RMetS
Royal Meteorological Society

Tropical cyclone track analog ensemble forecasting in the extended Australian basin: NWP combinations

Hans Langmack, Klaus Fraedrich and Frank Sielmann*
Meteorologisches Institut, Universität Hamburg, Germany

*Correspondence to: F. Sielmann, University of Hamburg, Meteorological Institute, Grindelberg 5, Hamburg 20144 Germany. E-mail: frank.sielmann@zmaw.de

Ensemble-Weighted Analogs (EWA) are introduced as an empirical forecast model for tropical cyclone (TC) tracks in the Australian basin extending to the Indian Ocean. The input requires only the 6 h positions of the preceding 24 hours to estimate the actual TC's future locations (up to 48 hours ahead) using optimally weighted ensemble members (analogs) from sections of historic cyclone tracks. They are determined as nearest neighbours of the actual track positions. After model calibration, independent forecasts are evaluated (2001–2007). The forecast analysis shows an error of about 140 km (for the 24 h prediction), which is better than the CLIPER reference forecast (161 km) and comparable to a numerical weather-prediction model (NWP of the UK Met Office). Intensity estimates and the forecast error-spread relation are discussed. Error-minimizing forecast combinations are also analysed: the EWA–NWP combination improves the individual 24 h NWP forecasts by about 25 km. Track cluster conditioned 24 h EWA forecasts are introduced providing probability estimates of hazardous areas. Finally, two cases of sudden track changes and their forecasts are presented with fields of position probabilities derived from the analog ensemble members and their clusters.

Key Words: metric adaption; ensemble forecast; error estimation; combination forecast; short-term prediction

Received 8 July 2011; Revised 16 January 2012; Accepted 6 February 2012; Published online in Wiley Online Library 6 March 2012

Citation: Langmack H, Fraedrich K, Sielmann F. 2012. Tropical cyclone track analog ensemble forecasting in the extended Australian basin: NWP combinations. *Q. J. R. Meteorol. Soc.* **138**: 1828–1838. DOI:10.1002/qj.1915

1. Introduction

In the last decade numerical weather-prediction (NWP) models have substantially improved the forecast accuracy of tropical cyclones tracks, while advances of intensity forecasts have been modest (for the west Pacific region, see Knaff *et al.*, 2003). The overall improvement is due to better models, assimilation techniques, and ensemble forecasts reducing the influence of initial uncertainty (Zhang and Krishnamurti, 1997). Thus empirical models, the majority of which are of the climatology–persistence type (for example, CLIPER: Neumann, 1981) and well suited for easy-to-implement applications, appear to fall out of favour. But, as demonstrated first by Leslie and Fraedrich (1990), it is the combination of independent forecast schemes, that is, dynamical (NWP) and statistical–empirical forecasts, which may still substantially improve forecast performance.

Empirical weather forecast schemes for tropical cyclone tracks are based on linear (Markov-type) approaches (Leslie *et al.*, 1992) or nonlinear schemes like analog methods, which provide weighted ensemble mean forecasts. As the latter have also gained theoretical interest analysing weather predictability based on physical (Lorenz, 1969) or time-delay coordinates (Fraedrich and Leslie, 1989), they appear to be particularly suitable for many kinds of weather- or climate-related prediction issues. Such empirical models using analog search are employed for tropical storm-track forecasting (Hope and Neumann, 1970; Neumann, 1972; HURRAN (hurricane analog techniques) and CLIPER),

and have been extended to chaotic nonlinear time series methods employing a self-adapting metric (Fraedrich and Rückert, 1998) which, after further improvements by statistical analysis techniques, were applied to tropical cyclone track forecasts (Bessafi *et al.*, 2002; Fraedrich *et al.*, 2003). Corresponding models to forecast tropical cyclone intensity based on time series are SHIFOR and SHIP (Jarvinen and Neumann, 1979; Knaff *et al.*, 2003); these models are limited to intensity forecasts on open sea, and landfall events leading to sudden intensity changes should be excluded.

The purpose of this paper is threefold: (i) an analog forecast method based on metric adaption is extended to an ensemble-weighted scheme by jointly optimising the state space metric (providing the past analogs) *and* the weights of the ensemble members (generating estimates of the future states and forecast error); (ii) an error-minimizing combination with the cyclone track predictions issued by the UK Met Office (UKMO) NWP model is used to obtain an improvement of tropical cyclone track forecasts and to compare the NWP predictions with a combination of two empirical schemes (weighted analog ensemble and the classical CLIPER); and (iii) the analysis focuses on the extended Australian cyclone basin, as it appears that for this region the forecasts have not reached the high level of accuracy as for other tropical storm basins, for example the North Atlantic. The outline is as follows: in section 2, database, model building, model calibration and method of forecasting are described. Section 3 evaluates forecast accuracy of position, intensity (surface pressure) and of the ensemble predicted forecast error (spread). Finally, in section 4, the empirical forecasts are further optimised by combining them with the numerical weather predictions and another empirical forecast scheme. Finally, areas spanned by clusters of the ensemble forecast tracks are compared to those of the numerical weather predictions to analyse the performance for hazardous areas. Section 5 summarizes the results.

2. Ensemble-weighted analog (EWA) forecasts: model building

Analog models can be successfully applied to arbitrary nonlinear dynamical systems represented by a long time series (Takens, 1981) where, for practical purposes, embedology (Sauer *et al.*, 1991) has been employed spanning a state space of time-delay coordinates. While adaption of the state space metric optimises utilization of the observed history (Fraedrich and Rückert, 1998), the weighting of forecast ensemble members allows further optimisation for prediction. That is, Ensemble-Weighted Analogs (EWA) combine these two features of utilizing the past (by metric adaption) *and* optimising the future (by ensemble weighting).

Empirical prediction models require a forecast function constructed by the ensemble members, a metric structure leading to weighted contribution of each ensemble member, and a cost function to perform the optimisation procedure. This leads to optimal model parameters for the state space metric and number and ranking of ensemble members. A short outline of the approach follows in four steps.

Step 1 (Dataset): Tropical cyclone tracks are embedded in a state space whose positions are given by time-dependent latitudes and longitudes. The occasionally supplemented central pressure and/or wind speed values are used only to attach intensity information *a posteriori* to the forecast ensemble selected. Transformation of the geographical tracks to relative ones involves normalization with the cosine of latitude. The data are binned into the following subsets: (i) learning set; (ii) independent forecast and NWP–combination set; and (iii) the final test set.

Step 2 (Model building): The forecast function (prediction system) is the core of the ensemble weighted analog scheme, which is calibrated in the learning period. A cost function (error in terms of tropical cyclone state space variables) is minimized comparing forecasts with the associated realizations to provide a state space metric for the forecast function (utilizing the past) and to determine the weights of the ensemble members (optimising the future ensemble mean).

Step 3 (Forecast and verification): Forecast performance is analysed in terms of ensemble mean errors in position and intensity, and error-spread relations for various lead times.

Step 4 (Combination forecasts, hazardous areas and case-studies): In section 4 we propose a method to combine this model with NWP forecasts for error reduction. Hazardous areas are marked by spatial ensemble track distributions, and case-studies illustrate the application.

