

Impact of Resolution and Design on the U.S. Navy Global Ensemble Performance in the Tropics

CAROLYN A. REYNOLDS, JUSTIN G. MCLAY, AND JAMES S. GOERSS

Naval Research Laboratory, Monterey, California

EFREN A. SERRA

DeVine Consulting, Inc., Fremont, California

DANIEL HODYSS AND CHARLES R. SAMPSON

Naval Research Laboratory, Monterey, California

(Manuscript received 22 June 2010, in final form 10 January 2011)

ABSTRACT

The performance of the U.S. Navy global atmospheric ensemble prediction system is examined with a focus on tropical winds and tropical cyclone tracks. Ensembles are run at a triangular truncation of T119, T159, and T239, with 33, 17, and 9 ensemble members, respectively, to evaluate the impact of resolution versus the number of ensemble member tradeoffs on ensemble performance. Results indicate that the T159 and T239 ensemble mean tropical cyclone track errors are significantly smaller than those of the T119 ensemble out to 4 days. For ensemble forecasts of upper- and lower-tropospheric tropical winds, increasing resolution has only a small impact on ensemble mean root-mean-square error for wind speed, but does improve Brier scores for 10-m wind speed at the 5 m s^{-1} threshold. In addition to the resolution tests, modifications to the ensemble transform initial perturbation methodology and inclusion of stochastic kinetic energy backscatter are also evaluated. Stochastic kinetic energy backscatter substantially increases the ensemble spread and improves Brier scores in the tropics, but for the most part does not significantly reduce ensemble mean tropical cyclone track error.

1. Introduction

A practical way to deal with inevitable uncertainties in atmospheric forecasting is through ensemble forecasting. Operational forecasting centers have been providing global ensembles with initial-state perturbations since the early 1990s (e.g., Toth and Kalnay 1993; Buizza and Palmer 1998). For tropical cyclone (TC) applications, previous research has shown that the mean track from an ensemble composed of different deterministic forecasts may have smaller error than the tracks associated with any one member of the “multimodel” ensemble (e.g., Goerss 2000; Williford et al. 2003). Other work shows utility in single-model ensemble mean TC track forecasts as well (e.g., Puri et al. 2001; Hamill et al. 2011).

As part of the National Oceanic and Atmospheric Administration (NOAA) Hurricane Forecast Improvement Project (HFIP), a series of experiments have been performed to examine the utility of the Navy’s global atmospheric ensemble forecast system for TC track prediction. In this study, the relative performance of different ensemble configurations is examined in regards to ensemble mean TC track error, as well as lower- and upper-tropospheric tropical winds, which are of strategic interest to the U.S. Navy.

Under practical computational constraints there will always be a trade-off between the resolution of the individual ensemble forecasts and the number of ensemble members. This trade-off is explored by comparing ensembles with more members run at a lower resolution to ensembles with fewer members run at a higher resolution. In the resolution versus member number experiments, the ensembles are all produced using the same initial perturbation methodology. This methodology is

Corresponding author address: Carolyn Reynolds, Naval Research Laboratory, 7 Grace Hopper Ave., Monterey, CA 93943-5502.
E-mail: carolyn.reynolds@nrlmry.navy.mil

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE JUL 2011	2. REPORT TYPE	3. DATES COVERED 00-00-2011 to 00-00-2011			
4. TITLE AND SUBTITLE Impact Of Resolution And Design On The U.S. Navy Global Ensemble Performance In The Tropics		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Research Laboratory, Monterey, CA, 93943		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES Monthly Weather Review, Washington: Jul 2011, Vol. 139, Iss. 7; pg. 2145, 11 pgs					
14. ABSTRACT The performance of the U.S. Navy global atmospheric ensemble prediction system is examined with a focus on tropical winds and tropical cyclone tracks. Ensembles are run at a triangular truncation of T119, T159, and T239, with 33, 17, and 9 ensemble members, respectively, to evaluate the impact of resolution versus the number of ensemble member tradeoffs on ensemble performance. Results indicate that the T159 and T239 ensemble mean tropical cyclone track errors are significantly smaller than those of the T119 ensemble out to 4 days. For ensemble forecasts of upper- and lower-tropospheric tropical winds, increasing resolution has only a small impact on ensemble mean root-mean-square error for wind speed, but does improve Brier scores for 10-m wind speed at the 5 m s⁻¹ threshold. In addition to the resolution tests, modifications to the ensemble transform initial perturbation methodology and inclusion of stochastic kinetic energy backscatter are also evaluated. Stochastic kinetic energy backscatter substantially increases the ensemble spread and improves Brier scores in the tropics, but for the most part does not significantly reduce ensemble mean tropical cyclone track error.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 12	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

the ensemble transform (ET) technique (Bishop and Toth 1999; McLay et al. 2008), where initial perturbations are produced from a global transformation of short-term ensemble forecast perturbations such that they are consistent with analysis error variance estimates.

There are aspects of the practical application of the ET described in McLay et al. (2008) that can be improved upon. Primarily, given a finite number of ensemble members, the initial ensemble perturbations are too small in the tropics and too large in the midlatitudes, when compared to the analysis error variance estimate produced by the Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System (NAVDAS; Daley and Barker 2001). McLay et al. (2007) illustrate how an archive of ensemble forecast perturbations can be used in a postprocessing approach to enhance the initial ensemble perturbation variance in regions where it is below the given estimate of analysis error variance. Most recently, McLay et al. (2010) describe how performing local ensemble transformations within latitude bands results in a better match to the initial analysis error variance estimate as well as enhanced ensemble performance. The impact of this new “banded” ET on ensemble performance in the tropics will be examined here. The banded ET ensembles considered here are run at a higher resolution and with a slightly different formulation than those described in McLay et al. (2010).

The problem of too little ensemble spread in the tropics has also been attributed to, in part, the neglect of model error in ensemble formulation, which is expected to be greater in the tropics than in the midlatitudes. Recent work on this issue includes the addition of a stochastic term to the physical parameterizations (e.g., Buizza et al. 1999; Shutts and Palmer 2004), adding stochastic perturbations to the tendencies from the convective parameterization (Teixeira and Reynolds 2008; Reynolds et al. 2008), using different parameterizations within the ensemble prediction system (e.g., Houtekamer et al. 1996), and stochastic kinetic energy backscatter (SKEB; Shutts 2005; Berner et al. 2009; Bowler et al. 2009; Charron et al. 2010). In this study the impact of adding SKEB to the banded ET ensemble on ensemble performance in the tropics is examined in the Navy global ensemble system.

This paper is organized as follows: section 2 describes the methodology, section 3 presents the results, and section 4 contains a summary and brief discussion of future work.

