RESEARCH ARTICLE

Key Points:
- The trade-offs between ensemble size and ensemble resolution for an ensemble-variational data assimilation system are evaluated
- Increasing either ensemble size or ensemble resolution can improve the forecast performance
- Increasing ensemble resolution is better than increasing ensemble size, particularly when considering errors at smaller scales

Correspondence to:
L. Lei,
lililei@nju.edu.cn

Evaluating the trade-offs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system

Lili Lei[1] [ID] and Jeffrey S. Whitaker[2]

[1]Key Laboratory of Mesoscale Severe Weather, Ministry of Education, School of Atmospheric Sciences, Nanjing University, Nanjing, China, [2]NOAA/Earth System Research Laboratory/Physical Sciences Division, Boulder, Colorado, USA

Abstract The current NCEP operational four-dimensional ensemble-variational data assimilation system uses a control forecast at T1534 resolution coupled with an 80 member ensemble at T574 resolution. Given an increase in computing resources, and assuming the control forecast resolution is fixed, would it be better to increase the ensemble size and keep the ensemble resolution the same, or increase the ensemble resolution and keep the ensemble size the same? To answer this question, experiments are conducted at reduced resolutions. Two sets of experiments are conducted which both use approximately four times more computational resources than the control experiment that uses a control forecast at T670 and an 80 member ensemble at T254. One increases the ensemble size to 320 but keeps the ensemble resolution at T254; and the other increases the ensemble resolution to T670 but retains an 80 ensemble size. When ensemble size increases to 320, turning off the static component of the background-error covariance does not degrade performance. When the data assimilation parameters are tuned for optimal performance, increasing either ensemble size or ensemble resolution can improve the forecast performance. Increasing ensemble resolution is slightly, but significantly better than increasing ensemble size for these experiments, particularly when considering errors at smaller scales. Much of the benefit of increasing ensemble resolution comes about by eliminating the need for a deterministic control forecast and running all of the background forecasts at the same resolution. In this "single-resolution" mode, the control forecast is replaced by an ensemble average, which reduces small-scale errors significantly.

1. Introduction

For an ensemble-based data assimilation and forecasting system, there is always a tension between increasing ensemble size and resolution—both can lead to improved performance with increased computational cost. Due to the multiscale nature of the dynamics, increasing model resolution can resolve new phenomena and account for nonlinear interactions of these phenomena with large-scale motions [e.g., *Navarra et al.*, 2010]. *Pellerin et al.* [2003] found that increasing ensemble resolution has a positive impact on the forecast skill for the ensemble forecasts at the Canadian Meteorological Centre (CMC). By increasing the horizontal resolution of the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble prediction system (EPS) from a spectral triangular truncation T95 to T799, *Buizza* [2010] found a strong impact on forecast skill at short lead times, a weaker impact in the medium range, and an undetectable impact at long leads.

Buizza and Palmer [1998] found that increasing ensemble size from 10 to 30 in the ECMWF EPS with spectral triangular truncation T63 and 19 vertical levels (T63L19) had a significant positive impact on ensemble mean forecast skill. Doubling the ensemble size of the UK Met Office Global and Regional EPS implemented at the Korean Meteorological Administration (KMA) with resolution T320L50 resulted in a slight positive impact on forecast skill [*Kay et al.*, 2013], especially for longer forecast lead times.

The trade-off between ensemble size and ensemble resolution has been investigated at several different operational centers. Using the ECMWF EPS, *Mullen and Buizza* [2002] showed that given approximately the same computational cost, and using a probabilistic measure of precipitation forecast, increasing ensemble size was more valuable than increasing ensemble resolution. *Ma et al.* [2012] used the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS) and concluded that increasing ensemble resolution is more (less) beneficial than increasing ensemble size for a short (long) forecast ranges. Based on the performance of the U.S. Navy Global Atmospheric EPS, *Reynolds et al.* [2011]

demonstrated that higher ensemble resolution with smaller ensemble size produced significantly smaller tropical cyclone track errors and slightly smaller tropical wind errors than larger ensemble size with lower ensemble resolution. These studies indicate that the trade-off between ensemble size and resolution may depend on the forecast skill metric, as well as other factors such as the range of ensemble sizes and resolutions used.

Ensemble Kalman filter (EnKF) data assimilation systems use an ensemble forecast to provide a flow-dependent background-error covariance (**B**) that determines the analysis increment given observations. Either increasing the ensemble size or ensemble resolution can result in a more accurate estimate of **B**. *Hamrud et al.* [2015a] showed that increasing ensemble size and increasing ensemble resolution both had a positive impact on forecasts initialized from an EnKF system. Similar results are obtained by *Houtekamer et al.* [2014], however they found that increasing resolution had a larger impact than increasing ensemble size, but this came at a greater computational cost.

The current NCEP operational hybrid four-dimensional ensemble-variational (4DEnVar) data assimilation system uses a dual-resolution configuration with a single control forecast at T1534 and an 80 member ensemble at T574 resolution. The operational hybrid 4DEnVAR system [*Wang and Lei*, 2014; *Kleist and Ide*, 2015b] uses a combination of time-invariant static **B** and flow-dependent **B** estimated from the ensemble that is updated using an EnKF [*Whitaker and Hamill*, 2002]. The EnKF analysis ensemble mean is then replaced by the 4DEnVar analysis. Experience has shown that the most important configuration parameters for the operational data assimilation system include the weight given to the static **B** component, the covariance localization length scales, the ensemble size, and the spatial resolution of the ensemble. There have been no studies that we are aware of that examine relative benefits of increasing ensemble size versus resolution in the context of a dual-resolution hybrid 4DEnVar system of the kind now operational at NCEP. Hence, this is our primary focus.