2.1. Dataset

Tropical cyclone (TC) best track data contain observations of centre positions. The central surface pressures, which reveal approximately 50% missing values (Australian Bureau of Meteorology (BoM), 1956 to 2007) are partially completed by data from Université de la Réunion (Figure 1). The mean lifetime of the total of 553 constituted TCs is about 151 h with a spread of 37 h, where TCs of less than 24 h life span are excluded. For model calibration we use the data up to 1991; the NWP combination (regression) is based on independent forecasts from 1992 to 2000 (104 cyclones with 174 observations), and the period 2001–2007 (64 cyclones with 378 six-hourly observations) serves as an independent test period (Figure 2). Using a learning dataset consisting of randomly sampled cyclone tracks, we utilize data of different observing technologies for model building. Note that the combination and test sets contain data based on recent technology (e.g. geostationary satellites). The deterministic NWP system of UKMO (global grid-point model solving the primitive equations at about 40 km resolution) provides forecasts whose daily verification statistics (Staniforth *et al.*, 2004), published on a monthly basis, is used for lead times up to 144 h. The tropical cyclone state space used here comprises (i) the cosine of latitude and (ii) the subsequent differences from the preceding value: latitude, longitude and central surface pressure. Missing position values and central pressure are linearly interpolated. Note that central pressure is not used in the cost-function analysis.

2.2. Model building: forecast function and cost-function optimisation

The forecast function $F_n(t, j)$ is the backbone of the ensemble-weighted analog forecasts (EWA). It provides the ranking of the weighted mean of the ensemble members (that is the historical analogs selected from the dataset) and, commencing at time t, it predicts the jth time step of each state space vector y_n consisting of $n = 1, \ldots, N$

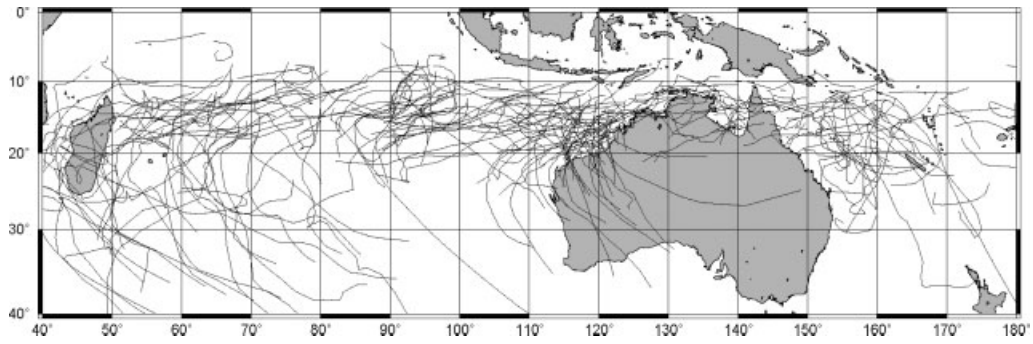

Figure 1. Best tracks (lines) in the extended Australian basin.

components.

$$F_n(t,j) = \sum_{k=1}^{K} r_k^2 \cdot y_n(t_k + j) / \sum_{k=1}^{K} r_k^2(t, t_k) \qquad (2.1)$$

Segments of historical tropical storm tracks represent vectors (points) in state space, whose (embedding) dimension is given by the number of variables, which includes their temporal evolution. That is, the state vector y_n comprises the state space variables (cosine of latitude, differences of latitude and longitude); the central pressure information does not enter as predictor but is used to supplement the forecasts in the output. Note that the Julian date, which represents seasonality, is not included in the state space, because no performance improvement was achieved. The time t_k denotes the start of the kth analog (reference time t). The ensemble size K (of the $k = 1, \ldots, K$ ensemble members) describes the analogs in terms of nearest neighbours of the base point, which is defined by the actual cyclone track segment. Finally, the ranking function r_k (Eq. (2.2)) characterizes the relevance of the kth ensemble member relative to the discarded $(K+1)$th analog or neighbour in terms of the weights, d_k and d_{K+1}:

$$r_k^2(t, t_k) = 1 - \frac{d_k^2(t, t_k)}{d_{K+1}^2(t, t_k)}. \qquad (2.2)$$

Thus, for example, extending the time segment of an analog state vector (which is the nearest neighbour to the base point) provides the forecast of the ensemble member of highest rank r_k. Note that the Euclidean distance metric, d_k, depends on weights λ_i:

$$d_k(t, t_k) = \sum_{n=1}^{N} \sum_{m=1}^{M} \lambda_l^2 \cdot \{y_n(t - m) - y_n(t_k - m)\}^2, \quad (2.3)$$

which are associated with squared differences between K pairs of track segments of length M (embedding dimension). A pair consists of (i) the actual TC track segment up to forecast step t, which represents the base point, and (ii) its kth neighbour representing a historic track segment commencing at t_k. The number of $n = 1, \ldots, N$ state space variables (state vector) and the time segment length (embedding) $m = 1, \ldots, M$ are preset; the total number of $k = 1, \ldots, K$ neighbours (ensemble members) determines ranking of the nearest neighbours in terms of λ_i -weights, indexed by $l = m + M \cdot (n - l)$.

The cost function (squared error) is used for optimisation as follows: metric weights λ_i and thus the ensemble member rankings r_k are determined by optimising a cost function $C(\lambda, r_k)$, which measures the forecast error in terms of the squared distance (error) between the weighted ensemble mean forecast set $F_n(t_i, j)$ up to lead time j (with $j = 1, \ldots, J$ prediction steps) and the respective verifying observations $y(t_i + j)$, all commencing at t_i; the errors are averaged over $i = 1, \ldots, I$ forecasts trials and $j = 1, \ldots, J$ lead times in 6 h steps:

$$C = \frac{1}{N \cdot I \cdot J} \cdot \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{j=1}^{J} |y_n(t_i + j) - F_n(t_i, j)|^2. \quad (2.4)$$

Now, the metric weights λ_l and rankings r_k of nearest neighbours can be optimised by minimizing the cost function (Eq. (2.4)) as described next.

The cost function is determined by the track forecast error defined by the great-circle distance d (Eq. (2.5)) between a cyclone's forecast position $P_2(\lambda_2, \phi_2)$ and its best-track position $P_1(\lambda_1, \phi_1)$ at the forecast verification time:

$$d(P_1, P_2) = \frac{R_{\text{Earth}} \cdot \pi}{180} \cdot \arccos\{\sin(\varphi_1) \cdot \sin(\varphi_2) + \cos(\varphi_1) \cdot \cos(\varphi_2) \cdot \cos(\lambda_1 - \lambda_2)\}. \qquad (2.5)$$

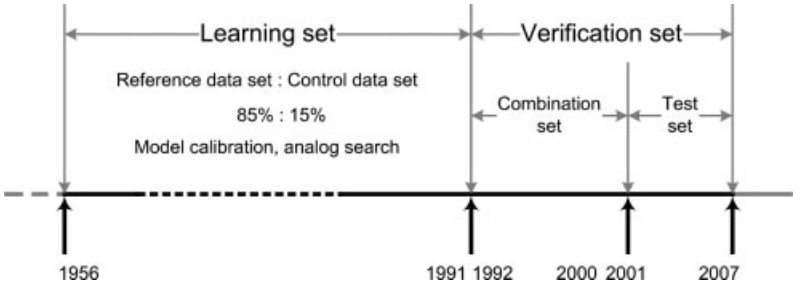

Figure 2. Data structure schematic.