2. Methodology

a. Ensemble transform initial perturbation scheme

The ET initial perturbation scheme applied over the globe (global ET) is described in Wei et al. (2006, 2008)

and McLay et al. (2007, 2008). A brief description is provided here, with details found in McLay et al. (2007). Let \mathbf{P}_g^a be an $N \times N$ matrix whose elements represent the best available guess of the error covariance of the initial condition for an NWP forecast, where N is the number of variables defining the model state. Let $\mathbf{Z}^f = [\mathbf{z}_1^f, \mathbf{z}_2^f, \dots, \mathbf{z}_K^f]$ be an $N \times K$ matrix whose i th column is given by

$$\mathbf{z}_i^f = \mathbf{x}_i^f - \overline{\mathbf{x}_i^f}, \quad (1)$$

where $\overline{\mathbf{x}_i^f} = 1/K \sum_{i=1}^K \mathbf{x}_i^f$, \mathbf{x}_i^f is the i th member of a given forecast ensemble, K is the number of ensemble members, and $\overline{\mathbf{x}_i^f}$ represents the mean of the forecast ensemble. In the ET, K analysis perturbations $\mathbf{Z}^a = [\mathbf{z}_1^a, \mathbf{z}_2^a, \dots, \mathbf{z}_K^a]$ are generated using the following formula:

$$\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T}, \quad (2)$$

which is subject to the following constraint:

$$\mathbf{Z}^{aT} (\mathbf{P}_g^a)^{-1} \mathbf{Z}^a = \mathbf{N}\mathbf{I}, \quad (3)$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K]$ is assumed to be a real symmetric $K \times K$ matrix of weighting coefficients. See the appendix of McLay et al. (2007) for further description of \mathbf{T} . Equation (3) insures that the perturbations are globally consistent with estimates of analysis error covariance. The mean of the initial perturbations is constrained to be zero, so $K - 1$ initial perturbations are independent. In the present study, \mathbf{P}_g^a is obtained from NAVDAS, and is diagonal. The analysis perturbations themselves do have spatial and multivariate correlations reflecting the fact that they are linear combinations of short-term forecast perturbations (e.g., the global ET perturbation total energy distribution peaks in synoptic-scale wavenumbers, see Fig. 3 from McLay and Reynolds 2009).

As noted in the introduction, because of limited ensemble size, the global ET described in McLay et al. (2008) results in initial perturbations that are too large in the extratropics and too small in the tropics when compared to the analysis error variance estimate. As discussed in McLay et al. (2010), localization of the transform operation in the ET scheme provides a closer fit to the analysis error estimate, and superior ensemble performance under a variety of metrics. The technique is conceptually similar to that used in the local ensemble transform Kalman filter (LETKF; Szunyogh et al. 2008), and the ensemble transform Kalman filter implemented at the Met Office (Bowler et al. 2009). McLay et al. (2010) consider different types of localization, including blocked (localization in latitude–longitude blocks) and

banded (localization in latitudinal bands). In these experiments, a formulation of the local ET based on five latitude bands (banded ET) is employed. Based on T119 16-member ensemble tests, McLay found the 5-banded ET to be superior to the global ET under a variety of metrics, including Brier scores for 10-m wind speeds exceeding 10 m s^{-1} , and ensemble mean root-mean-square error (RMSE) of 500-hPa heights and 925-hPa winds, in the tropics and Northern Hemisphere extratropics.

The implementation of the five-banded ET reported on here is modified from the implementation described in McLay et al. (2010) by the application of a scaling factor in the tropical band. The scaling factor is applied because inspection of the unmodified banded ET produced perturbations to the TC location that were unreasonably large at initial time and did not grow for the first few forecast days. Based on preliminary tests a scaling factor of 0.3 was chosen, although there was not strong sensitivity in ensemble mean track error to the application of the scaling factor. While this scaling reduces the size of the initial perturbations in the tropics, the wind field perturbations are still substantially larger in the banded ET than in the global ET, as will be shown in the results.

b. Stochastic kinetic energy backscatter

SKEB (Shutts 2005; Berner et al. 2009) is an algorithm that attempts to correct for the kinetic energy drain in NWP models introduced through horizontal diffusion as well as the parameterization of gravity wave drag and deep convection, through the incremental introduction of kinetic energy during the numerical integration. In addition, SKEB has also been shown to correct, in part, the tendency of NWP models to exhibit a deficient kinetic energy spectrum at the smaller resolved scales. The formulation of SKEB employed in this study closely follows the formulation of Shutts (2005) and Berner et al. (2009). The notable details are as follows:

- The SKEB forcing is spectral and autoregressive, and is introduced into the rotational component of the flow.
- The spectral power of the forcing is prescribed by the function $g(n) = (1 + n)^{-1.27}$, where n is the total wavenumber, as suggested by the experiments of Berner et al. (2009) with coarse-grained cloud-resolving simulations. This results in a forcing kinetic energy spectrum $E_n \approx n^{0.64}$. The kinetic energy injection is applied at all scales but is largest for large wavenumbers (small scales) as shown in Fig. 2b of Berner et al. (2009). Berner et al. found that when forcing only the wavenumbers closest to the truncation scale the errors did not cascade upscale fast enough. They

hypothesize that this might be due to not forcing the right multivariate structures, or it may reflect that certain model errors are associated with larger scales, such as organized convection in the tropics.

- The formulation used here does not employ a dissipation mask and is statistically uniform at all levels. Preliminary results testing the use of a convective dissipation mask that focuses the forcing on regions where deep convection is likely to be dissipating kinetic energy indicate that this type of mask leads to some enhancement of ensemble performance (discussed in more detail in section 3b).
- The autoregression parameter is chosen to give a forcing decorrelation time of roughly 30 min, based upon tuning experiments with NOGAPS. This contrasts with Berner et al. (2009), where the value of the autoregression parameter is chosen to give a longer forcing decorrelation time of 6 h.
- At each time step of the integration the global root-mean-square (rms) amplitude of the SKEB vorticity increment is constrained to be 1.5% of the unforced global rms amplitude of vorticity. This constraint on the global rms amplitude of the SKEB vorticity increment serves the purpose of the backscatter ratio tuning parameter of Berner et al. (2009). The global average RMS forcing on the wind field is 0.035 m s^{-1} per time step (200 s). This is consistent with the magnitude of the u increments, which vary between 0.06 and -0.05 m s^{-1} , shown in Fig. 3 of Bowler et al. (2009) for the implementation of SKEB in the Met Office ensemble system.

