Tellus **000**, 000–000 (0000) Printed 5 June 2013 (Tellus LaTeX style file v2.2)

The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment

By Nikki C. Privé[1,2*] and Ronald M. Errico[1,2]

[1] *Morgan State University, Baltimore, Maryland, USA;* [2] *National Aeronautics and Space Administration, Global Modeling and Assimilation Office, Greenbelt, Maryland, USA;*

(Manuscript submitted date)

ABSTRACT

A series of experiments that explore the roles of model and initial condition error in numerical weather prediction are performed using an observing system simulation experiment (OSSE) framework developed at the National Aeronautics and Space Administration Global Modeling and Assimilation Office (NASA/GMAO). The use of an OSSE allows the analysis and forecast errors to be explicitly calculated, and different hypothetical observing networks can be tested with ease. In these experiments, both a full global OSSE framework and an 'identical twin' OSSE setup are utilized to compare the behavior of the data assimilation system and evolution of forecast skill with and without model error. The initial condition error is manipulated by varying the distribution and quality of the observing network and the magnitude of observation errors.

The results show that model error has a strong impact on both the quality of the analysis field and the evolution of forecast skill, including both systematic and unsystematic model error components. With a realistic observing network, the analysis state retains a significant quantity of error due to systematic model error. If errors of the analysis state are minimized, model error acts to rapidly degrade forecast skill during the first 24-48 hours of forward integration. In the presence of model error, the impact of observation errors on forecast skill is small, but in the absence of model error, observation errors cause a substantial degradation of the skill of medium range forecasts.

1 Introduction

Forecast skill in numerical weather prediction is affected by two types of error: initial condition error and model error. The magnitudes of model error and initial condition error have changed over the decades as forecast models, data assimilation, and the global observing network have become more sophisticated (Simmons and Hollingsworth (2002), Compo et al. (2011)). Quantifying the relative impacts of these errors is of interest to determine where resources should best be expended in order to effect the greatest possible improvements in forecast skill.

Initial condition error is influenced by many factors, including the quality of the observations and the observational network and the handling of observation and background information by the data assimilation system (DAS). Some sources of initial condition error can be at least partially mitigated by techniques such as bias correction to remove persistent observation error and proper weighting of the background and observation error variances in the DAS. However, data voids and formulation deficiencies in the data assimilation algorithms are more difficult to rectify, and in some circumstances the very methods used to attempt to improve the analysis quality may instead result in a degradation. For example, bias correction may attribute persistent differences between observations and the background to observation biases when model bias is actually the root cause.

Early theoretical exploration of the roles of model and initial condition error involved simple representations of error growth. Leith (1978) assumed exponential growth of initial condition error and linear growth of model error for short term forecasts, while Lorenz (1982) and Dalcher and Kalnay (1987) included an additional quadratic growth term to account for saturation of initial condition error for longer forecasts. Simmons and Hollingsworth (2002) found reasonably good agreement between the theoretical and assumed error growth of operational forecasts at the European Centre for Medium-Range Weather Forecasts (ECMWF) when systematic forecast errors were also taken into account. However, these comparisons could only be made after the first day of forecast integration because the true error during the early forecast period could not be satisfactorily estimated.

While the growth of initial condition error can be estimated through a variety of means, the growth of model error is more difficult to determine. Comparison of ensemble forecasts

* Corresponding author.
e-mail: Nikki.Prive@nasa.gov

(Buizza, 2010) or 'perfect model' tests in which the differences between forecasts initialized on sequential days are examined (Lorenz, 1982) can be used to estimate the growth of initial condition errors. Statistics of model error are then estimated as residuals. Techniques such as restricted statistical correction (Schubert and Chang, 1996), and model drift (Orrell et al., 2001) have been used to attempt to quantify model error, but these methods have limitations.

