NCEP NOTES

A Comparison of Perturbations from an Ensemble Transform and an Ensemble Kalman Filter for the NCEP Global Ensemble Forecast System

XIAQIONG ZHOU

NOAA/NWS/NCEP/EMC/I.M. Systems Group, College Park, Maryland

YUEJIAN ZHU AND DINGCHEN HOU

NOAA/NWS/NCEP/EMC, College Park, Maryland

DARYL KLEIST

Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

(Manuscript received 3 June 2016, in final form 27 September 2016)

ABSTRACT

Two perturbation generation schemes, the ensemble transformation with rescaling (ETR) and the ensemble Kalman filter (EnKF), are compared for the NCEP operational environment for the Global Ensemble Forecast System (GEFS). Experiments that utilize each of the two schemes are carried out and evaluated for two boreal summer seasons. It is found that these two schemes generally have comparable performance. Experiments utilizing both perturbation methods fail to generate sufficient spread at medium-range lead times beyond day 8. In general, the EnKF-based experiment outperforms the ETR in terms of the continuous ranked probability skill score (CRPSS) in the Northern Hemisphere (NH) for the first week. In the SH, the ensemble mean forecast is more skillful from the ETR perturbations. Additional experiments are performed with the stochastic total tendency perturbation (STTP) scheme, in which the total tendencies of all model variables are perturbed to represent the uncertainty in the forecast model. An improved spread–error relationship is found for the ETR-based experiments, but the STTP increases the ensemble spread for the EnKF-based experiment that is already overdispersive at early lead times, especially in the SH. With STTP employed, an increase in the EnKF-based CRPSS in the NH is reduced with a larger degradation in both the probability and ensemble-mean forecast skills in the SH. The results indicate that a rescaling of the EnKF initial perturbations and/or tuning of the STTP scheme is required when STTP is applied using the EnKF-based perturbations. This study provided guidance for the replacement of ETR with EnKF perturbations as part of the 2015 GEFS implementation.

1. Introduction

During integration of a numerical weather prediction model, small errors in the initial conditions can amplify, leading to significant forecast error (Lorenz 1969, 1982). One method for increasing forecasting skill, sampling system uncertainty, and improving forecast reliability is to generate ensemble forecasts starting from perturbed initial conditions instead of relying on a single deterministic forecast (Epstein 1969; Leith 1974). The initial perturbations should sample the error probability density functions (PDFs) of the analyzed atmospheric state and represent the analysis uncertainty. Several schemes have been developed and implemented at operational forecast centers, including the singular vector method at ECMWF (Buizza and Palmer 1995; Molteni et al. 1996), the bred vector method at NCEP (Toth and Kalnay 1993), and the ensemble data assimilation method at the Canadian Meteorological Centre (CMC; Houtekamer et al. 1996; Houtekamer and Mitchell 1998). As a result of limited computing resources, a finite number of ensemble members are used to sample the analysis uncertainties, resulting in imperfect representation thereof (Buizza et al.

Corresponding author e-mail: Xiaqiong Zhou, xiaqiong.zhou@noaa.gov

DOI: 10.1175/WAF-D-16-0109.1

2005). Furthermore, even with sufficient computing resources to allow for large ensemble sizes, assumptions made within each of the various methods result in an imperfect representation of the initial uncertainties.

The breeding method is a computationally inexpensive technique for generating initial perturbations and was implemented as part of the NCEP Global Ensemble Forecast System (GEFS) in December 1992 (Toth and Kalnay 1993, 1997). The bred vector (BV) perturbations generated from 6-h breeding cycles are assumed to simulate the fastest-growing analysis errors. These fast-growing BVs are expected to replicate the forecast error growth and encompass the true forecast error. It has been suggested that for simple dynamic models, BVs would be an estimate of the leading Lyapunov vectors with rescaling and an infinite breeding time during the breeding cycling (Trevisan and Legnani 1995; Szunyogh et al. 1997; Toth and Kalnay 1993, 1997). However, BVs cannot accurately sample the true analysis uncertainty and tend to be too similar to each other (Anderson 1997).

Wei et al. (2006, 2008) introduced the ensemble transformation with rescaling (ETR) algorithm as a modification to the breeding method for the NCEP GEFS in 2005. The analysis perturbations are generated by multiplying the 6-h forecast perturbations from the previous cycle by a transformation matrix. This matrix is constructed such that the perturbations are consistent with an estimate for the analysis error covariance. The resultant ensemble perturbations span more directions than BVs. Wei et al. (2008) compared the probabilistic scores of forecasts initialized from the BV and ETR methods and found that the ETR had superior performance.

A hybrid variational–ensemble data assimilation system (hereafter, hybrid system) was successfully implemented as part of the NCEP GDAS using the GFS model in May 2012 (Whitaker and Hamill 2002; Whitaker et al. 2008; Wang et al. 2013; Kleist and Ide 2015). In the hybrid system, a flow-dependent background error covariance estimate based on the 6-h ensemble forecasts from an ensemble Kalman filter (EnKF) is combined with a static background error covariance from the Gridpoint Statistical Interpolation (GSI) deterministic data assimilation system (Wu et al. 2002; Kleist et al. 2009). The implementation of an EnKF as part of the NCEP GDAS provides an alternate method for generating ensemble initial perturbations for the operational GEFS. However, as the perturbations are designed to represent the short-range background error covariance, it is necessary to test whether the EnKF-based perturbations can or cannot perform as well as the ETR-based perturbations for medium-range forecasting in the GEFS.

The level of performance of the EnKF-based and BV initial perturbation generation schemes have been previously compared, but many of the conclusions were based on idealized numerical models or only are valid in terms of the performance for entire operational systems instead of the ensemble perturbation schemes alone. Idealized numerical studies showed that the EnKF performed better than the breeding method for generating initial perturbations (Bowler 2006; Descamps and Talagrand 2007). Buizza et al. (2005) compared the performance of the operational global ensemble forecast systems at three operational centers: NCEP, ECMWF, and CMC. These three ensemble forecast systems used the breeding of growing modes (BGM), singular vector (SV), and EnKF schemes, respectively, to generate the initial perturbations. The results indicated that ECMWF had the best overall performance. However, the purpose of the comparison was to evaluate the overall quality of the three ensemble systems rather than to focus solely on the performance of the initial perturbations. There were substantial additional differences between the three systems, such as the data assimilation algorithms, numerical modeling techniques, and model resolutions that were used in each.

The goal of this study is to perform a clean comparison between two perturbation generation schemes: the ETR-based scheme and the operational NCEP hybrid assimilation EnKF-based scheme. Section 2 describes the general experimental setup of the study. An overview of verification scores is presented in section 3. An evaluation of the two experiments covering two boreal summer seasons is presented in sections 4 and 5, and the verification of an ensemble tropical cyclone track is summarized in section 6. The conclusions and general discussions are presented in section 7.

2. Experimental setup and verification method

a. Experimental setup

The NCEP GEFS uses the stochastic total tendency perturbation (STTP) scheme to represent the effect of model uncertainties (Hou et al. 2006, 2008). In the STTP algorithm, the total tendency of the model prognostic variables (surface pressure, temperature, wind, and humidity) is perturbed stochastically by using the following equation:

$$x'_{i,t} = x_{i,t} + \gamma_t \sum_{j=1}^{N} w_{i,j}\{[(x_j)_t - (x_j)_{t-6h}] - [(x_0)_t - (x_0)_{t-6h}]\}, \quad (1)$$

where $x_{i,t}$ is the total tendency of the ith ensemble member at the forecast lead time t and x_0 is for the control. The temporal change of the total tendency for each ensemble member and the control is calculated

within a 6-h time interval. The differences in this temporal change between each ensemble member and the control are used to perturb the total tendency after a multiplication by a random number $w_{i,j}$ and the application of an additional rescaling factor γ_t. The rescaling factor is used to tune the perturbation amplitude as a function of region [Eqs. (5) and (6) in Hou et al. 2008] and lead time [Eq. (4) in Hou et al. 2008]. The regional rescaling is used to enforce larger (smaller) perturbations in the winter (summer) hemisphere. A slow decrease of the stochastic perturbation amplitude with lead time is empirically specified to ensure that the perturbation growth matches the error growth in the ensemble mean forecast.

Two groups of experiments are presented in this study using the ETR-based and EnKF-based schemes: 1) using solely initial perturbations (ETR_ONLY and EnKF_ONLY) and 2) initial and model perturbations (ETR_STTP and EnKF_STTP). The performance of ETR_ONLY and EnKF_ONLY is examined for the period of 1 July–25 October 2011 to study the differences that arise solely from changing the initial perturbation generation technique. The performance levels of ETR_STTP and EnKF_STTP are evaluated for the period of 1 July–30 September 2012 to explore the different initial perturbation generation schemes while attempting to account for model error. For each group of experiments, the same GFS model version and corresponding control analysis are used in order to ensure that the differences between the experiments are only due to the differences in the initial perturbations. Furthermore, the STTP-based experiments provide a proper benchmark for the replacement of ETR with EnKF perturbations within the context of the recent GEFS implementation (2015), given that the STTP is used to account for model uncertainty as a part of the operational GEFS.

The forecast model used in this study is the NCEP Global Forecast System (GFS) model, version 9.01 (http://www.emc.ncep.noaa.gov/?branch=GFS&tab=impl). The ensemble for each experiment consists of 1 control and 20 perturbed members. The control member is initialized with the hybrid GSI analysis (T574L64) interpolated to the lower ensemble model resolution (T254L42). The initial conditions for ensemble members are created by adding perturbations from either the cycled ETR or EnKF to the control analysis. The EnKF perturbations from the 2012 GDAS hybrid assimilation implementation are used, which was the operational version at the time that these experiments were performed. Both the EnKF and ETR cycling were run four times per day, but the medium-range ensemble forecasts in our experiments are run only once daily (0000 UTC) because of computational constraints. The model is integrated at T254L42 resolution for the first 8 days (0–192 h) and then at lower resolution (T190L42) for the second 8 days (192–384 h), as is done for the operational GEFS.

