOI: 10.1175/WAF-D-19-0005.1

In view of the vulnerability of the Philippines to TCs, the Met Office (UKMO) has built a partnership with the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA), the Philippine National Meteorological Service. A particular focus of this partnership is to improve the capability to predict regional weather extremes to prevent loss of life and mitigate damage. For TCs, the focus is on the Met Office high-resolution deterministic global forecast model, the Met Office Global Ensemble Prediction System (MOGREPS-G) and the Global Seasonal (GLOSEA) forecasting system, together with regional dynamical downscaling (Short and Petch 2018).

Recent studies have highlighted improved TC forecasts of location or position (often referred to as ''track'' in the literature) in global numerical weather prediction (NWP) models (Yamaguchi et al. 2017). TC intensity is still underestimated due primarily to a lack of resolution and the use of parameterized processes (DeMaria et al. 2014). Regional downscaling can provide more accurate TC forecasts. However, at the convective-scale resolutions required, with grid spacing less than 5 km, downscaling is often possible only over a rather restricted area that precludes forecasts beyond about 5 days (Short and Petch 2018). Studies have also shown that prediction errors, both of location (track) and intensity, can vary considerably between different ocean basins (Hodges and Emerton 2015, hereafter HE15). The WNP tends to have larger errors than most other TC regions.

Various tropical modes of variability can affect WNP TC genesis, path, and possibly intensity, for example, the El Niño–Southern Oscillation (ENSO; Kim et al. 2011), subtropical high variability (Wu et al. 2005), monsoon variability (Wu et al. 2012), and the Madden–Julian oscillation (MJO) (Klotzbach and Oliver 2015). However, studies of the impact of these modes of variability on the prediction of WNP TCs have so far been limited to seasonal (Lee et al. 2018; Vitart 2009; Camp et al. 2015) or longer-term prediction, with a frozen model. These studies emphasize TC occurrence rather than location and intensity errors of observed TCs, which is the focus of NWP. The lack of emphasis on NWP time scales is likely due to the lack of long-term datasets and the continual changes to the NWP systems. We attempt to address this issue in this paper in a limited way for the MJO and boreal summer intraseasonal oscillation (BSISO) (Lee et al. 2013). The MJO and BSISO are tropical modes of variability active on 30–70-day time scales. The MJO modulates the large-scale tropical environment and hence TC genesis (Klotzbach and Oliver 2015). The convectively active MJO phases are associated with enhanced TC activity and rapid intensification in all the main TC active regions (Klotzbach and Oliver 2015). The BSISO is considered since it is active during the main WNP TC season (May–October), whereas the MJO is generally more active during boreal winter. The BSISO also impacts the WNP TC genesis (Yoshida et al. 2014). How NWP forecast errors of TCs are affected by the MJO and BSISO is so far unknown.

The aims of this paper are to assess the accuracy of WNP TC predictions in the Met Office global model, document the improvements in TC predictions made in recent years with improvements to the forecasting system (data assimilation and model), and explore the impact of the subseasonal MJO and BSISO modes of variability on TC prediction. Heming (2016) found a dramatic improvement in TC track and intensity predictions in the global Met Office model, following upgrades to model resolution and physics in 2014, and in particular the introduction of a new TC bogusing scheme in 2015. However, this evaluation only covered forecasts conducted for a limited number of particular case studies. Here, a much longer record of forecasts are considered, encompassing several major model upgrades. We also contrast the performance of the high-resolution UKMO deterministic NWP system with that of the lower-resolution ensemble system and with the European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic and ensemble prediction systems.

2. Data and methodology

a. Forecast data

The primary data for this study are the operational Met Office forecasts from the global forecasting system, both the deterministic and ensemble (MOGREPS-G) forecasts, produced by the Unified Model (UM). The period covered is 2008–17. The deterministic system forecasts are produced twice a day at 0000 and 1200 UTC, resulting in over 7300 forecasts. The MOGREPS-G ensemble forecasts are produced twice a day at 0000 and 1200 UTC until November 2014 and four times a day after this (0000, 0600, 1200, and 1800 UTC) resulting in over 9300 forecasts. However, for consistency, only the 0000 and 1200 UTC MOGREPS-G forecasts are used. Note, not all forecasts will contain a TC in the WNP and PAGASA-PAR study areas.

During the 2008–17 period the assimilation system and model experienced several major upgrades, which are summarized in Table 1. The UM dynamical core is nonhydrostatic since 2002 (Davies et al. 2005), solving the deep atmosphere equations on a latitude–longitude horizontal grid with terrain following eta

TABLE 1. Summary of Met Office forecast configurations used in this study for the 2008–17 period (ETKF: ensemble transform Kalman filter; RP2: random parameters scheme; SKEB: stochastic kinetic energy backscatter; Det: deterministic; and Ens: ensemble). Stated resolutions are at the equator.

	Pre-2010		2010–14		2014–17		Post-2017	
Component	Det.	Ens.	Det.	Ens.	Det.	Ens.	Det.	Ens.
Resolution (H)	N320 (60km)	N144 (125km)	N512 (40km)	N216 (90km)	N768 (25km)	N400 (50km)	N1280 (17km)	N640 (30km)
Resolution (V)	L50/70	L38	L70	L70	L70	L70/85	L85	L85
Assimilation	Hybrid 4DVar							
Period	6 days	15 days	6 days	15 days	6 days	7 days	6 days	7 days
Start times	0, 12	0, 12	0, 12	0, 12	0, 12	0, 6, 12, 18	0, 12	0, 6, 12, 18
Ensemble size		24		24		12		18
Perturbations		ETKF		ETKF		ETKF		ETKF
Stochastic physics		RP2 + SKEB2/SKEB1		RP2 + SKEB2		RP2 + SKEB2		RP2 + SKEB2

levels in the vertical using a semi-implicit, semi-Lagrangian methodology. In 2014, significant improvements were made to both the dynamical core and physics packages (Walters et al. 2017). The effect of TC initialization (bogusing) on the forecasts of TCs in this new model setup was evaluated and contrasted with an older scheme by Heming (2016), which found significant improvements in both track and intensity errors. The single high-resolution deterministic forecasts are generally run at twice the resolution, in latitude and longitude, of the ensemble forecasts. Initial conditions are provided by a four-dimensional variational data assimilation scheme (4DVar).

For the ensemble forecasts, a single unperturbed control forecast is performed using the analysis from the higher -esolution deterministic system interpolated to the ensemble resolution. This constitutes one member of the ensemble. The other members are derived from a perturbed initial state obtained via the ensemble transform Kalman filter (ETKF) (Bowler et al. 2009). Additionally, the model is perturbed during the forecast integration using two stochastic physics schemes: the stochastic kinetic energy backscatter (SKEB) scheme (Shutts 2005) and the random parameters (RP) scheme (Bowler et al. 2008) that applies small perturbations to several parameters within the parameterizations. Perturbations are also applied to sea surface temperature (SST) and soil moisture.

The cyclone tracking scheme (section 2c) is applied to the 6-hourly data of all forecasts for the entire forecast; however, for verification only the common 0–7-day forecast lead times are used.

b. *Verification data*

To verify the TC forecasts two datasets are used. The first is the commonly used International Best Track Archive for Climate Stewardship (IBTrACS) dataset (Knapp et al. 2010), which is a postseason reanalysis of TC observations from all available agencies. In the WNP multiple TC operational agencies contribute data to IBTrACS, which must be quality controlled. However, there are considerable uncertainties in the data from the different agencies in terms of the frequency and intensity of WNP TCs (Ren et al. 2011; Barcikowska et al. 2012). These uncertainties affect TC verification, depending on which agency's data are used. Further discussion of the impact of observational uncertainty on forecast verification and the identification of TCs in model data can be found in the appendix of HE15 and in Hodges et al. (2017). Missing data in IBTrACS affect the verification of intensity measures of 10-m winds and surface pressure, both of which are used here. The original IBTrACS 10-m wind

speed data in knots are converted to meters per second. The World Meteorological Agency (WMO) winds from IBTrACS are used here, which are the Japan Meteorological Agency (JMA) 10-min-average sustained winds. These are converted to 1-min sustained wind speeds using a factor of 1.13 (Harper et al. 2010). Discussion of the uncertainties in the use of this conversion factor can be found in HE15. This conversion is retained here for consistency with that study.

While verification against IBTrACS data is important, a complementary perspective can be obtained by verifying against the unperturbed analyses, used as initial conditions for the deterministic forecasts, or a reanalysis product. This gives a broader scope for verification as more data are generally available than are present in IBTrACS. It also gives a clearer picture of the error growth, in particular for intensity, than obtained by verifying against best track, for which the initial errors swamp the error growth. This approach was used in HE15 for the ECMWF deterministic and ensemble predictions of NH TCs for 2008–12. The main uncertainty in this approach is that the same model is used for the deterministic analyses and the forecasts. The analyses and the forecasts have the same resolution and use parameterized physics, which results in weaker TCs than observed. However, the verification can be performed at a comparable resolution. Analyses are also sensitive to the observations that are assimilated and the data assimilation method. The operational analysis system changes over time with frequent upgrades, in particular to the background model (see Table 1). An alternative would be to use a reanalysis where the model and data assimilation system are frozen; however, current reanalyses have considerably coarser resolutions than the forecasts used here. Ideally, the Met Office 6-hourly operational analyses could be used. However, for the early part of the study period a considerable fraction of the analysis data is missing from the Met Office archive. Therefore, the ECMWF operational analyses are used instead. The ECMWF analysis system has also changed over the study period, as summarized in HE15, but has much higher resolution (between 25 and 10 km) than any global reanalysis with no missing time steps. Using an independent verification dataset, produced by a different model and data assimilation system, might also provide a fairer evaluation of the UKMO forecast errors than using the initial conditions used for the forecasts.

c. *Methods: Tracking and statistics*

The methods used here to track and identify the TCs in the Met Office forecasts and ECMWF analyses, and to produce the error diagnostics, are the same as used by HE15, where further details can be found. The tracking is performed using the scheme described in Hodges (1994, 1995, 1999), which has been applied in several TC related studies, as well as for other weather system types. First the vertical average of the relative vorticity between 850 and 600 hPa is obtained, which essentially uses the 850-, 700-, and 600-hPa levels. This field is then spatially filtered using spherical harmonics to a T63 resolution, at the same time removing the large-scale background by setting total wavenumbers $n \leq 5$ to zero. Vorticity maxima (in the NH) are then first determined on the T63 grid. These are then used as the starting points for determining the off-grid locations using a B-spline interpolation and a steepest ascent maximization method (Hodges 1995). In the first instance, all positive vorticity centers in the Northern Hemisphere (NH) are tracked in the filtered T63 vorticity field that exceed a threshold of $0.5 \times 10^{-5}\,\mathrm{s}^{-1}$ between 0° and 60°N. This approach produces the most coherent tracks for the full life cycles of the systems, including their pre-TC and post-TC stages. The tracking is performed over the full length of the forecasts and for the analyses over full years (January–January). After completing the tracking, other variables are added to the tracks, including the maximum 10-m winds and minimum mean sea level pressure (MSLP). This is done by searching for the maximum 10-m winds within a 6° geodesic radius, and for the true MSLP minimum within a 5° radius using the B-splines and minimization method.