This paper is organized as follows. Section 2 describes NCEP hybrid 4DEnVAR configuration and experimental design. Section 3 presents an analysis of the relative benefits of increasing either the ensemble size or ensemble resolution, including the sensitivities to the weight of static **B** and covariance localization length scales. The findings of this study are summarized in section 4.

2. Experimental Design

The current NCEP operational system uses a dual-resolution configuration with a single control forecast at T1534 and an 80 member ensemble at T574 resolution, with 64 vertical levels. Our control experiment (T254T670) is the same as the operational configuration but at a reduced resolution (T670 for the control forecast and T254 for the ensemble). Given approximately four times more computations than the control experiment T254T670, two sets of experiments are designed. One set of experiments (T254T670Ens320) is the same as the control experiment T254T670, but increasing ensemble size from 80 to 320. The second set of experiments (T670T670) is the same as the control experiment T254T670, but increasing the ensemble resolution from T254 to T670. Note that since the forecast model uses an unconditionally stable semi-Lagrangian time integration scheme, the time step remains the same when the ensemble resolution is increased.

In the dual-resolution experiments, the ensemble priors are first recentered around the control forecast at the beginning of each data assimilation cycle. The control and ensemble analyses are then computed separately at the beginning, middle, and end of the data assimilation window, using all observations in the window. The hybrid 4DEnVar algorithm [*Kleist and Ide*, 2015b] produces the control analysis, and the EnKF [*Whitaker and Hamill*, 2002; *NCAR Developmental Testbed Center*, 2015] produces the ensemble analyses. The EnKF analyses are then recentered around the hybrid 4DEnVAR analysis. For both the control and ensembles, the four-dimensional increment analysis update (4DIAU) [*Lei and Whitaker*, 2016] is performed to incorporate the 4-D analysis increments during the model forecast. There is no digital filter applied, and experiments by default are performed without the tangent-linear normal-mode constraint (TLNMC) [*Kleist et al.*, 2009; *Kleist and Ide*, 2015b]. The control and ensemble analyses are at last advanced to the end of next data assimilation window. The single-resolution experiments (T670T670) use the same procedure as the dual-resolution experiments, except that there is no separate control forecast—instead the ensemble

mean prior is used as background for the hybrid 4DEnVar analysis. Thus there is no need for recentering of the forecast ensembles, although the EnKF analyses are still recentered around the hybrid 4DEnVar analysis.

All experiments assimilate all of the observations used operationally in the NCEP global data assimilation system (GDAS) every 6 h, including conventional in situ observations, and remotely sensed satellite radiances, cloud-motion vectors, and global positioning system radio-occultation measurements. The observation error variances are the same as that used in the NCEP GDAS. To compute the observation prior ensemble $\mathbf{Hx}^b$ (needed by the EnKF) where $\mathbf{H}$ is the observation forward operator and $\mathbf{x}^b$ is the model ensemble background or prior, the "observer" portion of the Grid-point Statistical Interpolation system (GSI) [*Wu et al.*, 2002; *Kleist et al.*, 2009] is run for each ensemble member and the ensemble mean separately.

Hybrid 4DEnVAR is an extension of 3DEnVAR [*Wang et al.*, 2013] that utilizes 4-D ensemble perturbations but a time-invariant static $\mathbf{B}$ throughout the assimilation window. The amplitude of the time-invariant static $\mathbf{B}$ in the 4DEnVAR is controlled by a constant parameter β_0. The amplitude of the flow-dependent, ensemble estimated part of $\mathbf{B}$ is then $1 - \beta_0$. Covariance localization is used to remove spurious long-range covariance arising from sampling errors. The localization is applied separately in the horizontal and vertical, and the same horizontal and vertical localization length scales are used in both 4DEnVAR and the EnKF. The localization table utilized in NCEP operations has level-dependent horizontal localization scale (hl) and a constant vertical localization scale (vl). The level-dependent horizontal localization scale follows Figure 3 in *Kleist and Ide* [2015a], with a value hl_1 below 300 hPa, hl_2 between 56 and 14 hPa, and hl_3 between 5 hPa and the top of the model, with values varying linearly in the vertical in the transition zones. The localization scale hl_2 and hl_3 are fixed to 1000 and 1300 km as in *Kleist and Ide* [2015a], while a range of values for hl_1 is examined in our experiments. Multiplicative covariance inflation that relaxes posterior ensemble spread back to the prior ensemble spread (relaxation-to-prior spread) [*Whitaker and Hamill* 2012] is used with the relaxation coefficient set to 0.85. Stochastic parameterizations [*Palmer et al.*, 2009] are used to represent model uncertainty within the ensemble forecast step, and no additive inflation is applied.

