Predictability of frontal waves and cyclones

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Predictability of Frontal Waves and Cyclones

THOMAS H. A. FRAME AND JOHN METHVEN

University of Reading, Reading, United Kingdom

NIGEL M. ROBERTS

Met Office, Reading, United Kingdom

HELEN A. TITLEY

Met Office, Exeter, United Kingdom

ABSTRACT

The statistical properties and skill in predictions of objectively identified and tracked cyclonic features (frontal waves and cyclones) are examined in the 15-day version of the Met Office Global and Regional Ensemble Prediction System (MOGREPS-15). The number density of cyclonic features is found to decline with increasing lead time, with analysis fields containing weak features that are not sustained past the first day of the forecast. This loss of cyclonic features is associated with a decline in area-averaged enstrophy with increasing lead time. Both feature number density and area-averaged enstrophy saturate by around 7 days into the forecast. It is found that the feature number density and area-averaged enstrophy of forecasts produced using model versions that include stochastic energy backscatter saturate at higher values than forecasts produced without stochastic physics. The ability of MOGREPS-15 to predict the locations of cyclonic features of different strengths is evaluated at different spatial scales by examining the Brier skill (relative to the analysis climatology) of strike probability forecasts: the probability that a cyclonic feature center is located within a specified radius. The radius at which skill is maximized increases with lead time from 650 km at 12 h to 950 km at 7 days. The skill is greatest for the most intense features. Forecast skill remains above zero at these scales out to 14 days for the most intense cyclonic features, but only out to 8 days when all features are included irrespective of intensity.

1. Introduction

Ensemble forecasting aims to characterize forecast uncertainty associated with the growth of small uncertainties in the initial conditions of the forecast (Molteni et al. 1996). A natural interpretation of an ensemble forecasting system is as an estimator of the probability of occurrence of future weather events of interest (Leith 1974); for example, in the case of flood prediction, what is the chance of rainfall accumulation exceeding a particular critical threshold value in a given catchment? However, what is meant by future weather events of interest is highly dependent on the end user. This makes the identification of key predictands to be used in the assessment of forecast skill an important challenge (Morss et al. 2008). One rational choice may be to consider that in the extratropics the occurrence of significant weather such as intense rainfall or high winds is often associated with the development and passage of cyclonic weather systems (Bengtsson et al. 2005; Hawcroft et al. 2012). Therefore, the skill in predicting the location and intensity of these systems is of particular importance, their representation in models having implications for both weather and climate (Zappa et al. 2013).

Corresponding author address: Thomas H. A. Frame, Dept. of Meteorology, University of Reading, Reading RG6 6BB, United Kingdom.
E-mail: t.h.a.frame@reading.ac.uk

A relatively recent development in the prediction of extratropical weather systems has been the application of objective identification and tracking of cyclonic features to global ensemble forecasting systems. In particular the methodology of Hewson and Titley (2010) was introduced to run on the 15-day version of the Met Office Global and Regional Ensemble Prediction System (MOGREPS-15) (Bowler et al. 2009) forecasts in 2006, and has since been implemented in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System. The tracking methodology is one of those being used in the Intercomparison of Mid Latitude Storm Diagnostics (Neu et al. 2013). MOGREPS-15 was run as part of the THORPEX Interactive Grand Global Ensemble (TIGGE; Park et al. 2008) from 2006 to 2014, and has now been superseded by the higher-resolution 7-day MOGREPS-G ensemble. The Hewson and Titley (2010) cyclone tracking system will be implemented in the 7-day MOGREPS-G forecasts in 2015.

The feature tracking within these ensemble systems provides a compelling visual representation of the prediction uncertainty of ensemble forecasts (Swinbank et al. 2015) and the forecast products from the feature tracking have been used by Met Office operational forecasters for several years. For example, Fig. 1 is a sequence of "Dalmatian plots" from MOGREPS-15 at increasing lead time but with identical validity times. These forecasts correspond to a severe windstorm that struck the United Kingdom on 28 October 2013 causing severe disruption to transport and power supplies as well as resulting in several deaths. The contours show the mean sea level pressure from the control forecast and the various colored dots show the locations of all objectively identified cyclonic features from all ensemble members, colored according to the maximum wind speed at 1 km above the surface within 300 km of the feature center. The increasing uncertainty with lead time is apparent through the increasing scatter of the cyclonic features. At $T + 0$ h (Fig. 1a), the features are densely clustered into almost totally overlapping groups. As time progresses, the feature points become increasingly scattered, but at $T + 72$ h and $T + 120$ h (Figs. 1b,c) they are still in coherent and distinct groups of similar features, meaning that although the exact location and strength of the features becomes increasingly uncertain, it is still possible to be confident that there will be a severe weather system approaching the United Kingdom. In light of the forecasts, the Met Office issued a severe weather warning to the general public for high winds over southern England 5 days in advance of the storm. At $T + 168$ h, the features are sufficiently scattered that it is no longer possible to identify coherent
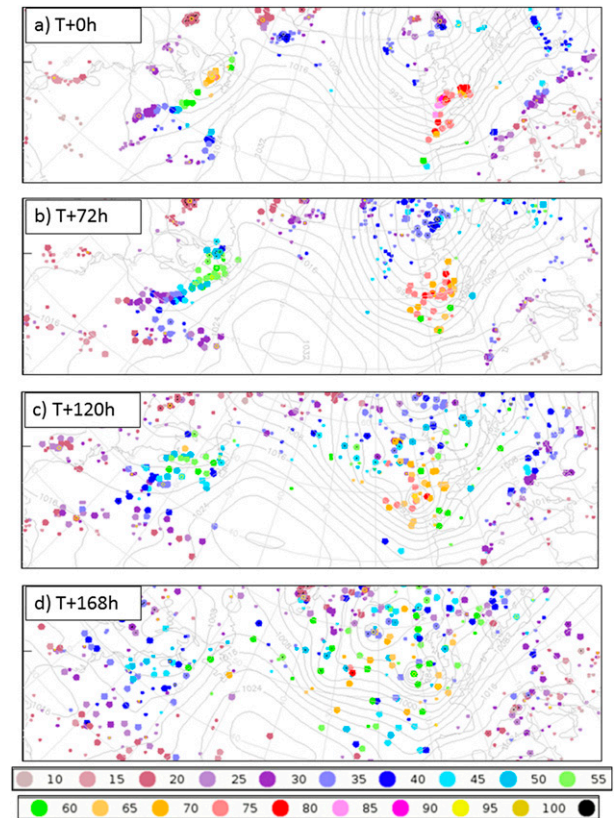

FIG. 1. "Dalmatian plots" showing (a) 0-, (b) 3-, (c) 5-, and (d) 7-day MOGREPS-15 forecast valid at 0000 UTC 28 Oct 2013. Contours show the mean sea level pressure from the control forecast. Colored dots show locations of cyclonic feature centers from all forecast members. Coloring indicates the max wind [knots (kt; where 1 kt = 0.51 m s^{-1})] within 300 km of feature center at 1 km ZAGL.

clusters of features of similar strength. For example, the red and orange dots are mixed together with blue and green dots. The large numbers of red and orange dots do indicate a high likelihood that a strong cyclone will occur, but the mixing with other colored dots indicates that the uncertainty in its location is larger than the distance separating it from adjacent cyclonic features. Despite this, the effect of the larger-scale flow is still very clear in, for example, the very low number of cyclonic features in the vicinity of the Mid-Atlantic Ridge seen in the control forecast mean sea level pressure.

