Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2011), 63A, 841–857
Printed in Singapore. All rights reserved

TELLUS

Examination of parameter variations in the U. S. Navy Global Ensemble

By CAROLYN A. REYNOLDS*, JAMES A. RIDOUT and JUSTIN G. MCLAY, *Naval Research Laboratory, 7 Grace Hopper Ave., Monterey, CA 94934–5502, USA*

(Manuscript received 4 February 2011; in final form 10 June 2011)

ABSTRACT

The impact of parameter variations on the Navy Operational Global Atmospheric Prediction System ensemble performance is examined, and subsets of ensemble members are used to identify the relative impact of the individual parameters. Two sets of parameter variations are considered. The first set has variations in the parametrization of cumulus convection only. The second set has variations in both convection and boundary layer parametrizations. In the tropics, parameter variations significantly increase ensemble spread in wind and temperature fields, and significantly reduce Brier scores for low-level wind speed and temperature, primarily through improvements to the resolution (the impact in the extratropics is negligible). There are also small but significant improvements in the ensemble mean tropical cyclone track forecasts. For the metrics considered here, the second set of parameter variations outperforms the first set. Examination of the spread within ensemble subsets suggests that the parameter with the biggest overall impact is one that helps to control the convective updraft parcel temperature deficit at cloud base level. Variations in the von Kármán constant significantly increase ensemble spread in the low-level tropical winds near the date line, and in the low-level temperature field throughout the tropics and subtropics.

1. Introduction

A practical way to deal with inevitable uncertainties in atmospheric forecasting is through ensemble forecasting. Ideally, both initial condition uncertainty and model uncertainty are accounted for in ensemble design. Operational forecasting centres have been providing global ensembles with initial-state perturbations since the early 1990s (e.g. Toth and Kalnay, 1993; Buizza and Palmer, 1998). More recently, several methods have been proposed to account for different aspects of model uncertainty and model error in ensemble design, including multimodel or multiparametrization ensembles (e.g. Houtekamer et al., 1996; Stensrud et al., 2000), stochastic forcing (e.g. Buizza et al., 1999, Berner et al., 2009) and parameter variations (e.g. Bowler et al., 2008; Hacker et al., 2011a, b). In this study, the focus is specifically on the impact of parameter variations in the Navy Operational Global Atmospheric Prediction System (NOGAPS) on ensemble performance. The parameter variations are chosen to reflect uncertainty in tune-able parameters that are known through prior experimentation to have an impact on forecasts. Two sets of parameter variations are considered. In the first set only parameters within the cumulus convection parametrization scheme are varied. In the second set a parameter in the boundary layer scheme is also varied. Subsets of ensemble forecasts are examined in order to investigate the relative impact of the individual parameters that are varied.

All ensembles are produced using the same initial perturbation methodology. This methodology is 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. Given a finite number of ensembles, the initial ensemble perturbations are too small in the tropics and too large in the mid-latitudes when compared to the analysis error variance estimate produced by the NRL Atmospheric Variational Data Assimilation System (NAVDAS, Daley and Barker, 2001). The problem of too little ensemble spread in the tropics is attributed, in part, to the neglect of model error in ensemble formulation, which is expected to be greater in the tropics than in the mid-latitudes. With the inclusion of model uncertainty, the aim is to both increase ensemble variance such that it is more consistent with ensemble mean error variance and to improve ensemble performance for other metrics of interest (e.g. Brier scores).

*Corresponding author.
e-mail: carolyn.reynolds@nrlmry.navy.mil
DOI: 10.1111/j.1600-0870.2011.00532.x

Different methods have been proposed to deal with different aspects of model error and model uncertainty. One broad way to deal with structural model error in ensemble design is to run multimodel ensembles or ensembles in which the forecast model contains different parametrization schemes for different ensemble members (e.g. Houtekamer et al., 1996; Stensrud et al., 2000; Bright and Mullen, 2002; Charron et al., 2010, Berner et al., 2011). While providing useful ensemble diversity, these ensembles may exhibit clustering by model (e.g. Alhamed et al., 2002) and this method also requires additional resources devoted to keeping different parametrizations or parametrization suites up-to-date. An alternative to multiparametrization techniques is to vary parameters within the parametrizations themselves (e.g. Bowler et al., 2008; Hacker et al., 2011a, b). This technique avoids the difficulties associated with maintaining different physical parametrization suites, but is also narrower in the type of model error considered, addressing only parametric uncertainties within the parametrizations themselves. In mesoscale model experiments over the continental United States, Hacker et al. (2011a) find little impact from parameter variations alone, but find the best results are obtained by combining parameter variations with multiparametrizations and stochastic forcing. Consideration of a global domain, as is done here, may highlight regional differences in the potential impact of parameter variations on ensemble performance.

Multiparametrization and parameter variations are not the only methods to account for model uncertainty. Other techniques employ stochastic methods of various types. Recent research on stochastic techniques includes the addition of a stochastic term to account for uncertainty in existing parametrizations due to, for example, grid-box sampling (e.g. Buizza et al., 1999; Shutts and Palmer, 2004), adding a stochastic perturbation to the convective available potential energy in the deep convective scheme (Lin and Neelin, 2000), adding stochastic perturbations to the tendencies from the convective parametrization to account for subgrid-scale uncertainties (Teixeira and Reynolds, 2008; Reynolds et al., 2008) and stochastic kinetic energy backscatter (SKEB, Shutts, 2005; Berner et al., 2009; Bowler et al., 2009; Charron et al., 2010; Berner et al., 2011), to account for aspects of structural uncertainty in conventional parametrizations.

Parameter variations may be varied stochastically. Li et al. (2008) describe accounting for model physics uncertainties by stochastically perturbing the threshold vertical velocity in the trigger function of the Kain-Fritsch deep convection parametrization, and the threshold humidity in the Sundqvist explicit scheme in an ensemble system using the limited-area version of the Canadian Global Environmental Multiscale model. They find that these stochastic parameter perturbations significantly enhance quantitative precipitation forecast performance for light rainfall rates. Bowler et al. (2008) describe the use of stochastic parameter variations in the design of the Met Office Global and Regional Ensemble Prediction System (MOGREPS). In that system, eight parameters (two each within their convection, boundary layer, gravity wave drag and large-scale precipitation parametrizations) are varied stochastically during the ensemble forecast integrations. These variations are meant to account for the somewhat arbitrary values of empirical, adjustable parameters and thresholds within the parametrization schemes. The parameters vary in time using a first order auto-regression model. They find that the addition of the parameter variations, as well as inclusion of stochastic convective vorticity (which is meant to represent anomalies associated with mesoscale convective systems), increased ensemble spread in mean sea level pressure by about 5% over ensembles that contained initial perturbations only.

In the study here, parameter variations are employed, but in contrast to Bowler et al. (2008), the modified parameter values are held constant throughout the forecast integration, with different parameter value combinations for each ensemble member. Although there are potential benefits to the use of temporally varying parameter perturbations, one also expects slowly varying (or even time-invariant) features in the distribution of optimal parameter values that may be best represented by the use of parameter variations that are fixed within a given ensemble member. For example, the use of different parameter values in the Emanuel convective scheme (Emanuel, 1991; Emanuel and Zivkovic-Rothman, 1999; Peng et al., 2004) over land compared with tropical warm pool regions is supported by some preliminary cloud-resolving simulation experiments (James Ridout, personal communication, 2010) with the Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS®; Hodur, 1997; Chen et al., 2003). Although the use of geographically varying parameters would appear to be the optimal solution in this regard, this approach has thus far proved unsuccessful in NOGAPS. Other related examples can be cited supporting the use of parameters that are fixed within a given model integration, but vary in different ensemble members. In general, this approach may be expected to be helpful in the representation of variability associated with strong contrasts in the nature of physical processes in slowly varying weather regimes, such as for example associated with different phases of the Madden Julian Oscillation (MJO) (e.g. Madden and Julian, 1971, 1972, 1994; Hayashi and Sumi, 1986; Hendon, 1988; Chen et al., 1996). Ideally this variability would be adequately represented by the model physics without the need for parameter variations, but such is not always the case. As with other methods of parameter variation (and to some extent parametrizations in general), this method is expected to result in situations where certain features of the flow are better represented at the expense of other features. In a careful study investigating the response of a mesoscale model to parameter variations for a mid-latitude domain, Hacker et al. (2011b) find that the lack of a domain-wide systematic response suggests time-dependent parameter variations may not be necessary for representing parameter uncertainty.

In order to focus on the improvement in ensemble behaviour that may be gained through accounting for parametric error, parameter variations are the only type of model uncertainty considered here. It is expected that, in the future, this method of accounting for model uncertainty may be profitably combined with other methods, as has been shown in, for example, Bowler et al. (2008), Charron et al. (2010) and Hacker et al. (2011a).

This paper is organized in the following manner. Section 2 describes the methodology; Section 3 presents the results comparing the full ensembles; Section 4 presents results from the examination of subensembles designed to evaluate the impact of the different parameters; and Section 5 contains a summary and brief discussion of future work. In this work, it is found that the parameter variations have little impact on ensemble performance in the extratropics, and therefore focus is placed on ensemble performance in the tropics.

2. Methodology

2.1. Initial perturbations

The ensemble transform (ET) initial perturbation scheme 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 $\mathbf{P}_\mathbf{g}^\mathbf{a}$ be an $N \times N$ matrix whose elements represent the error covariance estimate of the initial condition state, where N is the number of variables in the model state. Let $\mathbf{Z}^\mathbf{f} = [\mathbf{z}_\mathbf{1}^\mathbf{f}, \mathbf{z}_\mathbf{2}^\mathbf{f}, \ldots, \mathbf{z}_K^\mathbf{f}]$ be an $N \times K$ matrix whose ith column is

$$\mathbf{z}_i^\mathbf{f} = \mathbf{x}_i^\mathbf{f} - \overline{\mathbf{x}_i^\mathbf{f}}, \tag{1}$$

where $\overline{\mathbf{x}_i^\mathbf{f}} = \frac{1}{K}\sum_{i=1}^{K} \mathbf{x}_i^\mathbf{f}$, $\mathbf{x}_i^\mathbf{f}$ is the ith member of a given forecast ensemble, K is the number of ensembles and $\overline{\mathbf{x}_i^\mathbf{f}}$ represents the mean of the forecast ensemble. In the ET, K analysis perturbations $\mathbf{Z}^\mathbf{a} = [\mathbf{z}_\mathbf{1}^\mathbf{a}, \mathbf{z}_\mathbf{2}^\mathbf{a}, \ldots, \mathbf{z}_K^\mathbf{a}]$ are generated using

$$\mathbf{Z}^\mathbf{a} = \mathbf{Z}^\mathbf{f}\mathbf{T} \tag{2}$$

subject to the constraint

$$\mathbf{Z}^{\mathbf{a}\mathrm{T}} \left(\mathbf{P}_\mathbf{g}^\mathbf{a}\right)^{-1} \mathbf{Z}^\mathbf{a} = N\mathbf{I}, \tag{3}$$

where $\mathbf{T} = [\mathbf{t}_\mathbf{1}, \mathbf{t}_\mathbf{2}, \ldots, \mathbf{t}_K]$ is assumed to be a real symmetric $K \times K$ matrix of weighting coefficients. 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. Here, $\mathbf{P}_\mathbf{g}^\mathbf{a}$ is obtained from NAVDAS, and is diagonal.