For the ETR_ONLY and ETR_STTP experiments, the generation of the initial perturbations follows the procedure used in the NCEP GEFS that was operational as of 2012 [see Wei et al. (2006) and Wei et al. (2008) for a description of the ETR scheme in detail]. In the EnKF_ONLY and EnKF_STTP experiments, 6-h ensemble forecast perturbations are used instead of the EnKF analysis perturbations because of data availability issues in the NCEP operational environment. The GDAS EnKF is run as part of the late analysis (GDAS) cycle rather than the early one (GFS). The first analysis is run shortly after the synoptic time (2 h 45 min later) and is used as the analysis for the GFS/GEFS forecast. The later analysis, which is run 3 h later but with the same observation window, is presumably of slightly better quality, since it contains late-arriving observations. As a result of the configuration of the hybrid scheme for the GFS and GDAS, the EnKF is available only from the previous cycle at the time the GEFS starts following the early GFS analysis. Therefore, the EnKF prior (6-h forecast ensemble) is used instead of the posterior to generate the initial ensemble for the GEFS.

Another important component of the operational GEFS is a tropical cyclone (TC) relocation scheme (Liu et al. 2000, 2006). This scheme was originally developed to reduce TC position errors in the background fields (short-term forecast) in the NCEP data assimilation system. In this vortex relocation scheme, the TC vortex is separated from the environmental field using the methods of Kurihara et al. (1993, 1995) and then added back at the observed cyclone location. Liu et al. (2006) investigated the impact of the vortex relocation scheme on TC track forecasts in the GEFS and found that the scheme reduced TC track spread, but had little effect on the ensemble mean track forecast error. However, their study evaluated only a small number of cases. In the present study, the vortex relocation scheme is utilized in the ETR_ONLY and ETR_STTP experiments for all ensemble members and the control, as is done in operations for the GEFS. The EnKF-based experiments are evaluated with and without this relocation scheme in order to reexamine its impact on the TC track forecasts. It should also be noted that the vortex relocation scheme is also utilized as part of the process to generate the GSI hybrid analysis (Liu et al. 2000; Kleist 2011), which serves as the control analysis for all GEFS integrations.

b. *Verification metrics*

For all experiments, forecasts are verified against the GFS analysis. The forecast and analysis fields are interpolated to a 2.5° × 2.5° latitude–longitude resolution grid. The

performance of the experiments is evaluated with the NCEP ensemble verification system (Zhu et al. 1996; Toth et al. 2003, 2006; Zhu and Toth 2008). The verification system includes the calculation of traditional verification metrics such as root-mean-square error (RMSE) and pattern anomaly correlation (PAC) for the ensemble mean. It also computes metrics related to two important probabilistic attributes: the reliability and resolution (Toth et al. 2003, 2006), including the continuous ranked probability skill score (CRPSS) and the ranked probability skill score (RPSS). CRPSS is a skill score based on the ranked probability skill (CRPS) with a reference forecast based on the climatological distribution, measuring the difference of the full probability distribution [cumulative distribution functions (CDFs)] between the forecast and observations/analysis. The CRPSS has a maximum value of one, associated with a perfect forecast, while a value of zero is for forecasts that are no better than the reference.

A block bootstrap algorithm (Hamill 1999) is used to test the statistical significance of the differences between the ETR- and EnKF-based experiments. This technique generates multiple datasets by selecting random samples from the available data and can be applied to the statistical parameters regardless of the data probability distribution. In the current study, a 95% confidence level is derived from 20 000 random samples taken from more than 90 cases during the two experimental periods.

Tropical cyclone tracks are evaluated by comparing forecast-estimated tracks with the observed track from the best-track database for the Atlantic, western North Pacific, and eastern North Pacific basins (http://www.nhc.noaa.gov/data/). The ensemble mean track error is evaluated in addition to the ensemble track spread.

3. ETR and EnKF perturbations

The handling of model errors is also an important issue for EnKF data assimilation (Whitaker and Hamill 2002, 2012; Whitaker et al. 2008; Houtekamer et al. 2005, 2009; Zhang et al. 2004; Meng and Zhang 2008). Ideally, an EnKF ensemble should account for all sources of error, including sampling errors due to the limited ensemble size and errors in the forecast model. If the background error covariances are underestimated, the EnKF algorithm does not give sufficient weight to the observations. This problem can become progressively worse with time for cycled data assimilation systems, with the ensemble variance becoming unrealistically small as the filter gives increasing weight to its own forecast and ignores the information from observations, resulting in filter divergence (Anderson 2001; Evensen 2003).

To avoid filter divergence in EnKF data assimilation systems, multiplicative ensemble covariance (e.g., Whitaker and Hamill 2002, 2012) and additive noise (e.g., Whitaker et al. 2008; Houtekamer et al. 2005) inflation are often applied to the posterior ensemble to parameterize the various sources of the (under) represented error. The particular versions of multiplicative (relaxation to prior spread) and additive (using lagged forecast pairs) inflation used in the NCEP hybrid GSI–EnKF system[1] are based on Whitaker and Hamill (2012).

The ensemble spread in terms of total dry energy following Eq. (4) in Ma et al. (2014) is calculated for the ETR_ONLY and EnKF_ONLY analyses after multiplicative and additive inflation, as well as the EnKF_ONLY 6-h forecast, by

$$\mathrm{TE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{u_i'^2 + v_i'^2 + \kappa T_i'^2}, \tag{2}$$

where u_i', v_i', and $T_i'(i=1, 2, \ldots, N)$ are the deviations of the ith ensemble member from the mean for the wind components and temperature, with κ set to a value of $4.0\,\mathrm{J\,Kg^{-1}\,K^{-2}}$.

The vertical profiles of the period-averaged initial perturbation spread in the NH, SH, and tropics (Fig. 1) shows that the EnKF_ONLY ensemble spread decreases substantially in the first 6 h of model integration. While the two inflation methods have the effect of increasing the ensemble spread, they are also introducing imbalances to the analyzed states. Although the ad hoc methods of inflation act to stabilize the EnKF system and maintain appropriate amplitudes for spread, they are suboptimal in their representation of the real system errors. Also, while SV and BV perturbations explicitly target growing modes, perturbations based on analysis error covariances are expected to have large projections onto neutral or decaying modes. Without an explicit representation of the effects of model error, the decay of spread early in the forecasts is not unexpected.

The EnKF 6-h forecast perturbations are larger than the ETR initial perturbations (except at the tropical lower levels), which is consistent with the nature of these two techniques. The breeding cycling in ETR is explicitly designed to generate perturbations that contain fast-growing modes corresponding to the dynamics. As these are targeting fast-growing modes, their initial amplitudes are generally quite small and expected to grow quickly with forecast time (Toth and Kalnay 1993, 1997). In contrast, larger-amplitude perturbations are generally more favorable for the EnKF. EnKF methods that utilize inflation are typically somewhat overspread by

[1] As of January 2015, the EnKF that is part of the deterministic hybrid system uses stochastic physics instead of additive inflation.

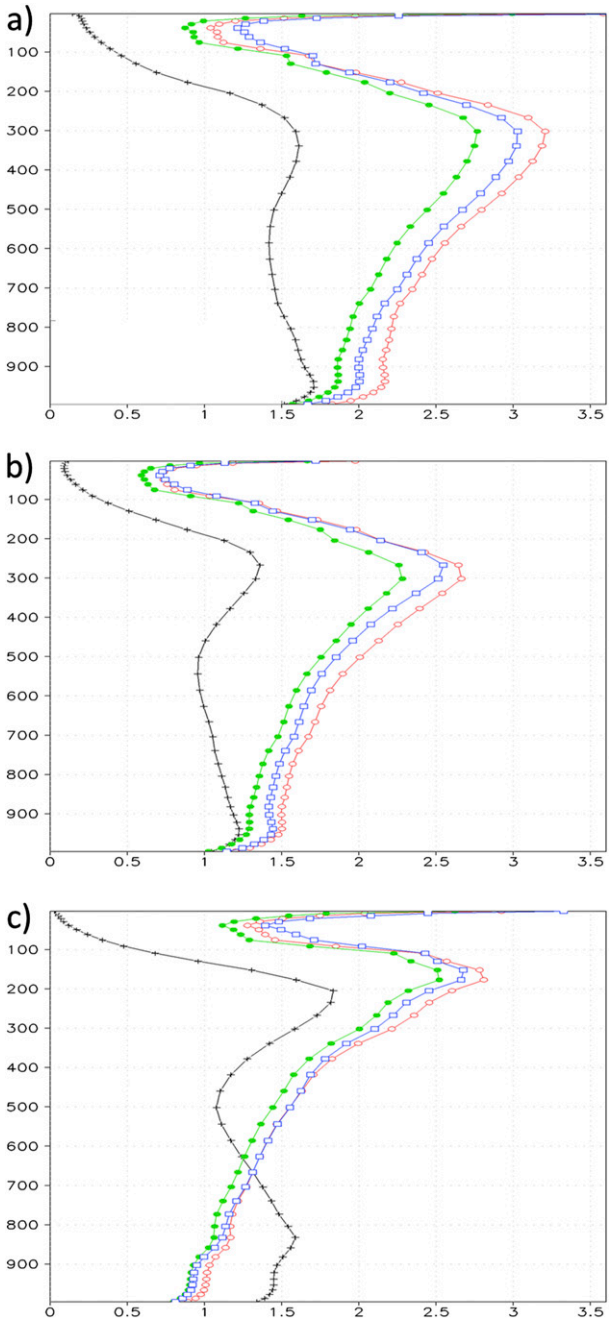

FIG. 1. The vertical profiles of the initial perturbation spread in terms of total dry energy in the ETR and EnKF experiments over the (a) NH, (b) SH, and (c) tropics. Three EnKF profiles represent the spread of EnKF perturbations after multiplicative inflation (green curves), additive inflation (red), and 6-h forecast (blue). The profiles are averaged from 1 Jul to 17 Oct 2011.

design to avoid eventual filter divergence. The EnKF posterior is designed to represent analysis errors, which are not necessarily fast-growing modes such as BV- and SV-based perturbations. Finally, the use of forecast perturbations instead of analysis perturbations is generally expected to result in larger amplitudes.