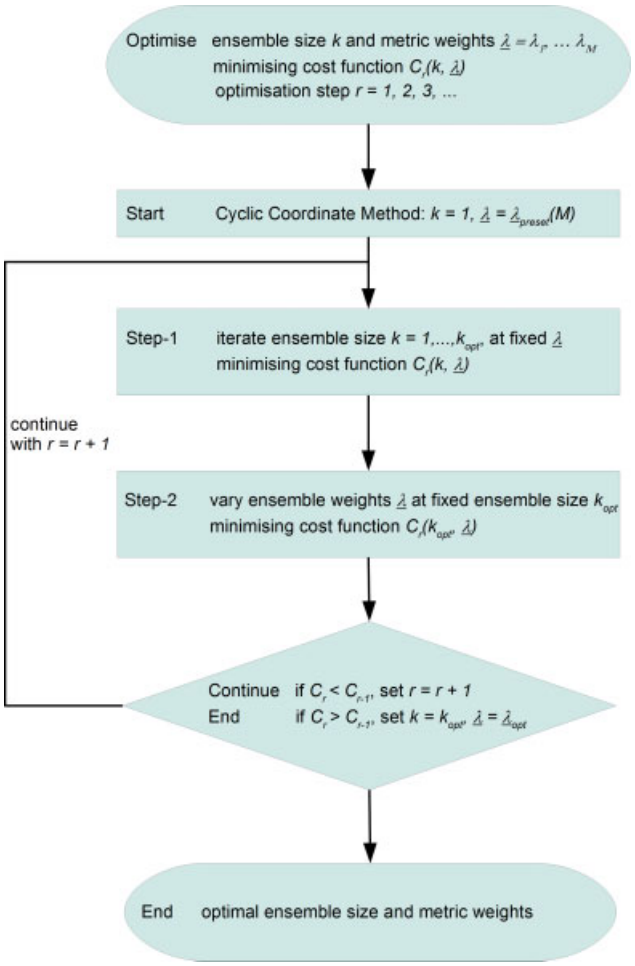

Figure 3. Flow chart of Cyclic Coordinate Method (CCM) for minimizing cost function with initial metric weights $\underline{\lambda}_{\text{preset}}$ and ensemble members k. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

The positions are $\lambda_i(t+j) = \lambda_i(t+j-1) + \Delta\lambda_i(t+j)$ and $\varphi_i(t+j) = \varphi_i(t+j-1) + \Delta\varphi_i(t+j)$.

A gradient method is applied to optimise the cost function C (Eq. (2.4)). The parallel tangent PARTAN gradient algorithm is used here (Bazaraa *et al.*, 1993, pp 334–340) in differentiating the cost function with respect to the distance weights. PARTAN is less sensitive to local minima than, for example, the Newton–Raphson method. The optimisation starts with an initial number of ensemble members k, and exponentially preset metric weights. This yields optimal metric weights and ranking of the ensemble members, which enter the actual ensemble forecasts. Iterative repetition of this procedure leads to the cyclic coordinate method (CCM) which optimises the (integer-valued) ensemble size K, that is defined as the maximum ensemble size, i.e. the used ensemble members. Here a generalization is applied combining cyclic optimisation with gradient-based optimisation (McNames, 2002). Given initial metric weights λ_{Start}, the number of ensemble members k starting with $k_{\text{Start}} = 2$ is increased to obtain the related optimum ensemble size K (at minimum cost function) with new metric weights λ. CCM is repeated with new metric weights until a quasi-saturation level is reached, i.e. the difference of C and C_{old} is less than min=0.01%, and the cost function approaches a minimum or an asymptotic value at final ensemble size K. A flowchart illustrates the procedure (see Figure 3).

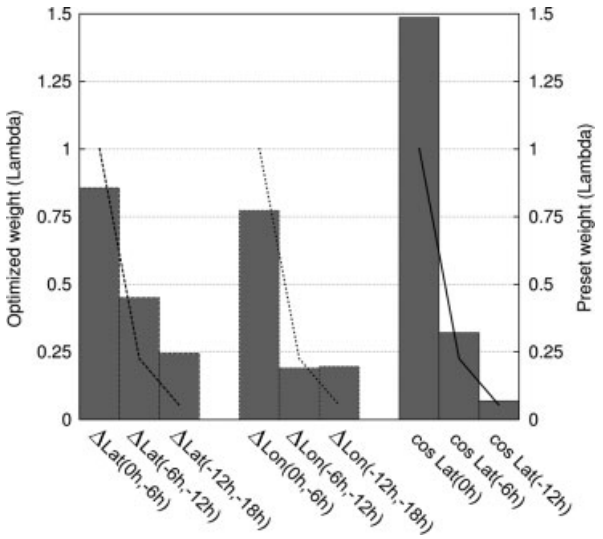

Figure 4. Metric weights of the the Ensemble-Weighted Analog model (EWA, 1956–1993) for the extended Australian basin: optimized (bars) and initially preset weights (lines).

The cost function C is optimised within a control period (1956–1993) with 321 tropical cyclones; 85% (reference dataset, Figure 2) provide the pool of analogs available for optimising the ensemble weights based on errors obtained by predicting the remaining 15% of the control period. After calibration within the control period, independent model predictions are performed using the control dataset as analog pool and the verification dataset (1994–2007) with 135 TCs (in section 3) for performance analysis. The metric weights (Figure 4) determined by the optimisation show only a weak deviation from the initially preset exponential distribution: the weight of the latest actual position (cosine of latitude at lead time 0 h) is dominant, indicating the influence of the mean large-scale atmospheric state at that latitude on the performance. The weights of latitude differences (between actual and preceding positions) are initially larger than those of longitude differences, also indicating the memory of the large-scale mean zonal flow. It must be emphasized that the gradient algorithm is sensitive to local minima, where a compromise between iteration number and a convergence criterion ends the minima search. The final metric weights and the maximum ensemble size, $K = 150$, are fixed for cost-function evaluation and independent forecasting (section 3). Larger ensembles did not lead to better results (due to overestimation). Note that the learning set contains many more historical tracks so there is no lack of analogs.

In summarizing, Figure 4 demonstrates the 12-dimensional vector spanning the tropical cyclone track phase space used for ensemble-weighted analog forecasts. It consists of a time-window of 0, 6, 12 and 18 h commencing with cos(lat) at 0 h, followed by the latitude and longitude differences, delta(lat) and delta(lon), between subsequent track locations at 0 to 6, 6 to 12 and 12 to 18 h in order to characterize the 6-hourly drifts of the tropical cyclone. Using cos(lat) instead of latitude accounts for the bias induced by the Coriolis force affecting the cyclone position, which is supported by optimisation tests (not shown). Likewise the initial longitude dependence (at 0 h) in the Australian basin did not lead to improved forecast accuracy. Note that all results have been obtained by metric weight optimisation.

Table 1. Tropical cyclone arithmetic mean forecast error (km), in the extended Australian basin and the Indian Ocean, changing with forecast lead time (hour) for the EWA model based on best track data with optimized parameters (1956–1993).

Lead time (*h*)	12	24	36	48
CLIPER	–	161	–	360
EWA	65	144	243	354
UKMO NWP	–	139	–	255
Combination: EWA + UKMO NWP	–	118 (116)	–	270 (262)
Combination: EWA + CLIPER	–	139 (139)	–	344 (346)
Combination: UKMO NWP + CLIPER	–	117 (117)	–	245 (245)
Combination: UKMO NWP, CLIPER + EWA	–	116 (114)	–	269 (260)

Reference forecasts of CLIPER and NWP are both provided by UKMO, combination forecasts of EWA + UKMO, EWA + CLIPER, UKMO + CLIPER and UKMO + CLIPER + EWA (values in brackets indicate the values for a quadratic regression), all over 2001–2007 at the same TC days forecasts.

3. Cyclone track ensemble forecasts: position error and spread

Independent cyclone track forecasts are evaluated in the extended Australian basin (including the Indian Ocean) using the optimised parameter set (see section 2). First, EWA performance is compared with the best track and compared with the UKMO NWP and CLIPER forecasts provided by the UKMO for 1994 to 2007 including a season-to-season varying verification climatology. Finally, the ensemble spread is determined to estimate the model's error-forecast capability.

3.1. Position error

The 7-year best-track forecast statistics of great-circle errors (Eq. (2.5)) are based on data provided by UKMO. They contain the NWP and CLIPER great-circle errors of forecasts starting at 0000 UTC and 1200 UTC for 24 h lead time as well as those from corresponding EWA forecasts calculated from Australian BoM best track positions.