2.2. Parameter variations

The first set of parameter variations in NOGAPS concerns only parameters within the moist convection scheme (Emanuel, 1991; Emanuel and Zivkovic-Rothman, 1999; Peng et al., 2004). The ensemble containing this set of parameter variations is referred to as PAR1. The four parameters that are varied within the forecast model between different ensemble members are *cu*, which is the coefficient that scales the computed convective momentum transport, *dtmax*, which represents the magnitude of small-scale temperature perturbations associated with rising updraft source-layer parcels, and *alpha* and *damp*, which are parameters that control the rate of approach to quasi-equilibrium. The parameter *cu* can assume values between '0' and '1', where a value of '0' corresponds to the maximum amount of convective momentum transport, and '1' corresponds to no convective momentum transport. The value used for *cu* in the control ensemble member in the current tests is 0.25. The other three parameters occur in the equation describing the temporal development and decay of cloud-base mass flux, M_b,

$$\Delta M_\mathrm{b} = 0.1 alpha\,(\Delta T_{LCL} + \Delta T_{sc} + dtmax) - damp\ \ M_\mathrm{b}, \tag{4}$$

where ΔM_b represents the change in M_b over the course of a single time-step, ΔT_LCL is the virtual temperature difference between a parcel (lifted from the updraft source-level) and the environment at the lifted condensation level (LCL), and ΔT_sc is the mean virtual temperature difference between the lifted parcel and the environment in the subcloud layer. In the current tests, the parameter values used for the control ensemble member are 0.5 kg m^{-2} K^{-1} s^{-1} for *alpha*, 0.8 K for *dtmax*, and 0.1 for *damp*. The treatment described by (4) tends to adjust M_b so that updraft parcel buoyancy in the subcloud layer remains approximately constant. The first term represents a source term for cloud-base mass flux when the sum of ΔT_LCL and ΔT_sc exceeds $-$ *dtmax*, and a sink term otherwise. The second term is always a sink, or 'damping' term. Typically, increases in $\mathrm{M_b}$ resulting from increases in parcel buoyancy tend to stabilize the profile, resulting in some degree of balance between convection and surface fluxes or other destabilizing factors. Thus on timescales of $\sim$12 h in tropical warm pool regions, (4) can be regarded (Emanuel and Zivkovic-Rothman, 1999) as an implication of the 'boundary layer quasi-equilibrium' hypothesis of Raymond (1995). In global numerical weather prediction (NWP) with NOGAPS the success of this relation has been shown in explicit simulations of deep convection to be consistent with an apparently more broadly observed 1–3 h 'cloud-base quasi-balance' (Ridout et al., 2005).

The selection of the four parameters for variation in PAR1 is based to some degree on a number of preliminary experiments, including parameter adjustments in response to various changes in the model physics. Adjustments to the value of *cu*, for example, have been shown to have a considerable impact on tropical cyclone track forecasts (Hogan and Pauley, 2007). The updraft parcel temperature perturbation parameter *dtmax* is of critical significance as well, exerting considerable control on the development of convection. Its value has recently been adjusted in response to several changes in the forecast model, including changes made in the computation of the updraft temperature perturbation in the convection scheme. Less testing has been addressed towards the *alpha* and *damp* parameters. In the

Table 1. Configuration of the ensemble experiments in PAR1. Member 0 has the control values

Member	*cu*	*dtmax*	*damp*	*alpha*
0	0.25	0.8	0.1	0.5
1	0.5	0.8	0.1	0.5
2	0	0.8	0.1	0.5
3	0.25	1.1	0.1	0.5
4	0.25	0.5	0.1	0.5
5	0.25	0.8	0.1333	0.5
6	0.25	0.8	0.08	0.5
7	0.25	0.8	0.1	0.625
8	0.25	0.8	0.1	0.375
9	0.5	1.1	0.1	0.5
10	0.5	0.5	0.1	0.5
11	0.5	0.8	0.1333	0.5
12	0.5	0.8	0.08	0.5
13	0.5	0.8	0.1	0.625
14	0.5	0.8	0.1	0.375
15	0	1.1	0.1	0.5
16	0	0.5	0.1	0.5
17	0	0.8	0.1333	0.5
18	0	0.8	0.08	0.5
19	0	0.8	0.1	0.625
20	0	0.8	0.1	0.375
21	0.25	1.1	0.133	0.5
22	0.25	1.1	0.08	0.5
23	0.25	1.1	0.1	0.625
24	0.25	1.1	0.1	0.375
25	0.25	0.5	0.1333	0.5
26	0.25	0.5	0.08	0.5
27	0.25	0.5	0.1	0.625
28	0.25	0.5	0.1	0.375
29	0.25	0.8	0.1333	0.625
30	0.25	0.8	0.1333	0.375
31	0.25	0.8	0.08	0.625
32	0.25	0.8	0.08	0.375

hypothetical case where an exact equilibrium is attained, the equilibrium value of cloud-base mass flux is proportional to the ratio of *alpha* to *damp*, suggesting the possibility of some degree of interdependence with respect to impacts of variations in these two parameters. In the PAR1 ensemble experiments, all four of these parameters are varied between the control value, and a high and a low value, as shown in Table 1. The ranges of variation with respect to the control member values are consistent with expected uncertainties as to the most appropriate values in the context of the convective parametrization. The combination of parameter values are unique to each ensemble member, and are held fixed throughout the ensemble forecast. Note that not all of the 81 possible parameter variation combinations are considered, as tests were designed to use an ensemble size that is currently operationally feasible (32 perturbed members plus one control). In the first eight perturbed members, only one parameter is perturbed at a time. Two parameters are perturbed at once in the remaining 24 members. Subensembles can be formed from the first 8 members to examine the impact of individual parameter variations, as described in Section 4.

The ensemble with the second set of parameter variations is referred to as PAR2. In PAR2, the set of parameter variations includes *cu* and *dtmax*, as in PAR1, but now *alpha* and *damp* are held to their control values. In their place, two new parameters are added. The first, *sigs*, is the fraction of precipitation falling outside of the cloud. Larger values of *sigs* result in greater evaporative cooling within the precipitating downdraft, with consequent impacts on the rate of stabilization of the atmospheric profile to further convection. It is important to note that in the current NOGAPS version of the Emanuel scheme, *sigs* does not assume the value of unity below cloud-base as should conceptually be the case (Emanuel, 1991). Enhanced sensitivity of the rate of low-level cooling by convective downdrafts to this parameter may thus be expected. The second new parameter, k, introduced to the set of varied parameters is the von Kármán constant. The von Kármán constant is a constant of the logarithmic wind profile in the surface layer,

$$\frac{\partial U}{\partial z} = \frac{u_*}{kz}, \tag{5}$$

where z is height, u_* is friction velocity and U is wind velocity. Although, as its name implies, k is generally taken to be a fixed constant (usually 0.4, which is the value used in NOGAPS), there have been a number of studies reported through the years supporting values in a range of about $\pm 10\%$ of 0.4 (see the discussion in Andreas, 2009). Recent work by Zhang et al. (2008) evaluated k under a range of stability conditions using an iterative analysis that minimized a cost function of the differences between computed and observed gradients of wind speed, temperature and humidity. The analysis, which adopted the Businger et al. (1971) flux-profile relationships, concludes that k decreases from 0.49 under stable conditions to 0.34 under unstable conditions, and is equal to 0.4 under neutral stratification. The boundary layer scheme currently employed in NOGAPS (Louis, 1979; Louis et al., 1982) is in part based on these same flux-profile relationships, particularly for unstable conditions. Although it can be argued that some of the reported variability might be more accurately interpreted as a reflection of deficiencies in the flux-profile relationships employed (Foken, 2008), the results do support variation of k within the framework of the NOGAPS ensemble. In this study, k is allowed to vary only within a small portion of this range (0.38–0.42). The regular variation of k with stability found by Zhang et al. (2008) suggests that this dependence is readily parametrizable (and an effort in this regard is currently underway). Nonetheless, the fact that such a variation of k with stability is not currently represented in NOGAPS is justification for its inclusion in our set of varying parameters in this study. The configuration of the second

Table 2. Configuration of the ensemble experiments in PAR2. Member 0 has the control values

Member	*cu*	*dtmax*	*sigs*	*k*
0	0.25	0.8	0.12	0.4
1	0.5	0.8	0.12	0.4
2	0	0.8	0.12	0.4
3	0.25	1.1	0.12	0.4
4	0.25	0.5	0.12	0.4
5	0.25	0.8	0.10	0.4
6	0.25	0.8	0.14	0.4
7	0.25	0.8	0.12	0.38
8	0.25	0.8	0.12	0.42
9	0.5	1.1	0.12	0.4
10	0.5	0.5	0.12	0.4
11	0.5	0.8	0.10	0.4
12	0.5	0.8	0.14	0.4
13	0.5	0.8	0.12	0.38
14	0.5	0.8	0.12	0.42
15	0	1.1	0.12	0.4
16	0	0.5	0.12	0.4
17	0	0.8	0.10	0.4
18	0	0.8	0.14	0.4
19	0	0.8	0.12	0.38
20	0	0.8	0.12	0.42
21	0.25	1.1	0.10	0.4
22	0.25	1.1	0.14	0.4
23	0.25	1.1	0.12	0.38
24	0.25	1.1	0.12	0.42
25	0.25	0.5	0.10	0.4
26	0.25	0.5	0.14	0.4
27	0.25	0.5	0.12	0.38
28	0.25	0.5	0.12	0.42
29	0.25	0.8	0.10	0.38
30	0.25	0.8	0.10	0.42
31	0.25	0.8	0.14	0.38
32	0.25	0.8	0.14	0.42

set of parameter variations, similar to that for PAR1, is shown in Table 2. As with the parameter variations in PAR1, the ranges of variation shown here with respect to the control member values are consistent with expected uncertainties.

2.3. Model description and experimental design

The results shown are based on ensemble forecasts using NOGAPS (Peng et al., 2004), the global spectral weather prediction model of the US Navy. The physical parametrizations 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 using the ET method with a 6-h cycling interval to produce the initial perturbations, with forecasts run at triangular truncation 119 (110 km) horizontal resolution and 30 levels (T119L30) with 33 members (32 perturbed members plus one member without initial perturbations). The control ensemble is run without any changes to the forecast model and is referred to as the CTL (control) ensemble. The second ensemble is run with the first set of parameter variations described above, and as noted previously, is referred to as PAR1. The third ensemble, PAR2, is run with the second set of parameter variations. All ensembles forecasts are run for 10 d starting at 00Z each day between 10 May and 12 September 2007.