Figure 2 shows that the ETR and EnKF perturbations have significantly different geographical distributions (Fig. 2). The amplitude of the ETR perturbations is largest in the polar region (Fig. 2a), consistent with the regional rescaling mask that is applied (Ma et al. 2014). Smaller spread is found over land than over the oceans (Wang and Bishop 2003; Ma et al. 2014), which is predominantly a result of the assumed geographical inhomogeneity in the density of non-satellite-based observations. The geographic distribution of the EnKF spread is similar to that of the forecast ensemble (Figs. 2b and 3), which is a direct result of the initial perturbations in the EnKF experiment coming from 6-h forecasts, as well as the posterior ensemble spread being relaxed to the prior spread as part of the multiplicative inflation scheme (Whitaker and Hamill 2012). The land–sea contrast is less evident in the EnKF spread (Fig. 2b). The initial spread is greater where the mean kinetic energy is high. For the NH, the maximum spread corresponds to the storm tracks over the Pacific and Atlantic. The EnKF initial perturbations are zonally symmetric in the SH with large amplitudes in the baroclinic regions around 60°S.

The difference between the EnKF and ETR perturbations exhibits a coherent zonal structure (Fig. 2c). The EnKF initial perturbations are larger in the midlatitude baroclinic zones but are generally smaller at polar latitudes. Initially, substantial differences in spread exist between the EnKF and ETR perturbations, but these differences decrease with forecast time, especially in the baroclinic midlatitudes, where the ETR perturbations grow rapidly. By 96 h, the spatial patterns of the ETR- and EnKF-based ensemble spread are similar, with the exception of the EnKF perturbations exhibiting larger amplitudes in the midlatitude baroclinic zones (Fig. 3).

4. Experiments without STTP

Forecast skill scores are computed for geopotential height at 500 and 1000 hPa, the wind at 10 m, 850 and 250 hPa, and the temperature at 2 m and 850 hPa. The NH (SH) region is defined as the area between 20°N (S) and 80°N (S). The tropics are defined as the area between 20°S and 20°N. Scores are weighted by the cosine of latitude ϕ to account for the change in area with latitude on a latitude–longitude grid.

In this study, the focus is on the forecast scores of 500-hPa geopotential height in the NH and SH, and the zonal wind component at 850 hPa in the tropics. However, results for all of the aforementioned parameters

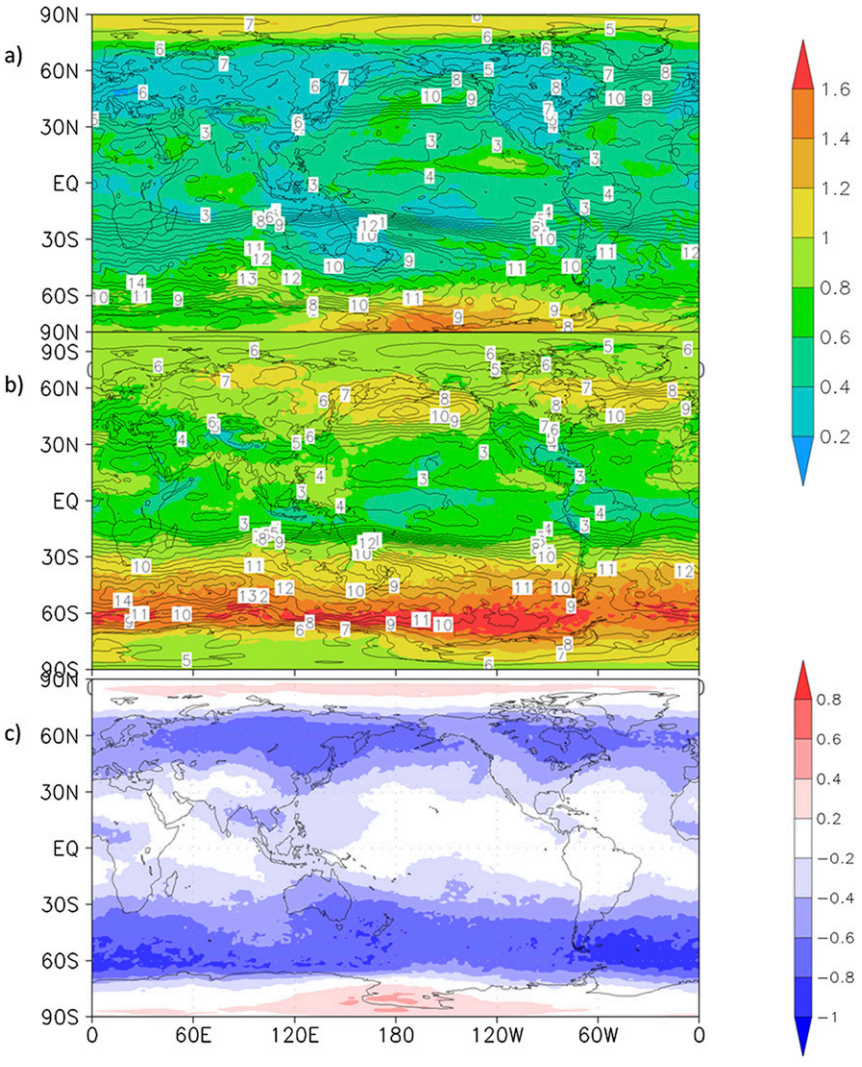

FIG. 2. Seasonally averaged kinetic energy at 500 hPa (contour) and the total energy ensemble spread (shading) at the initial time for (a) ETR, (b) EnKF, and (c) spread difference (ETR − EnKF) averaged from 1 Jul to 17 Oct 2011.

are included in a scorecard (Fig. 4). Overall, the EnKF_ONLY experiment outperforms the ETR_ONLY experiment in the NH and tropics, but exhibits some degradation in the SH in terms of RMSE and PAC (Fig. 4). The CRPSS of the EnKF_ONLY experiment is significantly higher in the NH in the first week. The improvement in the tropics is generally significant in the first 1–3 days in terms of PAC, RMSE, and CRPSS.

The skill of the ensemble mean can be taken as a first-order measure of the quality of the ensemble (Whitaker and Loughe 1998). Figure 5a shows the RMS error of the 500-hPa geopotential height ensemble mean forecast and ensemble spread for the NH of ENKF_ONLY and ETR_ONLY. The RMSE of both experiments rapidly increases over the first week and reaches its saturation point after 10 days. The RMSE is significantly smaller for ETR_ONLY for the first 2 days, but becomes similar at longer lead times in the NH. The EnKF_ONLY experiment is significantly better for the forecast of the low-level wind fields from day 1 to 5 and for the low-level temperature from day 1 to 3 in the NH (Fig. 4). There appears to be some advantage to using the ETR perturbations for the forecast skill in the SH (Fig. 5b), as the ETR_ONLY ensemble mean is significantly more accurate than the EnKF_ONLY ensemble mean for the first 4 days. Beyond 4 days, the ETR_ONLY ensemble mean remains more skillful, though the significance decreases with lead time. Similar to the 500-hPa geopotential height ensemble mean forecast, the skill degradation in EnKF_ONLY is generally

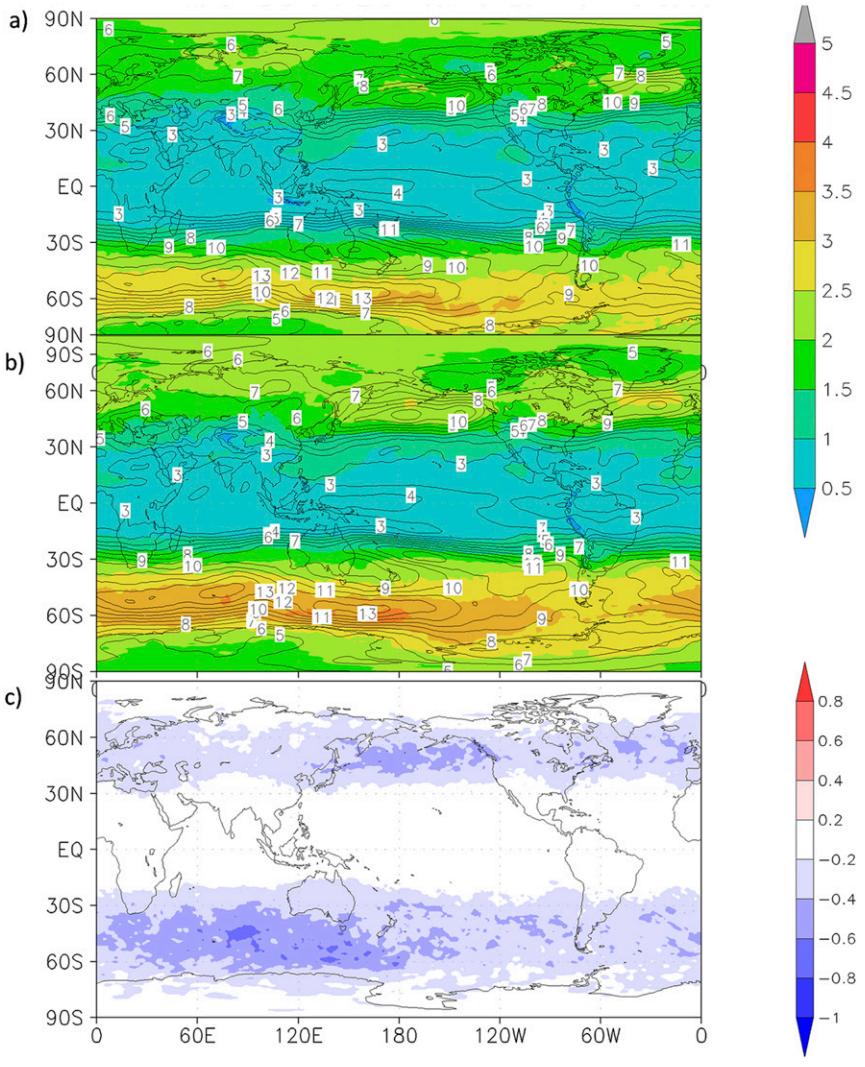

FIG. 3. As in Fig. 2, but for a 96-h forecast.

significant for all other variables in the first 1–3 days for the SH region.

Ideally, the spread of the ensemble forecast perturbations should be similar to the RMS error of the ensemble mean forecast at all lead times, representing the full forecast uncertainty (Whitaker and Loughe 1998; Zhu 2005). Having an ensemble that is either under- or overspread is not desirable for an ensemble forecasting system. Underdispersive (overdispersive) ensemble forecasts might lead to the underestimation (overestimation) of the probability of extreme events. Figure 5 shows that the growth of the ensemble spread is slower than that of ensemble mean forecast errors in both ETR_ONLY and EnKF_ONLY. For ETR_ONLY, the ensemble spread is generally underdispersive for 500-hPa geopotential height in both the NH and SH (Figs. 5a,b). The EnKF_ONLY experiment has larger spread than ETR_ONLY at the initial time, with overdispersion occurring for the first 3 (6) days in the NH (SH). With increasing lead time, this gradually transitions to underdispersion. The ensemble spread difference between EnKF_ONLY and ETR_ONLY increases with lead time for the first week. The decrease in spread difference in the second week is the result of the saturation of the ensemble spread and the forecast error.