Observing System Simulation Experiments (OSSEs) are pure simulation exercises commonly used to examine the potential implementation of future observing networks in numerical weather forecasting. In an OSSE, the real world is replaced with a long model simulation that captures the phenomena of interest. This simulation is referred to as the Nature Run (NR). Synthetic observations are generated by spatio-temporal interpolation of the NR fields for both the current and future networks of observing systems. These synthetic observations are then ingested into the DAS. The OSSE should be rigorously tested and calibrated to ensure that the behavior of the system is sufficiently similar to real-world behavior to give results pertinent to the latter.

In addition to evaluation of observing systems, an OSSE can be a powerful tool for investigating the behavior of data assimilation systems. Unlike the real world, in an OSSE, the 'true' state of the atmosphere is completely known. This allows the errors in model forecasts to be explicitly calculated, instead of the indirect methods required when working with real data. This is particularly advantageous during the analysis and early forecast periods, as the analysis and forecast errors are very difficult to quantify for real data at these times. An OSSE can also be used to determine how well the data assimilation process acts to improve the analysis state compared with the background.

A global OSSE has been developed at the National Aeronautics and Space Administration Global Modeling and Assimilation Office (NASA/GMAO; Errico et al. (2013), Privé et al. (2013b)) for use with investigations of DAS and forecast model performance. This OSSE includes both a well-calibrated configuration that emulates the real-world model performance, and in an 'identical twin' configuration, in which the NR and the forecast model are the same and there is no model error. The synthetic observation network may also be manipulated both in terms of the magnitude of the observation error and in terms of the frequency and location of observations. Thus, the GMAO OSSE may be used to investigate the roles of model and initial condition error in a more sophisticated framework than earlier idealized studies.

The intent of this work is to examine the relative effects of model error and initial condition error through a series of five experiments. In three cases, the OSSE is run in a configuration where model error is included. The initial condition error in these three cases is varied by manipulating both the observation error and the configuration of the observing network. In the two additional cases, the identical twin configuration is used to conduct perfect model tests with varying levels of initial condition error.

The configuration of the GMAO OSSE will be described in Section 2. Results of the experiments are examined in terms of analysis errors in Section 3, behavior of the DAS in Section 4, and forecast errors in Section 5. Discussion and conclusions are presented in Section 6.

2 Method

The GMAO OSSE framework includes code for the generation of synthetic observations for data types used in operational weather forecasting. Observation errors are added to the synthetic observations such that the variances of observation innovation and analysis increment in the OSSE are similar to those that occur when cycling the DAS with real observations. The forecast model and data assimilation system used for all cases are the Global Earth Observing System version 5.7.1 (GEOS-5, (Rienecker et al., 2008)) and the Gridpoint Statistical Interpolation (GSI, (Kleist et al., 2009)) data assimilation system, respectively.

Two Nature Runs were used for the experiments described herein. The baseline NR is a 13-month integration of the version c31r1 European Centre for Medium-Range Weather Forecasts operational forecast model, run at T511 horizontal resolution with 91 vertical sigma levels and 3-hourly output from 01 May 2005 to 31 May 2006. This integration was forced only with sea surface temperature and sea ice fields taken from 2005-6 archived datasets. No additional data was ingested into the NR.

A second NR was generated using a short, free run of the GEOS-5 forecast model in order to perform two experiments using an 'identical twin' setup with no model error. The initial state is taken as the operational analysis from 14 June 2011, and the model is integrated without observation ingestion until 11 August 2011.

Synthetic observations were generated from both NRs using archived data as a basis for the time and location of observations. For the ECMWF NR, the observational suite was

based on the real observations from June-August 2005, while for the identical twin NR, the observational suite was based on real observations from June-August 2011. The difference in the observing networks for the identical twin and ECMWF NR cases was not intentional, but merely a result of the generation of the identical twin cases at a considerably later time than the ECMWF NR cases, when the 2011 synthetic observations were newly available. The most significant differences between these two datasets are the inclusion of Quikscat, MSU, HIRS-2, and NOAA-15 for the ECMWF NR, and the inclusion of ASCAT, IASI, MHS, and the metop-a, NOAA-18, and NOAA-19 instruments in the identical twin cases. These differences are not expected to have a significant impact on the reported results of this study.