In this paper, we present an assessment of the changes in the climatological statistics of these features tracked over forecast lead times and an assessment of the skill in the probabilistic predictions of them.

The rest of the paper is divided into four sections. In section 2, we give a brief overview of the tracked data used in this paper and MOGREPS-15. In section 3, the statistical properties of the mesoscale features are

examined. In section 4, the predictive skill of the mesoscale features is assessed. A summary of our results and conclusions is presented in section 5.

2. Tracked feature data

a. Tracking methodology

The data used in this paper consist of objectively identified and tracked cyclonic features from MOGREPS-15 forecasts. The data consist of twice-daily forecasts produced at 0000 and 1200 UTC, covering a period of more than 6 yr from 1 December 2006 to 31 December 2012. To aid in the understanding of what is to follow, in this section we shall give a brief outline of the tracking method used. See Hewson and Titley (2010) for a full description of the objective feature identification and tracking methodology.

The tracking algorithm aims to identify and track developing cyclonic storms through their entire life cycle from small kinks in fronts (labeled diminutive frontal waves) through developed frontal waves to closed low pressure centers (labeled barotropic lows). The algorithm uses a hybrid of objective fronts, pressure minima, and vorticity maxima to identify features and define their locations. As in Hewson (1998), the fronts themselves are defined to coincide with sharp gradients in wet-bulb potential temperature. Warm and cold fronts are distinguished by the sign of the local geostrophic advection of wet-bulb potential temperature, with warm fronts defined for positive advection and cold fronts defined for negative advection. Following the terminology of Hewson and Titley (2010), the three classes of cyclonic feature are defined using data at 1-km geopotential height above ground level (hereafter 1 km ZAGL) as follows:

- a diminutive frontal wave is a maximum in the alongfront component of geostrophic vorticity,
- a frontal wave is the intersection of a cold and warm front at which the cross frontal geostrophic vorticity is positive, and
- a barotropic low is defined as any other isolated minimum in the 1000-hPa geopotential height.

The two classes of frontal wave are further subdivided into weak and standard depending on the strength of the frontal gradients, and warm and cold depending on the classification of the front. This makes a total of eight classes of frontal waves.

The tracking is performed on a 50-km-resolution equal-area grid within a subdomain of the full model domain, spanning the region of approximately 11.5°–85.5°N, 117.2°W–76.8°E. The methodology uses a
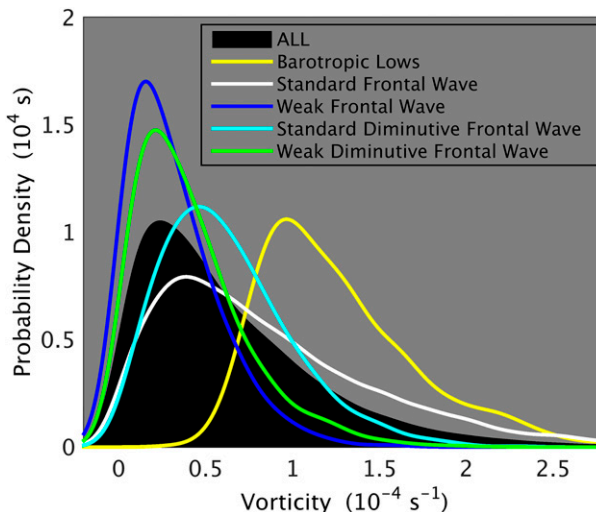

FIG. 2. Gaussian kernel smoothing probability density estimate of the relative vorticity at 1 km ZAGL of cyclonic features of different classes identified in analysis fields within the North Atlantic control region. A kernel bandwidth of $10^{-5}\,\mathrm{s}^{-1}$ was used to create the distributions.

contour intersection methodology so that the locations of feature centers are not restricted to grid locations but can vary continuously in space. The use of 1000-hPa geopotential height to identify barotropic lows has some disadvantages. One of these is that anomalously large numbers of barotropic lows are identified over high topography (particularly Greenland). The sensitivity of the statistical properties of tracked cyclones in analysis fields to the inclusion and exclusion of such orographic features has been examined by Rudeva et al. (2014). They found that bulk measures of cyclone behavior such as mean minimum central pressure and maximum deepening rate have only weak dependence on the inclusion or exclusion of orographic features. However, in the study presented here we are interested in the prediction of the location of mobile systems. For this reason, we shall restrict our analysis of the statistics of features to a North Atlantic control region. This control region is defined to lie within 30°–60°N, 50°–10°W, and covers a total area of $\sim 10^{7}\,\mathrm{km}^{2}$.

Within this work, we shall not consider the distinction between different classes of frontal waves. This choice is motivated by the observation that tracked features frequently switch between classes during their evolution and therefore the same feature can be classified differently in different ensemble members. We will instead use vorticity at the feature center 1 km ZAGL as a means of quantifying feature strength. For reference, Fig. 2 shows the pdf of 1 km ZAGL feature vorticity for different feature classes calculated from the analysis fields for the whole of the dataset.

b. MOGREPS-15

MOGREPS-15 consists of 23 perturbed members plus one unperturbed control forecast, produced using the global 33-km-resolution configuration of the Met Office Unified Model. Control and perturbed forecasts are run using the same model resolution and dynamics; however, perturbed members are run with stochastic physics and the control forecast without. The perturbed members are initialized from the control analysis using additive perturbations derived from an ensemble transform Kalman filter (ETKF; Wang and Bishop 2003). Because of the pseudo-operational nature of the data, the period under consideration spans 13 versions of MOGREPS-15. Details of the differences between these 13 versions, as well as descriptions of subsequent versions and the new MOGREPS-7, are available online (http://www.ecmwf.int/en/research/projects/tigge). Since our main interest is the generic predictability properties of cyclonic features, we shall not consider the different versions of the model in detail here. However, one change of interest is that starting in April 2012 the configuration of the ensemble was changed so that one of the perturbed members (member 23) was run directly from the control analysis without initial condition perturbations. The effect of stochastic parameterizations on the cyclonic features is discussed in the appendix.

3. Systematic differences between features in forecast and analysis fields

In this section, we shall highlight the evolution of the climatological statistics of cyclonic features with increasing lead time. We shall focus particularly on the differences between the analysis, control, and perturbed forecasts. Since the method of generation of member 23 changes part way through the dataset, we exclude it from the analysis in this section so that only 22 perturbed members are considered.