3. Results

3.1. Ensemble spread

One reason to include parameter variations is to enhance the diversity of model solutions through inclusion of some component of model uncertainty in ensemble design. The top panel of Fig. 1 shows the ensemble spread (measured as the ensemble standard deviation about the ensemble mean) for the day-5 forecasts of 850-hPa wind speed, averaged for the entire period, for the CTL ensemble. For the May through September time period considered here, the ensemble spread is greatest in the Southern Hemisphere mid-latitudes, and smallest in the tropics. The percentage difference between the ensemble variance of PAR1 and CTL [100×(PAR1 – CTL)/CTL] is shown in the middle panel of Fig. 1. The increase in variance is greatest in the tropics, and is significant at the 95% level only in this region. The percentage increase in variance is a maximum over the Bay of Bengal, and in general is large in regions of strong tropical convection. This is expected, as PAR1 ensemble parameters are varied in the convective parametrization scheme only. The percentage difference between PAR2 and CTL is shown in the bottom panel, and exhibits similar, somewhat larger, increases in variance than PAR1. Other fields examined, such as mid and upper-tropospheric wind speed and low, mid and upper-tropospheric temperatures, exhibit similar patterns, although the increase in spread in temperature at low levels is larger and extends further into the subtropics (Fig. 2). The percentage differences between the ensemble variance for PAR1 and PAR2 [100 × (PAR2 – PAR1)/PAR1] for 850-hPa wind speed and 850-hPa temperature are shown in Fig. 3. The parameter combination in PAR2 results in significantly higher variance than the parameter combination in PAR1 in the tropical central Pacific between 150E and 170W for the wind speed, and over broad areas covering the central Pacific, Indian Ocean and eastern Atlantic for 850-hPa temperature. An examination of what specific parameter variations primarily account for these differences is described in Section 4.

Consistent with these results, the ensemble standard deviations (spread) for the 850-hPa wind speed and 850-hPa

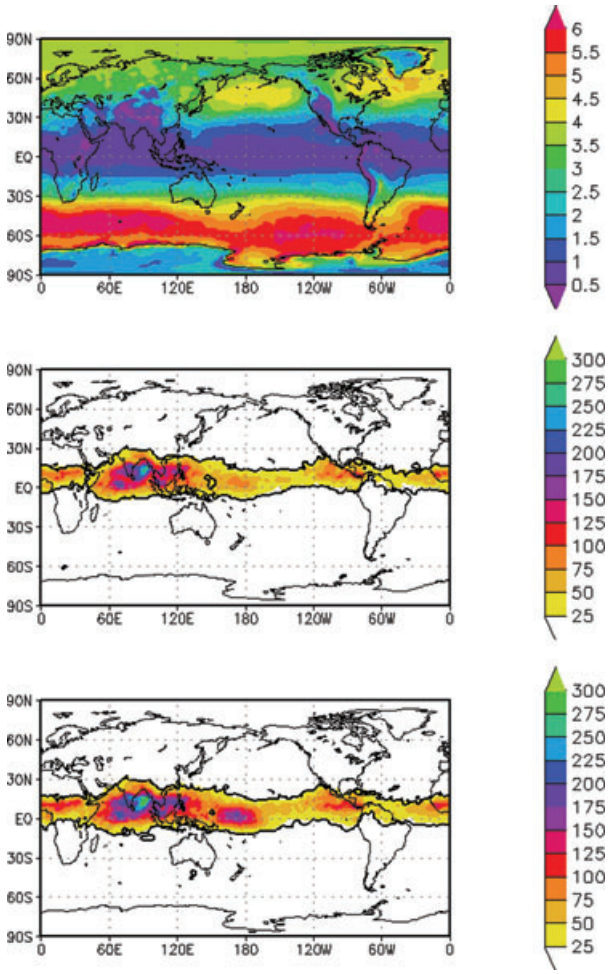

Fig. 1. Ensemble standard deviation for 850-hPa wind speed 5-day forecasts about the ensemble mean (m s^{-1}) for the control ensemble (top). Percentage difference between the PAR1 ensemble and CTL ensemble (middle), and percentage difference between PAR2 ensemble and CTL ensemble (bottom) for the 850-hPa wind speed ensemble variance. Black contour indicates 95% confidence level.

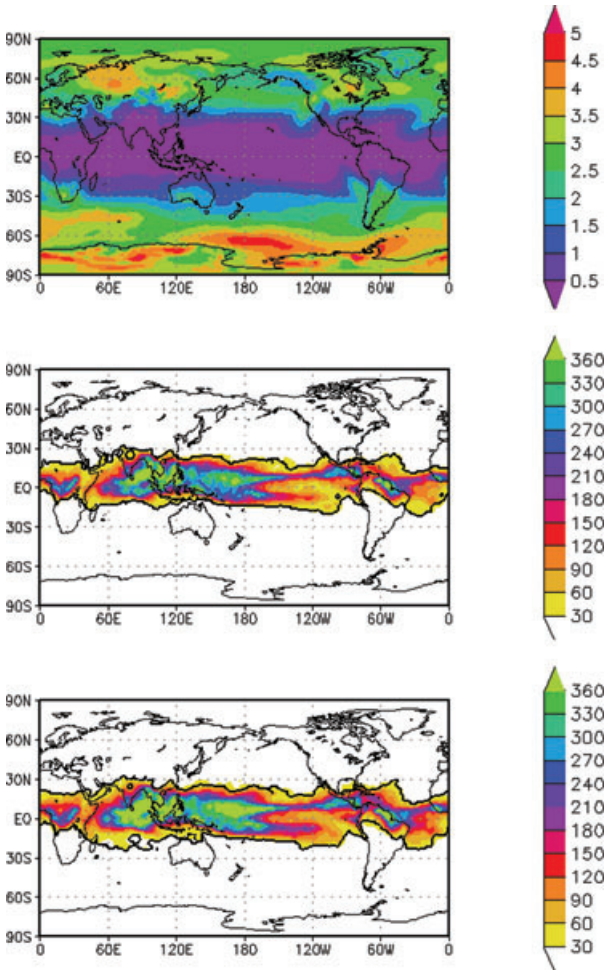

Fig. 2. Ensemble standard deviation for 850-hPa temperature 5-d forecasts about the ensemble mean (*K*) for the control ensemble (top). Percentage difference between the PAR1 ensemble and CTL ensemble (middle), and percentage difference between PAR2 ensemble and CTL ensemble (bottom) for the 850-hPa temperature ensemble variance. Black contour indicates 95% confidence level.

temperature, averaged from 20S-20N, as a function of forecast time, show significant increases when going from CTL to PAR1, and smaller increases when going from PAR1 to PAR2 (Fig. 4). For 850-hPa wind speed, the percentage increase in ensemble spread of PAR1 over CTL ranges from 23.4% at 24 h to 5.3% at 240 h. The percentage increase in ensemble spread of PAR2 over PAR1 is smaller, ranging from 5.4% at 24 h to 1.3% at 240 h. In the Northern and Southern Hemisphere extratropics, the change in ensemble spread (not shown) introduced through the parameter variations is never more than 2% and usually less than 1%. For the 850-hPa temperature, the increased spread obtained by going from CTL to PAR1 in the tropics ranges from 65% at 24h to 5.5% at 240 h. The additional spread gained by going to PAR2 from PAR1 ranges from 5.5% (at 24 h) to 1.2% (at 240 h). As with the wind speed fields, the change in spread for the temperature fields in the mid-latitudes is quite small (less than 1% in the Northern Hemisphere extratropics and less than 3% in the Southern Hemisphere extratropics).

To examine what portion of increased spread may be due to the development of different biases within the different PAR1 and PAR2 ensemble members, the time-averaged error (bias) for the entire period is subtracted from the ensembles, and the spread is recalculated. Through this de-biasing, the increase in spread is reduced, but not substantially. [The removal of the bias for the CTL case has no impact on spread, as all the members have the same bias]. For example, for the 850-hPa wind speed at 120-h, PAR1 (PAR2) increases spread over CTL by 9.86% (13.75%)

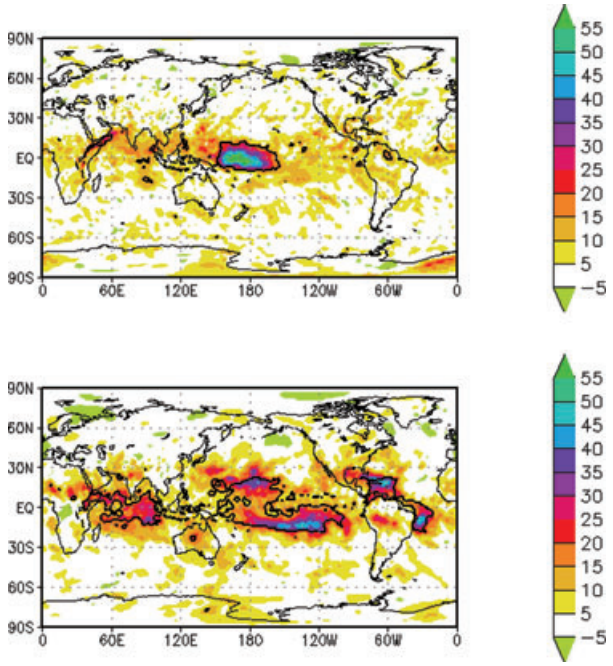

Fig. 3. Percentage difference between the PAR2 ensemble and PAR1 ensemble for the 850-hPa wind speed (top) and 850-hPa temperature (bottom) ensemble variance for the 5-d forecasts. Black contour indicates 95% confidence level.

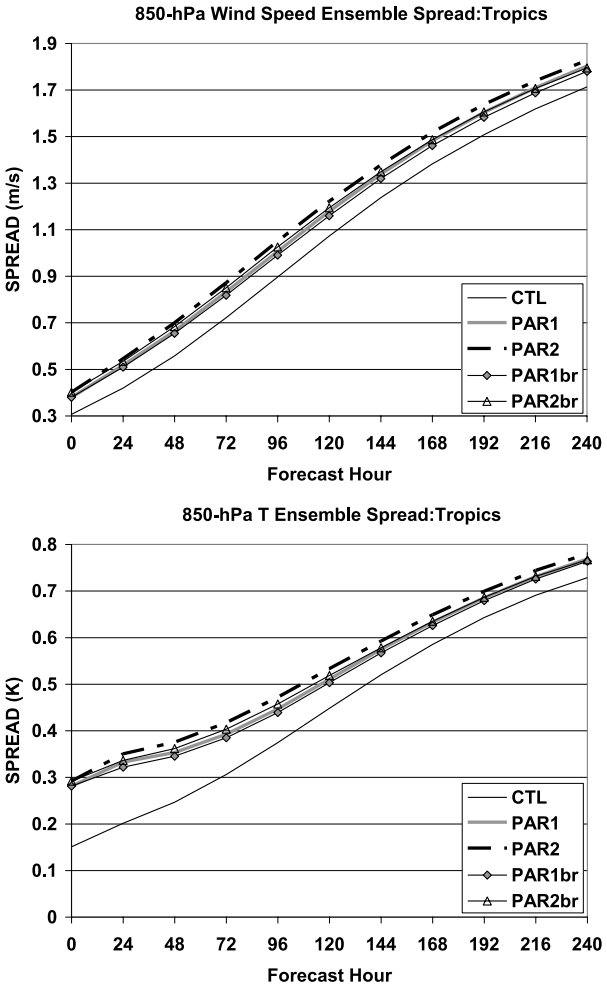

Fig. 4. Ensemble standard deviation about the ensemble mean in the tropics for the CTL (thin solid), PAR1 (thick solid grey) and PAR2 (thick dotted–dashed) ensembles for 850-hPa wind speed (top, m s^{-1}), and 850-hPa temperature (bottom, K). Results are also shown for PAR1 (thin line with diamonds) and PAR2 (thin line with triangles) when the time-mean bias is removed.

when using the raw ensembles, and 8.12% (11.33%) when using the de-biased ensembles. For the 850-hPa temperature at 120-h, PAR1 (PAR2) increases spread over CTL by 14.06% (18.97%) when using the raw ensembles, and 12.28% (15.85%) when using the de-biased ensembles.