3.1.1. Model comparison

Track forecast evaluations are presented in Figure 5 and supplemented by Table 1. EWA great-circle errors (378 forecasts in 2001–2007) for lead times up to 48 h are calculated in 6 h time steps, while UKMO great-circle errors are available in steps of 24 h (Figure 5). The following results are noted for single model predictions: (i) EWA track errors grow with increasing lead time from 144 to 354 km for 24 to 48 h; (ii) EWA track errors lie between UKMO NWP with 139 and 255 km at 24 and 48 h, compared to CLIPER growing from 161 to 360 km for 24 to 48 h lead time, respectively; (iii) EWA generally performs better or equal to UKMO NWP at lead times shorter than 24 h and approaches CLIPER errors beyond 48 h lead time; and (iv) UKMO NWP and EWA are always better than CLIPER forecasts up to 48 h. Note that EWA is optimised using one-day track segments for data assimilation prior to forecasting one and two days ahead. EWA forecasts with lead times between 48 and 72 h (and longer) attain errors which are of similar magnitude to those of regression-based statistical models like CLIPER. Therefore, we do not extend EWA forecasts beyond 48 h.

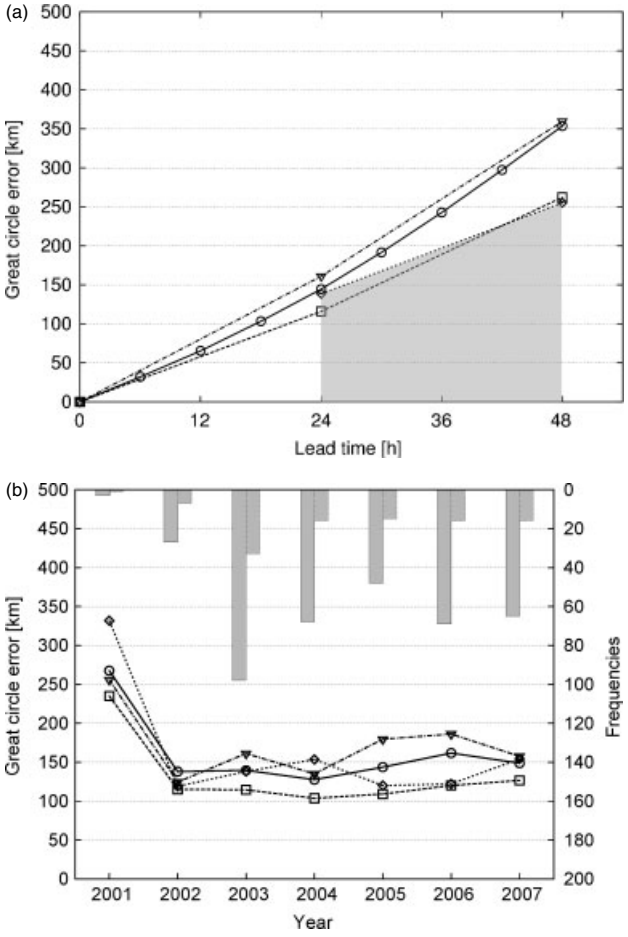

Figure 5. Australian tropical cyclone position forecast errors 2001–2007 (compared to the best track data, in km): (a) Mean great-circle errors changing with lead time (hours) are shown for the Ensemble-Weighted Analog (EWA) prediction model (solid line with circles; 6 h time step), the Australian CLIPER forecasts (dashed-dotted line with triangles; 24 h time step), and the UKMO NWP forecasts (shaded area below dotted line with diamonds). The error-minimizing combination EWA–UKMO forecast performance is also included (dashed line with squares; 24 h time step, see section 4). (b) Australian tropical cyclone position error climatology 2001–2007 (for 24 h lead time, in km) for the EWA prediction model (solid line with circles), the Australian CLIPER forecasts (dashed-dotted line with triangles), the UKMO NWP forecasts (dotted line with diamonds) and the error-minimizing combination EWA–UKMO (dashed line with squares). The seasonal forecast frequencies and storm counts (left and right columns) are also shown.

3.1.2. Forecast error climatology

The great-circle error distribution from 2001 to 2007 for 24 h lead time shows (Figure 5(b)) the following results: (i) except for 2001 there is almost no change in forecast performance and cyclone activity; the season 2001/2002 may be considered as an outlier with only one tropical cyclone being predicted and occurring in both the UKMO and

Table 2. Seasonal great-circle error differences (km) of 24 h and 48 h forecasts.

Season	EWA 24 h	EWA 48 h	CLIPER 24 h	CLIPER 48 h	Combination: EWA + UKMO NWP	Combination: EWA + UKMO NWP	Observations
1992–2000	**−52**	16	**124**	**171**	**−69**	−56	**174**
	(200)	(363)	**(200)**	**(363)**	**(200)**	(363)	
2000–2007	5	**98**	**22**	**104**	**−21**	**15**	378
	(139)	**(255)**	**(139)**	**(255)**	**(139)**	**(255)**	
2001							6
2002	19	**134**	5	59	−1	60	27
	(119)	**(212)**	(119)	(212)	(119)	(212)	
2003	3	**26**	**26**	**123**	**−21**	24	95
	(138)	**(234)**	**(138)**	**(234)**	**(138)**	(234)	
2004	**−26**	59	−18	53	**−47**	−18	68
	(154)	(267)	(154)	(267)	**(154)**	(267)	
2005	24	**171**	**60**	**186**	−14	−53	48
	(120)	**(204)**	**(120)**	**(204)**	(120)	(204)	
2006	**40**	**152**	**64**	**151**	4	49	69
	(122)	**(247)**	**(122)**	**(247)**	(122)	(247)	
2007	9	6	77	44	**−26**	54	65
	(154)	(336)	(154)	(336)	**(154)**	(336)	

For EWA, CLIPER, and combination EWA + UKMO NWP related to UKMO NWP (in brackets); bold numbers for >95% significance (two-sided *t*-test). Note, there are less than ten observations in 2001 and no test is performed.

Australian BoM provided data constituting the test set; (ii) the performance variability of EWA and UKMO fluctuates without demonstrated superiority of either one; and (iii) only CLIPER shows consistently less accurate forecasts and error fluctuations similar to EWA.

3.1.3. *Significance test of forecast performance*

We implemented a two-sided *t*-test to check the significance of EWA and EWA plus UKMO NWP combination forecasts compared to UKMO NWP predictions. Applying the test to the 24 h predictions (1992–2000) shows (Table 2) the UKMO NWP to be significantly worse (52 km) compared to EWA; for 2001 to 2007 NWP predictions improve, and the EWA–NWP combination forecasts are significantly better than NWP forecasts. Note that the 48 h forecasts of both EWA and CLIPER show significantly larger great-circle errors than UKMO NWP.

3.2. *Error-spread relation*

Ensemble spread is a measure of the ensemble members' dispersion and determines the expected forecast error (in a perfect model and perfect ensemble environment, see e.g. Elsberry and Carr, 2000). The error-spread relations for EWA forecasts are illustrated in Figure 6 (including a sample size histogram at top, grey bars): ensemble mean great-circle errors are sampled in ensemble spread bins of 5, 10 and 20 km for 12, 24 and 48 h predictions, respectively. The model output is calculated with weighted ensemble forecasts to determine the error-spread correlation (Grimit and Mass, 2007). The following results are noted (Figure 6): (i) error-spread scatter diagrams (Figure 6(a), (c), (e) or left panels) provide the database for the statistical analysis which show many cases with small spread and small error and fewer cases of large spread and large error; (ii) the error-spread relation is almost linear with a small quadratic contribution (not shown) and the linear slope (numbers in Figure 6(b), (d), (f) or right panels) increases with lead time; and (iii) the histogram (grey bars at top of Figure 6, right panels) of spread frequencies shows that most forecasts are associated with medium-range spread. High spread and small great-circle errors occur if many ensemble members head towards the correct direction, while the remaining members move more or less randomly in other directions. Vice versa, small spread and large great-circle errors occur if many ensemble members head to the same wrong direction. The upper and lower error-quantiles (Figure 6, right panels) are added to indicate the asymmetry of the error distribution and to provide a measure of the confidence in the error-spread relation.