Having an ensemble that is *slightly* overspread in the short-term forecast may actually be appropriate, as seen for the NH 500-hPa geopotential height. The ensemble mean forecast error is generally underestimated when self-analysis is used for verification. The 500-hPa geopotential height perturbations in the EnKF-ONLY experiment are overdispersive in the SH for the first week and are more evident and longer lasting than in the NH (Figs. 5a,b). This increased spread in the EnKF

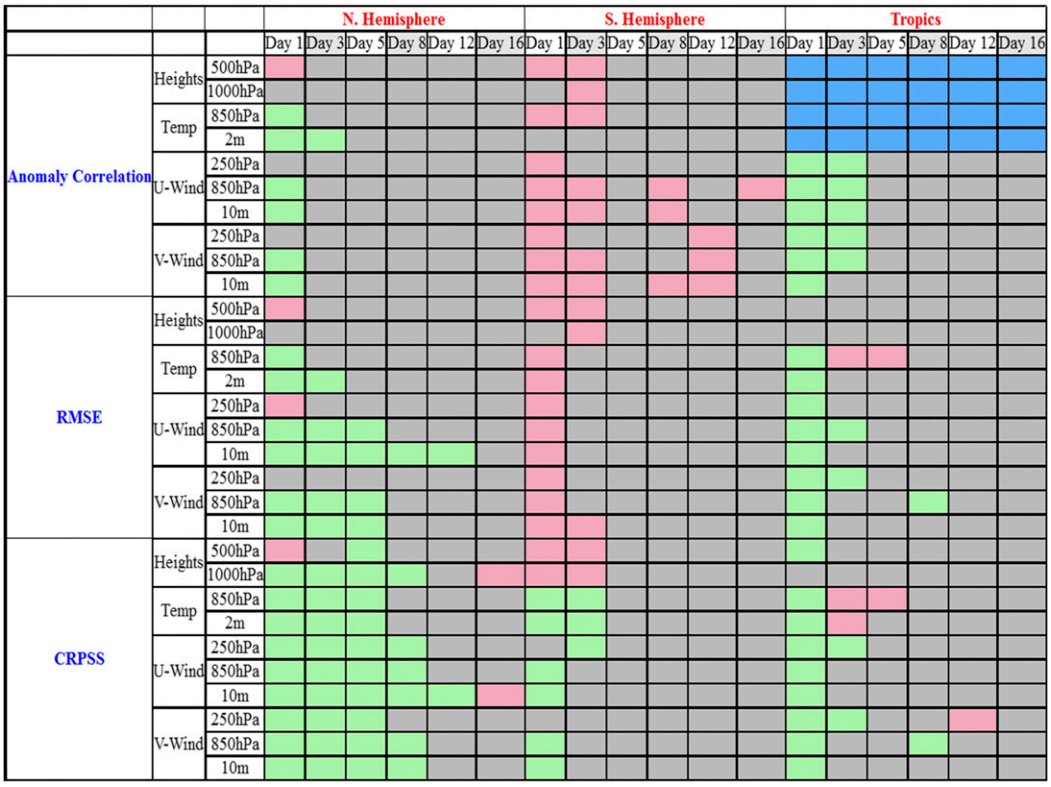

FIG. 4. Scorecards for the ETR_ONLY vs EnKF_ONLY experiments spanning the 2011 boreal summer. Green represents where the EnKF experiment is significantly better and red is significantly worse than the ETR experiment with the bootstrap significant test at 95% confidence levels. Gray means no significant difference.

perturbations is consistent with the degraded skill of the ensemble mean forecast for EnKF_ONLY. Overdispersion is also found in the short-range forecasts for other variables in the SH for this period (not shown), consistent with the corresponding forecast degradation (Fig. 2a). As shown in Fig. 2, the large spread of the EnKF initial perturbations is primarily located in the strongly baroclinic regions around 60°S. It may be possible to reduce the amplitude of these perturbations in this region at the initial time, assuming that small initial perturbations that project onto rapidly growing modes can result in larger spread with increasing forecast time, as in ETR_ONLY.

Figures 5c and 5d show the time evolution of the PAC for 500-hPa geopotential height in the NH and SH, respectively. The PAC score is close to the ideal value of 1 at the initial time and then decreases with increasing forecast lead time. There is little difference in the PAC scores of the geopotential height at 500 hPa for the first 5 days, although the scorecard shows a statistically significant degradation of EnKF_ONLY from day 1 to 3 in the SH (Figs. 4 and 5c,d).

The EnKF_ONLY experiment has statistically significantly higher CRPSS than ETR_ONLY, especially in the NH (Fig. 4). For 500-hPa geopotential height, the EnKF_ONLY CRPSS is significantly smaller in the NH than the ETR_ONLY CRPSS for the first 2 days, but it becomes larger at longer lead times (Fig. 6a). For other variables in the NH, the improvement of EnKF_ONLY in CRPSS is generally statistically significant in the first week (Fig. 4). For the SH, there is statistically significant degradation in the 500- and 1000-hPa geopotential height from day 1 to 3, but other parameters generally have significantly higher CRPSS in ENKF_ONLY.

The better performance for probability forecasts from EnKF_ONLY is also validated by performing a spectral analysis of the initial perturbations. The mean eigenvalue spectra of the ensemble covariance matrices are shown in Fig. 7. The sizes of the bars correspond to the eigenvalues of the covariance in terms of the normalized geopotential height at 500 hPa, with the sum of the eigenvalues corresponding to the ensemble spread. The majority of the ensemble variance for the both ETR and EnKF initial perturbations is contained in 19 uncorrelated orthogonal directions. However, the spectrum of the EnKF covariance is flatter than that of the ETR initial perturbations. The ETR ensemble has a much larger first eigenvalue, which indicates that the

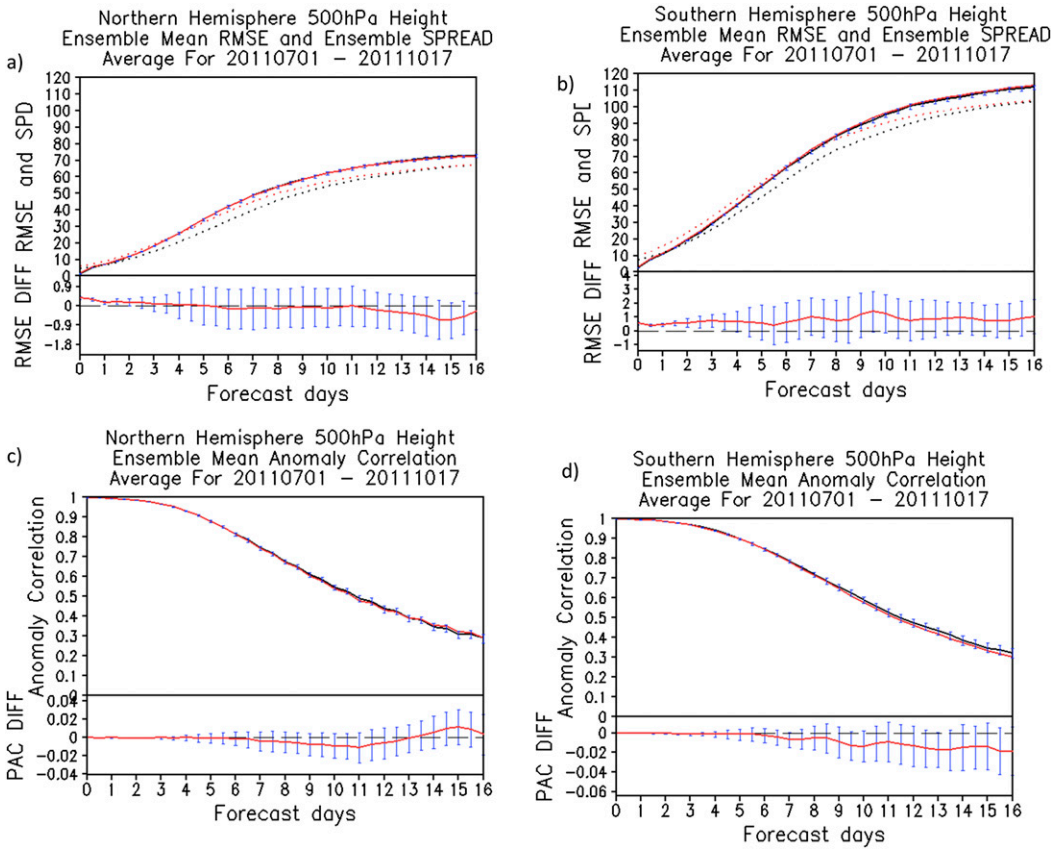

FIG. 5. Ensemble mean RMSE (solid) and ensemble standard deviation (dotted) for 500-hPa geopotential height over the (a) NH and (b) SH. The verification scores for ETR_ONLY (black) and EnKF_ONLY (red) are averaged during the period from 1 Jul to 17 Oct 2011. The bottom portion of each panel shows the difference in RMSE error (EnKF minus ETR). The blue bars display the 95% confidence intervals with the bootstrap significant test. (c),(d) As in (a),(b), but for the 500-hPa geopotential height ensemble mean predicted pattern anomaly correlations.

ensemble members are more similar to each other than the EnKF ensemble members. The spectra derived from the ETR ensemble are similar to those discussed by Wang and Bishop (2003) for the breeding method. The flat spectrum in the EnKF indicates a better estimation of the analysis error variance than in the ETR (Wang and Bishop 2003).