With ensemble experiments using the European Centre for Medium-Range Weather Forecasts (ECMWF) model, Berner et al. (2009) illustrate how SKEB can improve ensemble performance. Specifically, they show a better spread–error relationship, more realistic kinetic energy spectra, and better representation of forecast error growth, along with improved rainfall forecasts and better probabilistic skill with the addition of SKEB to the ECMWF ensemble. They also find these improvements were most pronounced in the tropics. It is therefore hoped that SKEB will likewise have a positive influence on the performance of ensembles examined here and may also have a positive impact on TC track errors.

Other weather prediction centers have also found improved performance through stochastic backscatter schemes. A stochastic backscatter scheme has been implemented in the global ensemble prediction system at the Meteorological Service of Canada (Charron et al. 2010), injecting energy primarily at scales in the total wavenumber range between 40 and 128. In their system, which also includes multiple parameterizations and

TABLE 1. Configuration of the ensemble experiments.

Name	Resolution	No. of members	Initial perturbations	Model perturbations
G119	T119 (110 km)	33	Global ET	None
G159	T159 (83 km)	17	Global ET	None
G239	T239 (55 km)	9	Global ET	None
B159	T159 (83 km)	17	Banded ET	None
S159	T159 (83 km)	17	Banded ET	SKEB

stochastic physical tendencies, they find that the stochastic forcing improves the reliability of the forecasts through impacts on ensemble dispersion. As noted above, a stochastic kinetic energy backscatter scheme has also been implemented in the global ensemble system at the Met Office in the United Kingdom (Bowler et al. 2009). In their formulation, SKEB compensates only for dissipation due to the horizontal diffusion and advection interpolation errors (not the physical parameterizations). The forcing peaks on scales that range from 1000 to 2000 km. They found that, at short lead times, the addition of SKEB significantly improves the skill for mean sea level pressure, 6-h accumulated precipitation, temperature at 250 and 500 hPa, and 250-hPa wind speed, but degraded skill for 2-m temperature.

c. Model description and experimental design

The results shown are based on ensemble forecasts using the Navy Operational Global Atmospheric Prediction System (NOGAPS; Peng et al. 2004), the global spectral weather prediction model of the U.S. Navy. The physical parameterizations include boundary layer turbulence (Louis et al. 1982), shallow and deep moist convection (Emanuel and Zivkovic-Rothman 1999; Peng et al. 2004), convective and stratiform clouds (Teixeira and Hogan 2002), and solar and longwave radiation (Harshvardhan et al. 1987). All ensembles are run with a 6-h cycling period, with 7-day forecasts produced at 0000 and 1200 UTC for the period 26 June–15 October 2008.

The details of the different ensemble configurations are shown in Table 1. The first ensemble configuration is the global ET ensemble run at triangular truncation 119 (approximately 110-km horizontal resolution) and 30 levels (T119L30) with 33 members (32 perturbed members plus 1 member without initial perturbations), and is referred to as G119. The current operational ensemble truncation is also T119L30. To examine the impact of the trade-off between resolution and number of members on ensemble performance, a global ET ensemble is run at a higher resolution, T159 truncation (approximately 83 km), but with only 17 members (16 perturbed members plus 1 member without initial

perturbations), referred to as G159. The G159 has a slightly higher total computational cost than G119 (20%), given that the higher computational cost for each forecast is largely offset by the smaller number of ensemble members. A global ET ensemble is also run at the same resolution as the operational deterministic model during 2008, T239 truncation (approximately 55 km), with 9 members (G239). This ensemble configuration has approximately twice the computational expense as the G119 configuration.

The impact of initial perturbation design and stochastic forcing on ensemble performance is also examined. An ensemble is run at T159 truncation with 17 members, but uses the 5-banded ET method (referred to as B159) to produce the initial perturbations. The banded ET results in significant improvements over the global ET for low-resolution (T119) ensembles (McLay et al. 2010) for a variety of measures, and has since been transitioned to operations at the Fleet Numerical Meteorology and Oceanography Center. To examine the impact of stochastic forcing on ensemble performance in the tropics, SKEB perturbations are added to the forecast integrations of the banded ET ensemble, again at the T159 truncation (referred to as S159).

3. Results

a. Experiments varying resolution and number of members

For the experiments considered here, increasing resolution while decreasing the number of members does not have a large impact on ensemble mean RMSE. Figure 1a shows the RMSE for the 850-hPa wind speed for the tropics (20°S–20°N) as a function of forecast time. Throughout the integration time, increases in resolution result in slightly improved (lower) RMSE (the control analyses are used for verification). Figure 1b shows the RMSE for the 200-hPa wind speed in the tropics. There are very small improvements seen in the G159 ensemble over the G119 ensemble, and basically no difference between G159 and G239.

The Brier scores (which may be interpreted as mean square errors of the ensemble-based probability of an event; Wilks 2006) for 10-m wind speed in the tropics at two thresholds (5 and 10 m s⁻¹) are shown in Fig. 2. The significance of the Brier score differences is determined using a moving block bootstrap technique as described in McLay and Reynolds (2009). At the 5 m s⁻¹ threshold (Fig. 2a), the lowest (best) scores are obtained from the high-resolution (G239) ensemble, and the improvement is statistically significant at all forecast times. Contrasting results are found for the 10 m s⁻¹ threshold

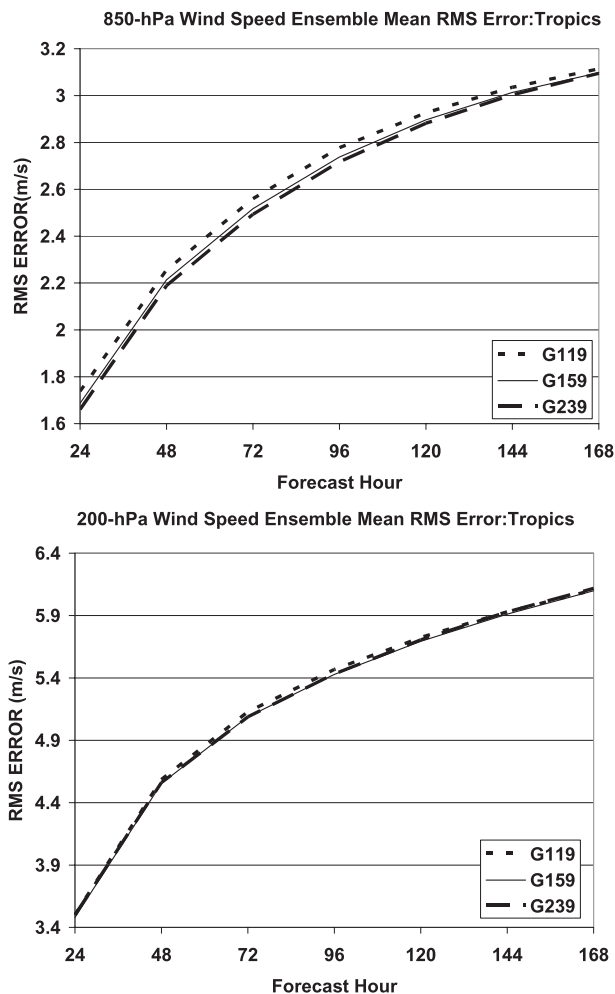


FIG. 1. RMSE ($m s^{-1}$) of the wind speed of the ensemble mean in the tropics for the G119 (dotted), G159 (thin solid), and G239 (thick dashed) ensembles for the (top) 850- and (bottom) 200-hPa level.