To identify the TCs from amongst all tracked features, the same matching methodology as in HE15 is used, which matches the forecast tracks against the verification tracks.

In the first instance, the identically same TC tracks are identified in the analyses as are in the IBTrACS data, to enable verification of forecast TCs against the identically same tracks for both verification datasets. Analysis and IBTrACS tracks are matched using the same methodology as in Hodges et al. (2011) for extratropical cyclones, so that two systems match if they overlap in time for at least 10% of their points and their mean separation distance is less than 4°. The small temporal overlap accounts for the disparate lifetimes of TCs in the IBTrACS data and the analyses. A summary of the number of TCs identified in the analyses for the whole NH is given in Table 2, which shows that nearly every IBTrACS TC can be found in the analyses; the small number missing are primarily weak and short lived. Also shown in Table 2 is the number of analysis and IBTrACS TCs for the two sampling regions described below. In this case, the number of analysis tracks can be greater than IBTrACS due to the longer tracks in the analyses.

TABLE 2. Number of IBTrACS TCs in the NH, WNP, and PAGASA PAR by year and the number identified in the analysis by direct matching.

Years	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
					NH					
IBTrACS	74	72	63	72	68	79	66	78	78	80
Analysis	73	70	62	70	66	77	65	77	77	75
					WNP					
IBTrACS	29	32	22	28	27	38	26	31	32	35
Analysis	32	32	24	28	31	38	25	30	32	33
					PAGASA PAR					
IBTrACS	21	20	13	20	21	25	22	16	17	23
Analysis	23	21	16	19	20	26	22	16	18	25

To identify and verify the forecast TC tracks, the forecast tracks are matched to the verification tracks using the same method and criteria as HE15, and originally used by Froude et al. (2007a) for extratropical cyclones. A forecast track is matched to a verification track if the mean spatial separation of its first four points (1 day) is less than or equal to 4° and is the track with the smallest separation for those four points. Only forecast tracks that have their first point within the first 3 days of the forecast are considered, to exclude matches by chance. This means that systems are included in the verification even if they are not present in the initial conditions. For the ensemble, this means that a match is possible from each ensemble member. The ensemble mean and ensemble spread can be computed for each TC, as long as a minimum of five members are present for each lead time. The matching approach is considered a simpler and more objective approach than applying objective detection criteria, such as the presence of a warm core and intensities above a chosen threshold to identify the TCs. The use of such criteria are often applied in climate model studies of TCs (Bengtsson et al. 2007; Strachan et al. 2013; Roberts et al. 2015; Manganello et al. 2012) but results depend strongly on the chosen thresholds and the ability of the modeling system to represent the TCs structure and intensity.

Statistics are computed for location error, absolute intensity error, intensity bias, and the ensemble spread for location and intensity. The 95% confidence intervals (CIs) for the error statistics are computed from the standard errors for each statistic. While TC intensity is usually measured by the surface winds, here intensity is also measured by MSLP. HE15 also used the T63 vorticity as an intensity measure, to focus on the large-scale aspects of the TCs, which may be more predictable than the smaller scales (Vukicevic et al. 2014); some results discussed here also use this measure.

HE15 considered both homogeneous and nonhomogeneous samples, but for the large sample sizes in that study, and used here, the nonhomogeneous samples (i.e., that use all of the data) are found to be acceptable. HE15 also considered the issue of serial correlation of the forecasts, which in principle could make the CIs too narrow. However, that study found (see also supplementary material of HE15) that the serial correlation decreases rapidly with forecast lead time. At shorter lead times, the sample sizes used in that study and here are so large that correcting for the serial correlation makes little difference to the CIs, hence this also has not been considered here. In general, the CIs are larger for smaller sample sizes as the standard error is proportional to $1/\sqrt{N}$, where N is the number of samples at a particular forecast lead time.

The whole of the NH is initially processed, but to focus on the WNP and in particular the Philippines region, two sampling regions are considered: the whole tropical/subtropical part of the WNP and a simplified version of the PAGASA-PAR (Fig. 1). A TC is considered only if the verifying track enters the sampling region; statistics are computed only for points on the verifying track in the sampling region. For the ensemble spread, statistics are computed for tracks for which the ensemble mean track is in the sampling regions.

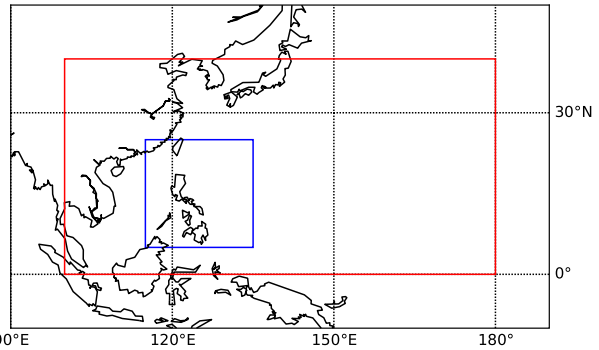

FIG. 1. Sampling regions for TC error statistics. Red box (0°–40°N, 100°E–180°); blue box (5°–25°N, 115°–135°E).

d. Methods: Modes of variability

To see how the TC prediction errors might depend on large-scale modes of variability, the error statistics are conditioned on indices that describe two large-scale modes. The period used here is too short to sample ENSO robustly, or other interannual modes associated with the subtropical high and the monsoons. We instead consider subseasonal modes of variability, of which the most obvious in the WNP is the MJO (Zhang 2005). We use the real-time multivariate MJO (RMM) indices of Wheeler and Hendon (2004) (obtained from http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt), based on the winds from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis and outgoing longwave radiation (OLR) from the National Oceanic and Atmospheric Administration (NOAA) Cooperative Institute for Research in Environmental Sciences (CIRES). Forecasts are partitioned according to whether the RMM amplitude is greater than 1 at the start of the forecast. Our approach is typical of many studies focused on the MJO, where the RMM phases are paired according to where the MJO has its greatest effect on the deep convection [i.e., 2–3 (Indian Ocean), 4–5 (Maritime Continent), 6–7 (Pacific Ocean), and 8–1 (Atlantic)]. The partitioning results in ~1000 forecasts for each phase, which contribute between 300 and 600 TC samples to the statistics in the WNP, depending on the MJO phase and lead time, for the deterministic and ensemble mean forecasts verified against IBTrACS. Using paired RMM phases provides a large enough sample size for reliable statistics, although for small regions such as the PAGASA-PAR the TC sample sizes are still limited (between 100 and 250). We also consider an alternative pairing suggested by Klotzbach and Oliver (2015) of 1–2, 3–4, 5–6, and 7–8. They argued "that TC activity is enhanced in the MJO phases associated with and immediately following the convective maximum in a specific basin, which causes an approximate one-phase shift from the maximum convective anomaly for a particular region" (p. 4201). Another consideration is the seasonal cycle of the MJO, which generally peaks in the boreal winter (December–March) (Zhang and Dong 2004), with a secondary maximum in the boreal summer (June–September). This seasonal cycle also affects the sample sizes of the statistics, as the peak TC season in the WNP is June–October. To examine the impact of this, we consider the BSISO1 indices (Lee et al. 2013) in the same way as for the MJO, i.e., as the two sets of pairings. The BSISO1 indices represent the northward 30–70-day propagating mode most comparable to the MJO (Lee et al. 2013). The BSISO1 index is computed from a multivariate empirical orthogonal function analysis of daily anomalies of OLR and zonal wind at 850 hPa (Lee et al. 2013), similar to the RMM index for MJO. The BSISO1 data are obtained from http://iprc.soest.edu/users/jylee/bsiso.

We do not analyze the MJO or BSISO conditional statistics for the PAGASA-PAR region as the sample sizes are too small for robust statistics at the longer lead times.

3. Results

a. Deterministic and ensemble error statistics

First, the general performance of the Met Office deterministic and ensemble systems are evaluated for the location and intensity errors in the WNP and PAGASA-PAR regions, verifying against both IBTrACS and the ECMWF analyses. Figure 2 shows the sample sizes and error statistics for verification against IBTrACS for the deterministic system (blue), the individual members of the ensemble (black), the control (green), and the ensemble mean (red); also shown is the ensemble spread (orange). In general, results are similar to those produced for the ECMWF system by HE15. The sample sizes (Fig. 2a) are large and greater than 2×10^3 throughout the forecast range for the deterministic, control, ensemble mean and spread and greater than 2×10^4 for the ensemble members. Location errors (Fig. 2b) are determined as the great circle distance and presented in units of degrees ($1° \sim 111$ km). The largest location errors occur for the perturbed ensemble members, considered as independent forecasts. The control and deterministic forecasts have the next lowest, but comparable errors. The ensemble mean has similar errors to the control and deterministic forecasts up to day 4, after which the errors become lower than for the other forecasts. At day 4 the control, deterministic and ensemble mean show one day more skill for location than the individual ensemble members; at day 7 the ensemble mean has just over 0.5 day more skill than the control and deterministic forecasts, though at this lead time errors are as large as 5° (~550 km). The ensemble spread indicates that the ensemble is progressively more underdispersive, as the control forecast error grows faster than the ensemble spread. This means as lead time increases, the observed location may no longer be within the spread of the ensemble. An ideal ensemble forecasting system should have a similar ensemble spread to the forecast error (Buizza 1997).