The EnKF initial perturbations have a limited, but generally positive, effect on forecasts in the tropics, as demonstrated by the RMSE for 850-hPa zonal wind (Fig. 8a). For EnKF_ONLY, the RMSE is significantly smaller from day 1 to 3 than for the ETR_ONLY, with the scores becoming similar between the two ensembles beyond day 4. Beyond 24 h, the spreads in both EnKF_ONLY and ETR_ONLY are generally much smaller than the RMSEs. This is observed despite the fact that the ETR_ONLY exhibits very large spread at the initial time. This initial overdispersion in ETR_ONLY is partially due to the perturbation amplitude for all vertical layers being rescaled with a factor based on a 500-hPa kinetic energy regional mask. For the lower levels, a tuning method that uses an ad hoc inflation factor is applied to obtain sufficient ensemble dispersion. This tuning strategy leads to an initial overdispersion at low levels in the tropics that decays rapidly with integration. Figures 8b and 8c show that the PAC and CRPSS scores for the 850-hPa zonal wind in the tropics for EnKF_ONLY are significantly higher for days 1–3 compared with ETR_ONLY, with the scores becoming more similar at longer lead times.

5. Experiments with STTP

The results in the previous section indicate that the spread growth for the ETR-based and the EnKF-based initial perturbations is not as fast as the growth of the ensemble-mean forecast error. Both EnKF_ONLY and ETR_ONLY are underdispersive for longer lead times. For EnKF-ONLY, the large initial spread does not propagate into the medium-range forecasts. This is to be

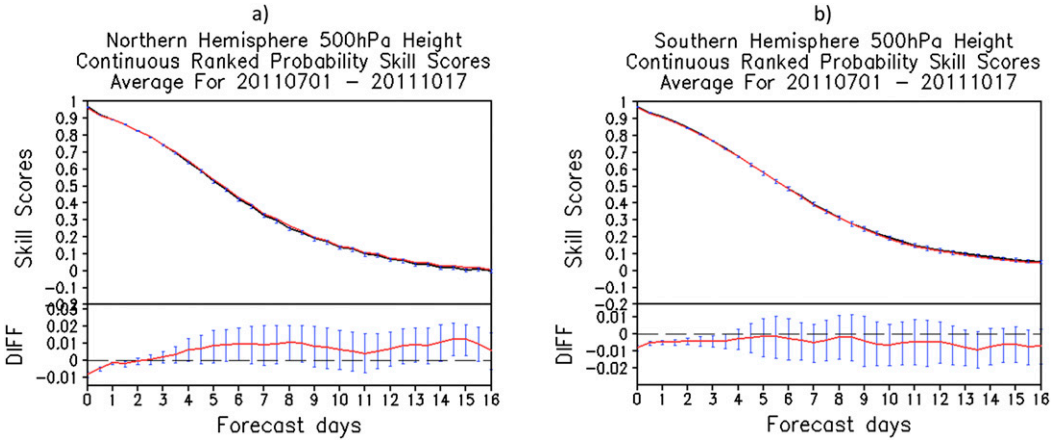

FIG. 6. CRPSSs for 500-hPa geopotential height over the (a) NH and (b) SH for EnKF_ONLY and ETR_ONLY. The bottom portion of each panel shows the difference of CRPSS (EnKF minus ETR). The blue bars display the 95% confidence intervals with the bootstrap significant test.

expected for a model configuration that does not take into account model uncertainty and error. In this section, the STTP model perturbation scheme, as used in the GEFS 2010 implementation, is included with both EnKF and ETR initial perturbations. The two experiments are named as EnKF_STTP and ETR_STTP, respectively. Note that ETR_STTP is identical to the operational GEFS that was implemented in 2012.

Consistent with the results of the experiments without STTP, the performance of EnKF_STTP is better than ETR_STTP in the NH and tropics. There is also a statistically significant degradation in the forecast skill in the SH for EnKF_STTP (Figs. 4 and 9). The degradation of the SH scores for EnKF_STTP is exacerbated for the 2012 period with the previously noted improvements in the NH scores lost.

As the forecast lead time increases, the STTP scheme adds spread regardless of the type of initial perturbations. The relationship between the ensemble spread and the RMS error is close to perfect for ETR_STTP (Figs. 10a,b). The ensemble spread in EnKF_STTP remains larger than that in ETR_STTP, with EnKF_STTP being overdispersive for the first 4 days. The STTP increases the ensemble spread for the EnKF_STTP experiment, which is already overdispersive at early lead times. It should be noted that the perturbations added by STTP, to some degree, are sensitive to the amplitude of the initial perturbation. Here, the parameters used in the STTP scheme have been carefully tuned for use with ETR and the associated amplitude of those initial perturbations. From the analysis presented later (Fig. 10), it can be concluded that the EnKF_STTP ensemble mean forecast is less skillful than ETR_STTP, with a larger 500-hPa geopotential height RMSE in the EnKF_STTP experiment. The difference is statistically significant at the 95% level through day 4 in the NH and day 8 in the SH. The aforementioned relationship between the ensemble spread and error for EnKF_STTP and ETR_STTP with STTP holds for other variables and levels (not shown).

To examine whether or not the degradation noticed for the EnKF experiment in the SH is due to seasonal effects, the EnKF experiment without STTP (EnKF_ONLY) is also performed for the 2012 boreal summer (Figs. 10a,b). The ensemble spread in EnKF_ONLY decreases significantly and there is no evident overdispersion in the SH. The RMSE of 500-hPa height in the EnKF experiment is comparable with that in ETR_STTP. However, with the decrease of the ensemble spread, the forecast skill of EnKF_ONLY is degraded in the second week in NH. In general, ETR_STTP (the operational GEFS) outperforms both EnKF_ONLY and EnKF_STTP in terms of the RMSE.

Similar to the RMSE, the difference in PAC between the ETR_STTP and EnKF_STTP experiments is more

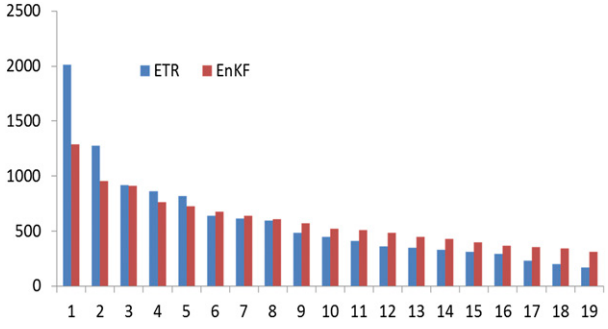

FIG. 7. Seasonal mean spectra of eigenvalues of ensemble-based initial spread of 500-hPa geopotential height for 2011 boreal summer in NH normalized by the global spread for ETR and EnKF.

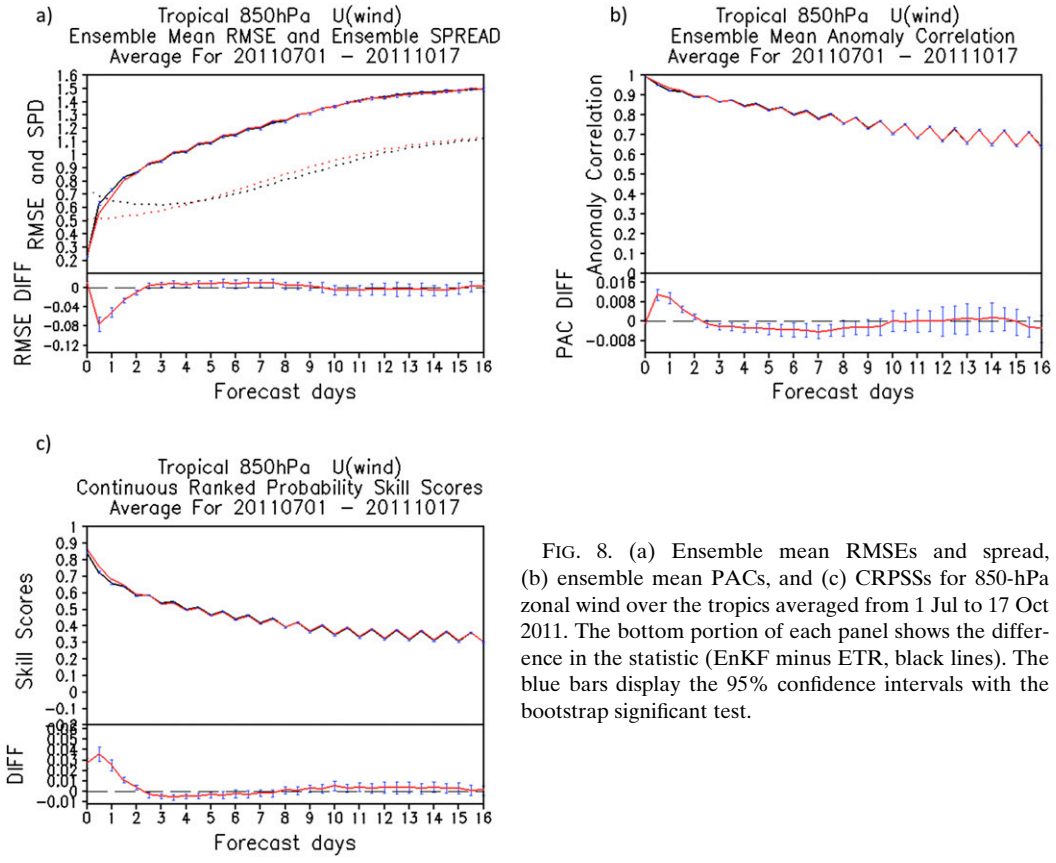

FIG. 8. (a) Ensemble mean RMSEs and spread, (b) ensemble mean PACs, and (c) CRPSSs for 850-hPa zonal wind over the tropics averaged from 1 Jul to 17 Oct 2011. The bottom portion of each panel shows the difference in the statistic (EnKF minus ETR, black lines). The blue bars display the 95% confidence intervals with the bootstrap significant test.

apparent than in the experiments without STTP. Figure 10c shows the 500-hPa geopotential height PAC becoming slightly greater with lead time in ETR_STTP than in EnKF_STTP for the NH. Similar results can also be found for the 1000-hPa geopotential height and for the wind field at various vertical levels (Fig. 9). The gap in performance between ETR_STTP and EnKF_STTP is especially pronounced in the SH, where the PAC of the 500-hPa geopotential height is statistically larger for ETR_STTP out to day 14 (Fig. 10b). The ETR_STTP experiment outperforms the EnKF_STTP experiment for the other variables except for 2-m temperature in the extratropics. In the tropics, EnKF_STTP outperforms ETR_STTP, particularly for forecasts from day 1 to 5 lead times (Fig. 9), similar to the experiments without STTP.