(Fig. 2b), where G119 appears to give the best performance. However, this benefit is not statistically significant, and is derived from the larger number of ensemble members in the G119 experiments, not through the skill of the individual ensemble members. Results not shown indicate that using only 17 ensemble members instead of the full set of 33 increases the Brier scores (decreases the quality) for the G119 experiment.

Figure 3 shows the homogenous comparison of ensemble mean track errors as a function of forecast time for all Northern Hemisphere TCs that occurred during the ensemble test period (26 June–15 October 2008). The number of verifying forecasts ranges from 356 at 12 h to 86 at 120 h. Track errors and statistical significance are computed using the Automated Tropical Cyclone Forecast (ATCF) system (Sampson and Schrader 2000). In addition to the results for the G119, G159, and

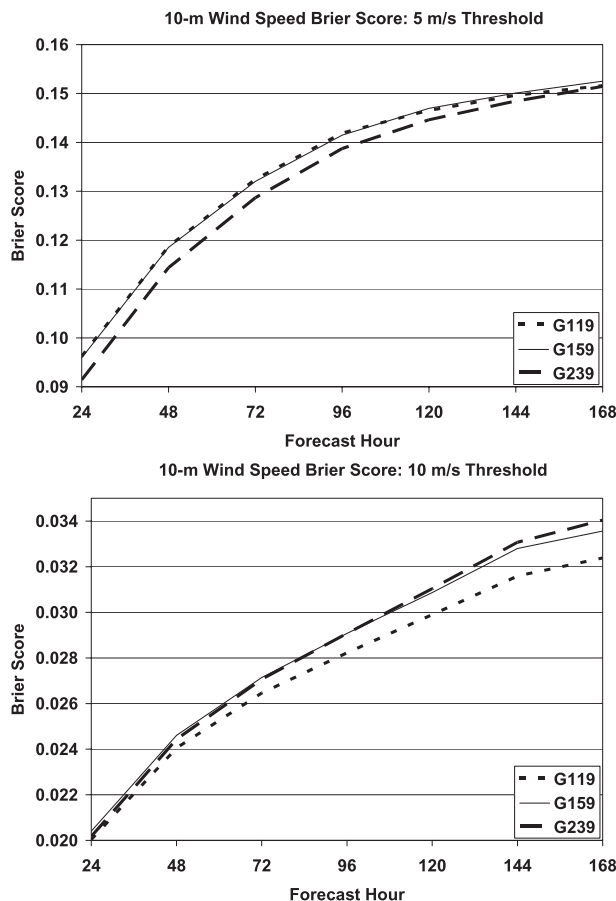


FIG. 2. Brier score for the G119 (dotted), G159 (thin solid), and G239 (thick dashed) ensembles for the (top) 5 and (bottom) 10 $m s^{-1}$ wind speed threshold. The differences between G239 and G119 at the 5 $m s^{-1}$ threshold are statistically significant. The differences at the 10 $m s^{-1}$ threshold are not statistically significant.

G239 ensemble means, the TC track error from the operational NOGAPS T239 deterministic forecast is also shown. The G159 and G239 ensemble mean track errors are both smaller than those of G119, and these differences are statistically significant at the 95% level (accounting for serial correlations) up to and including the 96-h lead time, and at the 90% level at the 120-h lead time. The gain from going to the G239 ensemble over the G159 ensemble is smaller than the gain from going to the G159 ensemble over the G119 ensemble, and this gain is significant at the 95% level only at 48 and 72 h (significant at the 90% level at 24 and 36 h). When compared to the T239 deterministic control, the G239 ensemble mean has comparable errors (no significant difference) after 12 h. While both G159 and G239 have smaller errors than the control at 96 and 120 h, these differences are not statistically significant.

Ensemble mean track error comparisons done separately for the Atlantic, western North Pacific, and eastern

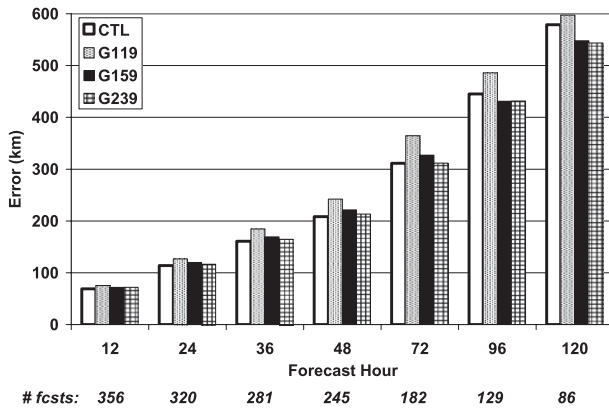


FIG. 3. Homogenous NH TC track forecast error (km), for G119, G159, and G239 ensemble mean tracks as denoted in key. Also shown is the average forecast error of the T239L30 NOGAPS operational deterministic forecast (CTL). The numbers of verifying forecasts are shown below the x axis. The differences between G119 and G159 are statistically significant at the 95% level out to 96 h. The differences between G159 and G239 are statistically significant at 48 and 72 h.

North Pacific basins give results very similar to the all-basin comparison shown in Fig. 3. For the Atlantic basin, G159 and G239 have significantly lower track errors than G119 for all forecast times from 12 to 96 h. For the eastern North Pacific basin, G159 and G239 ensembles have significantly lower forecast errors than G119 for forecast times from 36 to 120 h. Similar results are seen in the western North Pacific, with statistically significant improvements for forecasts times from 24 to 96 h. While G159 and G239 give better forecasts than G119 in all three basins, the relative improvement was largest in the western North Pacific. The western North Pacific also exhibited the most improvement for G239 over G159, with significant improvement at forecast times of 24, 48, and 72 h.

Experiments are also performed using only 9 or 17 members from the T119 ensemble to calculate ensemble mean track error (not shown). Comparison with the G119 (33 member) errors indicates only a slight reduction in ensemble mean track error as the number of ensemble members is increased. These results are consistent with the hypothesis that model error is an important component of ensemble mean TC track error, and increasing the number of ensemble members using the same forecast model does not account for this model error.