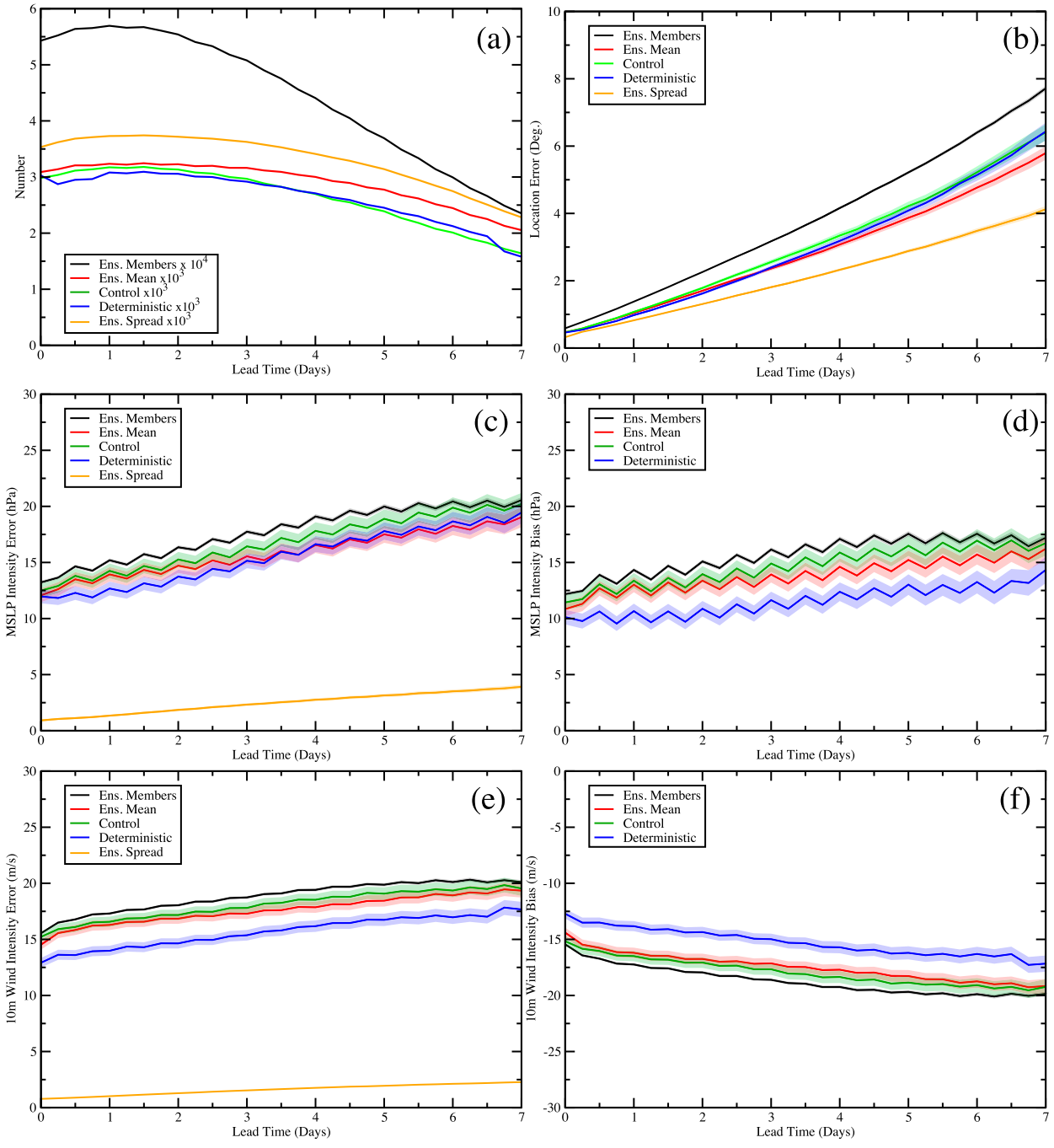

FIG. 2. Error statistics for TCs in the deterministic forecasts in the WNP using best track verification: (a) sample size for each forecast type, (b) location error, (c) MSLP intensity error, (d) MSLP intensity bias, (e) 10-m wind intensity error, and (f) 10-m wind intensity bias. Location error units are geodesic degrees, MSLP units are hPa, and 10-m wind units are m s^{-1}. Shading indicates the 95% confidence interval.

Figures 2c and 2d show the MSLP intensity error and bias, respectively. This shows the well-documented fact that current forecasting systems underestimate observed intensities due to coarse resolution and use of parameterized physics. There are quite large errors and biases, even in the initial states, of 10–15 hPa. The individual ensemble members again show the largest errors and biases; the deterministic forecasts show the lowest, with the control and ensemble mean between these. A positive bias here means the TCs are too

shallow. The difference between the deterministic and ensemble systems is likely to be mainly related to the difference in spatial resolution between the two systems (see Table 1). The control forecast shows systematically smaller errors and biases than the individual ensemble members, even though they are at the same resolution, a result also found for the ECMWF EPS by HE15. This is likely due to the use of stochastic physics during the forecast integration. Froude et al. (2007b) found similar results for extratropical cyclones in the ECMWF EPS. They suggested that if the only difference between the control and perturbed forecasts were the initial conditions, then the errors should converge at the longer lead times. They also suggested that the lack of convergence may be the consequence of using the stochastic physics that is only applied in the perturbed forecasts. The ensemble is severely underdispersive, with a difference between ensemble mean error and spread of similar magnitude to the error itself. This means the observations are highly unlikely to fall anywhere within the ensemble spread, at least over the observed TC life cycle.

An interesting feature of the MSLP intensity error and bias curves is the apparent semidiurnal oscillations. This is a tidal effect associated with the absorption of solar radiation by ozone and water vapor, and heating from the surface that manifests itself on the SLP via internal gravity waves (Dai and Wang 1999). Producing the mean intensities with lead time, for the forecasts and IBTrACS separately (Fig. S1 in the online supplemental material), shows that the oscillations in the MSLP errors and biases arise primarily from the forecast model, with a magnitude ~1 hPa. The oscillations in the observations are much smaller (subtracting the IBTrACS mean intensity from the forecast mean intensity recovers the biases exactly). The magnitude of the forecast oscillations are comparable to those found from direct surface observations, of 1 hPa in the tropics, with peaks just before midday and midnight (Dai and Wang 1999). This suggests that the IBTrACS intensities underestimate the magnitude of the pressure semidiurnal oscillation, probably due to uncertainties in the determination of the pressure minima in TCs related to the Dvorak technique, with greater error or uncertainty in Dvorak estimates for more intense TCs (Torn and Snyder 2012).

For the 10-m winds, results are shown in Figs. 2e and 2f for error and bias, respectively. These show results consistent with the MSLP with large errors, ~12.5–17.5 m s^{-1}, over the whole range of lead times. The different forecasts have a similar relationship in terms of largest and smallest errors, with again the deterministic forecasts having the lowest errors and bias. Again the spread indicates a severely underdispersive ensemble. The 10-m wind errors and biases also show the semidiurnal oscillations, but they are much less apparent than for the MSLP.

Error statistics are also computed using verification against the ECMWF analyses (Fig. S2). The sample sizes are somewhat larger (Fig. S2a) as a result of the longer life cycles of the analysis tracks. For location (Fig. S2b), the results are remarkably similar to those for verification against IBTrACS. This is not surprising, as HE15 showed that the mean difference in separation between the analysis TC tracks and those of IBTrACS is less than 1°, and less than 0.5° for most TCs, which is within the uncertainty in the observed location (Torn and Snyder 2012). Intensity error statistics for T63 vorticity, MSLP, and 10-m winds show similar results to those for verification against IBTrACS, but with generally lower errors and biases. This reflects the closer similarity in spatial resolution between the Met Office NWP system and ECMWF analyses. For the T63 vorticity intensity metric, the deterministic and ensemble mean errors and biases are comparable and lower than for the individual ensemble members and the control forecast. For both location and intensity the ensemble is still underdispersive, even though the resolutions of the forecasts and analyses are comparable. When verifying against the analyses the semidiurnal oscillation is still apparent in all three intensity measures to some extent, but the magnitudes, in particular for MSLP, are much smaller due to the analyses having a similar semidiurnal cycle magnitude to the forecasts.

For the PAGASA-PAR region (shown in Fig. 3 for verification against the IBTrACS), the sample sizes (Fig. 3a) are much smaller due to the smaller sampling region, ranging from $\sim 1 \times 10^3$ for the deterministic, control, ensemble mean and spread to $1–2 \times 10^4$ for the ensemble members. The location errors (Fig. 3b) are slightly smaller than those for the whole WNP, with ~0.25 day more skill for both the deterministic and ensemble mean forecasts. The ensemble spread is also smaller within the PAGASA-PAR region than in the WNP overall. This can also be seen in the spatial distribution of errors (Figs. S3a,b,e) for the ensemble mean and deterministic location errors and the location ensemble spread, respectively. These relatively small differences in location error between the PAGASA-PAR and the broader WNP are likely due to the more zonal propagation of the TCs through the PAGASA-PAR. The larger WNP errors are likely associated with a greater fraction of recurving TCs as discussed by HE15. This may also explain the smaller spread seen in the PAGASA-PAR region. This can also be seen in the much larger location errors and spreads to the

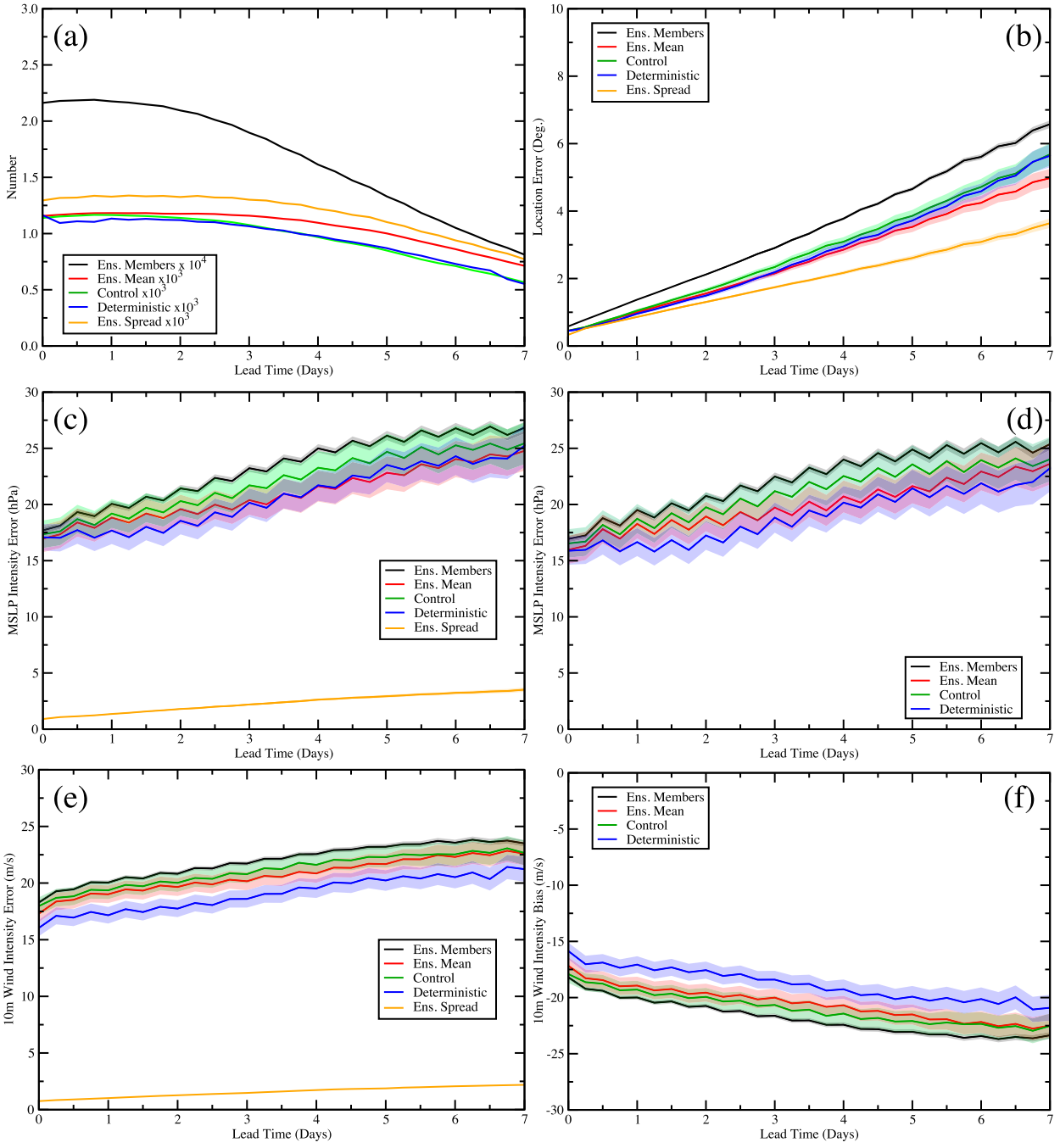

FIG. 3. As in Fig. 2, but for the PAGASA PAR.

north of the PAGASA-PAR, where recurving TCs are more likely (Figs. S3a,b,e).