All experiments are run from 00 UTC 1 April 2014 to 00 UTC 24 April 2014. The first 4 days of assimilation is discarded to avoid transient effects, and the remaining data are used for verification. We use the root-mean-square (RMS) observation increments, which are the RMS differences between the observations priors ($\mathbf{Hx}^b$) and all in situ observations, to evaluate the experiments in observation space. To evaluate the experiments in model space, forecasts from each experiment are postprocessed to 37 pressure levels between 1000 and 100 hPa on a $1° \times 1°$ grid, and then are compared to ECMWF $1° \times 1°$ gridded analyses. The RMS differences between the gridded forecasts and analyses are computed. The RMS differences among the experiments are converged in approximately 5 days. A spherical harmonics transform is applied to the latitude-longitude temperature difference, and the resulting 2-D wavenumber decomposition is summed over spherical harmonics with the same total spherical wavenumber to produce the 1-D spectrum of temperature error. To compute the error kinetic energy spectrum, a similar procedure is applied for the vector wind differences $\left[\left(\mathbf{u}^{EC} - \mathbf{u}^f\right)^2 + \left(\mathbf{v}^{EC} - \mathbf{v}^f\right)^2\right]/2$, where superscript EC denotes ECMWF analyses and f denotes the 6 h prior.

3. Results

3.1. Sensitivities to Data Assimilation Parameters

The RMS observation increments are used to tune the data assimilation parameters including β_0, hl_1, and vl for the three sets of experiments, and the best values of the parameters are shown in Table 1. When the ensemble size increases from 80 to 320, there is little sensitivity to changes in the horizontal and vertical localization scales, thus experiment T254T670Ens320 utilizes the same values of hl_1 and vl as experiment T254T670. Although a broader localization scale is expected with increasing ensemble size, the sensitivity to the localization scale is small when the localization scale is approaching its optimal value [*Lei et al.*, 2015], thus only very slight sensitivity is found for increasing ensemble size from 80 to 320. When

Table 1. The Optimal Values of the Weight of Static Background-Error Covariance (β_0), Horizontal Localization Scale Below 300 hPa (hl_1) and Vertical Localization Scale (vl) for the Three Sets of Experiment[a]

Experiment Name	β_0	hl_1 (km)	vl (Scale Height)
T254T670	0.125	350	0.5
T254T670Ens320	0	350	0.5
T670T670	0.125	263	0.375

[a]The horizontal and vertical localization scales are e-folding scales.
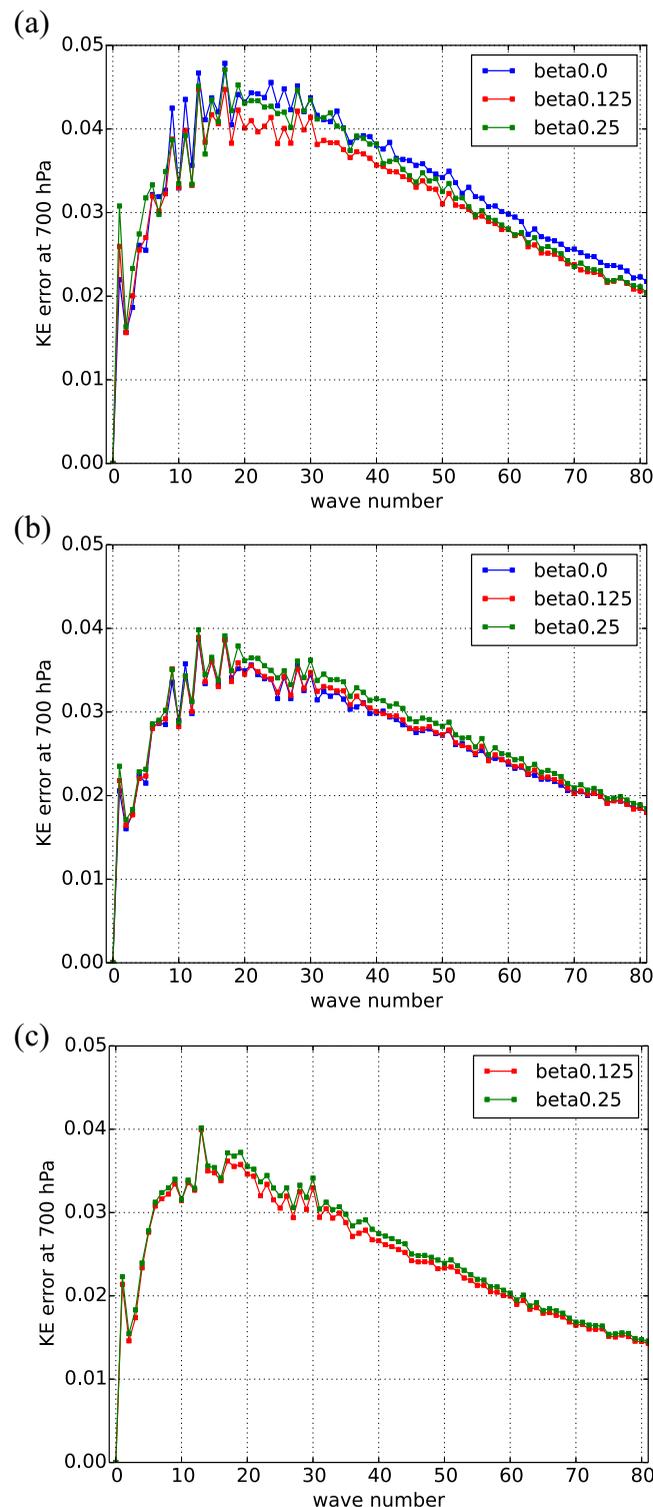
Figure 1. The power spectra of error kinetic energy at 700 hPa for experiments (a) T254T670, (b) T254670Ens320, and (c) T670T670 with different static background-error covariance amplitudes. The error kinetic energy is computed using the vector wind differences between 6 h forecasts and ECMWF analyses.

the ensemble resolution increases from T254 to T670 (experiment T670T670), slightly smaller horizontal and vertical localization scales are optimal.