As the ET is a cycling scheme, changes to the model formulation have both a direct effect on the long forecast *and* an indirect effect on the initial perturbations, as these are produced through a transformation of short (in this case, 6-h) ensemble forecasts. The significant increase in the initial-time ensemble spread in Fig. 4 with the addition of parameter variations (25–30% for low-level winds, almost 85–93% for low-level temperature) indicates that this indirect effect is substantial at short forecast lengths. Without running additional experiments, it is not possible to discern the relative contributions of these direct and indirect effects on the improvement in skill. A previous study (Reynolds et al., 2008) on stochastic convection indicated that 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. As this indirect effect can be quite significant, it is important to examine short-term forecasts to confirm that this effect results in reasonable perturbations. This issues is explored further in Section 3.5 in the context of precipitation forecasts.

3.2. *Ensemble mean error*

The increase in ensemble spread does not translate into a significant decrease in ensemble mean rms error. Figure 5 shows the ensemble-mean error for 850-hPa wind speed and 850-hPa temperature, as a function of forecast time. For the error calculations, analyses are used for verification. While the ensemble mean rms errors for PAR1 and PAR2 are slightly smaller than the control, the percentage decreases are always less than 1.3% and are not statistically significant. The curves for the corresponding errors in the extratropics (not shown) are basically indistinguishable from each other, with the percentage differences always less than 1%. The ensemble mean rms errors based on the de-biased ensembles are also included in Fig. 5. While de-biasing results in a substantial decrease in ensemble mean error for all three

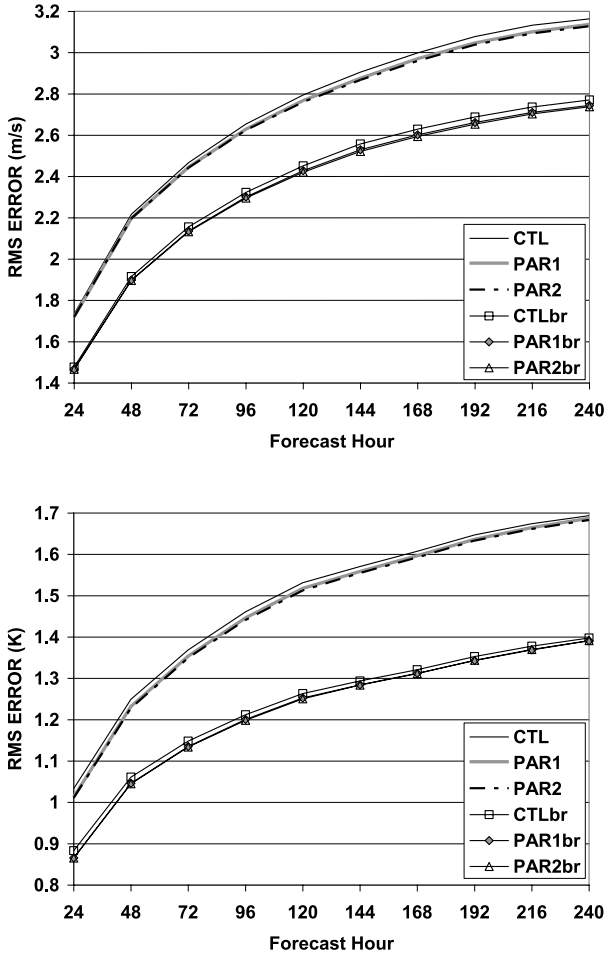

Fig. 5. Ensemble mean RMSE in the tropics for the CTL (thin solid), PAR1 (thick solid grey) and PAR2 (thick dotted–dashed) ensembles for 850-hPa wind speed (top, m s^{-1}), and 850-hPa temperature (bottom, K). Results are also shown for CTL (thin line with squares), PAR1 (thin line with diamonds) and PAR2 (thin line with triangles) when the time-mean bias is removed.

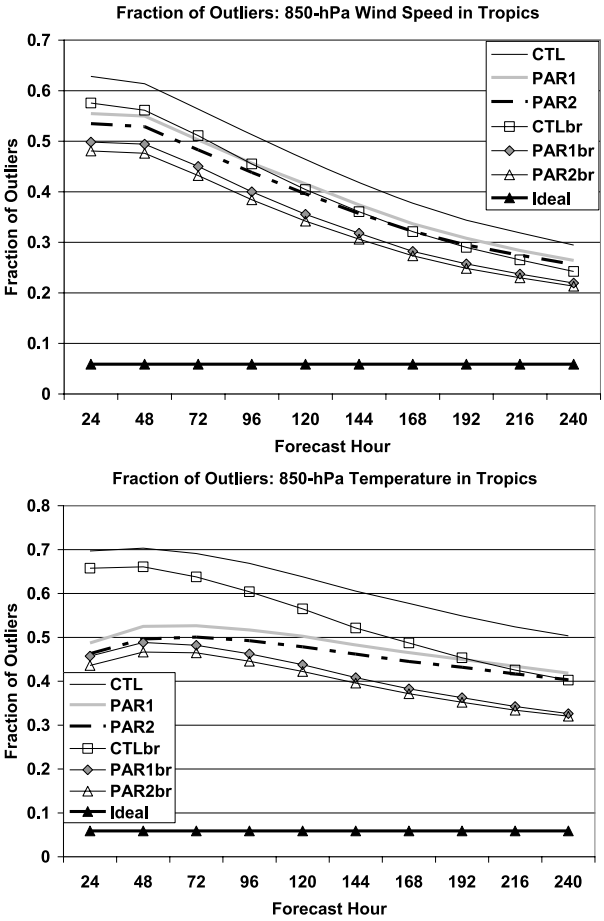

Fig. 6. Fraction of outliers in rank histogram diagrams in the tropics for the CTL (thin solid), PAR1 (thick solid grey) and PAR2 (thick dotted–dashed) ensembles for 850-hPa wind speed (top) and 850-hPa temperature (bottom). Results are also shown for CTL (thin line with squares), PAR1 (thin line with diamonds) and PAR2 (thin line with triangles) when the time-mean bias is removed. The ideal fraction of outliers is indicated by the solid black line with filled triangles.

ensembles, the relative skill of the three ensembles is approximately the same for both the raw and de-biased forecasts.

The RMSE is considerably larger than the ensemble spread (note the different *y*-axis values in Figs. 4 and 5). This indicates that even with the bias removed from the RMSE, the ensemble is still underdispersive. To quantify this, Fig. 6 shows the fraction of outliers in a rank histogram as a function of forecast time. The fraction of outliers is substantially too large at all forecast times for all three ensembles, with the largest number of outliers found for CTL, and the smallest number of outliers found for PAR2, for both the temperature and wind fields. Removal of the time-mean bias decreases the number of outliers for both the CTL and parameter variation ensembles, but the ensembles are still underdispersive.

As the ensemble is clearly underdispersive, it is desirable to increase ensemble spread, but in particular, it is desirable to increase the ability of the ensemble to (1) have ensemble variance reflect ensemble mean error variance (e.g. a positive 'spread-skill' relationship) and (2) have ensemble distributions that are both reliable and sharp. To investigate the first point, binned ensemble variance versus ensemble mean error variance plots, following Wang and Bishop (2003) are examined (not shown). Parameter variations slightly improve the spread-skill relationship (i.e. increase the range of error variances discernible through binned ensemble variances) for low-level wind speed (no impact on temperature). The impact on reliability and resolution are discussed in conjunction with the Brier scores in Section 3.3.

Latitude–longitude plots of the percentage differences in the ensemble mean RMS errors (not shown) indicate that the largest error reductions from PAR1 and PAR2 over CTL occur over Southeast Asia and the tropical western Pacific for both the wind and temperature fields. There are reductions of greater than 10%

in the 850-hPa temperature fields over Southeast Asia for 5-d forecasts, but these reductions are not statistically significant.

It is desirable that the parameter variations do not significantly affect the forecast skill of the individual ensemble members. Preliminary testing of parameter variation for the forecast model run in 'deterministic' mode provided a reasonable range of parameter values. The RMS errors of the individual ensemble members, averaged for the time period, do show some sensitivity to the parameter configurations in the tropics (with little impact in the extratropics). For example, for the 850-hPa wind speed, the ensembles with high values of *cu* have RMS errors that are, on average, 2.6% (2.9%) greater than those of the CTL ensemble members, at 120-h (240-h). The ensembles with low values of *cu* have RMS errors that are, on average, 0.5% (0.8%) lower than those of the CTL ensemble members, at 120-h (240-h). Sensitivity to the other parameters is smaller. For the 850-hPa temperature, the largest sensitivity is found to the values of *dtmax*. The ensembles with low values of *dtmax* have RMS errors that are, on average, 2.9% (3.4%) greater than those of the CTL ensemble members, at 120-h (240-h). The ensembles with high values of *dtmax* have RMS errors that are, on average, 1.6% (1.8%) lower than those of the CTL ensemble members, at 120-h (240-h). However, for the first few days of the forecast, the ensemble members with low values of *dtmax* actually perform better than the ensemble members with high values of *dtmax*. Sensitivity to *k* is comparable to that of the sensitivity to *dtmax*, with high values of *k* giving better performance at long lead times.

Despite these sensitivities, it is confirmed that the RMS error averaged over all the individual ensemble members does not significantly change with the addition of parameter variations. This is true despite the relatively large impact that the parameter variations have on initial-time ensemble variance. For the 850-hPa wind speed, the percentage difference between the average PAR1 or PAR2 individual ensemble forecast errors and the average CTL ensemble forecast errors are less than 0.7% at 24-h (smaller at longer lead times), and not statistically significant. For 850-hPa temperature, the percentage difference was 1.5% at 24-h (smaller for longer lead times). The small impact on short-term forecast errors suggests that the changes introduced to the initial perturbations and short term forecasts through the introduction of parameter variations are not unreasonable.

Other fields, such as upper-level (200-hPa) wind speed, and 500-hPa geopotential height, not shown, have also been examined. The impacts of PAR1 and PAR2 on these fields are qualitatively similar to those shown for the low-level temperature and wind fields. That is, there is a small (not significant) reduction in ensemble mean RMSE with the addition of the parameter variations in the tropics. There is also a substantial increase in tropical ensemble spread, slightly more so for PAR2 then PAR1. There is no significant impact on either the ensemble mean RMSE or the ensemble spread in the extra-tropics. The lack of impact in the mid-latitudes is consistent with the results of Hacker et al. (2011a), who find that parameter variations alone do not significantly impact ensemble performance for a mid-latitude domain.