The TC intensity forecast errors and biases in the PAGASA-PAR can be seen in Figs. 3c–f, for verification against IBTrACS for MSLP and 10-m winds. The general relationship between the errors and biases is similar to those for the WNP. The deterministic forecasts have the lowest errors and absolute biases and the ensemble members the largest. However, comparing Fig. 3 with Fig. 2 shows that the intensity errors and biases are larger in the PAGASA-PAR region by ~5 hPa for MSLP and ~2.5 $m\,s^{-1}$ for the 10-m winds at the start of the forecasts; these are more or less maintained over the forecast range. At a lead time of day 4 this is equivalent to ~4 days less skill (i.e., the error at day 4 in the WNP is similar to the error in the

PAGASA-PAR at day 0). Verifying against the analyses produces similar results (Fig. S4) but with ~1 day less skill at day 4 for the T63 vorticity, ~2 days less skill for MSLP and 1.5 days less skill for 10-m winds by comparing with the equivalent statistics for the WNP (Fig. S2). The spatial distribution of errors for intensity, based on the T63 vorticity, confirm these results (Figs. S3c,d,f for the ensemble mean and deterministic intensity errors and the intensity ensemble spread, respectively). These metrics reveal much larger errors in the PAGASA-PAR region than elsewhere in the WNP.

Similar results have been obtained for location and intensity errors in this region for other forecasting systems [e.g., the ECMWF deterministic and ensemble forecasting systems for the same 2008–17 period (not shown)]. The similarity of the intensity errors and biases between forecasting systems from different centers suggests systematic reasons for these errors. The underestimation of intensity, in particular in the PAGASA-PAR region, is likely related to both model resolution and physical parameterizations. Bender et al. (2017) showed that storm size had an important impact on the prediction of TC intensities, in particular leading up to maximum intensity and for rapid intensification. TCs tracking through this region typically attain their maximum intensity in the PAGASA-PAR region, which combined with the fact that models overestimate the size of the modeled TCs, may partly explain the intensity biases. Short and Petch (2018) found significant improvements in predicting WNP TC intensities using convection-permitting downscaling.

b. *Impact of forecast model changes*

Over the study period considered, 2008–17, many changes have been made to the forecast model and ensemble generation system; these are summarized in Table 1. To assess the impact of these changes, we perform a similar analysis to that in the previous section but for each model/ensemble configuration. These are shown in Fig. 4 for the deterministic system for the WNP region for verification with IBTrACS. Due to the different periods covered by each model configuration, the number of forecast samples varies considerably (Fig. 4a), which affects the width of the confidence intervals shown in Fig. 4. The sample sizes range from less than 250 for the shortest period to ~1500 for the longest period.

The location errors (Fig. 4b) show a steady decrease from one model configuration to the next, such that at day 4 there is a 1.5-day increase in skill between the earliest and latest NWP configuration. The latest changes in 2017 appear to have made little difference to the location errors compared with the preceding system configuration, but the forecast sample size is small (<250). Consequently the confidence intervals are broad and overlap with those of the preceding system.

For intensity, using either MSLP (Fig. 4c) or 10-m winds (Fig. 4e), there is a steady decrease in the error, in particular in the initial states (lead time 0). The error decreases by ~10 hPa for MSLP and ~10 $m\,s^{-1}$ for 10-m winds. This leads to smaller differences in error at longer lead times between the different model configurations. The reduced errors are also reflected in steadily reducing biases. For the latest system, the bias in MSLP (Fig. 4d) is relatively small compared to IBTrACS; intensity is even overpredicted at longer lead times, indicating that some TCs are deeper than observed. This propensity to overdeepen TC has also been seen for similar resolution forecasting systems in HE15 and climate models (Manganello et al. 2012). The bias has been suggested to be due to a lack of coupling to the ocean (Mogensen et al. 2017).

For the 10-m winds (Fig. 4f), while the biases become less negative in the later forecast system configurations compared to the earlier systems, the intensities are still underpredicted.

The wind–pressure relationship offers another perspective on the intensity errors. It is often used operationally to obtain winds from pressure or vice versa (Knaff and Zehr 2007; Brown et al. 2008). To compute this the cyclostrophic equation is written as $P = P_{\text{ref}} - (V_m/a)^c$, where $c = 2.0$ implies cyclostrophic balance (Knaff and Zehr 2007; Brown et al. 2008). This model is fit to the wind–pressure data using the nonlinear regression function in R (R project 2013). The results for the parameters are shown in Table 3 and the plots in Fig. 5. These show a gradual convergence toward the observations for both the parameters and the curves from the early to later model configurations. Substantial biases remain, however, primarily due to the underprediction of the wind intensities.

The MSLP errors and biases show semidiurnal oscillations for each model configuration (Fig. 4), but the oscillations seem much reduced in the absolute errors for the latest two model configurations. It is unclear why this should be the case, but may be due to differences in the cancellation of errors.

The ensemble forecasting system is generally upgraded at the same time as the deterministic system. In general, the results for the ensemble mean are similar to those for the deterministic system for location and intensity (Fig. 6). The most interesting results concern the ensemble spread. For location, while the ensemble is underdispersive for all forecast system configurations, differences between the ensemble mean location error and spread are definitely reduced in the latest

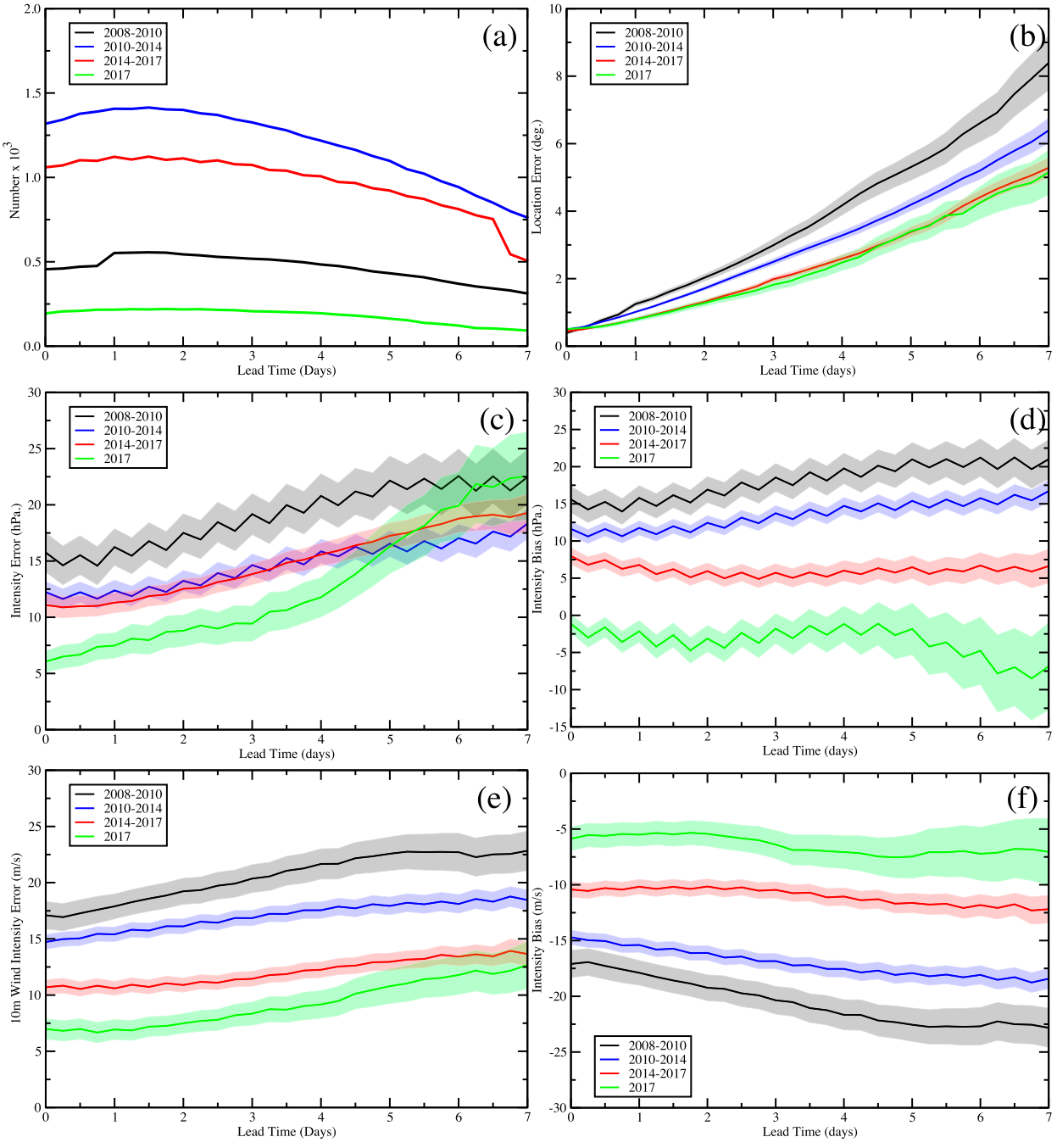

FIG. 4. Error statistics for different deterministic system configurations for TCs in the WNP using best track verification: (a) sample size for each forecast configuration, (b) location error, (c) MSLP intensity error, (d) MSLP intensity bias, (e) 10-m wind intensity error, and (f) 10-m wind intensity bias. Location error units are geodesic degrees, MSLP units are hPa, and 10-m wind units are m s^{-1}. Shading indicates the 95% confidence interval.

configurations. This appears to be due more to reduced errors than an increase in spread. For intensity, the error–spread relationship also improves with both reduced error and increased spread for both MSLP and 10-m winds. The MSLP in particular shows significant improvement.

Results for verification against the analyses give a similar picture (not shown) to those shown above for verification against IBTrACS.

For the PAGASA-PAR region, results for verification against IBTrACS, for both the deterministic (Fig. S5) and ensemble mean forecasts (Fig. S6), show a similar picture

TABLE 3. Parameters for the cyclostrophic equation obtained by fitting to the data for the different deterministic system configurations.

Parameters	Pre-2010	2010–14	2014–17	Post-2017	IBTrACS
a	6.86	5.13	5.44	4.93	3.53
P_{ref}	1010.0	1012.0	1012.0	1013.0	1015.0
c	2.63	2.22	2.14	1.99	1.54

to the results for the full WNP region, albeit with the higher intensity errors seen earlier. Results for the latest NWP configuration are rather noisy, with wide confidence intervals, due to the small available sample size. For the PAGASA-PAR region, results based on verification against the analyses (not shown) are consistent with those for verification against IBTrACS. Forecast error reduces with upgrades to the forecast model, in particular resolution changes, though these are less dramatic than those seen when verifying against the observations.

c. *Impact of Madden–Julian oscillation on forecast errors*

The error statistics are computed for each pair of MJO or BSISO1 phases. Results for the verification of the deterministic forecasts against IBTrACS are shown in Fig. 7 for the standard MJO phase pairings (2–3, etc.). Since forecasts are selected based on whether the index value is greater than 1 at the start of the forecast, the sample sizes are much reduced, even though paired phases are used. The sample sizes range from 300 to 600 depending on MJO phase (Fig. 7a). As with the analysis in the previous section, this results in wide confidence intervals with a strong overlap in places and for particular phases.