Figure 3 shows the mean areal density of features versus lead time for the control and perturbed forecasts. Figure 3a shows the density of frontal waves and Fig. 3b shows the density of barotropic lows. The thick solid line shows the control member, and the dashed solid line the average over all perturbed forecast members. For reference, the thin solid line shows the same statistic computed for the corresponding valid time of the analyses data. To give an illustration of how large differences between the control and perturbed forecasts would be anticipated to be through sampling, the yellow-shaded region illustrates the range of values of the statistic obtained from perturbed forecasts by selecting one perturbed member at random for each
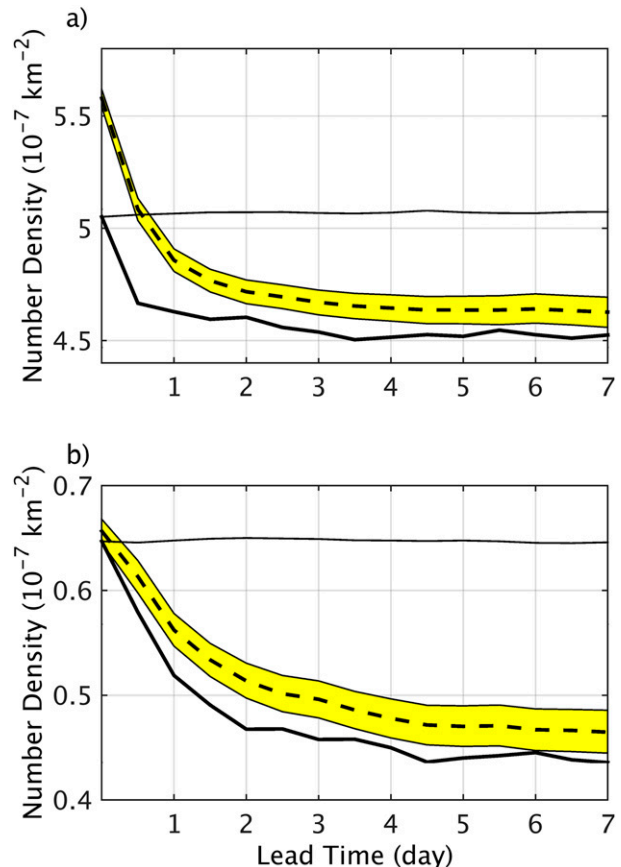

FIG. 3. The instantaneous mean number of (a) frontal waves and (b) barotropic lows per unit area as a function of lead time, calculated for the North Atlantic control region with the 6-yr dataset. Shown is the control forecast fields (thick solid line), perturbed forecast fields (dashed line), and concurrent analysis fields (thin solid line). Yellow shading indicates the 99th percentile region for resampling between perturbed ensemble members.

analysis time. This is plotted at the 99th percentile of the resampling distribution estimated from a sample of 10 000 realizations, but this is not intended to denote a 99th percentile confidence interval.

Several points can be made about the feature density. First, the number density is a decreasing function of lead time, with the mean density of all feature types below that of the analysis fields by 24-h lead time. Second, the expected number of features in perturbed forecasts is systematically larger than that of the control forecasts at all lead times. At the initial time, this must be attributable to the initial condition perturbations added to the perturbed members. At later lead times, this is a consequence of the stochastic parameterizations, which are applied to perturbed forecasts but not the control forecast (see the appendix). The third point to be made is that the rate at which the number of frontal waves
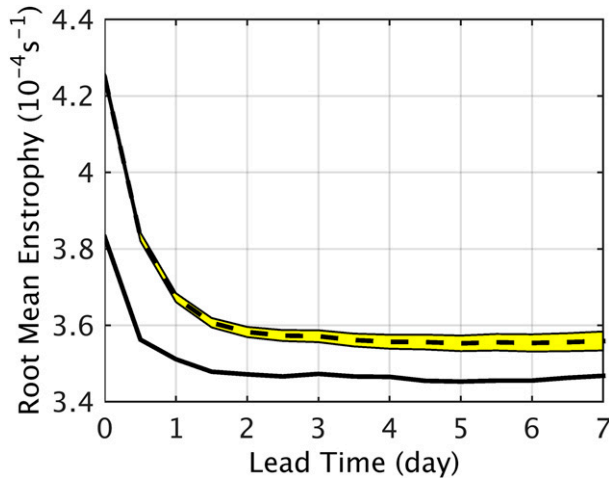
FIG. 4. The area-averaged 850-hPa enstrophy as a function of lead time, calculated for the North Atlantic control region from $1° \times 1°$ wind fields using all forecasts starting in January, April, July, and October between 2007 and 2012. Shown is the control forecast fields (thick solid line) and perturbed forecast fields (dashed line). Yellow shading indicates the 99th percentile region for resampling between perturbed ensemble members.

decreases with increasing lead time is significantly faster than that of barotropic lows.

For comparison, Fig. 4 shows the square root of the area-averaged 850-hPa enstrophy (vorticity squared) calculated for the North Atlantic region at different forecast lead times. There is a clear decay in enstrophy with increasing lead time (*e*-folding time of about 18 h), which is the same for both control and perturbed forecasts, although there is an approximately constant offset between the two. The exponential decay of enstrophy is very suggestive of numerical diffusion but may also relate to a spindown of unbalanced features in the initial conditions. At the initial time, the presence of the offset in enstrophy between the control and perturbed members must be attributable to the additive ETKF perturbations since this is the only difference between the two. After the first couple of days, however, we attribute the offset primarily to the additional enstrophy supplied to the perturbed members by stochastic parameterizations (see the appendix), most likely the stochastic kinetic energy backscatter scheme (Shutts 2005; Bowler et al. 2009). The 18-h *e*-folding time for enstrophy decay is comparable to the *e*-folding time scale of the frontal-wave feature number density (~17.5 h) shown in Fig. 3a. The number densities of barotropic lows (Fig. 3b) have a longer *e*-folding time scale ~24 h. This longer *e*-folding time scale in mean feature density reflects the higher vorticity of barotropic lows (Fig. 2); that is, there are fewer weak features that rapidly decay with diffusion. But it may also be interpreted as a reflection of the fact
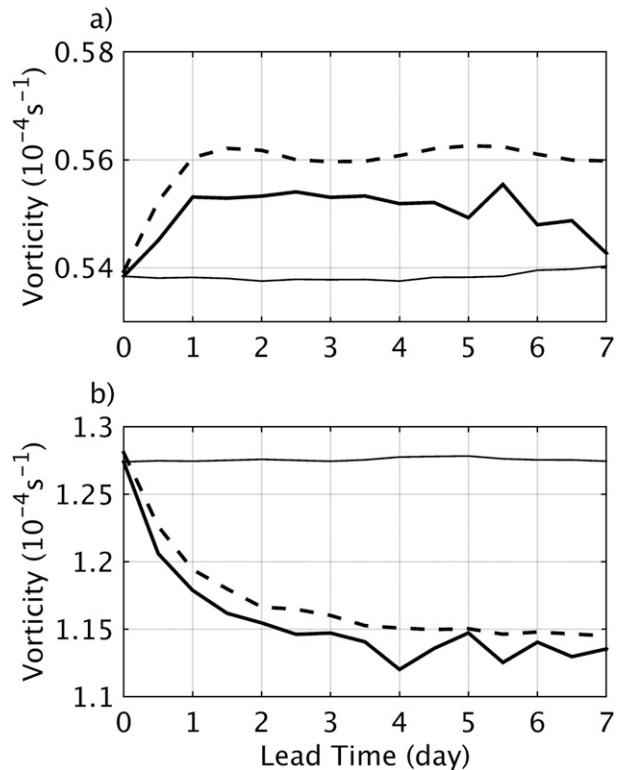

FIG. 5. The mean vorticity of (a) frontal waves and (b) barotropic lows at 1 km above the ground as a function lead time, calculated for features within the North Atlantic control region. Shown is the control forecast fields (thick solid line), perturbed forecast fields (dashed line), and concurrent analysis fields (thin solid line).

that barotropic lows are defined from pressure minima meaning that decay requires significant movement of mass, whereas frontal waves are defined from the wind or vorticity field at fronts and therefore have more rapid evolution (Neu et al. 2013).