Because of the coarse resolution of the ensembles, even at G239, intensity forecasts have not been considered here. Previous research (e.g., Sampson et al. 2008) has shown that global model forecasts do not provide skillful intensity forecasts as compared to statistical

forecasts, attributed to the fact that they cannot simulate the inner core of a tropical cyclone (e.g., Knaff et al. 2007). However, in the future, especially as computational resources allow for ever higher-resolution ensemble forecasts, the skill of global ensemble intensity forecasts should be assessed.

In summary, the results from the resolution experiments show that the optimal configuration will be a function of metric, particularly in regards to the Brier score for different threshold values. These results also indicate significant reductions in the TC mean track errors when going from the lower-resolution G119 ensemble to the G159 ensemble. Smaller improvements are seen when going from the G159 ensemble to the G239 ensemble, significant at 48 and 72 h. The promise of increased skill in TC track error for G159 over G119 prompts exploration of potential improvements in ensemble design at the T159 truncation, as discussed in section 3b.

b. Sensitivity to ensemble design

Figure 4 shows the ensemble mean RMSE for wind speed in the tropics at the 850- and 200-hPa levels for the three T159 ensembles (G159, B159, and S159). At both the 850- and 200-hPa levels, B159 shows some improvement over G159 at most forecast times out to 120 h, although these differences are small. The addition of SKEB improves things further between 96 and 168 h. Much larger differences can be found when examining the ensemble spread (Fig. 5). Note the difference in the values of the y axis between Figs. 4 and 5, indicating that for all ensembles considered here, the ensemble spread is underdispersive (smaller than the ensemble mean RMSE). B159 has more spread than G159 at early lead times, which is desirable given that the G159 ensemble is generally underdispersive in the tropics. However, this difference decreases as the forecast length increases, and is negligible after 120 h.

In contrast to the impact of the banded ET, the addition of SKEB increases the ensemble spread substantially at all lead times. Because the ET is a cycling scheme, the addition of SKEB changes not only the long forecast integrations, but also impacts the initial perturbations. For the 850-hPa wind speed, S159 ensemble spread at initial time is greater than that of B159 or G159. However, the S159 growth rate of ensemble spread, while larger than the B159 growth rate between 0 and 24 h, is smaller than the B159 growth rate after 24 h (Table 2). The growth rate for G159 is larger than B159 or S159 throughout the integration time. For the 200-hPa wind speed, S159 ensemble spread and B159 ensemble spread are approximately equal at initial time. The growth rate for S159 is larger than that of B159 from

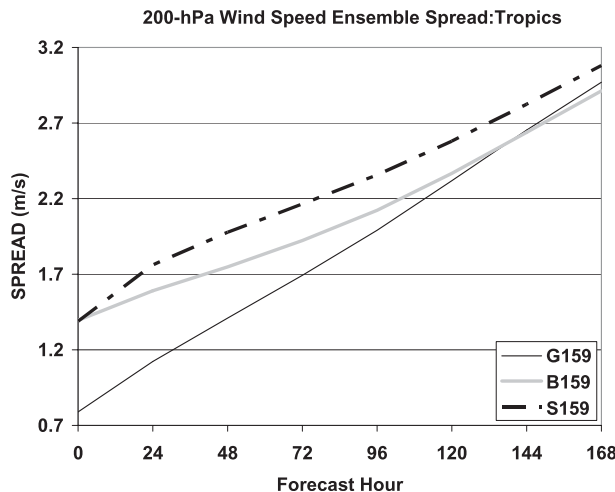
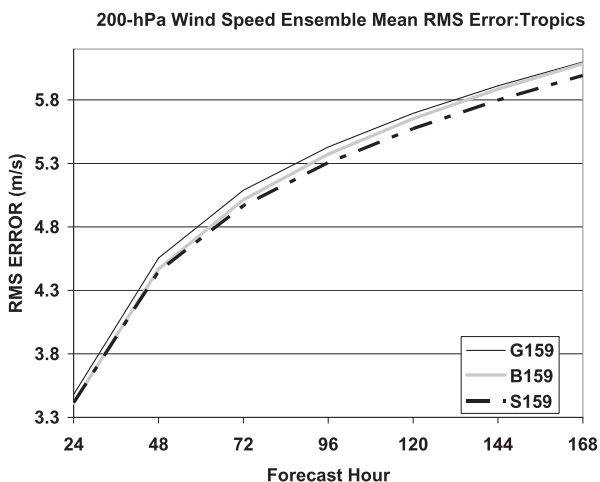
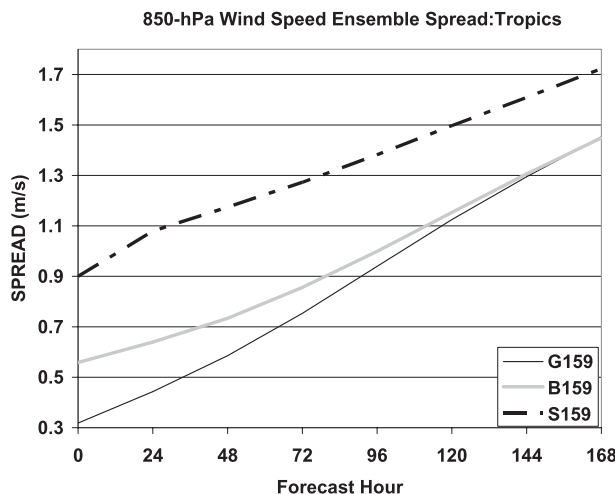
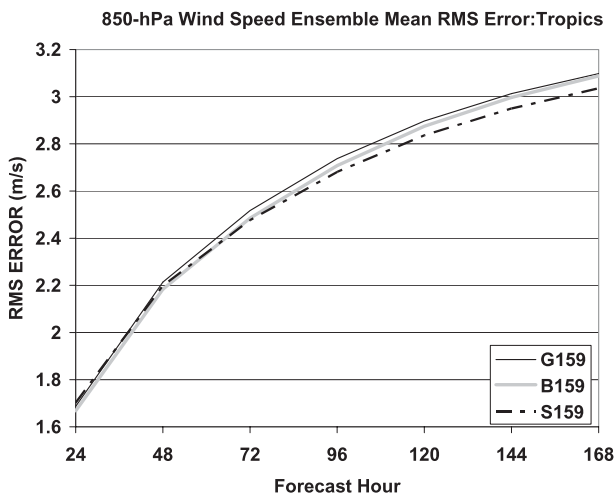


FIG. 4. RMSE ($m s^{-1}$) of the wind speed of the ensemble mean for the G159 (thin solid black), B159 (thick solid gray), and S159 (thick dash-dot) ensembles for the tropics at the (top) 850- and (bottom) 200-hPa level.