3.3. *Central pressure*

Note that an intensity (maximum wind speed) of tropical cyclones is almost never actually measured but inferred, so that empirical relationships may be used to derive the maximum sustained wind speed from the central pressure (Takahashi, 1939; Kraft, 1961). Instead, we analyse ensemble forecasts of tropical cyclone central surface pressure. These forecasts are obtained by the pressure values attached to each member selected for ensemble track prediction. The ensemble mean forecast error shows an almost linear growth with lead time (Figure 7). For completeness we include here the central pressure ensemble forecasts (Figure 8, Table 3) for two cases (*George* and *Wati*, see section 4). We include a set of forecasts up to 12 hours, which are initialized at 6-hourly time intervals; the diagrams show that weaker (stronger) intensification is well (hardly) predicted for *Wati* (*George*).

4. Combination forecasts, hazardous areas and case-studies

Combinations of NWP forecasts with empirical predictions improve, on average, short-term and long-range forecasts (Fraedrich and Smith, 1989; Leslie and Fraedrich, 1990; Metzger *et al.*, 2004). Applied to forecasts in the Australian tropical cyclone seasons 1984/1985 to 1987/1988, an improvement of 15% over the next most accurate individual

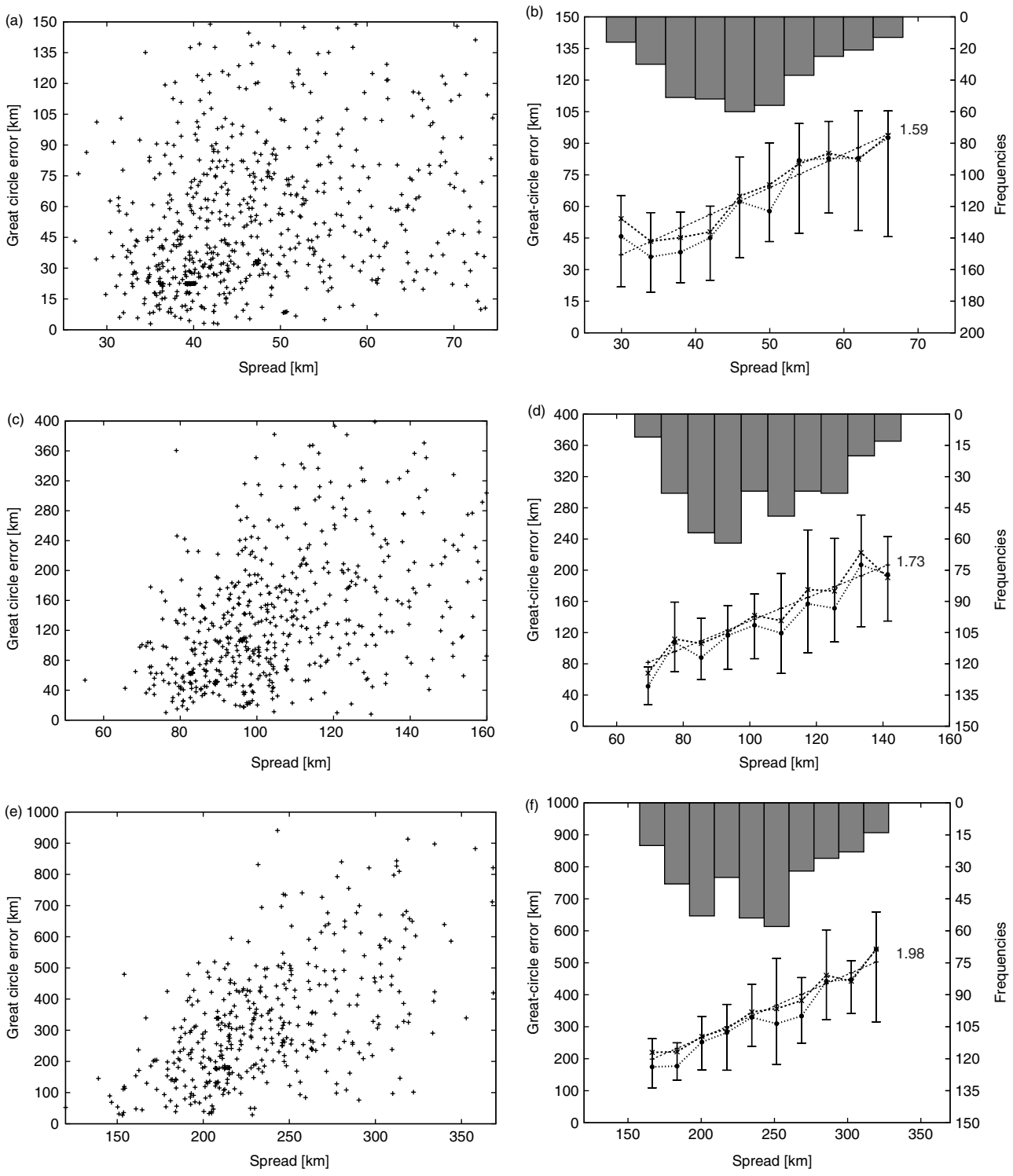

Figure 6. Relation between forecast error and ensemble spread (2001–2007) for 12, 24 and 48 h lead times. Scatter plots for (a) 12 h, (c) 24 h, and (e) for 48 h forecasts; great-circle error (dark grey dashed line) in spread bins of (b) 5 km for 12 h, (d) 10 km for 24 h, and (f) 20 km for 48 h forecasts: median (light grey dashed line) with 25% and 75% quantiles (vertical solid lines). The linear slope for the weighted mean (upper number) and median is indicated. The histogram (upper part of panel) shows the corresponding forecast spread distribution.

method was achieved (Leslie and Fraedrich, 1990). Similar results are demonstrated by superensemble forecasts based on linear combinations of NWP multimodels (Vijaya Kumar and Krishnamurti, 2003). Here the optimal combination of EWA with UKMO NWP is introduced, tested for independent forecasts, and compared with other model combinations, for example the combination of two empirical forecasts schemes, EWA and CLIPER.

4.1. *Combination forecasts*

The longitude $\Delta\lambda_{\mathrm{Comb}}$ and latitude $\Delta\varphi_{\mathrm{Comb}}$ displacement forecasts by UKMO NWP and EWA of tropical cyclone tracks are linearly combined (Leslie and Fraedrich, 1990):

$$\begin{aligned} \Delta\lambda_{\mathrm{Comb}} &= a_1 \cdot \Delta\lambda_{\mathrm{NWP}} + a_2 \cdot \Delta\lambda_{\mathrm{EWA}} + a_3 \\ \Delta\varphi_{\mathrm{Comb}} &= a_4 \cdot \Delta\varphi_{\mathrm{NWP}} + a_5 \cdot \Delta\varphi_{\mathrm{EWA}} + a_6. \end{aligned} \tag{4.1}$$

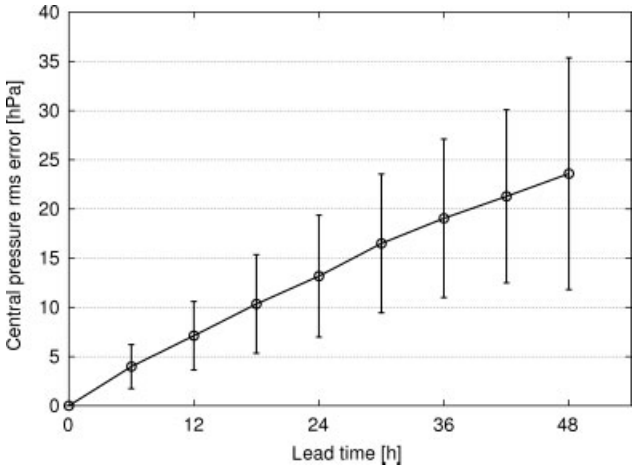

Figure 7. Australian tropical cyclone central pressure forecast errors (2001–2007) for the Ensemble-Weighted Analog (EWA) prediction model: mean forecast errors (in hPa, compared to the best track data) changing with forecast lead time (6 h time step).