Figures 5a and 5b show the mean vorticity of frontal waves and barotropic lows, respectively, as a function of forecast lead time. Considering first Fig. 5b, there is rapid decay in the mean vorticity of barotropic lows in both the perturbed and control forecasts. The vorticity saturates at ~90% of the value in the analysis for perturbed forecasts, with the control forecasts saturating at a slightly lower (~89%) value. By contrast, the mean vorticity of frontal waves in forecast fields increases. Comparing the vorticity tendencies along the tracks of features that are matched one to one between the control forecast and analysis, it is found that the average vorticity of both classes of features in the control forecasts decrease with lead time relative to the analysis field. At 24-h lead time, the vorticity of the frontal waves in the control forecast fields is on average $3.3 \times 10^{-6} \, \text{s}^{-1}$ lower than that of the corresponding frontal waves in the

analysis fields. For barotropic lows, the vorticity is on average $1.6 \times 10^{-5}\,\mathrm{s}^{-1}$ lower in the control forecast than the analysis. This implies that the increase in the mean vorticity of frontal waves in forecasts can be seen as an effect of the change in frontal-wave population; that is, the analysis fields contain a large number of weak features that are not sustained in the forecast.

4. Forecast skill of features

a. Definition of predictand and skill

In this section, we shall examine the forecast skill of the objectively identified features. We shall use the binary form (Wilks 2011) of the Brier skill score (BSS; Brier 1950) of forecast strike probabilities as a measure of skill and focus on how this varies with lead time, spatial scale, and feature vorticity. The objectively identified features in the analysis shall be used as the verifying observations and the climatology of these as the reference forecast. The skill in predicting the tracks of individual cyclones has been investigated using a different tracking method by Froude (2010) and will not be covered here. The notion of a strike probability (SP) originates in the tracking of tropical cyclones; however, the usage here is slightly different. Whereas for tropical cyclones strike probabilities are used to estimate the probability that a given tropical cyclone will pass over a given location, in this work we shall use the term to refer to the probability that at a given time a cyclonic feature is found within a given radius (the strike radius) of a given location. For a given latitude λ and longitude ϕ location, strike radius r, vorticity threshold ξ_0, analysis time t, and lead time τ, the strike probability can be expressed as an average over forecast members, indexed by i:

$$\mathrm{SP}(\lambda,\phi,r,\xi_0,t,\tau) = \frac{1}{N+1} \sum_{i=0}^{N} f(\lambda,\phi,r,\xi_0,x_{i,t,\tau}), \quad (1)$$

where $i = 0$ would indicate the control forecast and N is the total number of perturbed members. The indicator function $f(\lambda,\phi,r,\xi_0,x) = 1$ if field x has at least one feature with relative vorticity at 1 km ZAGL greater than ξ_0 within a distance r of the point (λ,ϕ), and $f(\lambda,\phi,r,\xi_0,x) = 0$ otherwise. Note that this definition does not include information from feature tracks and is, therefore, only a function of the instantaneous feature detection. To determine whether a feature lies within a given distance of a particular location, we use the great-circle distance between that location and the feature.

We shall use a climatological value of the strike probability to serve as a reference forecast for evaluating skill. This is defined as the time mean value:

$$\mathrm{SP}_{\mathrm{clim}}(\lambda,\phi,r,\xi_0) = \frac{1}{T} \sum_{t=1}^{T} f(\lambda,\phi,r,\xi_0,x_{0,t,0}). \quad (2)$$

The Brier score (BS) is defined as

$$\mathrm{BS} = E[(O - \mathrm{SP})^2], \quad (3)$$

where the outcome variable $O = f(\lambda, \phi, r, \xi_0, x_{0,t,0})$ takes the value $O = 1$ if a feature is observed in the analysis and $O = 0$ otherwise. The Brier skill score is normalized by the Brier score associated with climatology $\mathrm{BS}_{\mathrm{clim}}$. Here, $\mathrm{BS}_{\mathrm{clim}}$ is the Brier score obtained if the climatological value of the strike probability was always issued in place of the forecast. The Brier skill score can be written $\mathrm{BSS} = 1 - \mathrm{BS}/\mathrm{BS}_{\mathrm{clim}}$. The Brier score can be decomposed into three components: resolution (RES), reliability (REL), and uncertainty (UNC), defined as the discrete analogs of

$$\mathrm{RES} = E\{[E(O\,|\,\mathrm{SP}) - \mathrm{SP}_{\mathrm{clim}}]^2\}, \quad (4)$$

$$\mathrm{REL} = E\{[\mathrm{SP} - E(O\,|\,\mathrm{SP})]\}^2, \quad \text{and} \quad (5)$$

$$\mathrm{UNC} = \mathrm{SP}_{\mathrm{clim}}(1 - \mathrm{SP}_{\mathrm{clim}}). \quad (6)$$

The resolution term is a measure of the ability of the forecast to identify situations in which the observed relative frequency of the event differs from the climatological average. The reliability measures the discrepancy between forecast probabilities and the observed relative frequency of the event. For a perfectly reliable (perfectly calibrated) forecasting system, $\mathrm{REL} = 0$. The uncertainty is the variance of the outcome variable and is equivalent to the Brier score obtained by simply issuing the climatological probability of the event. With these definitions, the Brier skill score reduces to

$$\mathrm{BSS} = \frac{\mathrm{RES} - \mathrm{REL}}{\mathrm{UNC}}. \quad (7)$$

In the results that follow, REL is an order of magnitude or more smaller than RES. This is typical of modern ensemble prediction systems (Jolliffe and Stephenson 2012) and means that BSS may be written

$$\mathrm{BSS} \approx \frac{\mathrm{RES}}{\mathrm{UNC}} \quad (8)$$

and RES is bounded from above by UNC.