For the control experiment T254T670, the inclusion of a small contribution of static **B** ($\beta_0 = 0.125$) is beneficial, which is consistent with the results from *Kleist and Ide* [2015b]. To diagnose the role of the static **B** component, the error kinetic energy is computed using differences between the control forecasts and ECMWF analyses for different values of β_0, and the power spectra are analyzed at pressure levels of 1000, 700, 500, and 200 hPa. Generally consistent results are obtained for all pressure levels, so only the error kinetic energy power spectrum at 700 hPa is shown in Figure 1. Comparing to no static **B** ($\beta_0 = 0$), the inclusion of a small contribution of static **B** ($\beta_0 = 0.125$) decreases the kinetic energy error at nearly all scales except the largest scales with wavenumbers smaller than 9. This indicates that the flow-dependent **B** from the ensemble is able to well represent the larger-scale error statistics, but the ensemble estimated **B** suffers from sampling errors at small scales even with covariance localization. Thus from this point of view, the static **B** component can play a role similar to localization and improve the representation of error statistics at small scales. When the weight of static **B** is further increased ($\beta_0 = 0.25$), the error kinetic energy starts to increase at medium scales. Therefore, the optimal amplitude of the static **B** component in hybrid 4DEnVAR results from a balance between sampling error in the estimation of flow-dependent portion of the background-error covariance and the errors in the static covariance estimate at small scales.

When the ensemble size increases from 80 to 320, turning off the static **B** ($\beta_0 = 0$) has no negative impact on observation increments. Figure 1b shows that with 320 members, including a small amount of static **B**

($\beta_0 = 0.125$) does not reduce the error kinetic energy at small scales as it did with 80 members. When β_0 is increased to 0.25, the error kinetic energy error increases at small scales. Therefore, the flow-dependent,
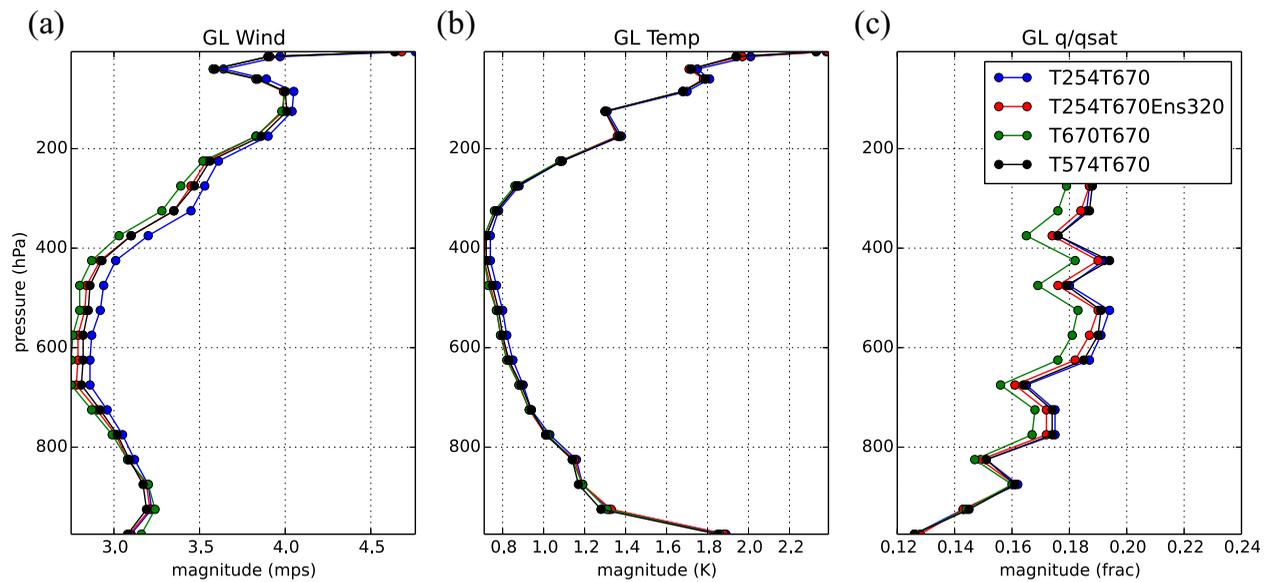
Figure 2. Globally and temporally averaged RMS observation increments for (a) vector wind, (b) temperature, and (c) normalized specific humidity (specific humidity divided by saturation specific humidity) for all in situ observations.

ensemble estimated **B** using 320 members with horizontal and vertical localization scales can represent the error statistics at all scales as well or better than the static estimate.