3.3. Brier scores

While the RMS error of the ensemble mean does not appear to be significantly impacted by the introduction of parameter variations, the Brier Score is. 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 m s^{-1}, and 10 m s^{-1}) are shown in Fig. 7. The statistical significance of the Brier score

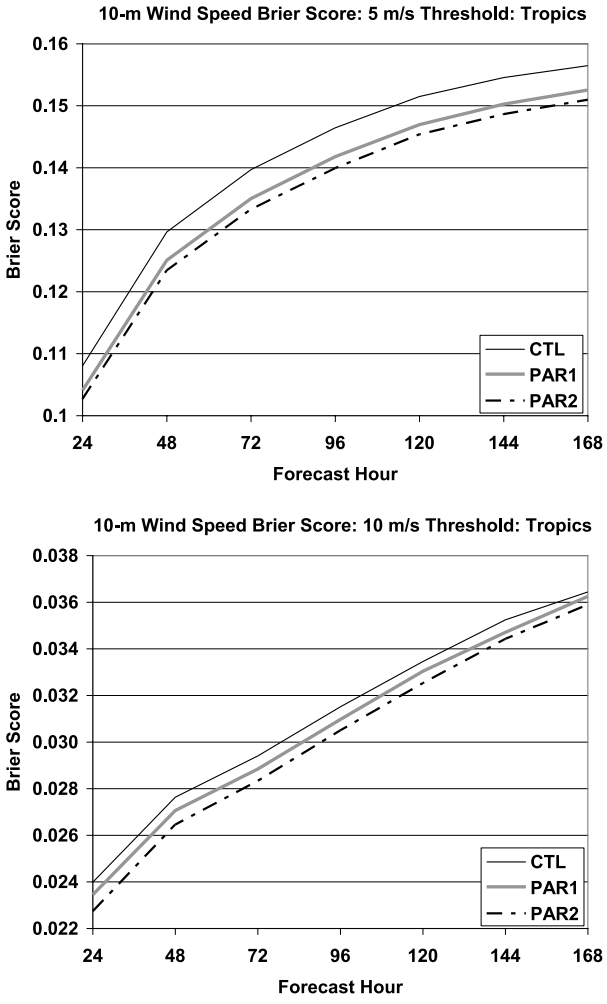

Fig. 7. Brier score for 10-m wind speed in the tropics (20S–20N) for CTL (thin black), PAR1 (thick grey) and PAR2 (thick dotted–dashed) ensembles for (top) 5 m s^{-1} wind speed threshold, and (bottom) 10 m s^{-1} wind speed threshold. For the 5 m s^{-1} threshold, the improvements of PAR1 and PAR2 over CTL are statistically significant at the 95% level for all forecast lengths. For the 10 m s^{-1} threshold, the improvements of PAR1 and PAR2 over CTL are statistically significant at the 95% level out to 96 hr.

differences is determined using a moving block bootstrap technique as described in McLay and Reynolds (2009). While the Brier scores in the mid-latitudes (not shown) are basically not sensitive to the parameter variations, there is significant improvement (reduction) in the Brier scores in the tropics at both the 5 and 10-m s^{-1} thresholds. For both thresholds, PAR2 produces better (lower) Brier scores than PAR1, and both are better than CTL. Specifically, for the 5-m s^{-1} threshold, both PAR1 and PAR2 are significantly better than CTL throughout the time period. PAR2 shows a statistically significant, though very small, improvement over PAR1. For the 10-m s^{-1} threshold, PAR1 and PAR2 are significantly better than CTL through the first half of the forecast period. The improvement of PAR2 over PAR1 is significant through 96-h, but the gains are very small.

Brier scores are also calculated for 850-hPa temperature in the tropics for temperatures greater than (less than) 1 standard deviation above (below) the time-mean value (Fig. 8). As with the Brier scores for 10-m wind speed, the temperature Brier scores are reduced (improved) through the inclusion of parameter variations, with the PAR2 ensemble performing better than the PAR1 ensemble. De-biasing (not shown) improves the scores for both positive and negative thresholds, but does not change the relative skill of the different ensembles.

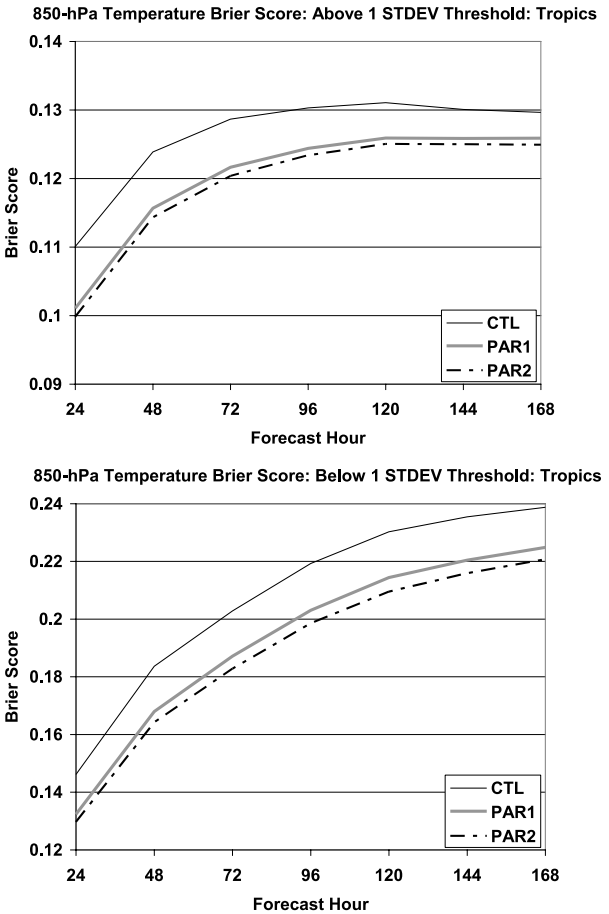

Fig. 8. Brier score for 850-hPa temperature in the tropics (20S–20N) for CTL (thin black), PAR1 (thick grey) and PAR2 (thick dotted–dashed) ensembles for (top) temperatures above 1 standard deviation above the mean, and (bottom) temperatures below 1 standard deviation below the mean.

Decomposition of the Brier scores into resolution and reliability (not shown) indicates that for both the wind and temperature fields, for both thresholds considered here, the resolution is improved through the addition of parameter variations, although the percentage improvement decreases as forecast length increases. The reliability results are more complex. For the low-level wind field, at the 5 m s^{-1} threshold, reliability is improved through the addition of parameter variations, with increased impact as forecast length increases. For the 10 m s^{-1} threshold, the addition of parameter variations actually degrades reliability at 24 and 48 h (improving reliability thereafter), but the degradation in reliability at short forecast lengths is small compared to the improvement in resolution. For temperature, the addition of parameter variations improves reliability for events less than 1 standard deviation below the time-mean value, and degrades reliability for events greater than 1 standard deviation above the time-mean value, although the degradation is small compared to the improvement in resolution. The fact that the improvements in Brier scores come primarily from improvements in the resolution, rather than reliability, suggests that the improvements in ensemble performance are not simply due to an increase in ensemble spread, but rather due to improvements in how well the system captures flow-dependent forecast uncertainties.

3.4. Tropical cyclone track errors

Previous work has shown that multimodel ensembles of tropical cyclone (TC) tracks may produce ensemble mean tracks of lower error than the individual ensemble members (Goerss, 2000). Therefore it is also plausible that a multiparameter ensemble, to the extent that it represents different forecast model formulations, may also improve ensemble-mean TC tracks over a single model ensemble. Figure 9 shows the ensemble mean TC track errors for all storms during the May 10 through September 12 2007 period for which the ensemble is run. The comparison is homogenous and the number of storms in each sample is shown below the *x*-axis. Track errors and statistical significance (accounting for serial correlation) are computed using the Automated Tropical Cyclone Forecast (ATCF) system (Sampson and Schrader 2000). In addition to the results for the CTL, PAR1 and PAR2 ensemble means, the TC track errors from the operational NOGAPS T239 deterministic forecast are also shown. The fact that the 12-h track errors are not increased in PAR1 and PAR2 over CTL suggests that parameter variations are not resulting in significant degradations to the initial perturbations or short-term forecasts. While the differences are relatively small, there are some statistically significant improvements gained by adding parameter variations. In particular, PAR2 offers small but

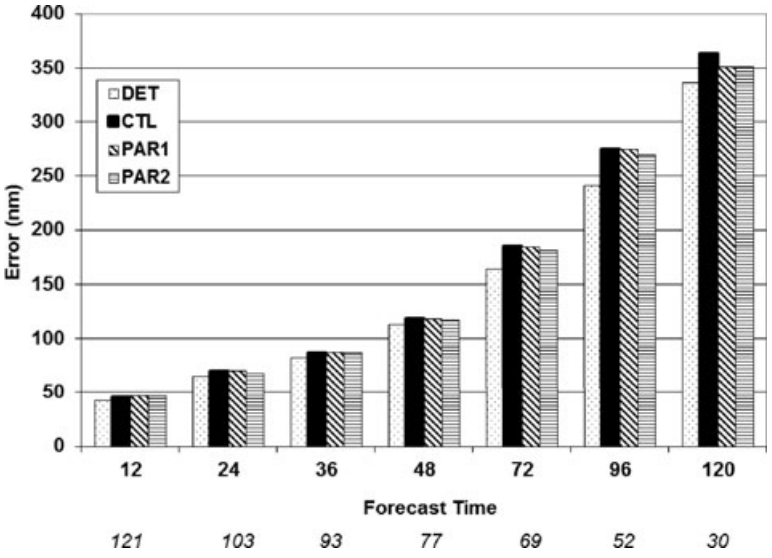

Fig. 9. Homogenous NH TC track forecast error (nm), for CTL, PAR1, and PAR2 ensemble mean track as denoted in key. Also shown is the average forecast error of the T239L30 NOGAPS operational deterministic forecast (DET). The numbers of verifying forecasts are shown below the *x*-axis. The improvements of PAR2 over CTL are statistically significant at the 95% level at 24, 72 and 120 h.

statistically significant improvements over CTL at the 95% level at 24, 72 and 120 h (and better at the 94% level at 96 h). However, these small improvements are not enough to make the T119 ensemble mean better than the T239 single high-resolution forecast. Recent work (Reynolds et al., 2011) suggests that going to T159 resolution can result in ensemble mean TC track forecasts that are better than the high-resolution control at longer lead times.

3.5. Precipitation forecasts

While precipitation is not one of the high-interest parameters to the Navy, an examination of the ensemble performance in terms of precipitation is useful to understand more fully how the parameter variations are impacting forecast evolution. This is particularly true as parameter variations are directly impacting the convective parametrization. Fig. 10 shows the ensemble spread and forecast RMS error for accumulated precipitation averaged over the tropics, for CTL, PAR1, and PAR2. 24-h accumulated precipitation (mm) is considered here, with the *x*-axis values denoting the final forecast time of the 24-h interval. For precipitation validation data we use daily rainfall accumulation fields from the Tropical Rainfall Measuring Mission[1] (TRMM) (e.g. Simpson et al., 1988; Schumacher and Houze, 2000).