The verification results are presented as area-averaged BSS. The method of averaging used is

$$\overline{\mathrm{BSS}} = 1 - \overline{\mathrm{BS}}/\overline{\mathrm{BS}_{\mathrm{clim}}}, \quad (9)$$

where the overbar denotes an area average. This method of averaging was proposed by Hamill and Juras

(2006) to avoid overestimation of forecast skill due to differing climatologies at different locations. In practice the results were not found to be sensitive to the choice of spatial averaging, and this form of averaging was chosen primarily because it has the advantage that the individual components of the Brier skill decomposition can be replaced by their area-averaged values. Therefore, Eqs. (7) and (8) hold for the area-averaged Brier skill score, uncertainty, resolution, and reliability: $\overline{\text{BSS}}$, $\overline{\text{UNC}}$, $\overline{\text{RES}}$, and $\overline{\text{REL}}$, respectively. We tested the sensitivity of the calculated forecast skill to seasonal variations in climatology by using separate climatological strike probabilities for each calendar month as a reference forecast. This produced similar results to those obtained using the climatological strike probability averaged over all months as a reference forecast. Therefore, for simplicity we shall use the latter as a reference forecast.

Results showing the variation in skill with vorticity threshold are presented using deciles of the vorticity distribution so that the figure axes are linear with respect to frequency. The deciles were calculated from the analysis data within the North Atlantic region and are the deciles of the distribution shown in Fig. 2. One possibility with this definition of vorticity threshold would be to calculate the strike probabilities using deciles of the vorticity distribution at differing lead times, so that some forecast recalibration would be included in the results. The results presented do not include such a recalibration.

b. Verification of strike probabilities

In this section we shall consider how the Brier skill score varies with lead time, strike radius, and vorticity threshold. The results are calculated for forecasts produced every 12 h between 1 December 2006 and 30 November 2012. This time range spans the change in the generation method of member 23; however, the results were found not to be sensitive to the inclusion or exclusion of these dates. Some forecasts had one or more individual forecast members, or verifying analyses, missing from the tracked dataset and these forecast dates have been excluded from the calculations. The forecasts with missing data were distributed randomly in time.

Figure 6 shows the Brier skill score (color shading) as a function of vorticity threshold and strike radius, for lead times of 1, 3, and 5 days. The white and dark blue contours show the resolution and uncertainty, respectively. These were all calculated as the average of the values calculated at single locations. For simplicity we used sample locations on a regular 2.5° grid. Figures produced for individual locations are
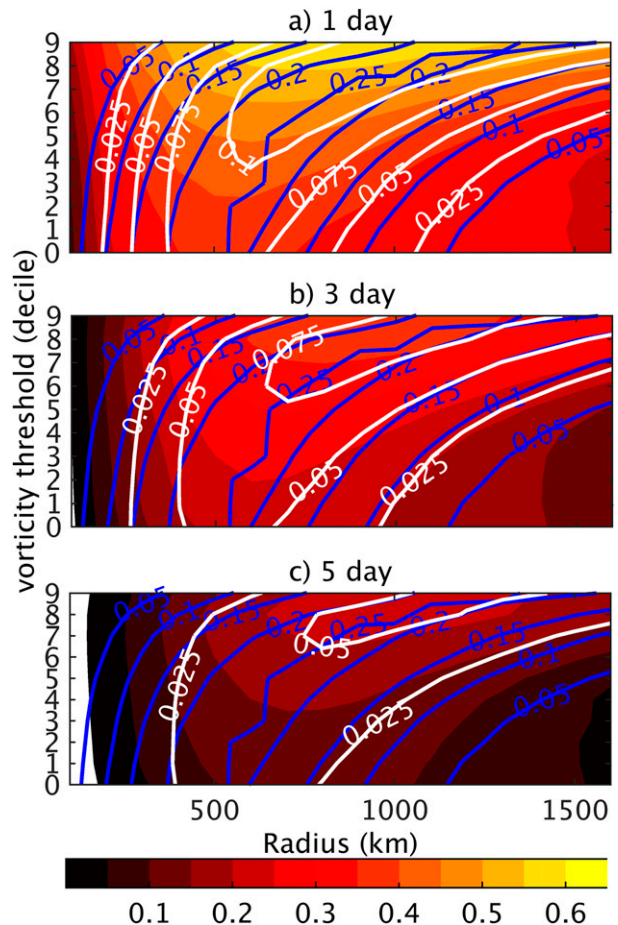

FIG. 6. Color shading showing BSSs for (a) 1-, (b) 3-, and (c) 5-day forecasts, plotted as a function of strike radius and vorticity threshold, and aggregated over multiple spatial locations. Shown is the resolution component (white contours) and the uncertainty component (blue contours) of the BS decomposition.

qualitatively similar but have some quantitative differences.

The skill is significantly less than 1.0 at 1-day lead time, with a maximum value of 0.65. The magnitude of the forecast skill lies somewhere between the values expected for small-scale predictands, such as rainfall, and large-scale predictands such as large-scale flow regimes. For example, Hamill (2012) find 1-day forecasts of 1°-resolution 24-h accumulated rainfall greater than 1 mm to be ~0.4, whereas Frame et al. (2011) find the skill in predicting movements of the North Atlantic jet to be ~1.0 at 1-day lead times. The strike probability depends on multiple scales of motion and can be associated with position uncertainties in the centers of large and perhaps fairly predictable features such as cyclones, and the existence uncertainty in smaller and perhaps inherently unpredictable features such as small kinks on frontal surfaces. The strike probability does
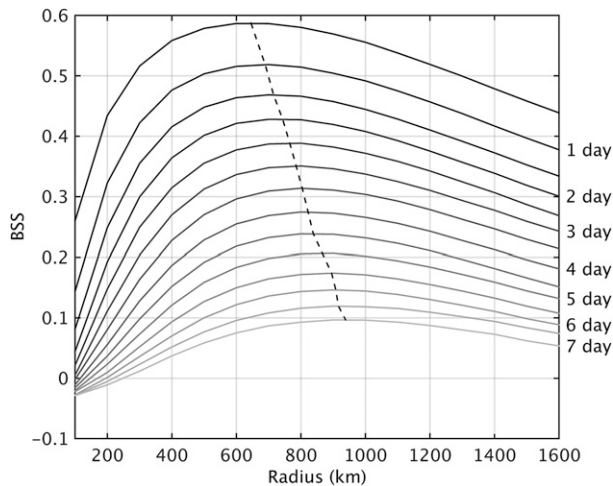
FIG. 7. BSS vs radius for strike probability forecasts of features with vorticity greater than the seventh decile. Each solid line shows a different lead time, with lighter shading indicating longer lead time. Plotted data have a resolution of 100 km. The dashed line indicates the radius at which the max skill occurs. The location of the max was obtained by cubic-spline fitting to the plotted data.

not distinguish between these two paradigms with all features being reduced to a single point in space by identifying turning points in the forecast fields. It may therefore be considered both a synoptic- and mesoscale predictand.