When the ensemble resolution increases from T254 to T670, including a small contribution of static **B** ($\beta_0 = 0.125$) produces very slightly smaller RMS observation increments than with no static **B** contribution ($\beta_0 = 0$), and produces slightly smaller error kinetic energy values for wavenumbers between 10 and 17 (not shown). Further increasing the weight of static **B** ($\beta_0 = 0.25$) leads to increased error kinetic energy for
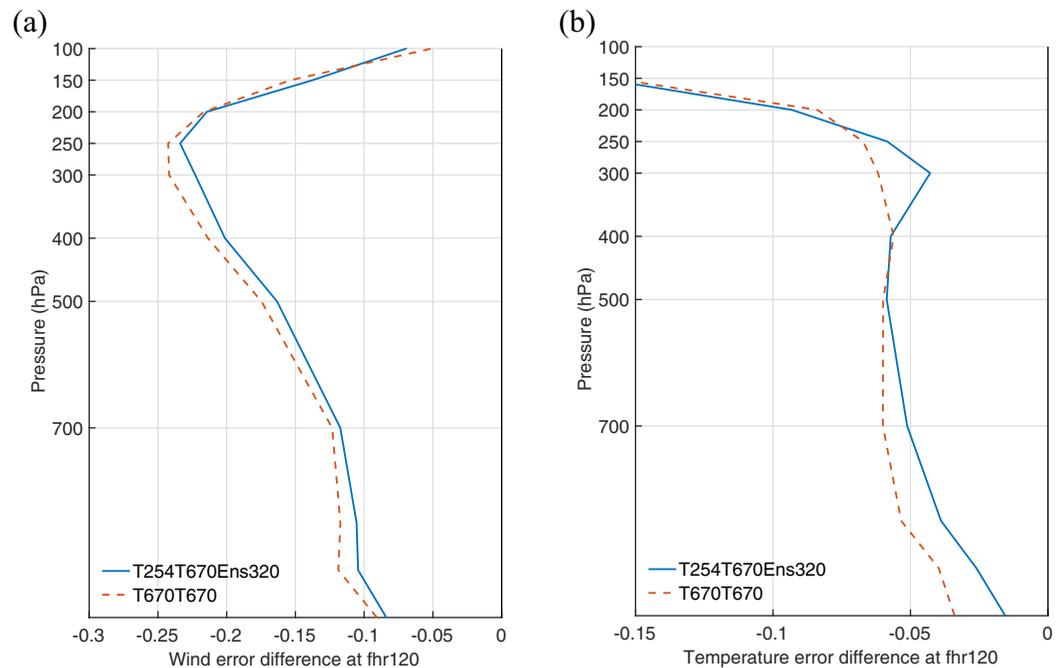
Figure 3. Globally and temporally averaged forecast error differences between the control experiment T254T670 and the other two experiments T254T670Ens320 and T670T670 for (a) vector wind and (b) temperature at 5 day forecast lead time. The error is computed by verifying the 120 h forecasts from each experiment against the ECMWF analysis. Negative (positive) values mean the control experiment has larger (smaller) forecasts errors than the compared experiment.
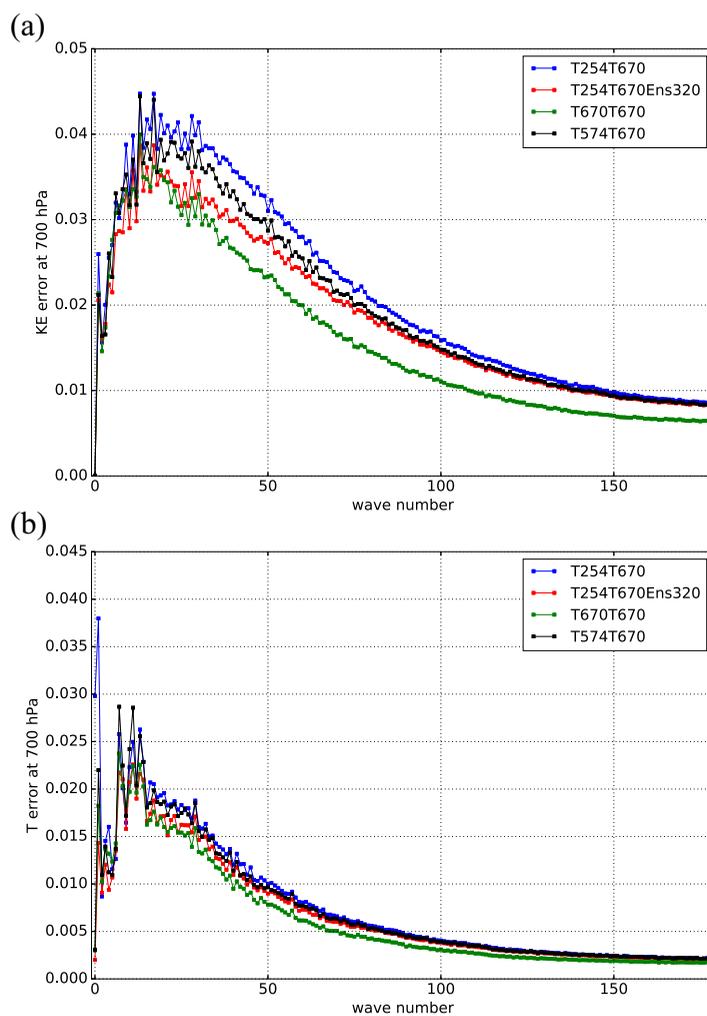
Figure 4. The power spectra of (a) error kinetic energy and (b) temperature forecast error at 700 hPa for the three experiments with tuned data assimilation parameters given in Table 1. The error kinetic energy and temperature forecast errors are computed using differences between 6 h forecasts and ECMWF analyses.

wavenumbers larger than 10. With increased ensemble resolution, the error statistics at smaller scales can be better resolved, and thus the impact of including a small contribution of static **B** becomes smaller, but including a small contribution of static **B** ($\beta_0 = 0.125$) is still beneficial to the experiment with resolution T670.