For the location errors (Fig. 7b), there appears to be little impact of the MJO. For intensities the lowest initial errors occur for the 6–7 and 8–1 MJO phases for both MSLP and 10-m winds (Figs. 7c and 7e). For phases 8–1, the error growth is more rapid and the error becomes similar to that of phases 2–3 and 4–5 by day 4; however, the error for phases 6–7 continues to be the lowest throughout the forecast. The same behavior is reflected in the intensity biases (Figs. 7d and 7f) with the lowest biases for phases 6–7 and 8–1. Considering the alternative MJO phase pairings (1–2, etc.), shown in Fig. 8, gives similar sample sizes (Fig. 8a) but a slightly different perspective. There is some indication that the MJO affects the location errors, in particular in phases 7–8 where errors are lower after day 4. However, decreasing sample sizes with lead time may make these results less reliable. Comparison with the results for the ensemble mean (not shown) and the ECMWF systems (not shown)

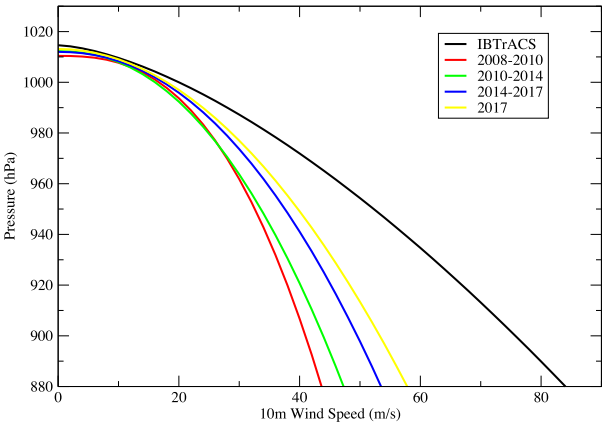

FIG. 5. Pressure–wind relationships for the different model configurations and IBTrACS.

confirm this, as they show no obvious differences between the MJO phases for location errors. The lowest intensity errors (Figs. 8c and 8e) are achieved for MJO phases 7–8, with considerably lower errors than for any of the other paired phases. The biases (Figs. 8d and 8f) show similar behavior. Results for the ensemble mean and for verification against the analyses generally confirm these results (not shown).

Phase 7 is common to the lowest errors for both sets of phase pairings. Phase 8 may also play a role in lower errors and biases at least in the initial part of the forecast. During MJO phases 6–8 the peak convection is in the WNP, which might influence TCs there either directly through cyclogenesis or by making the large-scale environment more conducive to TC development (e.g., reduced vertical shear and increased midlevel moisture). The MJO appears to exert a strong influence at the start of the forecasts, which then either continues to affect the forecast (e.g., phases 6–7 and 7–8), or quickly disappears with rapid error growth (e.g., phases 8–1). The influence of the MJO on the initial state errors is complicated by variable sample sizes with MJO phase, as well as the changing data assimilation/forecast system through the period, which substantially influences the initial intensity errors (cf. Figure 4). How the MJO affects TCs may also depend on the TC location and stage of the TC life cycle relative to the MJO propagation. The MJO propagates east, whereas TCs generally propagate west. A TC may move into an MJO active phase or trail behind it, and may be in any stage of development. This is explored by considering the mean intensities with lead time of the IBTrACS and forecast intensities separately for the different MJO paired phases (Fig. S7), for both MSLP and 10-m winds, and for both sets of MJO pairings. This shows that intensity errors and biases for each MJO phase depend strongly on the

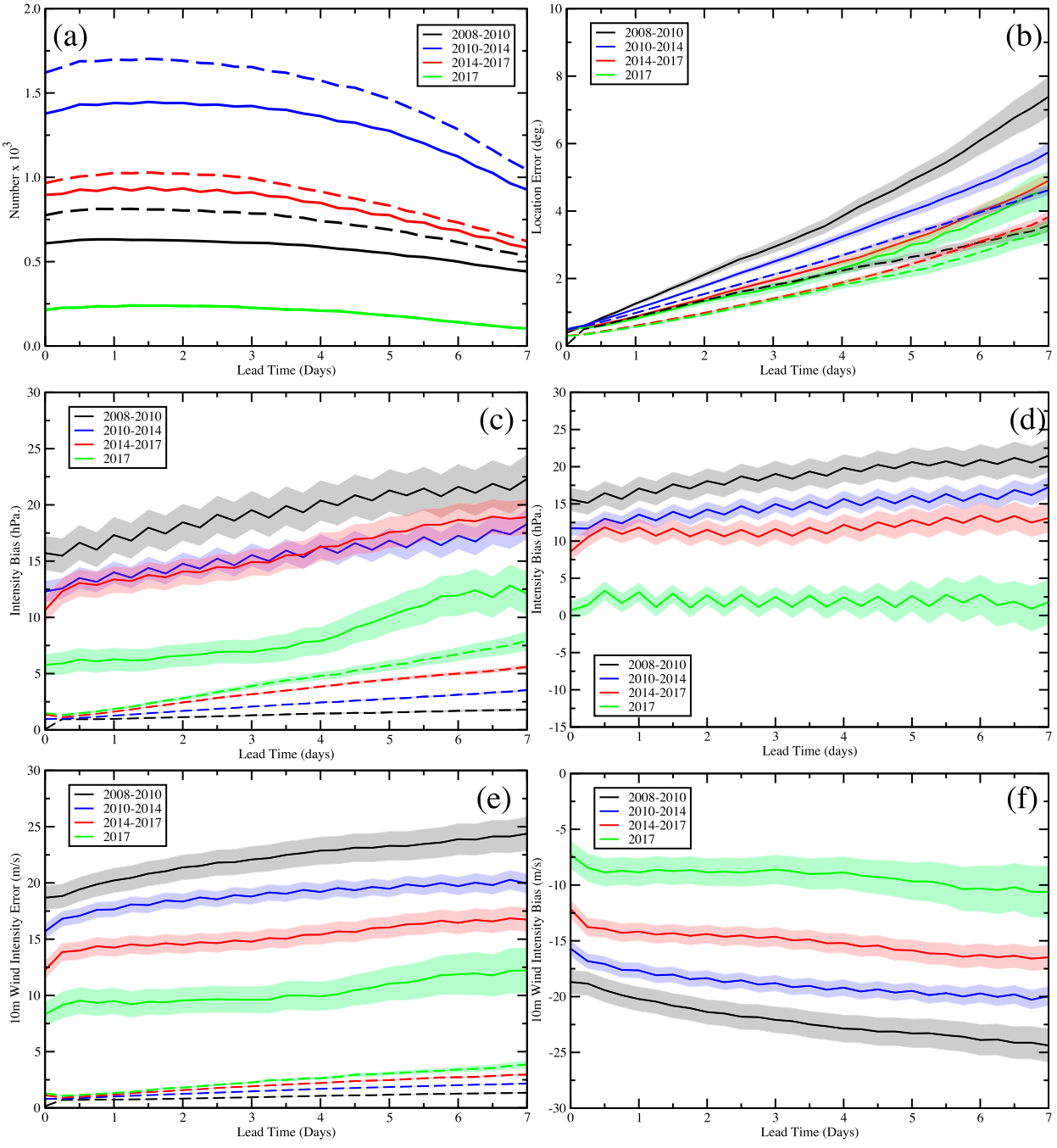

FIG. 6. As in Fig. 4, but for the ensemble mean (solid lines) and spread (dashed lines).

intensities at the start of the forecast. This is most obvious for the observations, but can also be seen in the forecasts. Contrasting the results in Figs. 7 and 8 with Fig. S7, shows that phases with lower TC intensities at the start of the forecasts often have lower errors throughout the forecast. Conversely MJO phases with higher initial intensities tend to have higher errors throughout the forecasts.

The results for the BSISO1 index for the phases 2–3, 4–5, 6–7, and 8–1 pairings are shown in Fig. 9 and for the phases 1–2, 3–4, 5–6, and 7–8) pairings in Fig. S8. The sample sizes for BSISO1 are more variable between the phases than for the MJO (Fig. 9a), with sample sizes as low as 200. For the location errors (Fig. 9b) there is little difference between the different phases, similar to that found for the MJO. However,

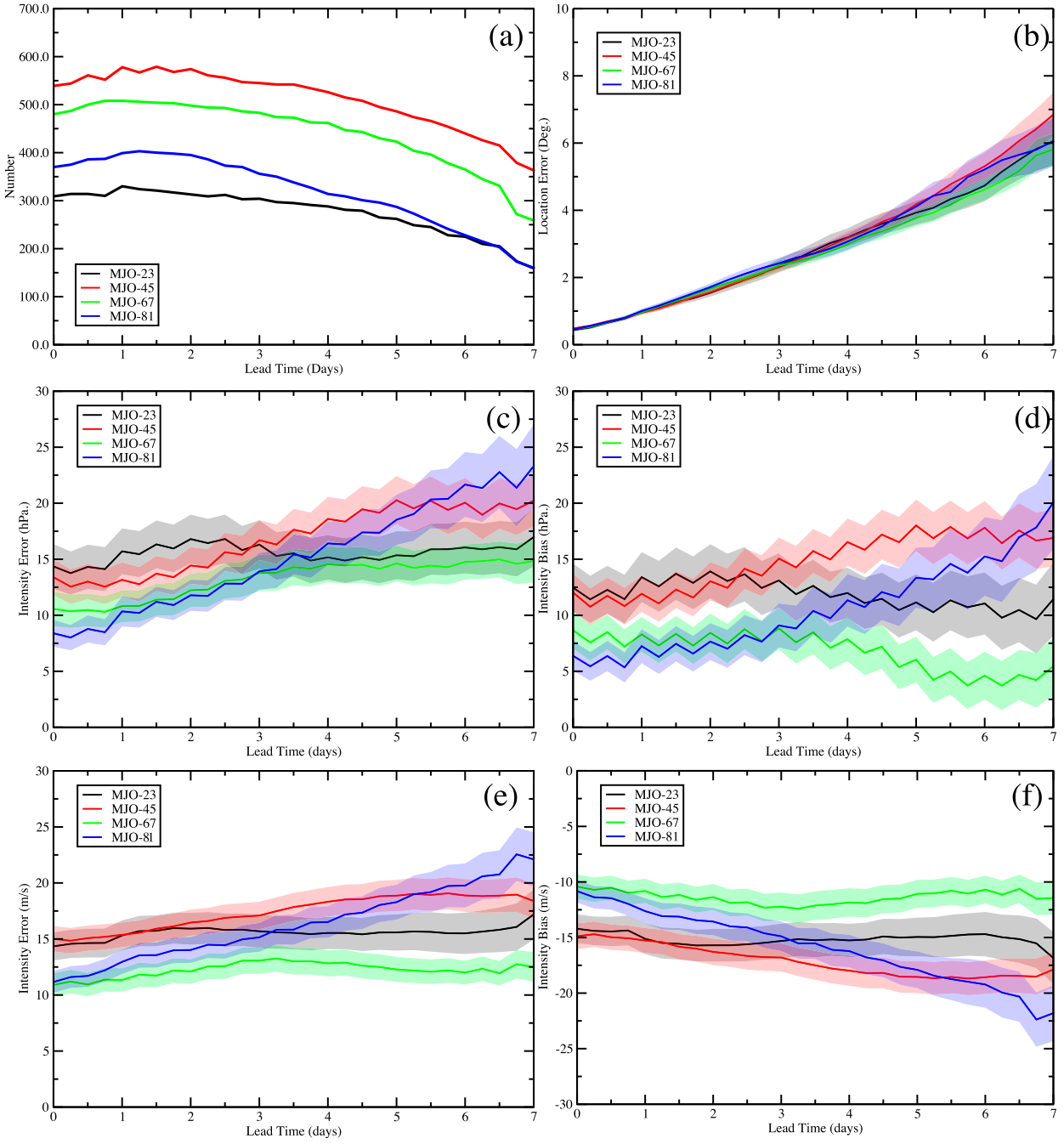

FIG. 7. Error statistics for different MJO phases (2–3, 4–5, 6–7, 8–1) for TCs in the deterministic forecasts in the WNP using best track verification: (a) sample size for each MJO paired phase, (b) location error, (c) MSLP intensity error, (d) MSLP intensity bias, (e) 10-m wind intensity error, and (f) 10-m wind intensity bias. Location error units are geodesic degrees, MSLP units are hPa, and 10-m wind units are m s^{-1}. Shading indicates the 95% confidence interval.