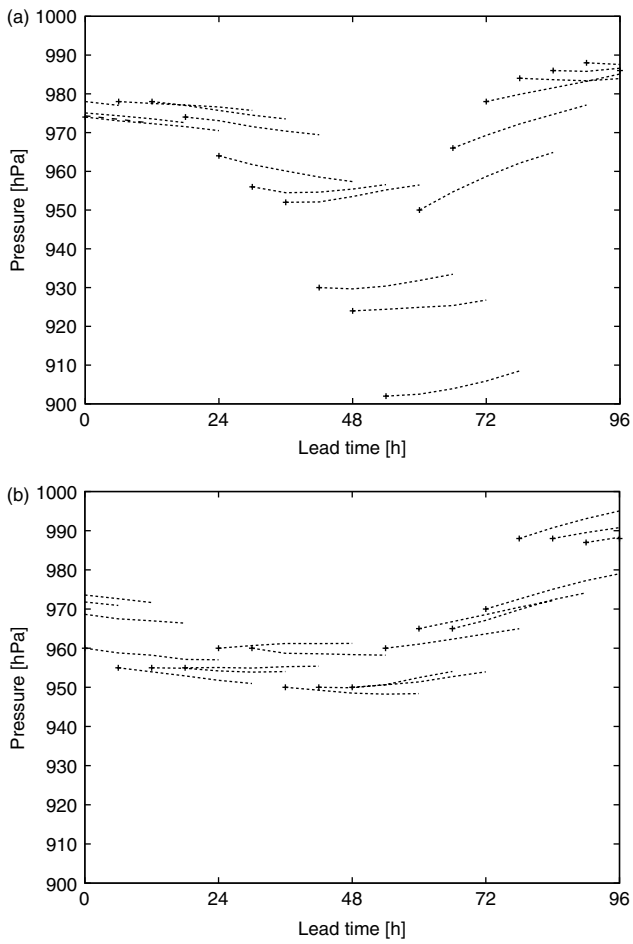

Figure 8. Tropical cyclone central pressure forecasts (up to 24 hours) at six-hourly intervals: (a) *George* forecasts commencing on 6 March at 0000 UTC; and (b) *Wati* forecasts issued on 22 March at 0000 UTC. The full dots denote the initial value identical to the (best track) verification; the dotted line connects the predicted central pressure values.

The coefficients $(a_1, a_2, a_3; a_4, a_5, a_6) = (0.293,\ 0.747,\ 0.202\ ;\ 0.465,\ 0.548,\ 0.081)$ are determined by multivariate regression conditioned by minimizing the squared forecast error. The regression coefficients used for independent consensus forecasts (2001–2007) are determined from the combination dataset (1992–2000, see Figure 2, section 2). Note that the regression coefficients are not equal, because the models contribute differently to minimizing the forecast error. That is, EWA dominates the zonal track component (a_1 vs. a_2). The EWA (NWP) contribution has a larger (smaller) influence on the zonal propagation. The joint bias of EWA and NWP on predicting the zonal (a_3) and lateral (a_6) displacements indicates an underestimation of the TC's propagation. Table 1 summarizes the linear combination forecasts error which is based on 378 independent UKMO NWP and CLIPER forecasts in 2001–2007: great-circle errors at 24 h and 48 h of 118 km (if quadratic terms are also considered: 116 km) and 270 km are achieved improving the best individual forecasts of the UKMO NWP of 139 km by about 21 km (at 24 h). In comparison, the NWP–CLIPER combination gives a comparable error of 117 km at 24 h. Note that the empirical forecast model combination, EWA and CLIPER, leads to a great-circle error of 139 km for 24 h predictions which is comparable to NWP.

Table 3. EWA intensity (pressure) forecasts compared with best track values for the case-studies in Figure 8: predicted pressures and best track pressure value (in brackets).

Lead time	0 h	6 h	12 h	18 h	24 h
George	976	975	974	973	972
6 March 2007	(976)	(974)	(978)	(978)	(974)
0000 UTC					
George	978	977	977	976	975
6 March 2007	(978)	(978)	(974)	(964)	(956)
1200 UTC					
George	974	973	971	970	969
7 March 2007	(974)	(964)	(956)	(952)	(930)
0000 UTC					
George	956	954	954	955	957
7 March 2007	(956)	(952)	(930)	(924)	(902)
1200 UTC					
Wati	960	959	958	957	957
22 March 2006	(960)	(955)	(955)	(955)	(960)
0000 UTC					
Wati	955	955	954	954	954
22 March 2006	(955)	(955)	(960)	(960)	(955)
1200 UTC					

Obviously, EWA forecasts contribute sufficient independent information to a mutual combination with the UKMO NWP, likewise CLIPER. Contributions with quadratic entries and cross-correlations are also studied and presented in brackets in Table 1. In summarizing, the combination forecast outperforms the individual dynamic UKMO NWP model and the statistic-based EWA model; best results are obtained by a combination of the three forecasts UKMO NWP–EWA–CLIPER with 116 km (114 km) at 24 h, for which, however, the CLIPER contributions are marginal. For lead times beyond 48 h, UKMO NWP always performs the best, while the combination with CLIPER or EWA improves the performance. The theoretical maximum gain using completely independent models would be $n^{1/2}$ (Goerss, 2000), where n is a measure of effective degrees of freedom and consequently less than or equal to the number of independent models. Note that combination schemes of higher-order regression may, at times, perform better than linear ones (see also Burton, 2006; Sampson *et al.*, 2007).

We tested the stability of the regression coefficients through a further calculation with a reduced combination dataset from 1992 to 1997 and a test set from 1998 to 2004. The resulting coefficients $(a_1, a_2, a_3; a_4, a_5, a_6) = (0.261, 0.759, 0.196 ; 0.423, 0.648, 0.065)$ are similar to those obtained from 1992 to 2000. Applying regression coefficients to the test set 2001–2007, the resulting coefficients $(a_1, a_2, a_3; a_4, a_5, a_6) = (0.664, 0.319, 0.140 ; 0.508, 0.440, 0.024)$ show dominance of UKMO NWP versus EWA in the latitude coefficients, which is due to the improvement of the UKMO NWP (see also Table 2 for individual hurricane seasons).

4.2. *Hazardous areas*

The EWA predicted ensemble tracks are subjected to a spatial cluster analysis to identify hazardous areas. Here we focus on track bifurcations which depend on atmospheric states that are not sufficiently considered by the forecast scheme. For example, midlatitude trough interaction (Sampson *et al.*, 2006) leading to track change may not be fully captured by an ensemble weighted forecast, if too few of such cases exist in the analog pool. However, different hazardous areas identified by the resulting cluster centroids may hint at ensemble bifurcations. We demonstrate that the average forecast track derived from the cluster of highest probability – that is the cluster whose centroid is *a posteriori* chosen as the one closest to the best track – outperforms the EWA ensemble means. The evaluation consists of statistics based on the verification dataset, comprising all lead times of 24 h within the period from 2001 to 2007 in the Australian region. Thus, selecting the best cluster from next best track observation, EWA probabilistic forecasts perform better than NWP with mean position errors of 136 km compared to 139 km of NWP. A method to determine the optimal number of clusters and considering further atmospheric states provides more-detailed information on the error-spread relation and may lead to better results; this is the subject of future research. Nevertheless bifurcations of highest probability point to potentially hazardous areas and attract the forecaster's interest, whereas the EWA average forecast track often leads to no hazardous areas in some cases.