3.2. Evaluating the Trade-Off Between Ensemble Size and Resolution

The globally and temporally averaged RMS observation increment profiles for the three sets of experiments with the tuned data assimilation parameters given in Table 1 are shown in Figure 2. Compared to the control experiment T254T670, experiments T254T670Ens320 and T670T670 have reduced wind and normalized specific humidity (specific humidity divided by saturation specific humidity) errors in the whole column except near the surface, and slightly reduced temperature errors between 825 and 325 hPa. Experiment T670T670 produces smaller wind and normalized specific humidity errors than T254T670Ens320, and similar temperature errors.

Similar results are obtained when validating the forecasts in model space relative to ECMWF analyses (not shown).

To examine whether the differences between the three experiments are representative of longer-lead forecasts, a single 5 day forecast is launched from each control analysis during the verification period. The globally and temporally averaged RMS error profiles (relative to ECMWF analyses) are computed for each of the three experiments. The forecast error differences at 5 day lead time are shown in Figure 3. Positive (negative) values indicate that the control experiment T254T670 has larger (smaller) RMS forecast error than the compared experiment. At 5 day lead time, experiments T254T670Ens320 and T670T670 produce smaller forecast errors of temperature and wind than T254T670, and T670T670 has smaller forecast errors than T254T670Ens320. Thus the observation increments differences between the experiments shown in Figure 2 is consistent with differences in forecast errors computed from longer-lead forecasts, except near the surface that is because results in Figure 2 is verified against the in situ observations while results in Figure 3 is verified relative to the ECMWF-gridded analyses.

Figure 4 shows the power spectra of error kinetic energy and temperature errors at 700 hPa computed from 6 h forecasts for all the three experiments. The error kinetic energy is reduced at nearly all scales by increasing either ensemble size or ensemble resolution, although the impact of increasing ensemble
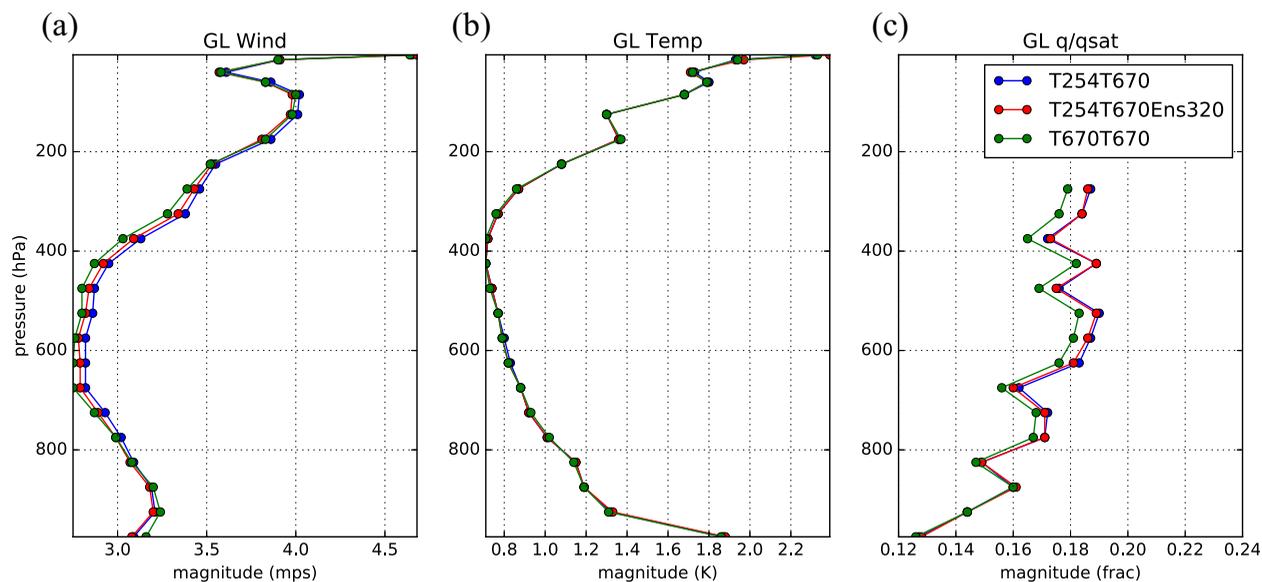
Figure 5. Same as Figure 2, except that the TLNMC is applied in the three experiments.

resolution is larger for wavenumbers greater than about 20. Similar, though somewhat smaller, sensitivities are seen in power spectra of temperature errors.

3.3. Impact of the Tangent-Linear Normal Mode Constraint

As a dynamic constraint, the TLNMC [*Kleist et al.*, 2009; *Kleist and Ide*, 2015b] applies a balance operator on the analysis increment, in order to mitigate imbalances that can be introduced by the analysis. The impact of the TLNMC on ensemble size and ensemble resolution sensitivities is examined here. Given the tuned data assimilation parameters shown in Table 1, the three sets of experiments are repeated with the TLNMC turned on.

The globally and temporally averaged RMS observation increment profiles for the three experiments with TLNMC are shown in Figure 5, which are similar to Figure 2 but the differences between the control experiment T254T670 and the other two experiments become smaller. Similarly, 5 day forecast error differences are similar with and without the TLNMC activated (figure not shown), although the magnitude of differences decreases when the TLNMC is on. Figure 6 shows the power spectra of temperature errors and error kinetic energy at 700 hPa for the three experiments with TLNMC, for comparison with Figure 4. The general conclusion still holds: increasing either ensemble size or ensemble resolution in hybrid 4DEnVar improves forecast skill, but increasing ensemble resolution has slightly more impact for a comparable increase in computational cost.