Observations were first generated by interpolating from the ECMWF NR and then testing ingestion into the GEOS-5/GSI. These initial tests were used to calibrate the added observation errors. For the identical twin dataset, a new dataset was then generated using interpolation of the GEOS-5 NR. The errors added to this new dataset were generated using the same statistics as calibrated for the former. Since no realistic analog of the identical twin experiments exist (ie., no perfectly realistic model exists), calibration of the identical twin cases against real data is not possible. Therefore, the added observation errors are not re-calibrated for the identical twin cases. No additional observation biases were added to the observations although there are small intrinsic biases in the radiance observations due to the handling of clouds and surface emissivity (Errico et al., 2013).

Five experiments were performed: three using the ECMWF NR and two using the identical twin GEOS-5 NR. An overview of these experiments is given in Table 1. For the ECMWF NR experiments, cycling began on 15 June 2005 and continued until 5 August 2005, with one forecast generated each day at 0000 UTC, for a total of 29 forecasts from 2 July to 30 July. In the first experiment, the ECMWF NR OSSE setup was employed, with calibrated synthetic observations and observation errors that mimic the operational data suite from 2005, denoted the Control case. In the second experiment (NE), the same synthetic observations as in the Control were ingested with no explicitly added synthetic errors. These two experiments were repeated in the identical twin framework, one case featuring synthetic observations with no added observation errors (Twin NE) and the other case having added observation errors with the same magnitudes and correlations as in the Control experiment (Twin Control).

In the third ECMWF NR case (DENSE), a global network of rawinsonde sounding

observations was generated, with one observation located at every other latitude, longitude, and vertical level of the NR grid at each cycling time, and with no added observation errors. No other data types were ingested. The rawinsonde observations were assumed to be taken instantaneously throughout the column from the surface to 1.5 hPa. The soundings were not extended to the top of the NR due to strongly incompatible differences in the dynamics of the upper atmosphere that result in numerical instability of the GEOS-5 forecasts when forced with observations from the ECMWF NR. The GSI-assumed error covariances for rawinsonde types were decreased by a factor of 10 in order to more strongly draw the background toward the observations in this experiment.

Some small but unspecified quantity of implicit representativeness error is present in the synthetic observations independent of the explicitly added observation errors. This arises because simulated radiance observations include cloud effects not accounted for by the assimilation system since the DAS instead attempts to remove cloud-contaminated radiance data through quality control. Small but still significant errors that are undetected by the quality control may thus remain. Also surface properties used to determine surface emissivity for observation simulations are not the same as used in the DAS. In the case of the ECMWF NR, simulated observations are determined by spatial interpolations on the NR grid that differs from the DAS grid. Thus, although the spatial interpolation techniques used for the simulations and DAS are the same, the results generally differ. For this reason, the implicit error in the Twin cases is smaller than that for the ECMWF NR cases, because identical grids are used for the Twin NR and DAS.

In the Control, NE, Twin Control, and Twin NE cases, the background and observation error covariances assumed by the DAS are not altered from the covariances used operationally in July 2011. For the Twin cases in particular, there is a significant mismatch between the actual and assumed background error covariances. A smaller mismatch between the actual and assumed background error covariances is expected for the ECMWF NR cases due to the change in the observational network between 2005 and 2011. Likewise, for the NE and Twin NE case, the actual observation error covariances are expected to be much smaller than the assumed covariances. Some of the ramifications of mismatched assumed and actual error covariances include the possible degradation of the analysis field in comparison to the background field, as discussed by Eyre and Hilton (2013) and Privé et al. (2013a).