As was found for the other variables considered, the addition of parameter variations both increases ensemble spread and decreases ensemble-mean RMS error. Fig. 10, bottom panel, also includes the average RMS error of the perturbed ensemble members, in addition to the skill of the ensemble member without initial perturbations or parameter variations (i.e. the unperturbed member). For CTL, the average skill of the perturbed members is very close to the skill of the unperturbed member (less than 0.7% difference, not shown). For PAR1 and PAR2, the range of forecast error (denoted by the vertical lines) among the individual members is larger, but they still differ from the unperturbed member by less than 3.5% after the first 24-h, and by less than 7.5% later in the integration. Some of the perturbed members are actually more skillful than the unperturbed member, and therefore the average of the PAR1 and PAR2 individual ensemble member errors is only slightly larger than the average of the CTL individual ensemble members. These relatively small changes in precipitation forecast skill suggest that the impacts of parameter variations on the initial perturbations and short-term forecasts are not unreasonable.

Figure 11 shows the Brier score for 24-h cumulative precipitation for the 5 mm threshold. As found with the low-level winds and temperature, the addition of the parameter variations provides a modest improvement (about a 5% decrease) to the Brier score throughout the forecast interval, with the improvements from PAR2 slightly larger than those from PAR1. Similar improvements were found for other thresholds (0, 3, 10 and 20 mm, not shown).

4. Evaluation of parameter impacts

The previous results indicate that the PAR2 parameter variations are more effective than the PAR1 parameter variations in improving ensemble performance. The parameter variations are based on experience gained within the ongoing effort to develop and modify parametrizations to improve the deterministic forecast model. Because of computational constraints, ensembles are not produced varying only one parameter at a time, and the PAR1 and PAR 2 ensembles have two varying parameters in common (*cu* and *dtmax*), and two different varying parameters

[1] The TRMM rainfall data were acquired using the GES-DISC Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni), a part of the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

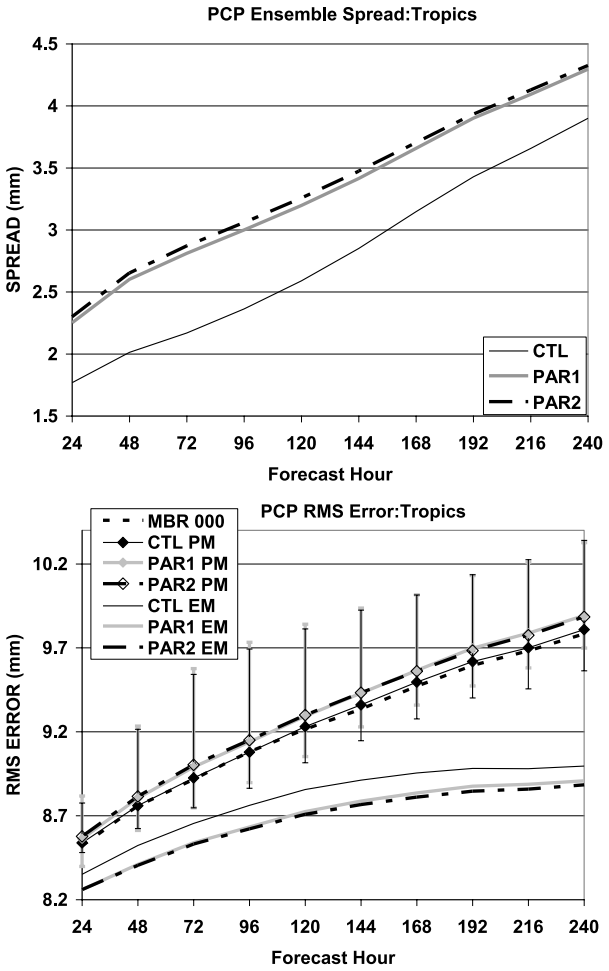

Fig. 10. Tropical (20S-20N) 24-h cumulative precipitation ensemble spread (top) and ensemble mean rms error (bottom) for CTL (thin black), PAR1 (thick grey) and PAR2 (thick dotted–dashed) ensembles. In the bottom panel, the ensemble-mean rms errors (EM) are denoted by the lines without symbols. The lines with symbols (PM) indicate the average rms error of the individual perturbed ensemble members. Vertical bars indicate the range of values for the PAR1 and PAR2 cases. The error of the unperturbed member (MBR 000) is indicated by the thick dotted line.

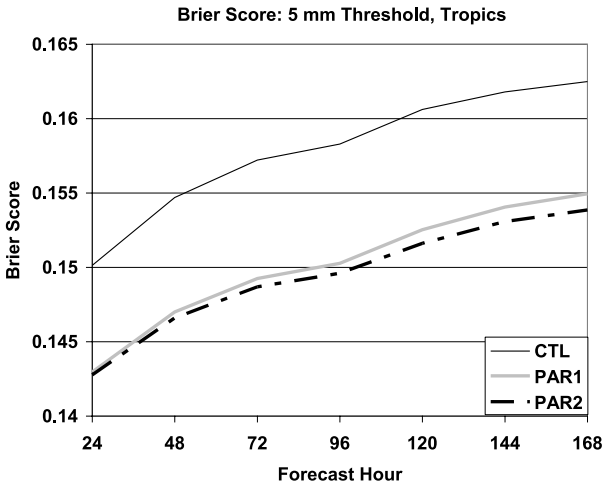

Fig. 11. Brier score for 24-h accumulated precipitation for the 5 mm threshold in the tropics (20S-20N) for CTL (thin black), PAR1 (thick grey) and PAR2 (thick dotted–dashed) ensembles as a function of forecast time.

(*alpha* and *damp* in PAR1, and *sigs* and *k* in PAR2). Despite the lack of independence in the parameter variations between the ensembles, it is still possible to gain some insight into the impact of individual parameters on ensemble spread by examining subsets of the ensemble forecasts.

For each parameter of interest there is one ensemble member with a high value of that particular parameter and all other parameters set to their control values. Likewise, there is one other member with a low value of that particular parameter and all other parameters set to their control value. 3-member subensembles are formed based on member 0 (all parameters set to the control value), and the two other members where only the particular parameter of interest is varied. These subensembles are referred to as CU, DTMX, DAMP, ALPH, SIGS, and VKRM for the *cu*, *dtmax*, *damp*, *alpha*, *sigs* and *k* parameters, respectively. As an example, for CU, this three member subensemble would be formed from members 0, 1, and 2. For DTMX this three member subensemble would be formed from members 0, 3, and 4. Because the ET is a cycling scheme, these subensembles will not produce the same results as if true 3-member ensembles are run, as each of the 33 6-h ensemble forecasts will contribute to the initial perturbations for the subsequent ensemble forecasts. Nevertheless, results from the subensembles are consistent with both the results from the evaluation of the full ensembles, and our understanding of the expected parameter impact. The spread of a 3-member control ensemble is calculated for comparison.

Figure 12 shows the ensemble spread of the 3-member subensembles for the 850-hPa wind speed (top), and 850-hPa temperature (bottom) in the tropics. The results for the PAR1 subensembles are shown in grey, and the results for the PAR2 subensembles are shown in black. A bar chart is chosen here because it is difficult to distinguish between 8 different curves in one figure panel. For the wind speed, the largest ensemble spread is seen for the DTMX subensemble (filled bars), followed by CU subensemble (diagonal lines), for both PAR1 and PAR2. The subensemble with the third largest spread (after DTMX and CU), is VKRM, followed by SIGS, both from the PAR2 ensemble. The subensembles with the least spread are ALPH and DAMP, both from the PAR1 ensemble. These results suggest that for wind speed variations, *k* and *sigs* are more effective parameters to choose to enhance ensemble spread than *alpha* and *damp*. For temperature, the largest subensemble spread by far is found in the DTMX subensembles from both PAR1 and PAR2. The VKRM subensemble has the second-most spread, followed by

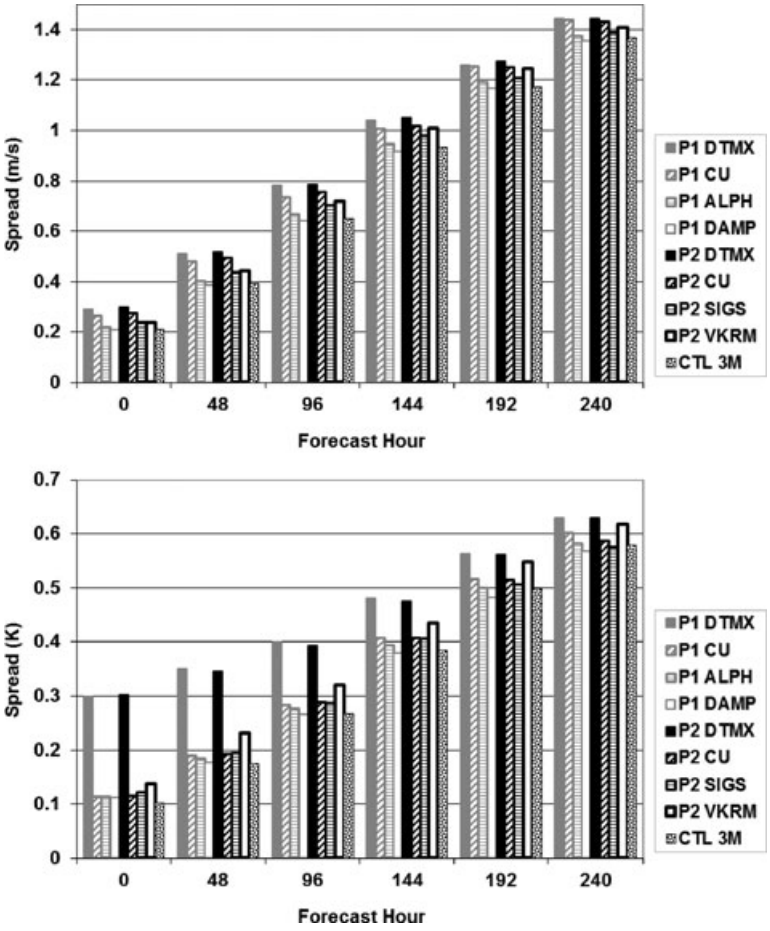

Fig. 12. Ensemble Spread (standard deviation about the ensemble mean) of the three member subensembles, as denoted in key, as a function of forecast time, for the 850-hPa wind speed (top, m s^{-1}), and the 850-hPa temperature (bottom, K) in the tropics.

the CU subensembles in PAR1 and PAR2. These results suggest that for temperature, in addition to varying *dtmax*, an effective way to introduce ensemble spread is through varying the von Kármán constant. For both the wind speed and temperature, the DAMP subensemble does not effectively increase the ensemble spread above the three-member control ensemble.