the intensities (Figs. 9c and 9e) show more obvious variations in errors with BSISO1 phase. In particular, for the first set of pairings the 6–7 phase has generally the lowest errors throughout the forecast range, though phases 2–3 have the lowest initial error, which then grows rapidly with lead time. However, phases 2–3 also have the lowest sample size. For the second set of phase pairings, the lowest errors occur for phases 5–6 and 7–8 throughout the forecast range. Though, phases 3–4 have the lowest initial error with a subsequent fast error growth, so that again sampling may be an issue. The BSISO1 results are

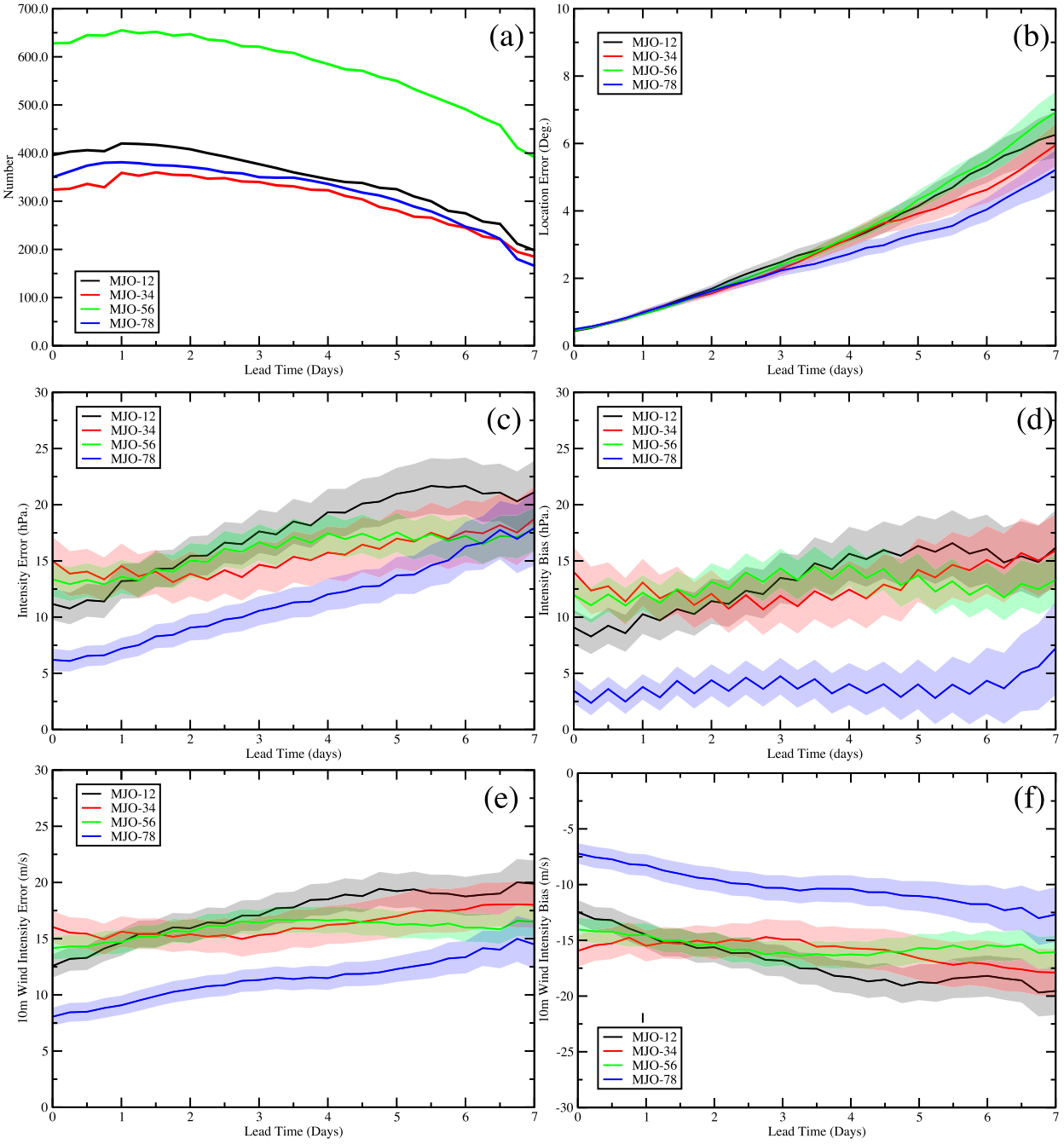

FIG. 8. Error statistics for different MJO phases (1–2, 3–4, 5–6, 7–8) for TCs in the deterministic forecasts in the WNP using best track verification: (a) sample size for each MJO paired phase, (b) location error, (c) MSLP intensity error, (d) MSLP intensity bias, (e) 10-m wind intensity error, and (f) 10-m wind intensity bias. Location error units are geodesic degrees, MSLP units are hPa, and 10-m wind units are m s^{-1}. Shading indicates the 95% confidence interval.

similar to those for the MJO, albeit with the same caveats on the sample sizes and changes in the forecasting system. The sample sizes are generally small for the BSISO1 statistics, so these results should be considered with caution until more reliable results can be obtained. Mean intensities for the BSISO1 phases with forecast lead time (Fig. S9) confirm the importance of the intensity of the storms at the start of the forecast, in particular in the observations.

To remove the dependence of the results on the changes in the forecasting system, the same analysis has been performed for periods, which minimize any

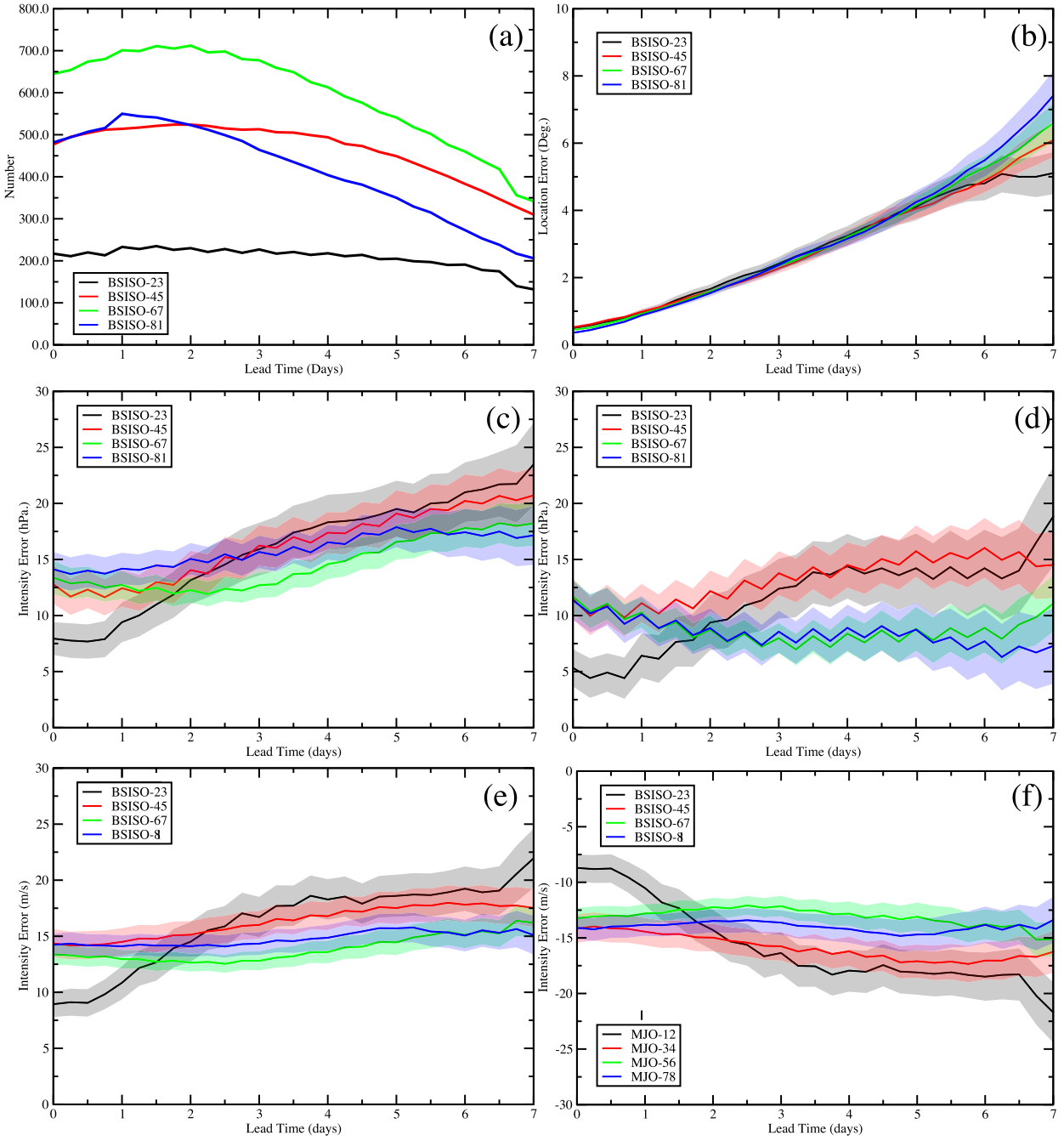

FIG. 9. As in Fig. 7, but for the BSISO1 index.

changes to the forecasting system (Table 1), in particular the 2014–17 period. The downside of this is the reduced number of forecast samples, so that the results are likely to be less reliable. Results (not shown) are found to be similar to the results discussed above for the full dataset.