The ECMWF NR and the GEOS-5 model use hybrid η vertical coordinates, although the ECMWF NR has 91 levels while the GEOS-5 model uses 72 levels. In the upper at-

mosphere (above 150 hPa in the GEOS-5 and above 80 hPa in the ECMWF), the η-levels follow pressure surfaces, while in the lower troposphere the η levels are dependent on the surface pressure in the fashion of σ-levels, with a blending in between (Untch et al. (1999), Rienecker et al. (2008)). Throughout this manuscript, the η levels will be referred to by the corresponding pressure that would occur if the surface pressure were 1000 hPa.

For the identical twin cases, verification of forecasts and analyses can be made directly on the native grid of the GEOS-5. However, for the ECMWF NR cases, the NR fields must be interpolated onto a compatible grid for comparison with the GEOS-5 output fields. For ease of validation, the ECMWF NR fields are interpolated onto the same grid as that used by the GEOS-5. Details of the interpolation method are given in Errico and Privé (2013).

3 Analysis Error

The analysis areal mean root-time mean-square error (RMSE) verified against the corresponding NR fields for July is shown in Figure 1 for temperature, humidity, and wind in the tropics and the extratropics of both hemispheres. The qualitative form of the analysis error in the extratropics is similar in both hemispheres, but the behavior in the tropics is somewhat different from that in the extratropics. The OSSE Control case has the greatest analysis error for all variables and all regions, as would be expected since that case has the most sources of analysis error (model error, observation error, and sub-optimal observational network). For the NE case where explicit observation errors are not added, there is a slight reduction in the analysis error compared with the Control, with greatest reduction seen for wind fields in the extratropics. This indicates that the observation errors have a relatively small contribution to the total error in comparison to other sources of error.

The smallest analysis RMSE were found in the DENSE case, except in the stratosphere and lowest levels of the troposphere where the TWIN NE case had the least analysis error. The DENSE case has less variation in analysis error with height compared with the Control case, the most striking example being the wind error in the tropics. The rawinsonde network in the DENSE case has consistent frequency and distribution of sampling throughout the troposphere and lower stratosphere, while the realistic observing network used in the other cases has very different distribution of observation sampling at different height levels. The analysis RMSE in the DENSE case is 50-60% smaller than in the Control in the extratropics, with 60-80% reduction in error in the tropics. The DENSE case estimates the limit of

improvement of the background state possible using the DAS and forecast model in question if the observing network were nearly ideal.

The Twin NE case has error close to that of the DENSE case for temperature and wind in the extratropics, but significantly greater error than the DENSE case in the tropics and globally for humidity. The larger errors seen in the tropics and for humidity are believed to be due to convection, which behaves mathematically nonlinearly and discontinuously and has a short timescale of error growth. A greater increase in analysis error is observed when observation errors are included in the Twin cases in comparison to the difference between the NE and Control cases. The analysis error increases by 20-25% in the extratropics and 10% in the tropics for temperature, and by 40% in the extratropics and 20% in the tropics for winds from the Twin NE to the Twin Control case.

The time mean analysis error field gives an indication of regions that experience a persistent source of error, which could stem from observation bias or systematic error of the forward model or the data assimilation system. In the OSSE, observation bias should be minimal as no explicit bias was added to the synthetic observations, so any time mean analysis error is likely to stem from model error or data assimilation processes. Time mean analysis error due to systematic model error would result from retention of model error in poorly observed areas, bias correction that incorrectly assumes observation bias in the presence of model error, or weighting of background error that does not account for the systematic error. Systematic errors in the data assimilation process may also result from improper balance assumptions.