To explore the regional impact of these parameter variations, latitude-longitude plots of subensemble variances relative to the 3-member control subensemble are shown for selected subensembles. Figure 13 shows the percentage difference for the 850-hPa wind speed ensemble variance between the PAR1 DTMX, CU, and ALPH subensembles and the 3-member control ensemble. DTMX substantially increases ensemble variance over the control ensemble through much of the tropics. CU also shows substantial increases in ensemble variance, although, with the exception of the Bay of Bengal and Arabian Sea area, these increases are not as large as in DTMX. In contrast, the increase in spread for ALPH over the control is quite small and for the most part not significant. The results for DAMP (not shown) are similar to those for ALPH. Results for the PAR2 DTMX, CU and VKRM subensembles are shown in Fig. 14. The increases in ensemble variance for DTMX and CU in PAR2 (top and middle panels of Fig. 14) are quite similar to those found for DTMX and CU in PAR1 (top and middle panels of Fig. 13). In contrast to ALPH, VKRM does show significant increases in variance over the control, particularly in the vicinity of the dateline (the SIGS subensemble results, not shown, are similar to those of VKRM). The increase in variance in this region exhibited by the PAR2 VKRM and SIGS subensembles helps explain the maximum in this region in Fig. 3 when comparing the full PAR2 and PAR1 ensemble variances.

In examining Figs. 13 and 14, the dominant impact of variations in *dtmax* on subensemble variance through most of the tropics is clear. Examination of the relation (4) governing the evolution of convective mass flux in the Emanuel scheme shows that in contrast to *damp* and *alpha*, *dtmax* directly impacts the timing and location of the onset of convection, and thus represents in a crude manner the 'convective trigger' in the parametrization (e.g. Kain and Fritsch, 1992; Li et al., 2008). The parameters *damp* and *alpha* share with *dtmax* a role in determining the magnitude of the convective mass flux, and thus the 'convective adjustment timescale' (e.g. Betts and Miller, 1986). Because convective schemes vary in substantial respects, it is not clear whether the relative sensitivities to variations in parameters

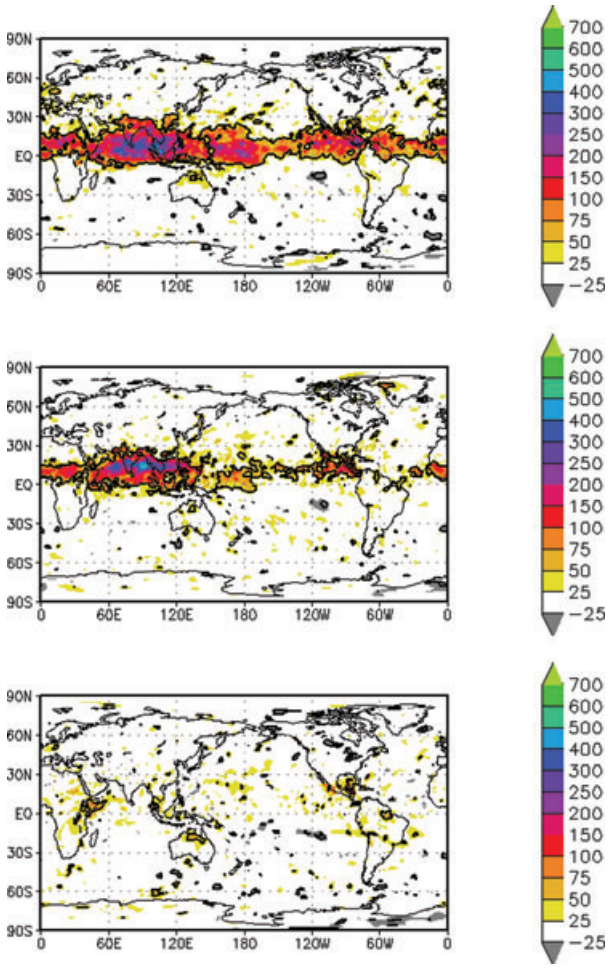

Fig. 13. Percentage difference between PAR1 subensembles DTMX (top), CU (middle) and ALPH (bottom) and a three-member control ensemble for the 850-hPa wind speed ensemble variance for the 5-d forecasts. Black contour indicates 95% confidence level.

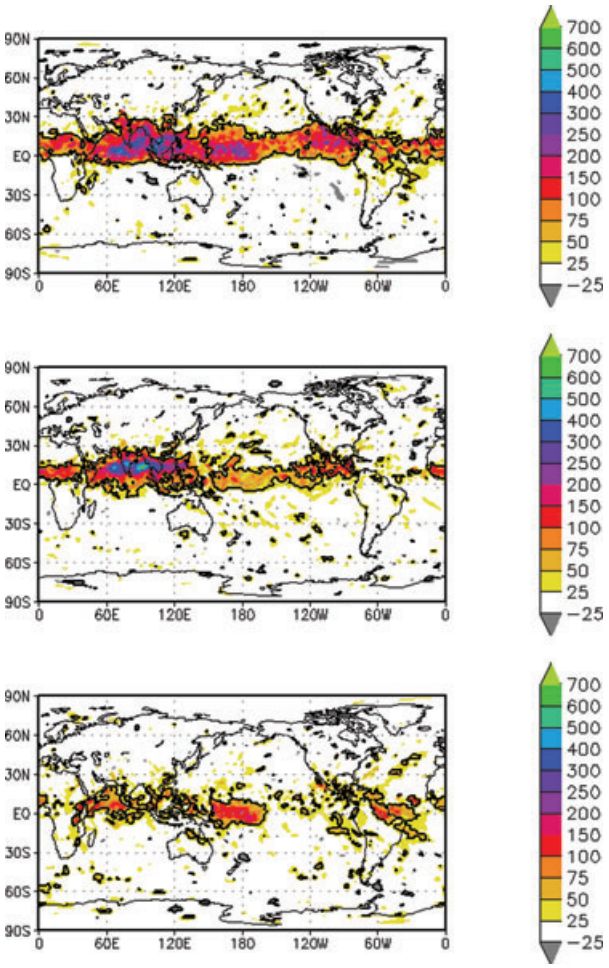

Fig. 14. Percentage difference between PAR2 subensembles DTMX (top), CU (middle) and VKRM (bottom) and a 3-member control ensemble for the 850-hPa wind speed ensemble variance for the 5-d forecasts. Black contour indicates 95% confidence level.

controlling these basic features observed here will hold for other schemes.

Similar plots are shown for the PAR2 subensembles for the 850-hPa temperature (Fig. 15). DTMX has significantly more ensemble variance than the control through much of the tropics and subtropics. In contrast to the wind fields, now VKRM has the second largest increase in ensemble variance, with smaller increases in CU. The PAR1 DTMX and CU subensembles, not shown, are very similar to the PAR2 DTMX and CU subensembles. The PAR1 ALPH and DAMP subensembles and PAR2 SIGS subensemble (not shown) do not indicate significant increases in ensemble variance over the control subensemble for the 850-hPa temperature field. This suggests that the addition of k to the suite of parameters that are varied is the essential ingredient in increasing the temperature ensemble variance in PAR2 over PAR1 (bottom panel of Fig. 3). It should be noted that comparison of individual ensemble members shows that varying k within the range adopted for PAR2 yields changes in the 850 hPa temperature bias (not shown) over broader scales (extending to the polar regions). The latitudinally constrained region of significant enhancement of the ensemble variance seen here may reflect the importance of interactions between the convection and boundary layer schemes for increasing ensemble spread, as suggested by the similarities between the regions of enhanced variance for k and *dtmax* (top and bottom panels of Fig. 15).

Comparison of the RMSE averaged for the subensemble perturbed members with the RMSE averaged for the CTL perturbed members confirms that the parameter variations have not significantly impacted the average skill of the individual members. The maximum percentage difference in the RSME between the perturbed subensemble members and the perturbed CTL members, for all subensemble, at all forecast lead time beyond 24 h, is less than 1.2% for the 850-hPa wind speed and less than 1% for the 850-hPa temperature.

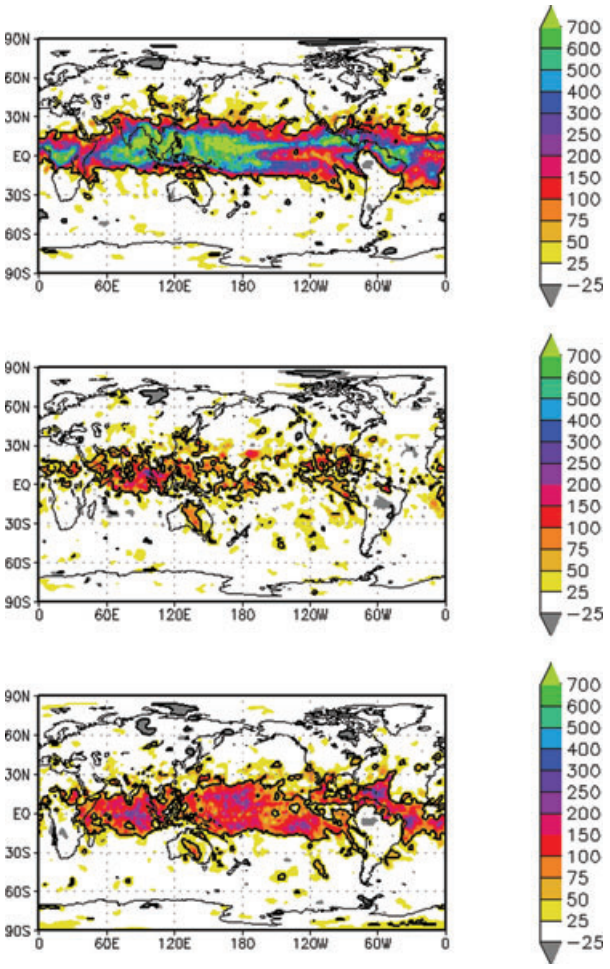

Fig. 15. Percentage difference between PAR2 subensembles DTMX (top), CU (middle) and VKRM (bottom) and a three-member control ensemble for the 850-hPa temperature ensemble variance for the 5-d forecasts. Black contour indicates 95% confidence level.

5. Summary

The impacts of two sets of parameter variations on ensemble performance for the U. S. Navy's global ensemble are examined. As the impact of the parameter variations is negligible in the extratropics, attention is focused on ensemble performance in the tropics. Both parameter variations result in increased ensemble spread (although the ensembles are still underdispersive) and improved (reduced) Brier scores for lower-level tropical winds and temperatures. The improvements to the Brier scores are obtained primarily through improvements to the resolution, rather than reliability, indicating that the improved ensemble performance is not just due to increased ensemble spread. Both also reduce ensemble mean RMSE, although this reduction is slight and not statistically significant. PAR2 has smaller ensemble mean TC track errors than CTL, and these differences are significant at the 95% level at 24, 72 and 120 h (and at the 94% level at 96 h). Improvements were also obtained for precipitation forecasts.

In general, PAR2, which contains parameter variations in the convective parametrization along with variations in the von Kármán constant in the boundary layer parametrization, outperforms PAR1, which contains parameter variations in the convective parametrization only. The calculation and examination of the variance of subensembles in which only one of the parameters is varied is used to identify those parameters which have a significant impact on ensemble variance as well as those that have little impact, without having to run additional ensembles. The variations in *dtmax* are found to dominate in both the PAR1 and PAR2 ensembles, for both temperature and wind fields, in the tropics. For the wind field, *cu* has the second largest impact in PAR1 and PAR2. The variations of *k* and *sigs* in PAR2 make significant contributions to ensemble spread in the wind field in the central Pacific, while the variations of *alpha* and *damp* in PAR1 do not. For the variations in the temperature field, *dtmax* dominates in both PAR1 and PAR2. The second largest impact is achieved through variations in *k*, resulting in significantly more variance in the PAR2 temperature field than in the PAR1 temperature field through most of the tropics and subtropics.