FIG. 5. Average standard deviation of the ensemble members about the ensemble mean, or spread ($m s^{-1}$) of the wind speed for the G159 (thin solid black), B159 (thick solid gray), and S159 (thick dash-dot) ensembles for the tropics at the (top) 850- and (bottom) 200-hPa level.

0 to 48 h, and slightly smaller thereafter. As with the 850-hPa wind speed, the G159 200-hPa wind speed ensemble spread growth rate is larger than the B159 growth rate or S159 growth rate throughout the integration time.

Without running additional experiments (e.g., running long forecasts with SKEB off of the initial perturbations created by the control or banded ensemble), it is not possible to discern which improvements are due to the direct effect of changing the long forecasts or the indirect effect of changing the initial perturbations. These experiments have been done for a different type of stochastic forcing (stochastic convection) in a previous study (Reynolds et al. 2008). In those experiments, the indirect effect of changing the initial perturbations accounted for most of the improvement during the first few days of integration, while the improvements seen at later integration times were achievable through just the

direct effect of adding stochastic forcing to the long forecasts. Berner et al. (2009) show that the improvements gained by using SKEB in the ECMWF ensemble were more pronounced against control ensembles with reduced initial perturbation amplitude (the same as used in the SKEB experiments), then against larger control initial perturbation amplitude, illustrating the complex relationship between changing model formulation and changing initial perturbation size.

Brier scores for the $5 m s^{-1}$ threshold for the tropical 10-m wind speed (Fig. 6a) show improvements for B159 over G159 throughout the integration period, significant at the 95% level. This indicates that changing the initial perturbation methodology is resulting in improved forecast performance in this metric throughout the integration time, although, as with the ensemble mean RMSE,

TABLE 2. Growth rate of ensemble spread for 850- and 200-hPa wind speed in the tropics (day^{-1}) for different forecast hour intervals.

	G159 850 hPa	B159 850 hPa	S159 850 hPa	G159 200 hPa	B159 200 hPa	S159 200 hPa
0–24	0.331	0.134	0.181	0.352	0.131	0.238
24–48	0.278	0.138	0.085	0.228	0.095	0.116
48–72	0.252	0.154	0.080	0.182	0.095	0.090
72–96	0.221	0.155	0.082	0.163	0.099	0.085
96–120	0.180	0.143	0.080	0.152	0.109	0.090
120–144	0.141	0.124	0.072	0.135	0.109	0.091
144–168	0.115	0.104	0.068	0.113	0.099	0.087

the relative improvements decrease as forecast time increases. The improvements gained by adding SKEB are even larger, resulting in lower Brier scores for S159 over B159 that are statistically significant at the 95% level over the entire forecast interval. The differences between the Brier scores for the 10 m s^{-1} threshold are qualitatively similar to those for the 5 m s^{-1} threshold, but smaller in magnitude and not statistically significant at the 95% level at 120 h and beyond. The improvements in RMSE and Brier scores with the banded ET are qualitatively similar to those seen for the T119 ensembles described in McLay et al. (2010). McLay et al. (2010) attribute the improved performance of the banded ET over the global ET to 1) the improved match between the initial-time ensemble variance and the estimated analysis error variance (in both the midlatitudes and the tropics); and 2) flatter forecast error covariance eigenvalue spectra, indicating more independence between ensemble members.

The differences in ensemble spread are also reflected in the fraction of rank histogram outliers, shown as a function of forecast time for the 10-m wind speed in the tropics (Fig. 7). All ensembles are underdispersive, that is, the verification lies outside the range of ensemble values more often than would be expected from a well-calibrated ensemble. (The ideal fraction of outliers is $\frac{2}{18}$, denoted by the solid line with filled diamonds.) While the banded ET reduces the number of extraneous outliers from the global ET at early forecast times, this difference decreases as forecast length increases. In contrast, the addition of SKEB substantially decreases the number of extraneous outliers throughout the integration period.

While our focus is on the tropics, we briefly describe ensemble performance in the extratropics as this is a global ensemble forecast system. Going from the global to banded ET significantly enhances ensemble performance in the Northern Hemisphere extratropics for the T159 ensembles (not shown), and has a negligible impact on the Southern Hemisphere extratropics, consistent with the T119 ensemble results discussed in McLay et al. (2010). The ensemble mean RMSE for 500-hPa

height and low-level wind, and Brier scores for low-level wind, are reduced in these experiments and in those of McLay et al. (2010) with the introduction of the banded ET. As discussed in that paper, much of the improvement in the Northern Hemisphere extratropics is attributed to the reduction of erroneously large initial-time ensemble spread in the midlatitudes, achieved through

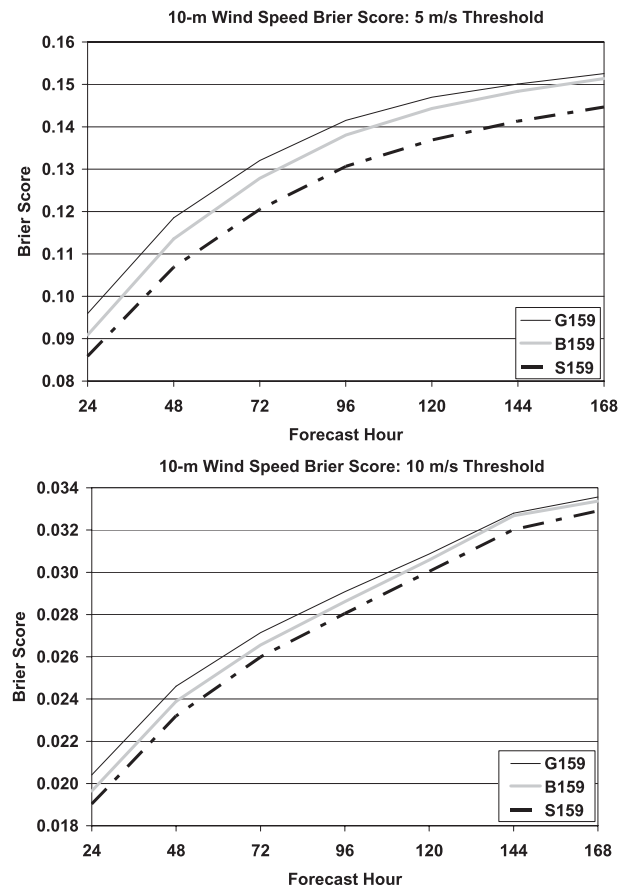


FIG. 6. Brier score for the G159 (thin solid black), B159 (thick solid gray), and S159 (thick dash-dot) ensembles for (top) 5 and (bottom) 10 m s^{-1} wind speed threshold. The differences for the 5 m s^{-1} threshold are statistically significant for the full time period. The differences for the 10 m s^{-1} threshold are statistically significant out to 120 h.