The variation in skill with strike radius can be interpreted as follows. At small radii, the skill will be low since the locations of the features are uncertain. For sufficiently large radii, the forecast strike probability will saturate at the climatological value and therefore provides no new information beyond climatology. The scales for which BSS is maximized provide an estimator of the transition point between these two paradigms. The Brier score is strongly linked to the information content of the forecasts (Weijs et al. 2010), so these scales may also be interpreted as those that contain the most forecast information. Figure 6 suggests that the strike radius at which skill is maximized is an increasing function of lead time. This can be seen clearly in Fig. 7, which shows the area-averaged BSS as a function of strike radius at different lead times for forecasts of features with vorticity above the seventh decile. The strike radius at which maximum skill occurs (shown by the dashed line) increases almost linearly at a rate of $43 \, \text{km day}^{-1}$ (estimated from a linear fit) from about 650 km at 12-h lead time to about 950 km at 7-day lead time. Similar figures for different vorticity thresholds between the first and ninth deciles have very similar results, with the rate of increase in strike radius of maximum skill in the range $38–43 \, \text{km day}^{-1}$. When no

threshold is applied, the strike radius of maximum skill increases more rapidly with lead time at $50 \, \text{km day}^{-1}$. We attribute this difference to the loss of weak features with increasing lead time noted in section 3 essentially leading to a bias in the forecast strike probability relative to the analysis. As can be seen in Fig. 6, for a given lead time, the strike radius at which maximum skill occurs is an increasing function of the vorticity threshold, so that the skill is maximized at smaller radius when lower vorticity thresholds are used. This may be due to the fact that there are larger numbers of small features, so that climatology becomes quite difficult to beat once the radius becomes large.

The fact that maximum skill occurs at such large radii even at very short lead time (e.g., 650-km scale at 12 h) may seem surprising considering that Froude (2010) finds mean position errors of one geodetic degree (~111 km) at day 1 and eight geodetic degrees (~888 km) at day 7; however, it must be remembered that the tracking in Froude (2010) was performed using a different methodology and much coarser–resolution data (T42, ~2.8°) than that used in the present paper, meaning that it focused on large synoptic-scale cyclones. More significantly the analysis performed by Froude (2010) was heavily conditioned on the verifying analysis; for example, all feature tracks that were unmatched to analysis tracks or had position error greater than four geodetic degrees before day 4 were removed. So while Froude (2010) provides useful information about the development of forecast errors in situations that are relatively well forecast, the use of verifying analysis to filter the forecast data would make it an overly optimistic assessment of the predictive capability of models. It is also worth noting that having maximum skill at 650 km does not imply that forecasts are not skillful for smaller radii; for example, Fig. 7 indicates there is still a small amount of skill for 300-km radii out to 7 days.

A more subtle feature of Fig. 6 is the way in which the relative skill at different r and ξ_0 changes with increasing lead time. For a given strike radius, the vorticity threshold at which maximum skill occurs decreases with increasing lead time, moving from the extremes to lower values; see, for example, $r = 600 \, \text{km}$. To see this more clearly, Fig. 8 shows the BSS versus lead time for a strike radius of 600 km and a varying vorticity threshold. It can be seen that for the first two days, the skill is a purely increasing function of the vorticity threshold. However, after the second day the skill of the ninth decile vorticity threshold strike probability forecast dips below that of the eighth decile. This feature may be related to the model drift in the vorticity of features with increasing lead time, with very high vorticity values being
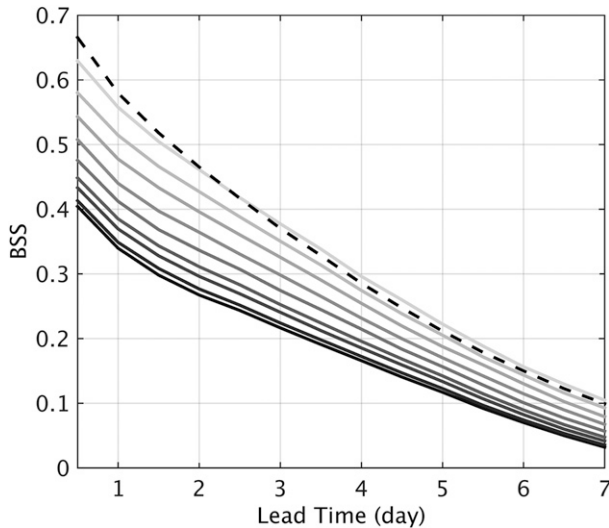
FIG. 8. BSS vs lead time for a strike radius of 600 km. The different lines correspond to different vorticity thresholds. Solid lines shows the zeroth to eighth deciles, with lighter shading indicating a higher threshold. Dashed line shows the ninth decile.

FIG. 9. The lead time (days) at which the BSS of strike probability forecasts reaches zero as a function of strike radius and vorticity threshold.

underrepresented by the forecast model relative to the analysis. Whether this should be viewed as a systematic underestimation of the vorticity of predicted features that could be corrected by recalibration or a failure to predict those features altogether is an open question.

The shapes of the BSS curves shown in Fig. 8 are typical of those calculated for all strike radii and vorticity thresholds, with the Brier skill decaying as a quadratic function of lead time. The lead time at which BSS reaches zero can be interpreted as the practical limit of the predictability of cyclonic features for forecasts produced by MOGREPS-15. This may of course be longer for other forecast models or increased ensemble sizes. Figure 9 shows the lead time at which the BSS reaches zero for different strike radii and vorticity thresholds. The longest lead time for which Brier skill remains above zero is 14 days for features with vorticity above the ninth vorticity decile and strike radius ~950–1000 km. The skill at these scales and intensities at such long lead times is likely due to constraints the larger, more predictable, scales place on the location and propagation of cyclones, as was illustrated in Fig. 1d. It may also be associated with the longer lifetime associated with strong cyclones.