The 500-hPa geopotential height CRPSS scores in the NH are similar for ETR_STTP and EnKF_STTP (Fig. 11). For the first week, EnKF_STTP has more skill in the troposphere for several variables (Fig. 9). However, the results are opposite for the SH, with the CRPSS values from ETR_STTP being greater than from the EnKF_STTP. The skill in ETR_STTP is generally statistically better for lead times from day 3 to day 12.

In the tropics, the STTP scheme does not significantly increase the ensemble spread. For reference, Figs. 8 and 12 are provided as examples showing the 850-hPa zonal wind. The performance of the ETR and EnKF perturbations is similar for the 2011 and 2012 experiments, demonstrating a lack of sensitivity to STTP in the tropics. There is again a general improvement in the tropical forecasts resulting from the use of EnKF perturbations even with STTP, particularly for the forecast lead times from day 1 to day 5 (Fig. 9). The relatively small impact from STTP on the ensemble spread is due to the fact that the tendency perturbation applied within the STTP scheme is proportional to the total tendency change on the 6-h time scales [Eq. (1)], which is generally small in the tropical troposphere. Three alternate stochastic schemes were implemented on the short-term EnKF-based forecasts used in the data assimilation cycling: 1) stochastic kinetic energy backscatter (SKEB; Shutts 2005), 2) perturbed boundary layer humidity (SHUM; Tompkins and Berner 2008), and 3) stochastically perturbed physics tendencies (SPPT; Buizza et al. 1999; Palmer 1997, 2001; Palmer et al. 2005). The impact of replacing the operational STTP scheme with a combination of these three

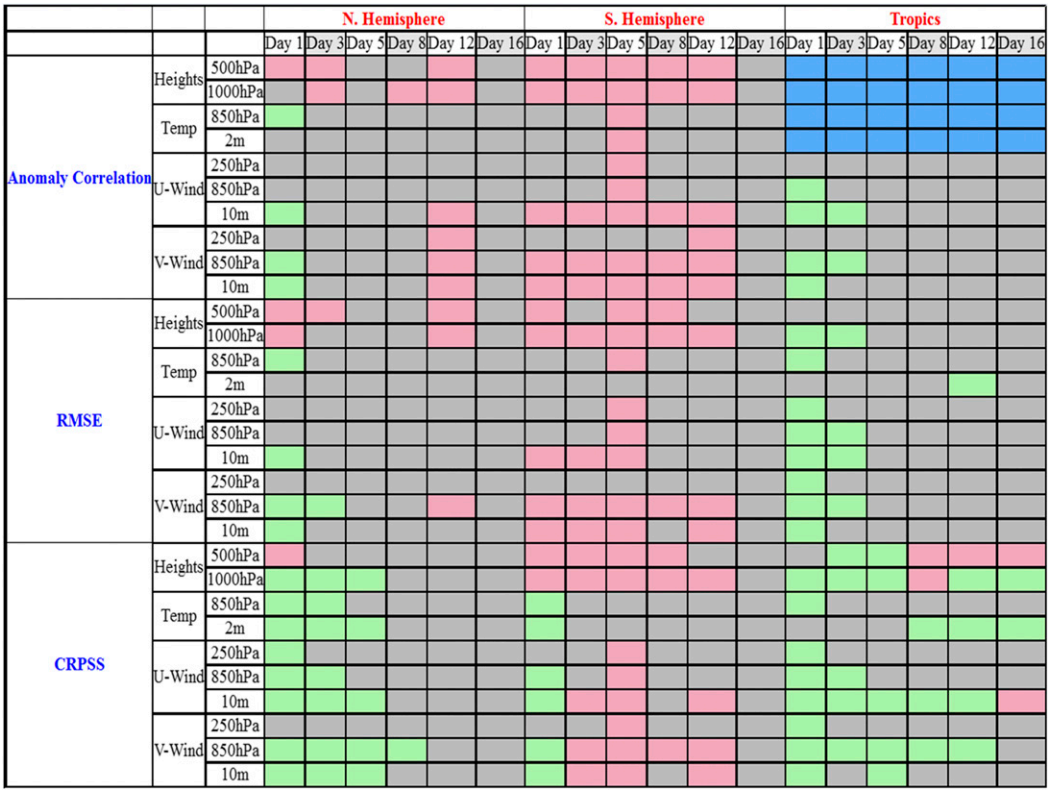

FIG. 9. Scorecards for the ETR_STTP vs EnKF_STTP experiments spanning the 2012 boreal summer. Green represents where the EnKF experiment is significantly better and red is significantly worse than the ETR experiment with the bootstrap significant test at 95% confidence levels. Gray means no significant difference.

stochastic schemes for use in the GEFS medium-range forecasts is currently under assessment.

6. Tropical cyclone track forecast

The ensemble-mean forecast of TC tracks is verified for two seasons across three basins: western North Pacific, eastern North Pacific, and North Atlantic. As in operations for the GEFS, a tropical cyclone relocation scheme (Liu et al. 2000, 2006) is used in the ETR experiments (ETR_ONLY and ETR_STTP) and the 2012 EnKF experiment with STTP. With the relocation scheme, the vortex for each member is mechanically relocated to the observed position dictated by the center responsible for advisories in the basin. For the 2011 EnKF experiment without STTP, two experiments are carried out to evaluate the impact of including the relocation scheme within the context of the EnKF initial perturbations (EnKF_ONLY and EnKF_ONLY_R, respectively). It is expected that the relocation scheme may reduce the initial TC spread as well as the track forecast error in the EnKF experiments, particularly for short forecast lead times.

First, the forecasts from the 6-h EnKF perturbations without relocation (EnKF_ONLY) are compared with the ETR with relocation (ETR_ONLY) for the 2011 experiments (Fig. 13a). The ensemble spread of the TC positions at the initial time is larger in EnKF_ONLY than in ETR_ONLY but becomes similar after 72 h as the spread growth in EnKF_ONLY is slower than in ETR_ONLY. Meanwhile, the EnKF_ONLY ensemble mean forecast exhibits larger forecast errors than ETR_ONLY at all forecast times, with the differences being statistically significant at 12-, 48-, 72-, and 120-h forecast times.

The degradation in the track forecasts for EnKF_ONLY is partially a result of the absence of TC relocation. When the EnKF_ONLY 6-h forecast perturbations are centered on the analysis, misalignment of some of the members can occur, exacerbated when the center of the EnKF_ONLY 6-h ensemble mean deviates from that of the control. The use of the TC relocation scheme in EnKF_ONLY_R reduces the track spread at the short forecast range up to 3 days (Fig. 13a). The track error is generally reduced in the EnKF_ONLY_R experiment. The addition of the TC relocation algorithm

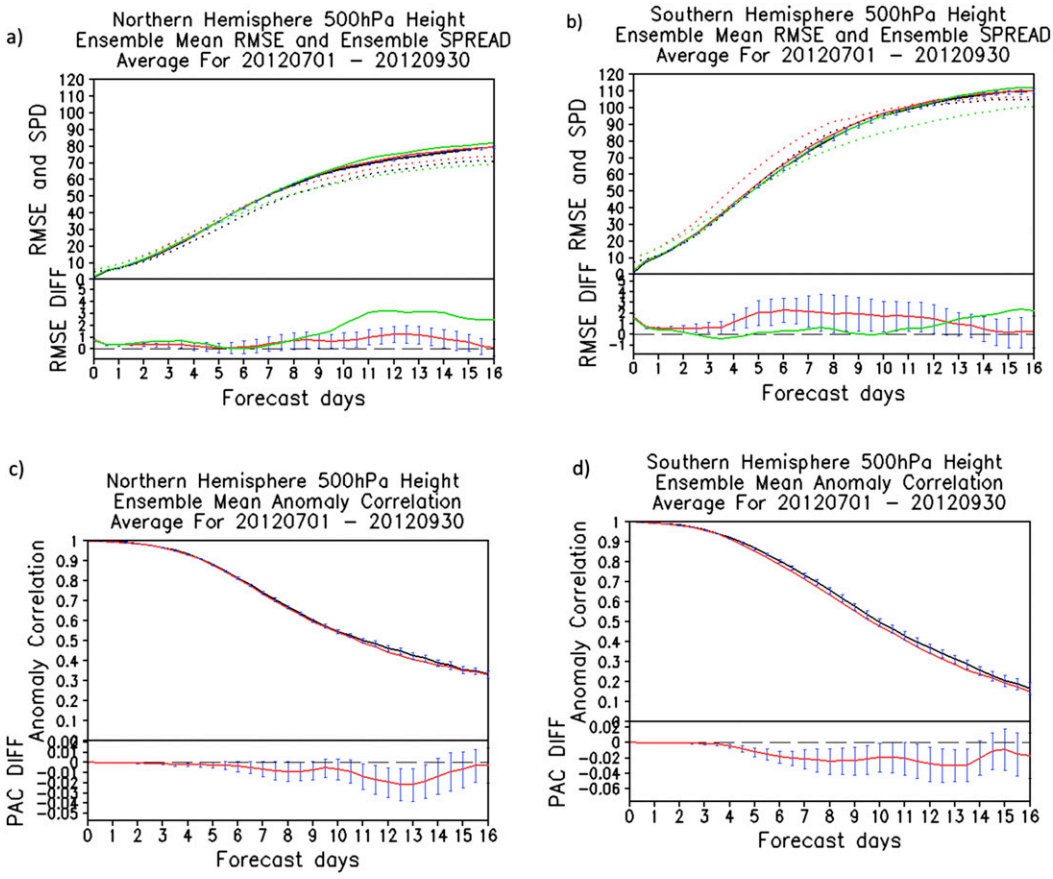

FIG. 10. Ensemble mean RMSE (solid) and ensemble standard deviation (dotted) for 500-hPa geopotential height over the (a) NH and (b) SH. The verification scores for ETR_STTP (black), EnKF_STTP (red), and EnKF_ONLY (green) are averaged during the period from 1 Jul to 30 Sep 2012. The bottom portion of each panel shows the difference in RMSE (EnKF_STTP minus ETR_STTP and EnKF_ONLY minus ETR_STTP). The blue bars display the 95% confidence intervals with the bootstrap significant test. (c),(d) As in (a),(b), but for the 500-hPa geopotential height ensemble mean predicted pattern anomaly correlations for ETR_STTP and EnKF_STTP.

does cause EnKF_ONLY_R to behave more similarly to ETR_ONLY, and the differences in the track error between the ETR_ONLY and EnKF_ONLY_R (black and green in Fig. 13a) are not statistically significant.