The comparison of Figures 2 and 5 and Figures 4 and 6 indicates that the application of the TLNMC significantly reduces the error of the control experiment T254T670 at nearly all levels and scales, but only has very slight impact on the experiments T254T670Ens320 and T670T670. We would expect that the impact of TLNMC be reduced with increasing ensemble size (T254T670Ens320), since the imbalance introduced by localization decreases. To examine why the TLNMC has limited impact on experiment T670T670, two single-resolution experiments at T254, with and without the TLNMC, are performed, which confirm that the TLNMC has very little impact on single-resolution configurations of hybrid 4DEnVar that do not include a high-resolution control forecast. A second dual-resolution T254T670 experiment was then performed in which the recentering of the EnKF ensemble around the control analysis was turned off, thus there is no feedback from the control to the ensemble. This experiment performs better than experiment T254T670 without the TLNMC, and is only slightly worse than experiment T254T670 with the TLNMC (not shown). This indicates that much of the impact of the TLNMC in the dual-resolution configuration comes from mitigating the imbalances introduced in the EnKF ensemble priors by the recentering procedure. This is also consistent with the fact that the impact of the TLNMC is relatively small in the single-resolution configuration.

(a)

(b)
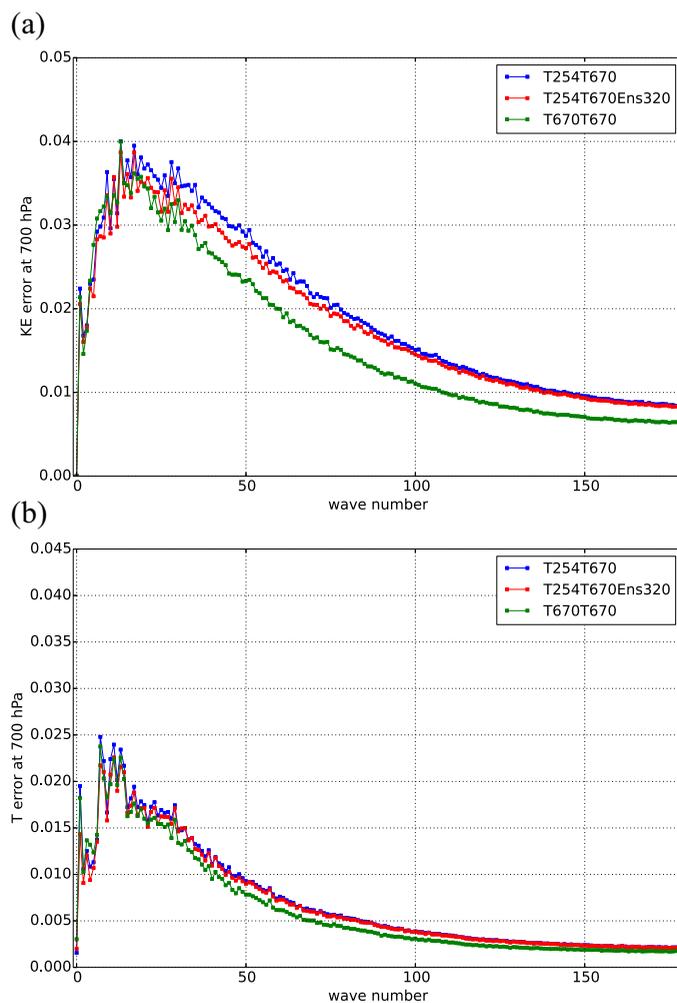
Figure 6. Same as Figure 4, except that the TLNMC is applied in the three experiments.

3.4. Impact of Ensemble Averaging

Unlike experiments T254T670 and T254T670Ens320, the single-resolution experiment T670T670 does not utilize a deterministic control forecast but instead utilizes the ensemble mean prior as background for the hybrid 4DEnVAR analysis. Thus, experiment T670T670 may also benefit from ensemble averaging. To examine the impact of ensemble averaging, two dual-resolution experiments T574T670, with and without the TLNMC, are performed.

The globally and temporally averaged RMS observation increment profiles for the experiment T574T670 without TLNMC are shown in Figure 2. Experiment T574T670 produces errors comparable to T254T670Ens320 but larger than T670T670, especially for wind and specific humidity. Consistent with the RMS observation increments, experiment T574T670 has similar power spectra of temperature errors and error kinetic energy to T254T670Ens320 (Figure 4), except that T574T670 has larger errors than T254T560Ens320 for wavenumbers between 25 and 50. Similar results, though somewhat smaller differences, are obtained when TLNMC is applied (figure not shown). Therefore, we conclude that ensemble averaging in the single-resolution experiment T670T670 reduces errors significantly, particularly at smaller scales.

4. Conclusions

The current NCEP operational system uses a T1534 resolution control forecast and a T574 80 member ensemble. Given an increase in computer resources and assuming the control resolution remains fixed, the trade-off between increasing either ensemble size or ensemble resolution is investigated in this study. The control experiment (T254T670) is the same as the current operational system except at a reduced resolution. With approximately four times more computation cost than the control experiment, two experiments are conducted, one (T254T670Ens320) increases the ensemble size but keeps the ensemble resolution the same and the other (T670T670) increases the ensemble resolution but keeps the ensemble size the same.

There are slight sensitivities for varying the horizontal and vertical localization scales with increasing either ensemble size or ensemble resolution. While experiment T670T670 still benefits from an inclusion of a small contribution of static **B**, turning off the static **B** has no impact for the experiment with 320 members. The error power spectra show that the static **B** can play a role similar to localization and improve the representation of error statistics at small scales for an 80 member ensemble. The impact of static **B** decreases with increasing ensemble size, since the error statistics at all scales can be more accurately represented. The

impact of static **B** is also reduced with increasing ensemble resolution due to the improved representation of error statistics at small scales.