A similar analysis has been performed on the ECMWF deterministic and ensemble forecasts (not shown) with similar results, lending confidence to the results for the Met Office model. Ideally, a study using high-resolution hindcasts over a longer time period, to increase sample sizes, and with a consistent forecasting system could resolve some of the uncertainties concerning the MJO and BSISO1 influence on the TC forecast errors. Current hindcast systems are at much lower resolutions than the operational systems used

here, however, and so would have larger intensity biases.

d. Comparison with other forecasting systems

To compare the UKMO forecasting system to another modern deterministic and ensemble forecasting system, the ECMWF and UKMO forecasts are compared over the same period. The configuration of the ECMWF forecasting system over 2008–12 is given in HE15. Since then further enhancements include an increase to 137 (91) vertical levels in the deterministic (ensemble) system in 2013, an increase in horizontal resolution to 9 km (18 km) in 2016, as well as changes to model physics and the data assimilation/ensemble perturbation subsystems, including the introduction of the ensemble of data assimilations (EDA) (see https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model). For a fair comparison, only verification against IBTrACS is performed, as this is independent of both forecasting systems. The comparison of the deterministic and ensemble systems for the WNP is shown in Fig. 10. Both deterministic systems have comparable sample sizes (Fig. 10a). However the ECMWF ensemble mean and spread sample sizes are larger at the longer lead times, possibly related to the larger ensemble size used by ECMWF (50 members). The initial location errors are similar. The error growth for the ECMWF system is slower than for the UKMO for both deterministic and ensemble forecasts, such that at day 4 the ECMWF system has just under a day more skill for both deterministic and ensemble systems. The error–spread relationship is also better for the ECMWF system, with smaller differences at every lead time compared to the UKMO ensemble system. Interestingly the ECMWF ensemble has less spread than that for the UKMO, even though the ECMWF system has more ensemble members (50) and both use stochastic physics, albeit with different implementations.

The initial intensity errors in the ECMWF systems, deterministic and ensemble, are much lower than for the UKMO systems (Fig. 10). The largest differences in errors occur in the earliest part of the forecasts. However, the ECMWF errors grow more rapidly, so that beyond day 5 the errors are similar to those for the UKMO for MSLP and begin to converge for 10-m winds. The larger ECMWF MSLP errors at the longer lead times may be due to the ability of the ECMWF system to predict deeper storms than in the Met Office deterministic system, which may then lead to larger errors if the timing is wrong. The smaller initial errors for ECMWF may partly be related to the relative resolutions of the two systems. The ECMWF deterministic system has been higher resolution than the UKMO system over most of the period, until 2017 when they became more comparable (~10 km). Differences in data assimilation may also have played a role. For the intensity ensemble spreads, the two systems appear to give similar results, but with large differences between the error and spread, which are marginally better for ECMWF due to the lower errors. Intensity biases are consistent with the errors.

A similar analysis has been conducted for the PAGASA-PAR region (Fig. S10). This shows similar results to the WNP region, albeit with larger intensity errors and biases, reflecting the typically larger errors found there that have already been discussed.

4. Summary and conclusions

An analysis of the forecast errors associated with TCs found in the WNP and PAGASA-PAR has been conducted for the UKMO global deterministic and ensemble forecasting systems, over an extended period of 10 years (2008–17). A summary of the main results and discussion follows:

- For location, errors are comparable for the deterministic, control and ensemble mean forecasts; the ensemble is underdispersive. For intensities, in terms of pressures and winds, there are large biases relative to observations, which are smallest for the deterministic system; the ensemble is severely underdispersive.
- The PAGASA-PAR region has larger intensity errors and biases and larger intensity ensemble spread compared with the broader WNP region.
- The forecast errors for location and intensity have reduced significantly with system upgrades over the period studied. For location there is a 1.5-day increase in skill, at day 4, between the earliest and latest NWP configuration for the deterministic system; there is a similar improvement for the ensemble system. For intensity, the error of the latest configuration at day 4 is below the initial error of all the earlier configurations for both pressure and winds.
- The MJO affects the intensity forecast errors, but does not significantly affect the location errors. Intensity errors are lower at the initiation of the forecasts in phases 6–7 and 7–8, when the MJO is active in the WNP, which can persist throughout the forecasts. Results for the BSISO1 are similar. The forecast errors depend strongly on the observed intensities in the different phases.
- Over the studied period the ECMWF deterministic and ensemble systems have lower errors and biases for both location and intensity than the UKMO forecast systems.

Global forecasting systems are improving rapidly in their ability to predict TCs, in particular for location or

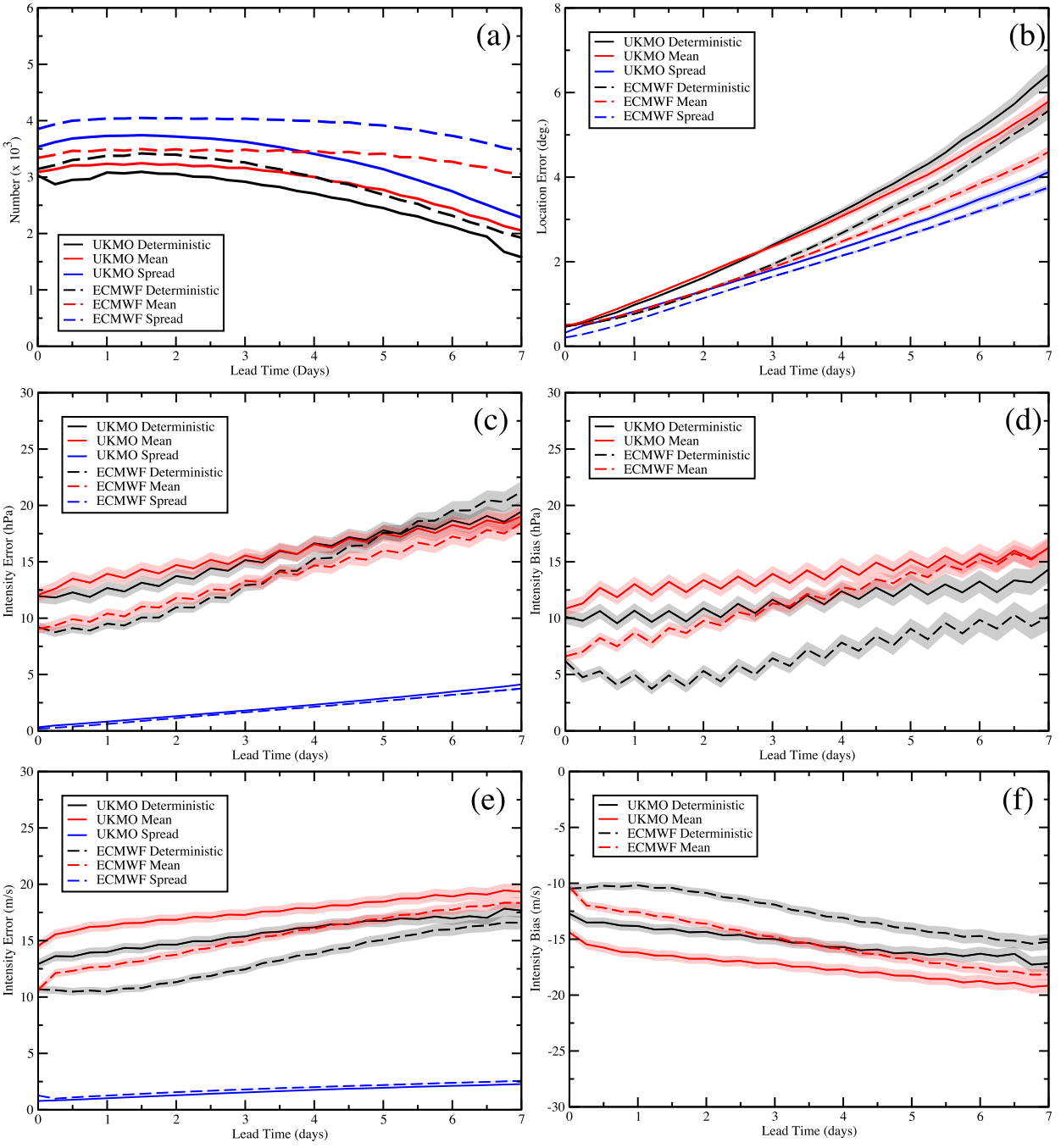

FIG. 10. Error statistics for the UKMO and ECMWF deterministic and ensemble forecasting systems for TCs in the WNP using best track verification: (a) sample size, (b) location error, (c) MSLP intensity error, (d) MSLP intensity bias, (e) 10-m wind intensity error, and (f) 10-m wind intensity bias. Location error units are geodesic degrees, MSLP units are hPa, and 10-m wind units are m s^{-1}. Shading indicates the 95% confidence interval.

path (often termed "track"), such that skill for the UKMO deterministic forecasting system in the WNP studied here has improved by ~1.5 days at a lead time of 2–4 days over the 10-yr period. Predictions of intensity have generally lagged behind those for location, though even here the UKMO forecasting system has improved, such that pressures are now much more realistic in the latest deterministic system, but winds are still underpredicted. Short and Petch (2018) highlighted that explicit convection-permitting models are required to

realistically simulate intensities and rapid intensification of TCs in the PAGASA-PAR region, but this requires resolutions < 5 km, which are still computationally prohibitive for global models. While the ensemble remains underdispersive for both location and intensity, these metrics have also improved with upgrades to the forecasting system in terms of resolution, perturbation methods and model formulation. The difference between error and spread has reduced since the earliest 2008 system, mainly due to the error reduction.

The smaller PAGASA-PAR region has similar location errors to that of the full WNP region, but larger intensity errors for both deterministic and ensemble systems as well as the ensemble spreads. The reasons for this are not clear, but the inability of the global model to capture rapid intensification may play an important role (Short and Petch 2018). This requires further evaluation.

The impact of modes of subseasonal tropical variability on the forecast errors and biases were considered. The results are complicated by the changing forecasting systems over the study period, although results for a period with a relatively fixed forecasting system (2014–17) show strong similarities with the results over the whole period, as do results for the ECMWF forecast system. Neither the MJO or BSISO appear to substantially affect the location errors, though the MJO 7–8 phase has some impact after day 4. However, intensity errors are lower at the start and in the early part of the forecast when the active phase of the MJO or BSISO1 is in the WNP (phases 6, 7, and 8). This appears to derive predominantly from the dependence of the IBTrACS intensities on the MJO phase. The effect of the MJO and BSISO on the initial forecast states is a secondary impact. This may be related to the location of TCs relative to the active MJO phase and the TC life cycle stage. Another interesting aspect is that of rapid intensification. Na et al. (2018) found that official forecast errors, issued by the operational agencies, are anticorrelated with 24-h intensity changes. A further study could examine whether rapid intensification affects the intensity errors in the different MJO and BSISO1 phases.

Further study of the dependence of TC life cycles and modes of variability is required with larger TC sample sizes to get a more robust view of their interdependence. Understanding the impact of modes of variability (e.g., MJO and ENSO on the detailed predictability of TCs) requires studies using frozen NWP forecasting systems at operational resolutions, or explicit convection-permitting resolutions, over multidecadal periods. Such datasets would provide both a good representation of TC properties and large enough sample sizes for robustness of forecast error statistics. Existing hindcasts or reforcasts that cover such periods (e.g., Hamill et al. 2013) are generally too coarse to represent observed TC intensities and are therefore not currently suitable for this type of study.