4.3. *Two cases of sudden change in track direction*

The tracks of tropical cyclone *George* (2007) in the western Australian basin and *Wati* (2006) in the eastern Australian basin show a sudden change of direction (Figure 9). Their track forecasts are analysed supplementing the error statistics presented so far by two case-studies. Note that most of the tropical cyclones occur in the western part of the Australian region but the threat of landfall in densely populated areas is higher in the east. Therefore, we focus on the ability of EWA to predict change of track direction in both, utilizing the 48 h forecast ensemble expressed probability of occurrence in an area covered by 1° grid. Their results presented in Figure 8 are briefly described.

Western Australian basin (*George* 2007; (a)–(d)). At the track bifurcation, issued at 7 March (1200 UTC), the 48 h probability prediction (issued at that time) shows two regions of equal shadowing which reveal the same number of tropical cyclone tracks at these locations. The preceding 48 h forecasts issued on 6 March (1200 UTC) and 7 March (0000 UTC), however, show darker shading (larger probability) where both the EWA mean and the NWP forecast tracks point in the wrong direction, following the dominant first cluster. However, lobes of lighter shading (directed towards the coastline) already emerge, marking another cluster and giving landfall a considerable probability.

Eastern Australian basin (*Wati* 2006; (e)–(f)). At the track bifurcation, issued at 22 March (0000 UTC), the 48 h EWA prediction (issued at that time) dislocates the future cyclone position to the south-west of the best track. The NWP forecast also follows that route; it changes to the correct south-eastward direction, however, 24 h later. Although EWA mean and probabilistic forecasts (with darkest shadowing) point south-westward, a bifurcation into the correct direction is indicated by the structure of the probability field (one lobe in the darker shaded area). Note that the EWA–NWP combination follows the best track direction, but with too slow propagation; 12 h later, NWP, the EWA mean, the combination and the closest, dominant lobe of shaded areas point into the best track direction, but EWA and combination forecasts reveal too slow propagation.

A further comment appears in order here: EWA forecasts perform best in the western and eastern parts of the Australian basin where most tropical cyclones occur. Track forecasts in the Coral Sea or the Gulf of Carpentaria region, however, appear to be more erratic with sudden track changes and strong stationarity. Thus, the analog pool contains too few adequate ensemble members supporting a reliable forecast ensemble for such cases. Here, the NWP combination forecast gains much better results (see also Leslie and Fraedrich, 1990).

5. Summary and conclusion

An analog forecast scheme with self-adapting metric has been developed which, using local averaging optimisation, does not include previous forecasts in the analog search, thus avoiding inflation of the analog track ensemble. The forecast scheme comprises two optimisations: (i) adapting the state space metric weights to select optimal ensemble members from the history; and (ii) identifying ensemble member weights to obtain optimal ensemble averaged

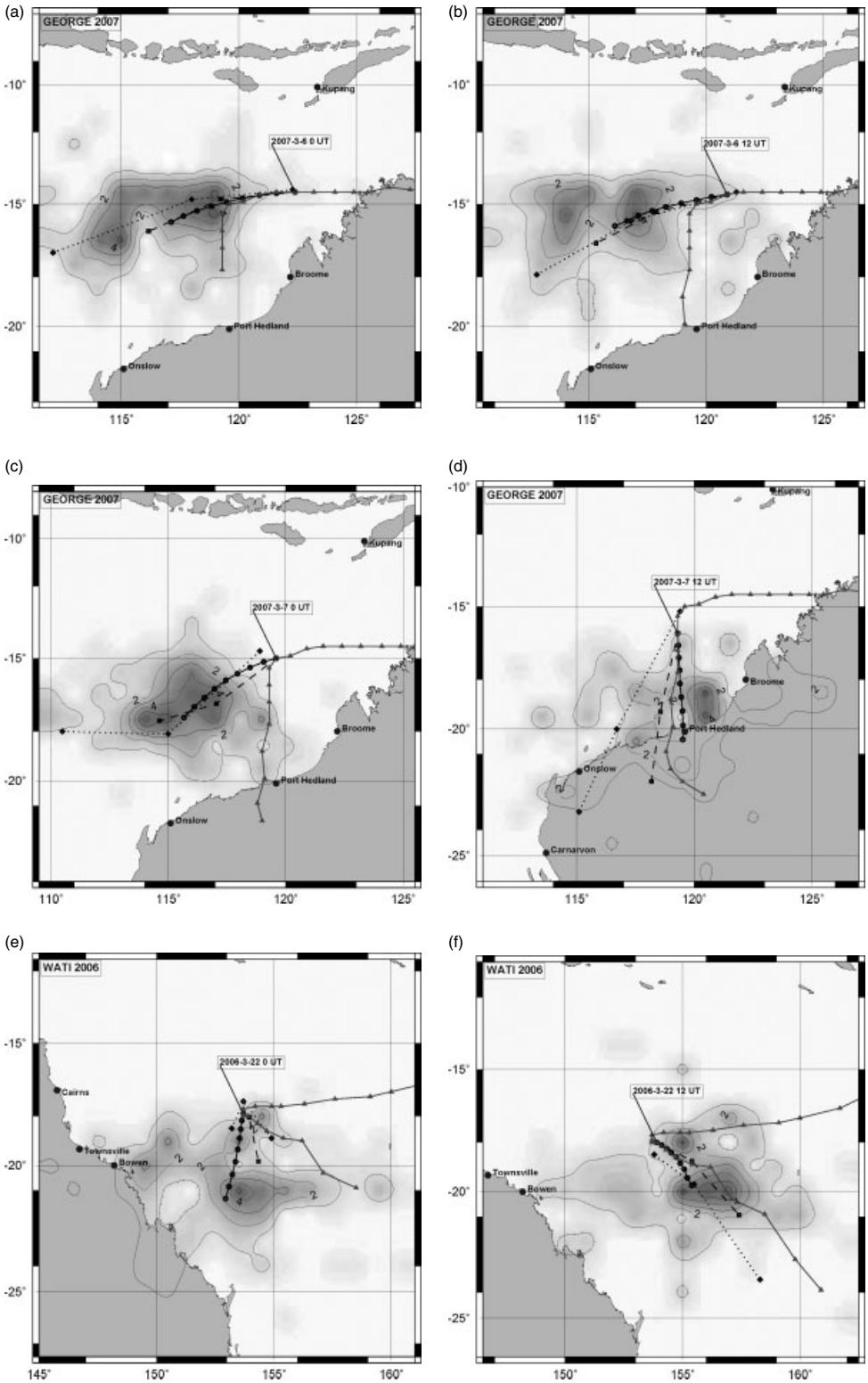

Figure 9. Independent tropical cyclone track forecasts. For *George* (western Australian basin 2007, 6 and 7 March at 0000, 0600 and 1200 UTC), forecasts are issued on 6 March at 0000 UTC (a) and 1200 UTC (b), on 7 March at 0600 UTC (c) and 1200 UTC (d); for *Wati* (eastern Australian basin 2006) forecasts are issued on 22 March at 0000 UTC (e) and 1200 UTC (f). EWA 6 h weighted ensemble mean forecasts (solid line with circles), UKMO NWP forecasts (dotted line with diamonds), combination forecasts of EWA mean and UKMO NWP (dashed line with squares), and the best-track positions (solid line with triangles). The shaded areas quantify the 48 h EWA ensemble forecast probability of the tropical cyclone occurring in ($1^\circ \times 1^\circ$) grids.

forecasts. These weights are determined by a metric adaption algorithm minimizing a forecast error representing the cost function. Independent forecasts are performed in the Australian basin extending over the Indian Ocean (2001–2007). Standing alone, the Ensemble-Weighted Analog scheme demonstrates some advantages:

- Forecasts are substantially better than other, simpler statistical techniques (e.g. CLIPER). Track position errors of the ensemble-weighted mean forecasts are comparable to a numerical weather-prediction model (NWP of the UKMO) for the 24 h predictions. The forecast analysis shows an error of about 140 km;

intensity estimates and the relation of forecast error to ensemble spread are also discussed.