Figure 2 shows the monthly mean analysis error for temperature and zonal wind at 500 hPa and 250 hPa respectively. The DENSE case has very little time mean analysis error, indicating that the extensive observational network successfully removes any systematic sources of error. The NE and Control cases both show significant regions of cold biased analysis temperature, especially in the deep tropics. In these two cases, systematic model error of the GEOS-5 in comparison to the ECMWF NR is suspected to account for much of this temperature bias. The zonal wind time mean analysis error features strong easterly biases in the eastern Pacific and Atlantic equatorial basins, and westerly bias in the northern Indian Ocean. The easterly biases are due to a known issue with the cross-correlation of background errors of wind and temperature (and therefore to radiance data) by this version of the GSI/GEOS-5. The westerly bias over the Indian Ocean may be due to model error in representation of the upper tropospheric Asian monsoon circulation.

The Twin cases also show a cold bias in the deep tropics, but of a much smaller magnitude than in the ECMWF NR cases. Since there is no model error in the Twin cases, the low tropical temperatures in the analysis field may be due to destabilization of the vertical column during data assimilation, possibly resulting in excessive convection and mid-tropospheric cooling during the initial forecast period as the forward model physics attempts to adjust the unbalanced initial state.

4 Analysis Increments

The analysis increment, or analysis minus background, is a measure of the work performed by the data assimilation system in modifying the background field to produce an analysis field $\mathbf{x}_a$. The analysis increment can be expressed as

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{K}\left[\mathbf{y}_o - H(\mathbf{x}_b)\right] \tag{1}$$

where the background state $\mathbf{x}_b$ is adjusted by the ingestion of observations $\mathbf{y}_o$ using the operation operator H and the Kalman gain $\mathbf{K}$. The time mean of the analysis increment indicates the amount of persistent modification of the background. The left column of Figure 3 shows the zonal mean time mean analysis increment for temperature for the five experimental cases, and Figure 4 illustrates the same fields for zonal wind.

The analysis increment gives a measure of work done by the observations, but does not indicate whether this work is beneficial, harmful, or neutral to the analysis quality. The difference between the absolute value of analysis error and absolute value of background error, denoted here as $(|A| - |B|)$, gives an indication of whether the analysis increment is performing useful work or not, with negative values indicating an improvement of the analysis compared to the background. The zonal mean time mean distribution of $(|A| - |B|)$ for temperature is shown on the right columns of Figures 3 and 4. For most regions, $(|A| - |B|)$ is negative, with the analysis having less error than the background.

The largest time mean positive analysis increments for temperature are found in the tropics in areas of deep convection. Large negative time mean $(|A| - |B|)$ is seen in the tropics, indicating that the time mean analysis increments are acting to remove errors from the background. These time mean analysis increments correspond to the regions of time mean analysis error seen in Figure 2. The Twin cases have smaller increments than the ECMWF NR cases (note the different contour intervals).

Figure 5 shows the zonal mean temporal variances of the analysis increment for the five experimental cases. While the time mean analysis increment primarily illustrates the removal of systematic background error by the observations, the variance of the analysis increment illustrates the role of non-systematic errors. Variance of the analysis increment is influenced both by the observation error directly (as in Eq. (1)) and through the growth of ingested observation errors from previous cycles. The Control case has larger analysis increment variance than the NE case as a result of these two factors.

While the Control case has larger time mean and variance of analysis increments than the NE case, the $|A| - |B|$ field shows that the useful work done by the observations in both cases is nearly the same. In a stable data assimilation system, the work done by ingestion of observations should be equal to the growth of errors between cycle times. This error growth is a function of the chaotic nature of the model dynamics and physics, the model error, and the initial analysis error; the first two are identical in the two experiments. As seen in Figure 1, there is also little difference in the analysis error between the NE and Control cases for temperature, so it is not surprising that the error growth rate in the two cases should be nearly the same.

The NE and Control cases feature regions of wind field quality degradation by the assimilation process on the equator in the middle and upper troposphere, due to improper balancing of radiance observations (a known issue with this version of GEOS-5/GSI). This degradation is not observed in the DENSE case, where only rawinsonde observations with paired temperature and wind observations are ingested. The Twin NE case also shows some degradation of the analyzed wind field at the equator, although this is predominantly in the lower troposphere.