With tuned data assimilation parameters, results show that forecast skill can be improved by increasing either ensemble size or ensemble resolution, and the advantages of increasing either ensemble size or ensemble resolution persist out to a 5 day forecast lead time. For a comparable computational cost, we find increasing ensemble resolution is superior to increasing ensemble size, and the extra improvement comes mainly from reducing errors at smaller scales through ensemble averaging in a single-resolution configuration. These results hold when the tangent-linear normal model constraint is applied, although the differences among the experiments become smaller, primarily because the TLNMC improves the dual-resolution control experiment by mitigating the impacts of imbalances associated with small ensemble size and ensemble recentering.

The question remains whether these results can be extrapolated to the higher resolutions now run in NCEP operations (T574 for the ensemble and T1534 for the control forecast). Preliminary experiments using the NCEP operational resolutions but without tuning the optimal localization scales and weight of static **B** show consistent results with increasing ensemble size (R. Mahajan, personal communications, 2016), and work is ongoing to fully answer this question.

References

Buizza, R. (2010), Horizontal resolution impact on short- and long-range forecast error. *Q. J. R. Meteorol. Soc.*, *136*, 1020–1035.

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

Hamrud, M., M. Bonavita, and L. Isaksen (2015), EnKF and hybrid gain ensemble data assimilation. Part I: EnKF implementation, *Mon. Weather Rev.*, *143*, 4847–4864.

Houtekamer, P. L., X. Deng, H. L. Mitchell, S.-J. Baek, and N. Gagnon (2014), Higher resolution in an operational ensemble Kalman filter, *Mon. Weather Rev.*, *142*, 1143–1162.

Kay, J. K., H. M. Kim, Y.-Y. Park, and J. Son (2013), Effect of doubling the ensemble size on the performance of ensemble prediction in the warm season using MOGREPS implemented at the KMA, *Adv. Atmos. Sci.*, *30*, 1287–1302.

Kleist, D. T., and K. Ide (2015a), An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-Hybrid results, *Mon. Weather Rev.*, *143*, 433–451.

Kleist, D. T., and K. Ide (2015b), An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DEn-VAR and hybrid variants, *Mon. Weather Rev.*, *143*, 452–470.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, R. M. Errico, and R. Yang (2009), Improving incremental balance in the GSI 3DVar analysis system, *Mon. Weather Rev.*, *137*, 1046–1060.

Lei, L., and J. S. Whitaker (2016), A four-dimensional incremental analysis update for the ensemble Kalman filter, *Mon. Weather Rev.*, *144*, 2605–2621.

Lei, L., J. L. Anderson and G. S. Romine (2015), Empirical localization functions for ensemble Kalman filter data assimilation in regions with and without precipitation, *Mon. Weather Rev.*, *143*, 3664–3679.

Ma, J. H., Y. J. Zhu, R. Wobus, and P. X. Wang (2012), An effective configuration of ensemble size and horizontal resolution for the NCEP GEFS, *Adv. Atmos. Sci.*, *29*, 782–994.

Mullen, S. L., and R. Buizza (2002), The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF ensemble prediction system, *Weather Forecast.*, *17*, 173–191.

Navarra, A., J. L. Kinter III, and J. Tribbia (2010), Crucial experiments in climate science, *Bull. Am. Meteorol. Soc.*, *91*, 343–352.

NCAR Developmental Testbed Center (2015), NOAA ensemble Kalman Filter (beta release v1.0 compatible with GSI community release v3.3) User's Guide, 55 pp., NCAR, National Center for Atmospheric Research, Boulder, Colo. [Available at http://www.dtcenter.org/com-GSI/users/docs/enkf_users_guide/EnKF_UserGuide_v1.0Beta.pdf.]

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer (2009), Stochastic parameterization and model uncertainty. ECMWF Tech. Memo. 598, 44 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K. [Available at http://www.ecmwf.int/sites/default/files/elibrary/2009/11577-stochastic-parametrization-and-model-uncertainty.pdf.]

Pellerin, G., L. Lefaivre, P. Houtekamer, and C. Girard (2003), Increasing the horizontal resolution of ensemble forecasts at CMC, *Nonl. Processes Geophys.*, *10*, 463–468.

Reynolds, C. A., J. G. McLay, J. S. Goerss, E. A. Serra, D. Hodyss, and C. R. Sampson (2011), Impact of resolution and design on the US Navy global ensemble performance in the tropics, *Mon. Weather Rev.*, *139*, 2145–2155.

Wang, X., and T. Lei (2014), GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for the NCEP Global Forecast System, *Mon. Weather Rev.*, *142*, 3303–3325.

Wang, X., 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. Weather Rev.*, *141*, 4098–4117.

Whitaker, J. S., and T. M. Hamill (2002), Ensemble data assimilation without perturbed observations, *Mon. Weather Rev.*, *130*, 1913–1924.

Whitaker, J. S., and T. M. Hamill (2012), Evaluating methods to account for system errors in ensemble data assimilation, *Mon. Weather Rev.*, *140*, 3078–3089.

Wu, W. S., R. J. Purser, and D. F. Parrish (2002), Three-dimensional variational analysis with spatially inhomogeneous covariances, *Mon. Weather Rev.*, *130*, 2905–2916.