- Ensemble forecast techniques have been employed to provide:
  (i) additional information to forecast possible track recurvature; and
  (ii) useful quantitative forecasts of the forecast error which shows a linear error-spread relation, which does not deteriorate with increasing forecast lead time.
- Intensity (or central pressure) forecasts may also be performed, appearing as a skilful forecast by-product comparable to NWP if sufficient observations are available.
- Combining EWA forecasts with UKMO NWP yields the best performance up to a lead time of 48 h, if compared to UKMO NWP alone. EWA and the combination forecast errors are smaller in the regions to the western part of the basin and higher in the Coral Sea east of Australia, where sudden changes of direction occur frequently.
- Ensemble weighted-average forecasts improve over their unweighted counterparts by extending the good performance up to 48 h. That is, the ensemble-weighted analog predictions are useful beyond 24 h.
- Probability fields of predicted tropical cyclone positions are derived from a cluster analysis dividing all analog ensemble members into two clusters, whose forecasted probability fields describe ensemble dispersion structures which, even for the 48 h forecasts, quantify possible future changes in track direction (and/or speed).

In summarizing, we try to demonstrate the usefulness of empirical nonlinear schemes for ensemble forecasts of tropical cyclones. These schemes appear to be of particular value when attaching weights to each ensemble member to determine future ensemble mean positions (and intensities) after being selected from a state space of optimal metric. As these optimisations lead to an extended forecast range, the model may provide an optimal partner for NWP combination forecasts and, due to the ensemble spread, also for forecast error predictions. In particular for longer lead times, possible quantification of the chances of changing track direction is suggested. Thus clusters of EWA ensemble forecasts support the forecasters well by pointing to the hazardous regions. Adding information on midlatitude troughs to the model in future may reduce the bifurcations of forecasted tracks and give the forecasters more significant support.

Acknowledgements

Thanks are due to Julian Heming (UKMO) for providing tropical cyclone NWP and CLIPER track data from 1994 until 2007 and to the two referees for their valuable comments. We also acknowledge support by Max Planck Society (K.F.) and by CliSAP (F.S.).

References

Bazaraa MS, Sherali HD, Shetty CM. 1993. *Nonlinear Programming: Theory and algorithms*, 2nd edition. John Wiley & Sons: New York, NY.

Bessafi M, Lasserre-Bigorry A, Neumann CJ, Pignolet-Tardan F, Payet D, Lee-Ching-Ken M. 2002. Statistical prediction of tropical cyclone motion: An analog–CLIPER approach. *Weather and Forecasting* **17**: 821–831.

Burton A, Caroff P, Franklin J, Fukada E, Lee TC, Sampson B, Smith T. 2006. 'Sharing experiences in operational consensus forecasting.' Pp 424–441 (Special Focus Topic 3a) in *Sixth International Workshop on tropical cyclones, San José, Costa Rica.* WWW/CAS/WMO 305. Available at http://severe.worldweather.org/iwtc/document/Topic_3a_Andrew_Burton.pdf

Elsberry RL, Carr III LE. 2000. Consensus of dynamical tropical cyclone track forecasts – Errors versus spread. *Mon. Weather Rev.* **128**: 4131–4138.

Fraedrich K, Leslie LM. 1989. Estimates of cyclone track predictability. I: Tropical cyclones in the Australian region. *Q. J. R. Meteorol. Soc.* **115**: 79–92.

Fraedrich K, Rückert B. 1998. Metric adaption for analog forecasting. *Physica A* **253**: 379–393.

Fraedrich K, Smith NR. 1989. Combining predictive schemes in long-range forecasting. *J. Climate* **2**: 291–294.

Fraedrich K, Raible CC, Sielmann F. 2003. Analog ensemble forecasts of tropical cyclone tracks in the Australian region. *Weather and Forecasting* **18**: 3–11.

Goerss JS. 2000. Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Weather Rev.* **128**: 1187–1193.

Grimit EP, Mass CF. 2007. Measuring the ensemble spread–error relationship with a probabilistic approach: Stochastic ensemble results. *Mon. Weather Rev.* **135**: 203–221.

Hope JR, Neumann CJ. 1970. An operational technique for relating the movement of existing tropical cyclones to past tracks. *Mon. Weather Rev.* **98**: 925–933.

Jarvinen BR, Neumann CJ. 1979. '*Statistical forecasts of tropical cyclone intensity for the North Atlantic basin.*' NOAA Tech. Memo., NWS NHC-10, 22 pp. NOAA: Miami, Florida, USA.

Knaff JA, DeMaria M, Sampson CR, Gross JM. 2003. Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Weather and Forecasting* **18**: 80–92.

Kraft RH. 1961. The hurricane's central pressure and highest wind. *Mar. Weather Log* **5**: 155.

Leslie LM, Fraedrich K. 1990. Reduction of tropical cyclone position errors using an optimal combination of independent forecasts. *Weather and Forecasting* **5**: 158–161.

Leslie LM, Hess GD, Holland GJ, Morison RP, Fraedrich K. 1992. Predicting changes in intensity of tropical cyclones using a Markov chain technique. *Aust. Meteorol. Mag.* **40**: 41–46.

Lorenz EN. 1969. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.* **26**: 636–646.

McNames J. 2002. Local averaging optimization for chaotic time series prediction. *Neurocomputing* **48**: 279–297.

Metzger S., Latif M, Fraedrich K. 2004. Combining ENSO forecasts: A feasibility study. *Mon. Weather Rev.* **132**: 456–472.

Neumann CJ. 1972. '*An alternate to the HURRAN (Hurricane Analog) tropical cyclone forecast system.*' NOAA Tech. Memo., NWS SR 62, 32 pp. NOAA: Fort Worth, Texas, USA.

Neumann CJ. 1981. Trends in forecasting the tracks of Atlantic tropical cyclones. *Bull. Am. Meteorol. Soc.* **62**: 1473–1485.

Sampson CR, Goerss JS, Weber HC. 2006. Operational performance of a new barotropic model (WBAR) in the western North Pacific basin. *Weather and Forecasting* **21**: 656–662.

Sampson CR, Knaff JA, Fukada EM. 2007. Operational evaluation of a selective consensus in the western North Pacific basin. *Weather and Forecasting* **22**: 671–675.

Sauer T, Yorke JA, Casdagli M. 1991. Embedology. *J. Stat. Phys.* **65**: 579–616.

Staniforth A, White A, Wood N, Thuburn J, Zerroukat M, Cordero E. 2004. '*Joy of U.M. 6.0 – model formulation.*' Unified Model Documentation Paper, No. 15. Met Office: Exeter, UK.

Takahashi K. 1939. Distribution of pressure and wind in a typhoon. *J. Meteorol. Soc. Jpn* **17**: 417–421.

Takens F. 1981. 'Detecting strange attractors in turbulence.' Pp 366–381 in *Dynamical Systems and Turbulence – Warwick 1981, Lecture Notes in Mathematics 898,* Rand DA, Young LS (eds). DOI: 10.1007/BFb0091924.

Vijaya Kumar TSV, Krishnamurti TN, Fiorino M, Nagata M. 2003. Multimodel superensemble forecasting of tropical cyclones in the Pacific. *Mon. Weather Rev.* **131**: 574–583.

Zhang Z, Krishnamurti TN. 1997. Ensemble forecasting of hurricane tracks. *Bull. Am. Meteorol. Soc.* **78**: 2785–2795.