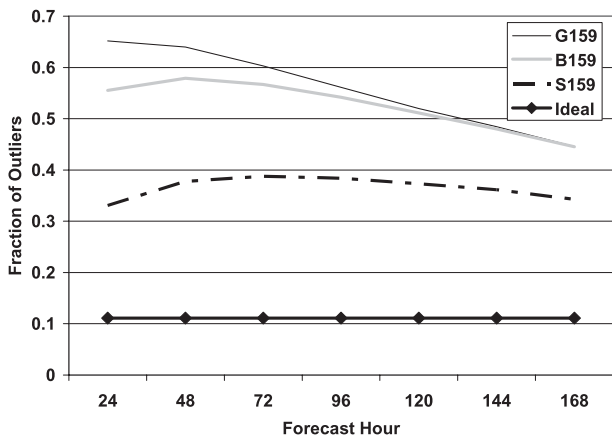


FIG. 7. Fraction of outliers in rank histograms for tropical 10-m wind speed for G159 (thin solid black), B159 (thick solid gray), and S159 (thick dash-dot). Ideal number of outliers is indicated by solid line with diamonds.

the banded localization of the transform calculation. The addition of SKEB has a small impact on the Brier scores, and slightly increases the spread and RMSE in the Northern Hemisphere extratropics.

As noted in section 2b, the implementation of SKEB evaluated here follows the implementation used in Berner et al. (2009) closely, except for the fact that a dissipation mask is not used. Preliminary experiments using a flow-dependent mask reflecting dissipation due to the convective parameterization indicate that additional improvements are possible through the refinement of the SKEB implementation. The addition of the convective dissipation mask has no significant impact on the Brier scores in the tropics. It does, however, remove the slight increase in RMSE observed in the Northern Hemisphere extratropics with the first SKEB implementation. For example, for low-level winds, the ensemble mean RMSE for SKEB without the dissipation is between 0.5% and 2% larger than the ensemble mean RMSE for the control ensemble in the mid-latitudes. With the inclusion of the convective dissipation mask, the RMSE for SKEB is comparable or up to 2.5% lower than the RMSE of the control ensemble mean, depending on lead time. The addition of the convective mask also resulted in very small improvements to the RMSE in the tropical low-level winds. Shutts (2005) notes that as there is considerable uncertainty in observational estimates of atmospheric energy dissipation rates, estimating local upscale energy transfer rates will also be prone to error. This uncertainty may limit the amount of improvement obtainable through the use of dissipation masks.

The improvements found because of SKEB in this study are consistent with those found in Berner et al.

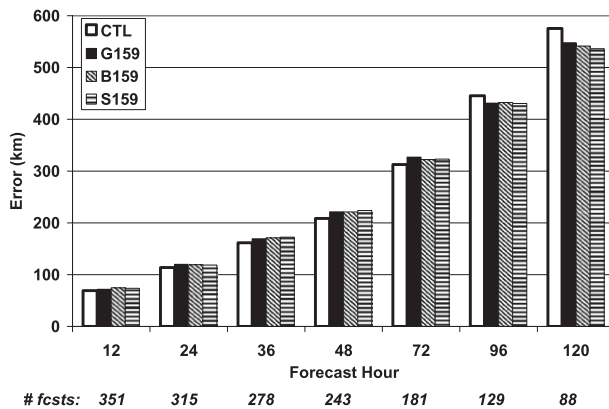


FIG. 8. Homogenous NH TC track forecast error (km) for G159, B159, and S159 ensemble mean tracks, as denoted in key. The average forecast error of the T239L30 NOGAPS operational deterministic forecast (control or CTL) is also shown. The numbers of forecasts are shown below the *x* axis. The differences are not statistically significant.

(2009) for the ECMWF ensemble system, where the most pronounced improvements were found in the tropics. Shutts (2005) provides physical justification for the use of kinetic energy backscatter in forecast models based on dissipation arising both from the numerical schemes and the physical parameterizations. He proposes that a contribution of kinetic energy from deep convection may arise because of the fraction of the kinetic energy released by buoyancy forces being injected back onto partially resolved scales. This may account, in part, for the improved ensemble performance in the tropics with the addition of SKEB.

The ensemble mean TC track errors for the ensemble design experiments are shown in Fig. 8 for a homogeneous sample, and include errors from the NOGAPS T239 deterministic operational forecast (control or CTL). While all three T159 ensemble mean track errors are smaller than the control errors at 96 and 120 h, the differences are not statistically significant. Nor are there statistically significant differences between the different T159 ensemble formulations (except at 72 h, when the improvement of S159 over G159 is significant at the 90% level). Comparison of ensemble mean track errors done separately for the different basins did not indicate systematic differences between the different ensembles. The ensemble method with the lowest mean error is different for different forecasts times, but the differences are small and for the most part not significant.

These results indicate that certain metrics such as Brier score show significant improvement in the ensemble performance through the adoption of the banded ET and the addition of SKEB. The ensemble mean TC track errors, while suggesting that the three T159 ensemble

means are more skillful than the T239 control at days 4 and 5, do not show statistically significant improvements given the current sample size.

4. Summary

In this study the impact of different ensemble design formulations on ensemble performance in the tropics is explored, focusing on tropical winds and TC track forecast errors. The resolution versus the number of member experiments indicate that increasing resolution from T119 to T159 and T239, while also reducing the number of ensemble members, has only a small impact on tropical wind RMSE. For the Brier score based on 10-m wind speeds, at the 5 m s^{-1} threshold, the G239 ensemble performs the best, but the differences between the ensembles decreases with increasing forecast lead time. At the 10 m s^{-1} threshold, the G119 ensemble appears to have the best performance; primarily due to the larger number of ensemble members (i.e., 33) in the low-resolution ensemble, but the differences are not statistically significant. For ensemble mean TC track errors, G159 and G239 ensembles perform better than the G119 ensemble, with the differences statistically significant out to 96 h at the 95% level, and at 120 h at the 90% level. The gain from going to G239 from G159 is smaller, and significant at the 95% level only at 48 and 72 h. When compared to the T239 deterministic control forecast, G159 and G239 have smaller errors at 96 and 120 h, but the differences are not statistically significant given the current sample size.