This work has shown that significant improvements in ensemble forecast performance are obtainable through varying model physics parameter values. The lack of impact in the mid-latitudes is consistent with the results of Hacker et al. (2011a), who find that parameter variations alone do not significantly impact ensemble performance for a mid-latitude domain. While the ensemble formulations tested here will hopefully provide insights with respect to application of the method, the realization of similar results in other models with different physical parametrizations cannot necessarily be expected. In addition, questions remain regarding some basic features of the method, including how varying parameters in time compares with holding parameter variations fixed throughout the integration. It is possible that strong temporal and spatial heterogeneities in the extra-tropics necessitate the need for temporal and spatial heterogeneities in the parameter variations in order to see a positive impact in those regions. Parameter variations within the physical parametrization schemes not considered here may also prove beneficial.

As this method only accounts for parametric uncertainty, it may complement approaches that address other aspects of model error. Several previous studies have found it profitable to combine different methods of accounting for model uncertainty (e.g. Bowler et al., 2008; Charron et al., 2010; Berner et al., 2011; Hacker et al., 2011a). Following the promising results from these other researchers, future experiments will investigate the impact of including both stochastic forcing and parameter variations in ensemble design, as the implementation of stochastic kinetic energy backscatter into the Navy's global ensemble system is also currently under development (Reynolds et al., 2011).

6. Acknowledgments

We gratefully acknowledge the support of 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. We thank Efren Serra and Buck Sampson for help with the tropical cyclone track statistics. We thank Josh Hacker and two anonymous reviewers for their thoughtful reviews, which have helped us improve the paper. COAMPS© is a registered trademark of the Naval Research Laboratory.

References

Alhamed, A., Lakshmivarahan, S. and Stensrud, D. J. 2002. Cluster analysis of multimodel ensemble data from SAMEX. *Mon. Wea. Rev.* **130**, 226–256.

Andreas, E. L. 2009. A new value of the von Kármán constant: implications and implementation. *J. Appl. Met. Clim.* **48**, 923–944.

Berner, J., Schutts, G. J., Leutbecher, M. and Palmer, T. N. 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.

Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A. and Snyder, C. 2011. Model uncertainty in a mesoscale ensemble prediction system: stochastic versus multi-physics representations. *Mon. Wea. Rev.* **139**, 1972–1995.

Betts, A. K. and Miller, M. J. 1986. A new convective adjustment scheme. Part II: single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.* **112**, 693–709.

Bishop, C. H. and Toth, Z. 1999. Ensemble transformation and adaptive observations. *J. Atmos. Sci.* **56**, 1748–1765.

Bowler, N. E., Arribas, A., Beare, S. E., Mylne, K. R. and Schutts, G. J. 2009. The local ETKF and SKEB: upgrades to the MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.* **135**, 767–776.

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

Businger, J. A., Wyngaard, J. C., Izumi, Y. and Bradley, E. F. 1971. Flux profile relationships in the atmospheric surface layer. *J. Atmos. Sci.* **28**, 181–189.

Bright, D. R. and Mullen, S. L. 2002. Short-range ensemble forecasts of precipitation during the southwest monsoon. *Wea. Forecast.* **17**, 1080–1100.

Buizza, R. and Palmer, T. N. 1998. Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.* **126**, 2503–2518.

Buizza, R., Miller, M. and Palmer, T. N., 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.* **125**, 2887–2908.

Charron, M., Pellerin, G., Spacek, L., Houtekamer, P. L., Gangon, N. and co-authors. 2010. Toward random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.* **138**, 1877–1901.

Chen, S. S., Houze, Jr., R. A. and Mapes, B. E. 1996. Multiscale variability of deep convection in relation to large-scale circulation in TOGA COARE. *J. Atmos. Sci.* **53**, 1380–1409.

Chen, S., Cummings, J., Doyle, J., Hodur, R., Holt, T. and co-authors. 2003. COAMPS version 3 model description—general theory and equations. Naval Research Laboratory Technical Report, NRL/PU7500–04-448. 141 pp.

Daley, R. and Barker, E. 2001. NAVDAS: formulation and diagnostics. *Mon. Wea. Rev.* **129**, 869–883.

Emanuel, K. A., 1991. A scheme for representing cumulus convection in large-scale models. *J. Atmos. Sci*. **48**, 2313–2335.

Emanuel, K.A. and Zivkovic-Rothman, M. 1999. Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.* **56**, 1766–1782.

Foken, T. 2008. Interactive comment on "The von Kármán constant retrieved from CASE-97 dataset using a variational method" by Y., Zhang *et al*. *Atmos. Chem. Phys. Discuss.* **8**, S67655–S6659, www.atmos-chem-phys-discuss.net/8/S6655/2008/.

Goerss, J. S. 2000. Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.* **128**, 1187–1193.

Harshvardhan, Davies, R., Randall, D. A. and Corsetti, T. G. 1987. A fast radiation parameterization for atmospheric circulation models. *J. Geophys. Res.* **92**, 1009–1016.

Hayashi, Y.-Y. and Sumi, A. 1986. The 30–40 day oscillations simulated in an "aqua-planet" model. *J. Meteorol. Soc. Japan* **64**, 451–466.

Hacker, J. P., Ha, S.-Y., Snyder, C., Berner, J., Eckel, F. A. and co-authors. 2011a. The U.S. Air Force Weather Agency's mesoscale ensemble: scientific description and performance results. *Tellus* **63A**, 625–641.

Hacker, J. P., Snyder, C., Ha, S.-Y. and Pocernich, M. 2011b. Linear and non-linear response to parameter variations in a mesoscale model. *Tellus* **63A**, 429–444.

Hendon, H. H. 1988. A simple model of the 40–50 day oscillation. *J. Atmos. Sci.* **45**, 569–584.

Hodur, R. M. 1997. The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.* **125**, 1414–1430.

Hogan, T. F. and Pauley, R. L. 2007. The impact of convective momentum transport on tropical cyclone track forecasts using the Emanuel cumulus parameterization. *Mon. Wea. Rev.* **135**, 1195–1207.

Houtekamer, P.L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.* **124**, 1225–1242.

Kain, J. S. and Fritsch, J. M. 1992. The role of the convective "trigger function" in numerical forecasts of mesoscale convective systems. *Meteorol. Atmos. Phys.* **49**, 93–106.

Li, X., Charron, M., Spacek, L. and Candille, G. 2008. A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. *Mon. Wea. Rev.* **136**, 443–462.

Lin, J. W.-B. and Neelin, J. D. 2000. Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophys. Res. Lett.* **27**, 3691–3694.

Louis, J. F. 1979. A parametric model of vertical eddy fluxes in the atmosphere. *Bound.-Layer Meteor.* **17**, 187–202.

Louis, J. F., Tiedtke, M. and Geleyn, J. F. 1982. A short history of the operational PBL parameterization at ECMWF. In *Proceedings of the ECMWF Workshop on Planetary Boundary Parameterizations*, Shinfield Park, Reading, UK, 59–79.

Madden, R. and Julian, P. 1971. Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.* **28**, 702–708.

Madden, R. and Julian, P. 1972. Description of global scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.* **29**, 1109–1123.

Madden, R. and Julian, P. 1994. Observations of the 40–50 day tropical oscillation—a review. *Mon. Wea. Rev.* **122**, 814–837.

McLay, J. G. and Reynolds, C. A. 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.

McLay, J. G., Bishop, C. H. and Reynolds, C. A. 2008. Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.* **136**, 1093–1108.

McLay, J. G., Bishop, C. H. and Reynolds, C. A. 2007. The ensemble transform scheme adapted for the generation of stochastic perturbations. *Quart. J. Roy. Meteor. Soc.* **133**, 1257–1266.

Peng, M. S., Ridout J. A. and Hogan, T. F. 2004. Recent Modifications of the Emanuel Convective Scheme in the Navy Operational Global Atmospheric Prediction System. *Mon. Wea. Rev.* **132**, 1254–1268.

Raymond, D. J. 1995. Regulation of moist convection over the west Pacific warm pool. *J. Atmos. Sci.* **52**, 3945–3959.

Reynolds, C. A., McLay, J. G., Goerss, J. S., Serra, E. A., Hodyss, D. and co-authors. 2011. Impact of resolution and design on the U.S. Navy Global Ensemble Performance in the Tropics. *Mon. Wea. Rev.* **139**, 2145–2155.

Reynolds, C. A., Teixeira, J. and McLay, J. G. 2008. Impact of stochastic convection on the ensemble transform. *Mon. Wea. Rev.* **136**, 4517–4526.

Ridout, J. A., Jin, Y. and Liou, C.-S. 2005. A cloud-base quasi-balance constraint for parameterized convection: application to the Kain-Fritsch cumulus scheme. *Mon. Wea. Rev.* **133**, 3315–3334.

Sampson, C. R. and Schrader, A. J. 2000. The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bull. Am. Meteor. Soc.* **81**, 1231–1240.

Schumacher, C. and Houze Jr. R. A. 2000. Comparison of radar data from the TRMM satellite and Kwajalein oceanic validation site. *J. Appl. Meteor.* **39**, 2151–2164.

Shutts, G. and Palmer, T. N. 2004. The use of high-resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. In *ECMWF/CLIVAR Proc. Simulation and Prediction of Intra-seasonal Variability with Emphasis on the MJO*. ECMWF, Reading, United Kingdom, 83–102.

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

Simpson, J., Adler, R. F. and North G. R. 1988. Proposed Tropical Rainfall Measuring Mission (TRMM) satellite. *Bull. Am. Meteor. Soc.* **69**, 278–295.

Stensrud, D. J., Bao, J.-W. and Warner, T. T. 2000. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.* **128**, 2077–2107.

Teixeira, J. and Hogan, T.F. 2002. Boundary layer clouds in a global atmospheric model: simple cloud cover parameterizations. *J. Climate*, **15**, 1261–1276.

Teixeira, J. and Reynolds, C. A. 2008. Stochastic nature of physical parameterizations in ensemble prediction: a stochastic convection approach. *Mon. Wea. Rev.* **136**, 483–496.

Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteor. Soc.* **74**, 2317–2330.

Wang, X. and Bishop, C. H. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**, 1140–1158.

Wei, M., Toth, Z., Wobus, R. and Zhu, Y. 2008. Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operation forecast system. *Tellus* **60A**, 62–79.

Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. H. and co-authors. 2006. Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus*, **58A**, 28–44.

Wilks, D. S. 2006. *Statistical Methods in the Atmospheric Sciences*. 2nd Edition. Academic Press, Burlington, MA, USA, 627 pp.

Zhang, Y., Ma, J. and Cao, Z. 2008. The von Kármán constant retrieved from CASES-97 dataset using a variational method. *Atmos. Chem. Phys.* **8**, 7045–7053.