Acknowledgments. Kevin Hodges was supported by the Met Office Weather and Climate Science for Service Partnership (WCSSP) Southeast Asia, as part of the Newton Fund. Nicholas Klingaman was supported by the U.K. Natural Environment Research Council (NE/L010976/1). This work is part of the Forecasting Air–Sea Coupled Interactions in NWP for Atmospheric Tropical Extremes (FASCINATE) project. We thank the three anonymous reviewers for their helpful comments that have improved the paper.

REFERENCES

Barcikowska, M., F. Feser, and H. Von Storch, 2012: Usability of best track data in climate statistics in the western North Pacific. *Mon. Wea. Rev.*, **140**, 2818–2830, https://doi.org/10.1175/MWR-D-11-00175.1.

Bender, M. A., T. P. Marchok, C. R. Sampson, J. A. Knaff, and M. J. Morin, 2017: Impact of storm size on prediction of storm track and intensity using the 2016 operational GFDL Hurricane Model. *Wea. Forecasting*, **32**, 1491–1508, https://doi.org/10.1175/WAF-D-16-0220.1.

Bengtsson, L., K. I. Hodges, M. Esch, N. Keenlyside, L. Kornblueh, J.-J. Luo, and T. Yamagata, 2007: How may tropical cyclones change in a warmer climate? *Tellus*, **59A**, 539–561, https://doi.org/10.1111/j.1600-0870.2007.00251.x.

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, https://doi.org/10.1002/qj.234.

——, ——, S. E. Beare, K. R. Mylne, and G. J. Shutts, 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **135**, 767–776, https://doi.org/10.1002/qj.394.

Brown, D. B., J. L. Franklin, and C. Landsea, 2008: A fresh look at tropical cyclone pressure-wind relationships using recent reconnaissance based "best-track" data (1998-2005). *27th Conf. on Hurricanes and Tropical Meteorology*, Orlando, FL, Amer. Meteor. Soc., 3B.5, https://ams.confex.com/ams/27Hurricanes/techprogram/paper_107190.htm.

Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99–119, https://doi.org/10.1175/1520-0493(1997)125<0099:PFSOEP>2.0.CO;2.

Camp, J., M. Roberts, C. MacLachlan, E. Wallace, L. Hermanson, A. Brookshaw, A. Arribas, and A. A. Scaife, 2015: Seasonal forecasting of tropical storms using the Met Office Glosea5 seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 2206–2219, https://doi.org/10.1002/qj.2516.

Cinco, T. A., and Coauthors, 2016: Observed trends and impacts of tropical cyclones in the Philippines. *Int. J. Climatol.*, **36**, 4638–4650, https://doi.org/10.1002/joc.4659.

Dai, A., and J. Wang, 1999: Diurnal and semidiurnal tides in global surface pressure fields. *J. Atmos. Sci.*, **56**, 3874–3891, https://doi.org/10.1175/1520-0469(1999)056<3874:DASTIG>2.0.CO;2.

Davies, T., M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N. Wood, 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **131**, 1759–1782, https://doi.org/10.1256/qj.04.101.

DeMaria, M., C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, https://doi.org/10.1175/BAMS-D-12-00240.1.

Froude, L. S. R., L. Bengtsson, and K. I. Hodges, 2007a: The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. *Mon. Wea. Rev.*, **135**, 315–333, https://doi.org/10.1175/MWR3274.1.

——, ——, and ——, 2007b: The prediction of extratropical storm tracks by the ECMWF and NCEP ensemble prediction systems. *Mon. Wea. Rev.*, **135**, 2545–2567, https://doi.org/10.1175/MWR3422.1.

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAAs second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, https://doi.org/10.1175/BAMS-D-12-00014.1.

Harper, B., J. Kepert, and J. Ginger, 2010: Guidelines for converting between various wind averaging periods in tropical cyclone conditions. WMO/TD-1555, World Meteorological Organization, 64 pp., https://www.wmo.int/pages/prog/www/tcp/documents/WMO_TD_1555_en.pdf.

Heming, J. T., 2016: Met Office Unified Model tropical cyclone performance following major changes to the initialization scheme and a model upgrade. *Wea. Forecasting*, **31**, 1433–1449, https://doi.org/10.1175/WAF-D-16-0040.1.

Hodges, K. I., 1994: A general-method for tracking analysis and its application to meteorological data. *Mon. Wea. Rev.*, **122**, 2573–2586, https://doi.org/10.1175/1520-0493(1994)122<2573:AGMFTA>2.0.CO;2.

——, 1995: Feature tracking on the unit sphere. *Mon. Wea. Rev.*, **123**, 3458–3465, https://doi.org/10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2.

——, 1999: Adaptive constraints for feature tracking. *Mon. Wea. Rev.*, **127**, 1362–1373, https://doi.org/10.1175/1520-0493(1999)127<1362:ACFFT>2.0.CO;2.

——, and R. Emerton, 2015: The prediction of Northern Hemisphere tropical cyclone extended life cycles by the ECMWF ensemble and deterministic prediction systems. Part I: Tropical cyclone stage. *Mon. Wea. Rev.*, **143**, 5091–5114, https://doi.org/10.1175/MWR-D-13-00385.1.

——, R. W. Lee, and L. Bengtsson, 2011: A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *J. Climate*, **24**, 4888–4906, https://doi.org/10.1175/2011JCLI4097.1.

——, A. Cobb, and P. L. Vidale, 2017: How well are tropical cyclones represented in reanalysis datasets? *J. Climate*, **30**, 5243–5264, https://doi.org/10.1175/JCLI-D-16-0557.1.

Kim, H.-M., P. J. Webster, and J. A. Curry, 2011: Modulation of North Pacific tropical cyclone activity by three phases of ENSO. *J. Climate*, **24**, 1839–1849, https://doi.org/10.1175/2010JCLI3939.1.

Klotzbach, P. J., and E. C. J. Oliver, 2015: Variations in global tropical cyclone activity and the Madden–Julian Oscillation since the midtwentieth century. *Geophys. Res. Lett.*, **42**, 4199–4207, https://doi.org/10.1002/2015GL063966.

Knaff, J. A., and R. M. Zehr, 2007: Reexamination of tropical cyclone wind–pressure relationships. *Wea. Forecasting*, **22**, 71–88, https://doi.org/10.1175/WAF965.1.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS) unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

Lee, C.-Y., S. J. Camargo, F. Vitart, A. H. Sobel, and M. K. Tippett, 2018: Subseasonal tropical cyclone genesis prediction and MJO in the S2S dataset. *Wea. Forecasting*, **33**, 967–988, https://doi.org/10.1175/WAF-D-17-0165.1.

Lee, J.-Y., B. Wang, M. C. Wheeler, X. Fu, D. E. Waliser, and I.-S. Kang, 2013: Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. *Climate Dyn.*, **40**, 493–509, https://doi.org/10.1007/s00382-012-1544-4.

Manganello, J., and Coauthors, 2012: Tropical cyclone climatology in a 10-km global atmospheric GCM: Toward weather-resolving climate modeling. *J. Climate*, **25**, 3867–3893, https://doi.org/10.1175/JCLI-D-11-00346.1.

Mogensen, K. S., L. Magnusson, and J.-R. Bidlot, 2017: Tropical cyclone sensitivity to ocean coupling in the ECMWF coupled model. *J. Geophys. Res. Oceans*, **122**, 4392–4412, https://doi.org/10.1002/2017JC012753.

Na, W., J. L. McBride, X.-H. Zhang, and Y.-H. Duan, 2018: Understanding biases in tropical cyclone intensity forecast error. *Wea. Forecasting*, **33**, 129–138, https://doi.org/10.1175/WAF-D-17-0106.1.

R project, 2013: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, accessed 24 May 2019, http://www.R-project.org/.

Ren, F., J. Liang, G. Wu, W. Dong, and X. Yang, 2011: Reliability analysis of climate change of tropical cyclone activity over the western North Pacific. *J. Climate*, **24**, 5887–5898, https://doi.org/10.1175/2011JCLI3996.1.

Roberts, M. J., and Coauthors, 2015: Tropical cyclones in the UPSCALE ensemble of high-resolution global climate models. *J. Climate*, **28**, 574–596, https://doi.org/10.1175/JCLI-D-14-00131.1.

Short, C. J., and J. Petch, 2018: How well can the Met Office Unified Model forecast tropical cyclones in the western North Pacific? *Wea. Forecasting*, **33**, 185–201, https://doi.org/10.1175/WAF-D-17-0069.1.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, https://doi.org/10.1256/qj.04.106.

Strachan, J., P. L. Vidale, K. Hodges, M. Roberts, and M.-E. Demory, 2013: Investigating global tropical cyclone activity with a hierarchy of AGCMs: The role of model resolution. *J. Climate*, **26**, 133–152, https://doi.org/10.1175/JCLI-D-12-00012.1.

Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. *Wea. Forecasting*, **27**, 715–729, https://doi.org/10.1175/WAF-D-11-00085.1.

Vitart, F., 2009: Impact of the Madden Julian Oscillation on tropical storms and risk of landfall in the ECMWF forecast system. *Geophys. Res. Lett.*, **36**, L15802, https://doi.org/10.1029/2009GL039089.

Vukicevic, T., E. Uhlhorn, P. Reasor, and B. Klotz, 2014: A novel multiscale intensity metric for evaluation of tropical cyclone intensity forecasts. *J. Atmos. Sci.*, **71**, 1292–1304, https://doi.org/10.1175/JAS-D-13-0153.1.

Walters, D., and Coauthors, 2017: The Met Office Unified Model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geosci. Model Dev.*, **10**, 1487–1520, https://doi.org/10.5194/gmd-10-1487-2017.

Wheeler, M., and H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Wu, L., B. Wang, and S. Geng, 2005: Growing typhoon influence on East Asia. *Geophys. Res. Lett.*, **32**, L18703, https://doi.org/10.1029/2005GL022937.

——, Z. Wen, R. Huang, and R. Wu, 2012: Possible linkage between the monsoon trough variability and the tropical cyclone activity over the western North Pacific. *Mon. Wea. Rev.*, **140**, 140–150, https://doi.org/10.1175/MWR-D-11-00078.1.

Yamaguchi, M., J. Ishida, H. Sato, and M. Nakagawa, 2017: WGNE intercomparison of tropical cyclone forecasts by operational NWP models: A quarter century and beyond. *Bull. Amer. Meteor. Soc.*, **98**, 2337–2349, https://doi.org/10.1175/BAMS-D-16-0133.1.

Yoshida, R., Y. Kajikawa, and H. Ishikawa, 2014: Impact of boreal summer intraseasonal oscillation on environment of tropical cyclone genesis over the Western North Pacific. *SOLA*, **10**, 15–18.

Zhang, C., 2005: Madden–Julian Oscillation. *Rev. Geophys.*, **43**, RG2003, https://doi.org/10.1029/2004RG000158.

——, and M. Dong, 2004: Seasonality in the Madden–Julian oscillation. *J. Climate*, **17**, 3169–3180, https://doi.org/10.1175/1520-0442(2004)017<3169:SITMO>2.0.CO;2.