There is little difference between the Twin NE and Twin Control for either time mean analysis increment or $|A| - |B|$ for temperature, but a significant difference in $|A| - |B|$ for zonal wind. While the Twin NE case shows improvement of the background state due to the DAS, the Twin Control case shows degradation of the background state in the mid and lower troposphere due to the presence of observation errors. The Twin Control case has larger variances of analysis increment than the Twin NE case, although the variances of both Twin cases are an order of magnitude smaller than for the ECMWF NR cases. The increase in error variance relative to the Twin NE case is much greater than the relative change for the ECMWF NR cases. In the Twin cases, the observation errors and their growth are

large compared to other sources of error, while in the ECMWF NR cases model error is a significant additional source of error.

The DENSE case has the largest time mean analysis increments and $|A| - |B|$ of all experiments, due to the quantity and distribution of observations. The low analysis errors in the DENSE case are maintained by persistent ingestion of observations that work to counteract the growth of model and initial condition errors. The observations have no added errors and are so dense that the simplified statistically defined background and observation error covariances that determine the spread of information in the DAS do not degrade the accurate information provided by the observations. Also, the use of reduced observation error covariances by the GSI in the DENSE case causes the analysis to draw more strongly to the observations.

The $|A| - |B|$ field for temperature in the DENSE case shows horizontal striations that are not present in the analysis increment field. These striations are due to differences in the vertical interpolation of the rawinsonde observations in comparison to the NR verification field interpolation, illustrating an aspect of implicit representativeness error in the synthetic observations. The larger $|A| - |B|$ in the lower troposphere in the DENSE case compared to the NE and Control cases results because the observations remove systematic model error from the analysis in the DENSE case that is retained in the NE and Control cases.

The greatest variances of analysis increment in the DENSE case occur in different regions compared to the other cases. For the DENSE case, the variance of the analysis increment indicates the region with the fastest growing errors - in this case, near the intertropical convergence zone (ITCZ) and near the surface. In the Control and NE cases, there are not always sufficient observations in these regions to incur a large analysis increment. Instead, errors in these poorly-observed regions may be retained by the DAS in the Control and NE cases, with greater analysis increments in other regions that have additional observations.

There is less difference in magnitude of the monthly mean background RMSEs between the Control and NE cases than the analysis RMSEs. This implies that during the initial forward integration period, damping of the initial condition errors introduced from the synthetic observation errors is dominant over growth of these errors. It is anticipated that a large fraction of the initial condition errors will be rapidly damped during forward integration, while only a portion project onto growing modes (Errico et al., 2001). Spatially uncorrelated errors in particular are preferentially damped by filtering processes of the forecast model.

5 Forecast Errors

Forecast error can be calculated explicitly using the NR as verification. The RMS forecast errors at 24 hours and 120 hours are shown for the month of July in Figures 6 and 7, for temperature and zonal wind respectively. Although the DENSE case has lowest error at the analysis time, by 24 hours the error in the DENSE case has increased significantly and generally matches or exceeds the error in both Twin cases. At 120 hours there is a significant reduction in RMSE for the DENSE case compared to the Control and NE cases, but the Twin cases have the smallest forecast RMSE. The wind and temperature fields show greater forecast RMSE in the Southern hemisphere extratropics compared to the Northern hemisphere extratropics, including for the DENSE case in which there is no hemispheric difference in the observing network. The greater errors in the Southern hemisphere may be due to the seasonal differences or topographic differences between the Northern and Southern hemispheres.

In the tropics, physical processes play a large role in forecast error. The wind and temperature RMSEs show relatively little spread between the five experimental cases at 120 hours although there is large spread at the analysis time. The lack of spread of error at 120 hours suggests that the forecast errors of wind in the tropics are not dominated by differences in model climatologies but by more intermittent processes such as convection. The temperature RMSEs in the tropics shows more spread between the Twin cases and ECMWF NR cases in the lower troposphere, indicating that model error plays a larger role at those levels.