For the 2012 period with STTP, the TC relocation scheme is used for both the EnKF and ETR configurations (EnKF_STTP_R and ETR_STTP, respectively). There is no statistically significant difference between EnKF_STTP_R and ETR_STTP in the ensemble mean TC track forecast error. The spread–error relationship for the TC tracks is improved in the 2012 experiments relative to those in 2011 (Fig. 13). This improvement may be attributable to seasonal variation. It is unclear whether the STTP plays a role in increasing the TC track spread. As previously mentioned, STTP does not increase the spread for the other verified forecast parameters in the tropics. For EnKF_ONLY_R and EnKF_STTP_R, which use the TC relocation scheme, the track spread growth rate is larger than in all of the ETR experiments, regardless of the use of STTP or the season evaluated (2011 or 2012).

7. Summary and discussion

Two schemes available to generate initial ensemble perturbations for the GEFS in the NCEP operational environment are compared. An EnKF was implemented as part of the NCEP data assimilation system in 2012 to accommodate the hybrid ensemble variational development for the GFS. The EnKF short-range ensemble forecasts provide flow-dependent ensemble covariances for the data assimilation system. One question that came about from the implementation of the EnKF is whether the ensemble perturbations generated by the EnKF can replace the ensemble transform technique (ET) perturbations for the operational GEFS or not. A comprehensive comparison of experiments with the ETR-based and EnKF-based schemes was performed

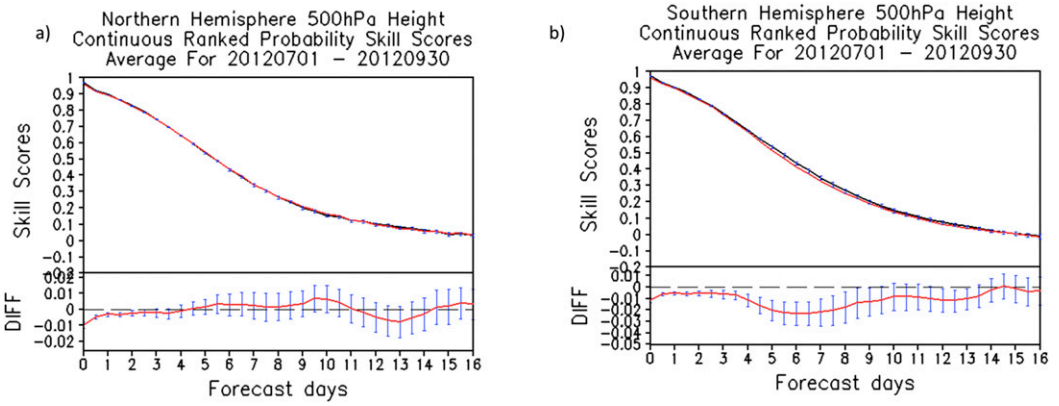

FIG. 11. As in Fig. 6, but for the experiments with STTP (EnKF_STTP and ETR_STTP) averaged from 1 Jul to 30 Sep 2012.

to help investigate this question and inform future implementations.

Currently at NCEP, the STTP scheme is used in the GEFS to account for a portion of the model uncertainties. In addition to initial perturbation methodology experiments, a comparison of EnKF and ETR initial perturbations is performed with the use of the STTP scheme.

The comparison shows that the performance of the EnKF perturbations is generally comparable with that of the ETR perturbations when the ensemble forecasts are run without STTP. The EnKF_ONLY experiment outperforms ETR_ONLY in the NH for the trial period. In the SH, the ensemble mean forecast in EnKF_ONLY is slightly degraded, but EnKF_ONLY performs better

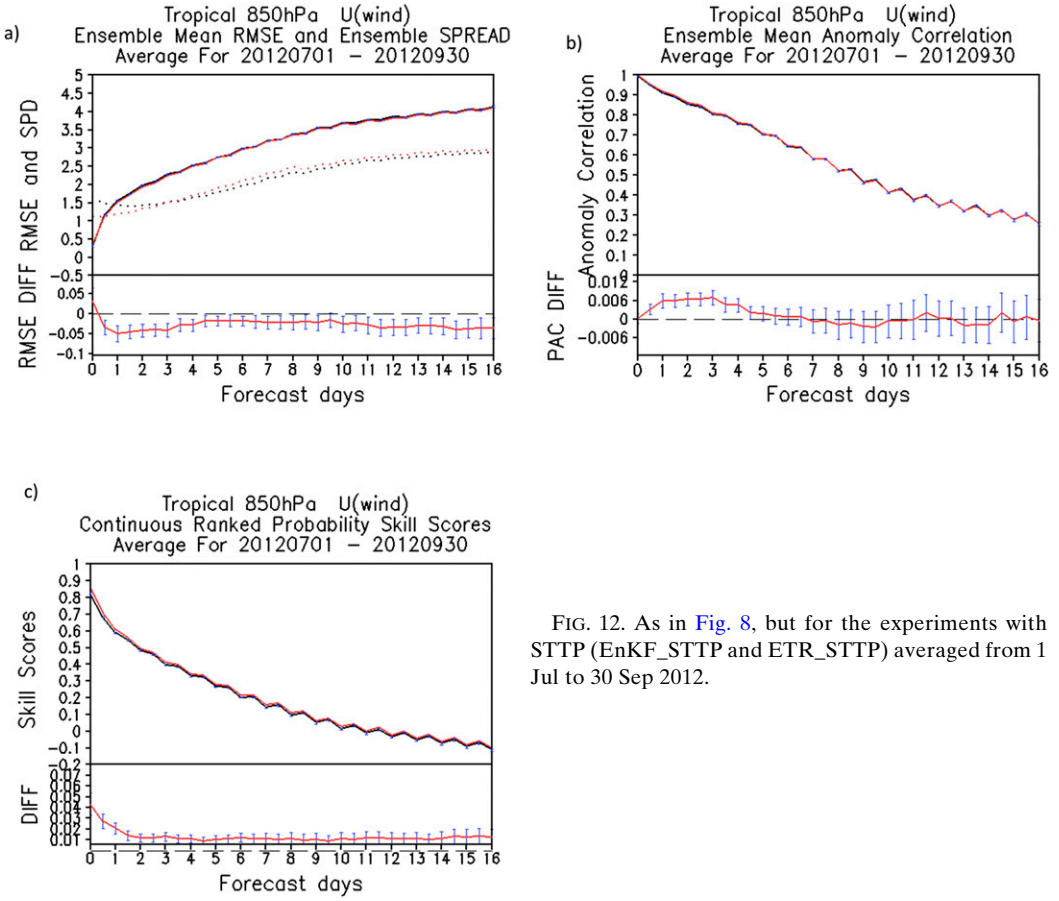

FIG. 12. As in Fig. 8, but for the experiments with STTP (EnKF_STTP and ETR_STTP) averaged from 1 Jul to 30 Sep 2012.

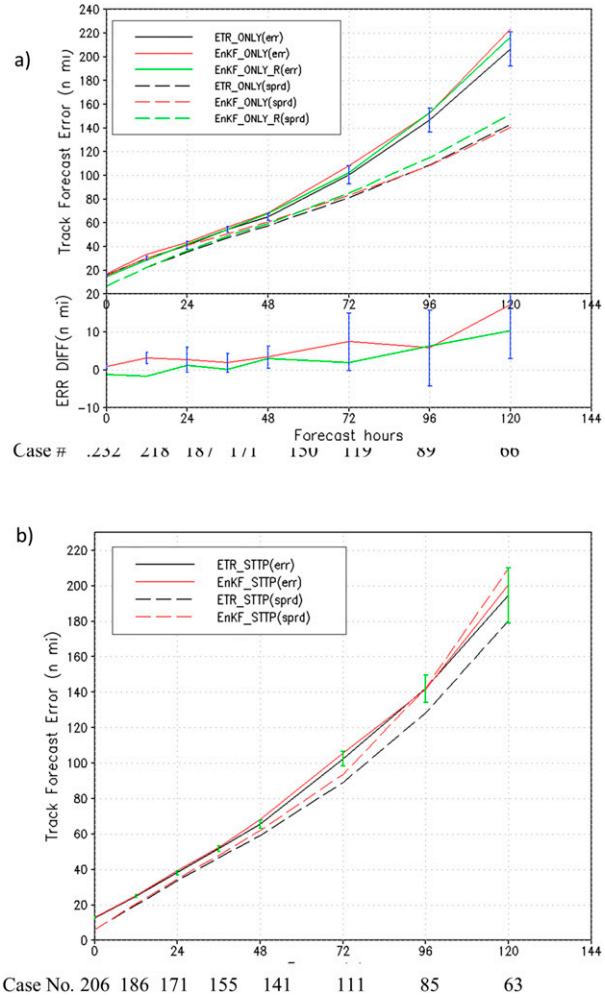

FIG. 13. (a) Ensemble mean tropical cyclone track error (solid) and spread (dashed line) for ETR_ONLY (black), EnKF_ONLY (red), and EnKF_ONLY_R (green) for 2011. The bottom portion in (a) shows the difference (EnKF_ONLY and EnKF_ONLY_R minus ETR_ONLY; green and red lines, respectively). (b) As in (a), but for 2012. A TC relocation scheme is applied for both ETR and EnKF for this season. The green bars display the 95% confidence intervals with the bootstrap significant test.

in terms of CRPSS and ensemble-mean RMSE for the first week. The ensemble spread is found to be slightly overdispersive in EnKF_ONLY for the first 3 days, which is probably related to some degradation of the ensemble forecast. Although the amplitude of the initial perturbations in EnKF_ONLY is much larger than in ETR_ONLY, especially in the SH, the spread growth remains slower than the growth of the mean forecast error. For longer lead times, the ensemble is generally found to be underdispersive for both EnKF_ONLY and ETR_ONLY when model uncertainty and error goes unaccounted for.

The EnKF experiment is found to be superior in terms of CRPSS to ETR in the NH when STTP is included, similar to the group of experiments without STTP. However, the application of STTP increases the ensemble spread considerably in the SH, which results in significant degradation for short-range forecasts. The STTP scheme seems to be exacerbating an already overdispersive ensemble for these regions and lead times when EnKF initial perturbations are used.