Figures 10 and 11 (color shading) show the radius at which maximum skill occurs and the density of features with vorticity above the seventh decile at different geographic locations calculated from the analysis fields. For reference, the black contours in Fig. 10 show the corresponding maximum BSS. We show the radius at which skill is maximized as opposed to the radius at
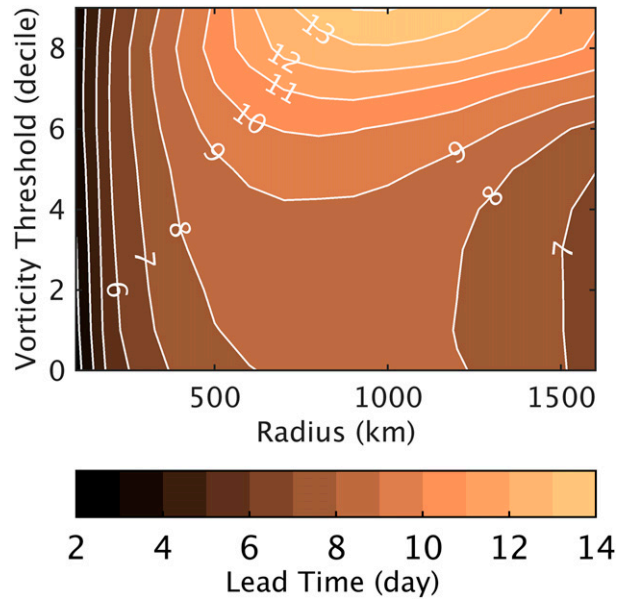
which skill meets a particular threshold since this provides an upper bound on the skill. For strike radii larger or smaller than this, the skill will be lower. There is a clear, if imperfect, inverse proportionality between cyclone density and the radius at which skill is maximized, which becomes stronger as lead time increases. We attribute the existence of this proportionality to the strength of the constraint the large-scale flow has on the evolution of cyclones. For example, the high density of cyclones in the northwestern region of Fig. 11 is associated with the jet stream, with a region of particularly high density near the tip of Greenland. This second feature is likely due to lee cyclogenesis at the tip of Greenland. Similar density maxima are seen in studies such as that of Hoskins and Hodges (2002). The jet region provides both the baroclinicity needed for cyclone development and the strong steering that constrains the cyclone's trajectory to remain within it. We attribute the high skill at relative small radius to the strong steering by the large-scale and hence more predictable flow. This region is also associated with larger cyclones, which have longer lifetimes (Rudeva and Gulev 2007) and may therefore be more predictable. In contrast, the southeastern half of the domain has greatest skill at much larger radius. We attribute this to the lack of strong steering from the large-scale flow, and to the predominance of secondary cyclones or frontal waves in this area of the domain. Figures 6–9 show area-averaged quantities, so they will be either over- or underestimates for many specific locations. Regions where the
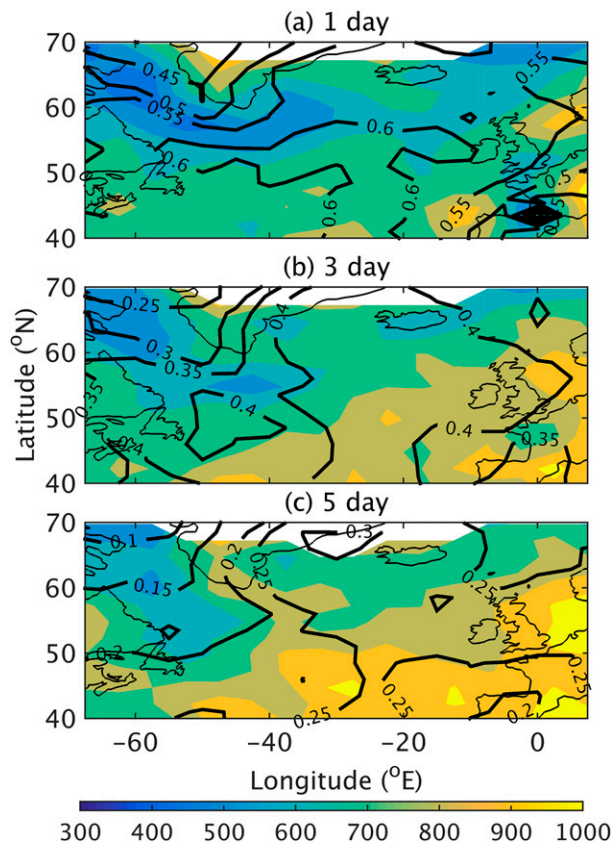
FIG. 10. Color shading with the radius (km) at which the max BSS of a forecast of features with vorticity above the seventh decile occurs. Values are calculated on a 2.5° × 2.5° grid and smoothed for plotting using a 2 × 2 gridpoint–averaging kernel. Black contours show the corresponding values of BSS.

FIG. 11. Areal number density of features $(10^{-7}\,km^{-2})$ with vorticity above the seventh decile, estimated using 300-km-radius flat circular kernel.

maximum skill is low are expected to reach the zero skill limit of predictability sooner. For example, the maximum skill is noticeably lower in the northwestern region of Fig. 10c so that we expect skill in this region to reach zero at shorter lead times.

5. Summary and conclusions

This paper has summarized the statistical properties and forecast skill of objectively tracked cyclonic features in MOGREPS-15. It has been shown that there is a noticeable reduction in the areal density of features in the forecast relative to the analysis, which saturates at about 7 days lead time for barotropic lows and 5 days for frontal waves. The perturbed forecasts saturate at a slightly higher feature density than do those of the control, which is attributable to the presence of stochastic parameterization in the perturbed forecast. This stochastic effect may simply be that the increased vorticity in the stochastically forced forecast members puts
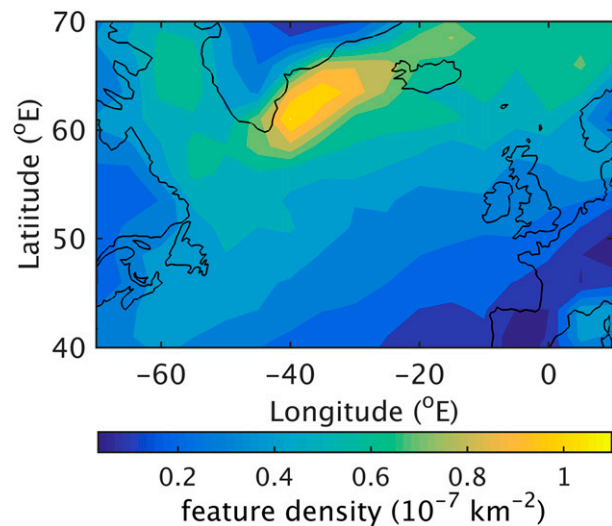
larger numbers of features over thresholds built into the objective identification algorithm, or it may be a more subtle (and perhaps more physical) effect.

The forecast skill of strike probabilities has been assessed. It is found that the maximum skill is largely uniform in the Atlantic to the east of Greenland; however, the strike radius at which this occurs is much smaller along the region associated with the North Atlantic storm track, indicating that it is easier to forecast the locations of cyclones in this region. Furthermore, it is found that for a given vorticity threshold as lead time increases, the radius at which forecast skill is maximized increases linearly with time. The longest lead time at which the skill of strike probability forecasts remains greater than zero is found to be ~14 days but only for cyclonic features with vorticity greater than the ninth decile of the vorticity distribution and for strike radii ~1000 km. When all features are accounted for, the Brier skill score falls to zero within 9 days on average, for any strike radius.
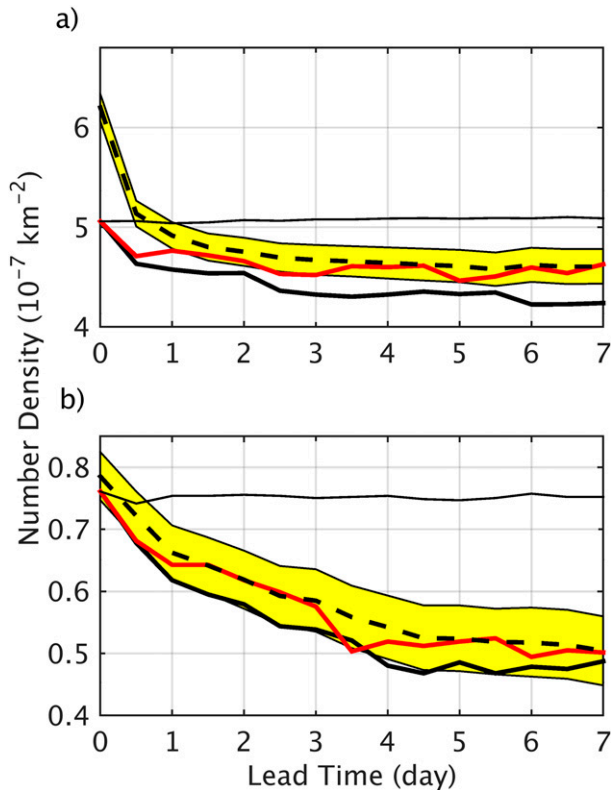
a)

b)

FIG. A1. As in Fig. 3, but for a subset of analysis times from 1 Apr to 1 Dec 2012 in which one stochastically forced member is run from unperturbed initial conditions. Red line shows values computed for this member.