The ensemble design tests performed at T159 show small improvements in ensemble mean tropical wind speed RMSE going from the global to banded ET (G159 to B159) at intermediate forecast lead times. Adding stochastic perturbations (S159) further improves things slightly at longer lead times. B159 has larger ensemble spread than G159 at early lead times for the tropical wind field. The addition of SKEB, which impacts the initial perturbations as well as the long forecast integrations, results in substantially increased spread and decreased number of extraneous outliers at all lead times. For both 5 and 10 m s^{-1} thresholds, the Brier score is improved by going from the global ET to the banded ET, and is improved even further with the addition of the stochastic perturbations. These differences are larger at the 850-hPa level than at the 200-hPa level. However, the differences in ensemble design do not have a significant impact on TC track errors. The T159 ensemble mean track errors for all configurations are smaller than the T239 deterministic forecast track error at 96 and 120 h, but these differences are not statistically significant.

The implementation of SKEB evaluated here did not employ an energy dissipation mask. Preliminary results with the addition of a mask based on convective dissipation estimates did not significantly impact the tropical Brier scores, but did result in small improvements to the ensemble mean RMSE in the midlatitudes. It is possible that further refinements to the SKEB implementation (as well as a more complete accounting for model uncertainty in ensemble design) may result in even larger improvements, and current work is under way to develop and test these refinements.

The current research suggests that an upgrade in resolution to the Navy Operational Global Atmospheric Prediction System from the current T119 truncation to T159 truncation will result in significantly reduced ensemble mean TC track errors. Evaluation of case studies indicates substantial increases in the spread of the ensemble TC tracks with the addition of SKEB. Recent work (Yamaguchi et al. 2009; Majumdar and Finocchio 2010) has shown utility in using single model ensemble spread to produce confidence information on ensemble mean TC track predictions and TC track probabilities. Future work will investigate the impact of ensemble design on TC track ensemble spread and its utility for confidence information and track probabilities.

Acknowledgments. We gratefully acknowledge the support of the NOAA Hurricane Forecast Improvement Project and the Office of Naval Research (ONR) through Program Element 0602435N and 0601153N. The DoD High Performance Computing Program at NAVO MSRC provided part of the computing resources.

REFERENCES

- Berner, J., G. J. Schutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626.
- Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765.
- Bowler, N. E., A. Arribas, S. E. Beare, K. R. Mylne, and G. J. Schutts, 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **135**, 767–776.
- Buizza, R., and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503–2518.
- , M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.
- Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.*, **138**, 1877–1901.
- Daley, R., and E. Barker, 2001: NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.*, **129**, 869–883.

- Emanuel, K. A., and M. Zivkovic-Rothman, 1999: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.*, **56**, 1766–1782.
- Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187–1193.
- Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 668–688.
- Harshvardhan, R. Davies, D. A. Randall, and T. G. Corsetti, 1987: A fast radiation parameterization for atmospheric circulation models. *J. Geophys. Res.*, **92**, 1009–1016.
- Houtekamer, P. L., L. Lefavre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.
- Knaff, J. M., C. Guard, J. Kossin, T. Marchok, B. Sampson, T. Smith, and N. Surgi, 2007: Operational guidance and skill in forecasting structure change. *Proc. Sixth WMO Int. Workshop on Tropical Cyclones (IWTC-VI)*, San José, Costa Rica, WMO Tech. Doc. 1383. [Available online at <http://severe.worldweather.org/iwtc/>]
- Louis, J. F., M. Tiedtke, and J. F. Geleyn, 1982: A short history of the operational PBL parameterization at ECMWF. *Proc. ECMWF Workshop on Planetary Boundary Parameterizations*, Reading, United Kingdom, ECMWF, 59–79.
- Majumdar, S. J., and P. M. Finocchio, 2010: On the ability of global ensemble prediction systems to predict tropical cyclone track probabilities. *Wea. Forecasting*, **25**, 659–680.
- McLay, J. G., and C. A. Reynolds, 2009: Two alternative implementations of the ensemble-transform (ET) analysis-perturbation scheme: The ET with extended cycling intervals, and the ET without cycling. *Quart. J. Roy. Meteor. Soc.*, **135**, 1200–1213.
- , C. H. Bishop, and C. A. Reynolds, 2007: The ensemble transform scheme adapted for the generation of stochastic perturbations. *Quart. J. Roy. Meteor. Soc.*, **133**, 1257–1266.
- , —, and —, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093–1108.
- , —, and —, 2010: A local formulation of the ensemble transform (ET) analysis perturbation scheme. *Wea. Forecasting*, **25**, 985–993.
- Peng, M. S., J. A. Ridout, and T. F. Hogan, 2004: Recent modifications of the Emanuel convective scheme in the Navy Operational Global Atmospheric Prediction System. *Mon. Wea. Rev.*, **132**, 1254–1268.
- Puri, K., J. Barkmeijer, and T. N. Palmer, 2001: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. J. Roy. Meteor. Soc.*, **127**, 709–731.
- Reynolds, C. A., J. Teixeira, and J. G. McLay, 2008: Impact of stochastic convection on the ensemble transform. *Mon. Wea. Rev.*, **136**, 4517–4526.
- Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240.
- , J. L. Franklin, J. A. Knaff, and M. DeMaria, 2008: Experiments with a simple tropical cyclone intensity consensus. *Wea. Forecasting*, **23**, 304–312.
- Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102.
- , and T. N. Palmer, 2004: The use of high-resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. *Proc. ECMWF/CLIVAR Simulation and Prediction of Intra-seasonal Variability with Emphasis on the MJO*, Reading, United Kingdom, ECMWF, 83–102.
- Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter for the NCEP global model. *Tellus*, **60A**, 113–130.
- Teixeira, J., and T. F. Hogan, 2002: Boundary layer clouds in a global atmospheric model: Simple cloud cover parameterizations. *J. Climate*, **15**, 1261–1276.
- , and C. A. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach. *Mon. Wea. Rev.*, **136**, 483–496.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- Wei, M., Z. Toth, R. Wobus, Y. Zhu, C. H. Bishop, and X. Wang, 2006: Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus*, **58A**, 28–44.
- , —, —, and —, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operation forecast system. *Tellus*, **60A**, 62–79.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.
- Williford, C. E., T. N. Krishnamurti, R. C. Torres, S. Cocke, Z. Christidis, and T. S. Vijaya Kumar, 2003: Real-time multimodel super-ensemble forecasts of Atlantic tropical systems of 1999. *Mon. Wea. Rev.*, **131**, 1878–1894.
- Yamaguchi, M., R. Sakai, M. Kyoda, T. Komori, and T. Kadowaki, 2009: Typhoon ensemble prediction system developed at the Japan Meteorological Agency. *Mon. Wea. Rev.*, **137**, 2592–2604.