It is worth noting that the results presented in this manuscript are based on the performance of the ensemble forecasts during two boreal summer seasons. Similar experiments with STTP were carried out for a Northern Hemisphere winter season, but were excluded from this manuscript for the sake of brevity. In general, the conclusions from that NH cold season experiment were similar to the findings for the summer season. The lone exception to this was for the SH, where the degradation in the EnKF_STTP was not found to be as pronounced.

Tropical cyclone track forecast errors were verified for two boreal summer seasons. Tropical cyclone relocation, in which the TC vortices as represented in the model are relocated to the same position (observed location), was found to be beneficial for improving ensemble mean track forecasts. With the application of the tropical cyclone relocation scheme, tropical cyclone track forecasts are found to be similar for experiments that use ETR and EnKF initial perturbations. It was noted that the spread grew faster and to larger amplitudes in the EnKF-based experiments. The use of mechanical relocation tends to help avoid unreasonable TC structure when TC perturbations calculated from the ensemble forecast fields with the large spread in TC location are added to the analysis. However, the initial TC position spread is close to zero with the relocation scheme, which cannot represent the uncertainty of the TC initial location. Further study is under way to improve the TC relocation scheme for the ensemble forecast through direct assimilation of tropical storm information (position, intensity, and size) in EnKF.

The STTP scheme that is used in the operational GEFS is well tuned for use with the ETR initial perturbations, which undoubtedly has an impact on the spread–error relationship for the ETR. Including the STTP scheme without additional tuning proved to be detrimental for EnKF forecasts at short lead times. Consistent with the nature of these two schemes, the amplitude of the initial perturbations is generally smaller in ETR than EnKF. The inclusion of STTP in the EnKF experiments results in an overdispersive ensemble for short-term lead times. However, the forecast ensemble that uses EnKF perturbations cannot alone generate sufficient ensemble spread in the medium-range

forecasts without model perturbations. The use of STTP or other stochastic schemes to account for model uncertainty remains necessary in order to generate sufficient spread at the medium range.

Overall, the performance of the EnKF initial perturbations with and without STTP is comparable to the ETR initial perturbations except for the degradation in the SH due to overdispersion. To fix this issue, a possible solution is to rescale the EnKF initial perturbation amplitude when used for initialization of the GEFS. A set of experiments was performed and results are presented in a separate paper (Ma et al. 2014) to address this option. It was found that the EnKF initial perturbations with rescaling outperformed the unmodified counterparts. Another potential solution is to reduce the EnKF perturbation amplitudes that are used in the hybrid data assimilation cycles while mitigating any negative impact on the data assimilation. In a recent GDAS implementation (January 2015), the EnKF perturbation methodology was changed significantly, using three stochastic physics schemes, namely SKEB, SPPT, and SHUM, to account for system uncertainty instead of additive inflation on the posterior ensemble. An assessment of the performance of the GEFS by using these newly formulated EnKF perturbations is a topic for future work and will be discussed in a follow-up study.

For the present study, the EnKF background (prior) ensemble (i.e., 6-h forecasts) is used as the initial perturbations in the EnKF experiments instead of the EnKF posterior ensemble. The former represents the uncertainties of the background instead of the analysis after assimilation. Better performance would be expected if the posterior ensemble were used, but this is not practical in the current NCEP operational environment because of timing constraints. A potential alternative to the current configuration is to move the EnKF to the GFS cycle or to run a separate EnKF as part of the early cycle with a reduced set of observations. However, a corresponding cost–benefit analysis will be required to assess the practicality of this approach. This will likely be pursued in an attempt to further consolidate the efforts for generating appropriate initial perturbations for use both in the data assimilation *and* the medium-range ensemble forecasting.

Acknowledgments. This work was completed as part of EMC/NCEP/NWS/NOAA regular work duties. We thank Jun Du and Zhan Zhang for their careful internal reviews and Christopher Melhauser and Walter Kolczynski for aiding in an editorial review of the manuscript. We are grateful for three anonymous reviewers' careful reading and their many insightful comments and suggestions.

REFERENCES

Anderson, J. L., 1997: The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Wea. Rev.*, **125**, 2969–2983, doi:10.1175/1520-0493(1997)125<2969:TIODCO>2.0.CO;2.

——, 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

Bowler, N., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus*, **58A**, 538–548, doi:10.1111/j.1600-0870.2006.00197.x.

Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456, doi:10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2.

——, 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, doi:10.1002/qj.49712556006.

——, P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, doi:10.1175/MWR2905.1.

Descamps, L., and O. Talagrand, 2007: On some aspects of the definition of initial conditions for ensemble prediction. *Mon. Wea. Rev.*, **135**, 3260–3272, doi:10.1175/MWR3452.1.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739–759, doi:10.1111/j.2153-3490.1969.tb00483.x.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, doi:10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

Hou, D., Z. Toth, and Y. Zhu, 2006: A stochastic parameterization scheme within NCEP Global Ensemble Forecast System. Preprints, *18th Conf. on Probability and Statistics*, Atlanta, GA, Amer. Meteor. Soc., 4.5. [Available online at https://ams.confex.com/ams/Annual2006/techprogram/paper_101401.htm.]

——, ——, ——, and W. Yang, 2008: Impact of a stochastic perturbation scheme on NCEP Global Ensemble Forecast System. Preprints, *19th Conf. on Probability and Statistics*, New Orleans, LA, Amer. Meteor. Soc., 1.1. [Available online at https://ams.confex.com/ams/pdfpapers/134165.pdf.]

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

——, L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242, doi:10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2.

——, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, doi:10.1175/MWR-2864.1.

——, ——, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2126–2143, doi:10.1175/2008MWR2737.1.

Kleist, D. T., 2011: Assimilation of tropical cyclone advisory minimum sea level pressure in the NCEP Global Data Assimilation System. *Wea. Forecasting*, **26**, 1085–1091, doi:10.1175/WAF-D-11-00045.1.

——, and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433–451, doi:10.1175/MWR-D-13-00351.1.

——, D. F. Parrish, J. C. Derber, R. Treadon, W. S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, doi:10.1175/2009WAF2222201.1.

Kurihara, Y., M. A. Bender, and R. J. Ross, 1993: An initialization scheme of hurricane models by vortex specification. *Mon. Wea. Rev.*, **121**, 2030–2045, doi:10.1175/1520-0493(1993)121<2030:AISOHM>2.0.CO;2.

——, ——, R. E. Tuleya, and R. J. Ross, 1995: Improvements in the GFDL hurricane prediction system. *Mon. Wea. Rev.*, **123**, 2791–2801, doi:10.1175/1520-0493(1995)123<2791:IITGHP>2.0.CO;2.

Liu, Q., T. Marchok, H.-L. Pan, M. Bender, and S. J. Lord, 2000: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. NWS Tech. Procedures Bull. 472, 7 pp. [Available online at http://www.nws.noaa.gov/om/tpb/472.htm.]

——, S. Lord, N. Surgi, Y. Zhu, R. Wobus, Z. Toth, and T. Marchok, 2006: Hurricane relocation in global ensemble forecast system. Preprints, *27th Conf. on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc., P5.13. [Available online at https://ams.confex.com/ams/pdfpapers/108503.pdf.]

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

Lorenz, E. N., 1969: The predictability of a flow which possess many scales of motion. *Tellus*, **21**, 289–307, doi:10.1111/j.2153-3490.1969.tb00444.x.

——, 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513, doi:10.1111/j.2153-3490.1982.tb01839.x.

Ma, J., Y. Zhu, D. Hou, X. Zhou, and M. Peña, 2014: Ensemble transform with 3D rescaling initialization method. *Mon. Wea. Rev.*, **142**, 4053–4073, doi:10.1175/MWR-D-13-00367.1.

Meng, Z., and F. Zhang, 2008: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Wea. Rev.*, **136**, 522–540, doi:10.1175/2007MWR2106.1.

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

Palmer, T. N., 1997: On parametrizing scales that are only somewhat smaller than the smallest resolved scales, with application to convection and orography. *Proc. Workshop on New Insights and Approaches to Convective Parametrization*, Reading, United Kingdom, ECMWF, 328–337.

——, 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304, doi:10.1002/qj.49712757202.

——, G. J. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung, and M. Leutbecher, 2005: Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.*, **33**, 163–193, doi:10.1146/annurev.earth.33.092203.122552.

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

Szunyogh, I., E. Kalnay, and Z. Toth, 1997: A comparison of Lyapunov and optimal vectors in a low-resolution GCM. *Tellus*, **49A**, 200–227, doi:10.1034/j.1600-0870.1997.00004.x.

Tompkins, A. M., and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *J. Geophys. Res.*, **113**, D18101, doi:10.1029/2007JD009284.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

——, and ——, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, doi:10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

——, O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley and Sons, 137–163.

——, ——, and Y. Zhu, 2006: The attributes of forecast system. *Predictability of Weather and Climate*, T. N. Palmer and R. Hagedorn, Eds., Cambridge University Press, 584–595.

Trevisan, A., and R. Legnani, 1995: Transient error growth and local predictability: A study in the Lorenz system. *Tellus*, **47A**, 103–117, doi:10.1034/j.1600-0870.1995.00006.x.

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

——, D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, doi:10.1175/MWR-D-12-00141.1.

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, doi:10.1111/j.1600-0870.2006.00159.x.

——, ——, ——, and ——, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global ensemble forecast systems. *Tellus*, **60A**, 62–79, doi:10.1111/j.1600-0870.2007.00273.x.

Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292–3302, doi:10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2.

——, and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

——, and ——, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, doi:10.1175/MWR-D-11-00276.1.

——, ——, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482, doi:10.1175/2007MWR2018.1.

Wu, W., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, doi:10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability. *Adv. Atmos. Sci.*, **22**, 781–788, doi:10.1007/BF02918678.

——, and Z. Toth, 2008: Ensemble based probabilistic forecast verification. Preprints, *19th Conf. on Probability and Statistics*, New Orleans, LA, Amer. Meteor. Soc., 2.2. [Available online at https://ams.confex.com/ams/88Annual/techprogram/paper_131645.htm.]

——, G. Iyengar, Z. Toth, M. S. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP Global Ensemble Forecasting System. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., J79–J82.