APPENDIX

Impact of Stochastic Parameterizations

Starting in April 2012 MOGREPS-15 was constructed from one unperturbed control member without stochastic parameterizations, 22 perturbed members with stochastic parameterizations, and one unperturbed member with stochastic parameterizations. This configuration provides a convenient means of determining whether stochastic parameterization is the cause of differences in the statistics of cyclonic features between the control and perturbed members described in section 3. Figure A1 shows the evolution of the mean areal density of features over the forecast lead time for the subset of data in which MOGREPS-15 has been configured with an unperturbed stochastic member. The lines and shading are as in Fig. 3, with the addition of the red line, which shows values for the unperturbed stochastic member. There is clear evidence that while the control member is distinct from the perturbed members the unperturbed stochastic member is not. This is more evident when considering the area-averaged enstrophy
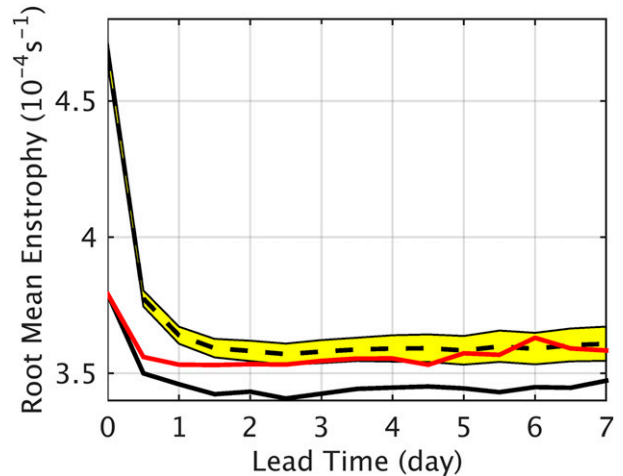

FIG. A2. The area-averaged 850-hPa enstrophy as a function of lead time, calculated for the North Atlantic control region from $1° \times 1°$ wind fields using all forecasts starting in April, July, and October 2012. Shown is the control forecast fields (thick solid line) and the perturbed forecast fields (dashed line). Yellow shading indicates the 99th percentile region for resampling between perturbed ensemble members. Red line shows values for the ensemble member run with stochastic parameterizations but without initial condition perturbation.

shown in Fig. A2. The enstrophy of the control members remains distinct from that of the perturbed members throughout the forecast, whereas the enstrophy of the members with stochastic parameterizations but without initial condition perturbations converges rapidly toward the enstrophy of the perturbed members. Within 2 days it is closer in terms of enstrophy to the perturbed members than the control, and by 5 days, at the latest, it is indistinguishable from the perturbed members.

REFERENCES

Bengtsson, L., K. I. Hodges, and L. S. R. Froude, 2005: Global observations and forecast skill. *Tellus*, **57A**, 515–527, doi:10.1111/j.1600-0870.2005.00138.x.

Bowler, N. E., A. Arribas, 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, doi:10.1002/qj.394.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Frame, T. H. A., M. H. P. Ambaum, S. L. Gray, and J. Methven, 2011: Ensemble prediction of transitions of the North Atlantic eddy-driven jet. *Quart. J. Roy. Meteor. Soc.*, **137**, 1288–1297, doi:10.1002/qj.829.

Froude, L. S., 2010: TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. *Wea. Forecasting*, **25**, 819–836, doi:10.1175/2010WAF2222326.1.

Hamill, T. M., 2012: Verification of TIGGE multimodel and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States. *Mon. Wea. Rev.*, **140**, 2232–2252, doi:10.1175/MWR-D-11-00220.1.

——, and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905–2923, doi:10.1256/qj.06.25.

Hawcroft, M., L. Shaffrey, K. Hodges, and H. Dacre, 2012: How much Northern Hemisphere precipitation is associated with extratropical cyclones? *Geophys. Res. Lett.*, **39**, doi:10.1029/2012GL053866.

Hewson, T. D., 1998: Objective fronts. *Meteor. Appl.*, **5**, 37–65, doi:10.1017/S1350482798000553.

——, and H. A. Titley, 2010: Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution. *Meteor. Appl.*, **17**, 355–381, doi:10.1002/met.204.

Hoskins, B. J., and K. I. Hodges, 2002: New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**, doi:10.1175/1520-0469(2002)059<1041:NPOTNH>2.0.CO;2.

Jolliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* 2nd ed. J. Wiley and Sons, 292 pp.

Leith, C., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418, doi:10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, doi:10.1002/qj.49712252905.

Morss, R. E., J. K. Lazo, B. G. Brown, H. E. Brooks, P. T. Ganderton, and B. N. Mills, 2008: Societal and economic research and applications for weather forecasts: Priorities for the North American THORPEX program. *Bull. Amer. Meteor. Soc.*, **89**, 335–346, doi:10.1175/BAMS-89-3-335.

Neu, U., and Coauthors, 2013: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bull. Amer. Meteor. Soc.*, **94**, 529–547, doi:10.1175/BAMS-D-11-00154.1.

Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: Preliminary results on comparing and combining ensembles. *Quart. J. Roy. Meteor. Soc.*, **134**, 2029–2050, doi:10.1002/qj.334.

Rudeva, I., and S. K. Gulev, 2007: Climatology of cyclone size characteristics and their changes during the cyclone life cycle. *Mon. Wea. Rev.*, **135**, 2568–2587, doi:10.1175/MWR3420.1.

——, ——, I. Simmonds, and N. Tilinina, 2014: The sensitivity of characteristics of cyclone activity to identification procedures in tracking algorithms. *Tellus*, **66A**, 24961, doi:10.3402/tellusa.v66.24961.

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

Swinbank, R., and Coauthors, 2015: The THORPEX Interactive Grand Global Ensemble (TIGGE) and its achievements. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-13-00191.1, in press.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158, doi:10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.

Weijs, S. V., R. Van Nooijen, and N. Van De Giesen, 2010: Kullback–Leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition. *Mon. Wea. Rev.*, **138**, 3387–3399, doi:10.1175/2010MWR3229.1.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.

Zappa, G., L. C. Shaffrey, and K. I. Hodges, 2013: The ability of CMIP5 models to simulate North Atlantic extratropical cyclones. *J. Climate*, **26**, 5379–5396, doi:10.1175/JCLI-D-12-00501.1.