The Twin cases demonstrate a marked impact of observation error on the RMSE forecast error that persists from the analysis time to the 120 hour forecast. This is in contrast to the Control and NE case pair, where the initial difference in the two cases at the analysis time diminishes so that there is nearly identical RMSE at 120 hours in the tropics and Northern hemisphere extratropics (the lines in Figure 6 are nearly completely overlain), and only minimal difference in the Southern hemisphere extratropics.

5.1 Error growth rates

The forecast error variances are shown in Figure 8 for 250 hPa wind and 500 hPa temperature as a function of time, with semi-log scaling of the variance to allow easy comparison of growth rates. If exponential growth of error variance is assumed (see (2) in the Appendix), the slope of the temporal curve of the logarithm of variance will give the growth rate. All five

cases have error variance curves that are roughly parallel after the first 48 hours of forecast; however, the error variance growth during the first 24 hours differs significantly between cases.

The fastest initial growth of error variance is observed for the DENSE case, particularly in the tropics. The second fastest initial error growth is seen in the NE case, but the DENSE case growth rates are considerably higher than the NE growth rates. Maps of differences between the 12-hour forecast and analysis error variances (Figure 9) illustrate the spatial distribution of error growth. In the DENSE case, error variances grow rapidly over areas of deep convection in the tropics and summer hemisphere as well as the winter hemisphere storm track. A test was performed using the observational network from the DENSE case in the Twin OSSE framework for comparison, and rapid error growth was not observed during the first two days of the forward integration. These test results in the Twin setup imply that the primary cause of the rapid initial error growth in the DENSE case is model error.

In the Control and NE cases, large regions degraded by model error remain in the analysis state. Compared with the DENSE case, the pattern of error variance growth during the initial forecast period seen in Figure 9 is more spatially uniform and not focused in convective regions. Instead, regions of fast error variance growth include the eastern Pacific basin and the South American and African continents.

By day two of the forecast, the initial condition errors that project onto damped modes have been greatly reduced and similar dynamics is governing the dominant error growth. The rapid initial error growth in the DENSE case slows dramatically after 24-36 hours of forward integration. The similarity in growth rates during this period implies that additional model error is not contributing significantly to error variances during this time.

After day four, the growth rates for the Control and NE cases begin to slow in the extratropics, while there is little change in the DENSE or Twin cases. After a period of exponential error growth, the error begins to saturate at longer forecast times (Lorenz, 1982). The slight decrease in error variance growth rate in the Control and NE cases may be the beginning of the error saturation regime. Because the Twin and DENSE cases have error variances that lag behind the Control and NE cases, the Twin and DENSE cases are further from error saturation at day five and do not show a reduction in error growth rate.

The improvement in forecast skill is equivalent to a reduction of forecast lead time of approximately 12 hours for the Twin NE case compared to the Twin Control case. This improvement is maintained through the 120 hour forecast period. Likewise, the DENSE

case shows an improvement of approximately 24 hours in the extratropics, and 2-3 days in the tropics compared to the Control case, which is also maintained through the entire forecast period. There is very little improvement in the NE case compared to the Control case after the first 24 hours of forecast period.

Doubling times for the RMS forecast errors of temperature and zonal wind are calculated using the 48-96 hour forecast period to estimate the error growth rate. These doubling times are listed in Table 2. The doubling times were calculated using the simple relation in equation (2) in the Appendix. This relation neglects model error and error saturation terms, assuming a simple exponential growth of RMSE. For the forecast period from 48 to 96 hours, this assumption of exponential growth gives a relatively good fit to the actual error growth.

The doubling times are very similar for the two Twin cases and the DENSE case, with slightly longer doubling times for the Control and NE cases for wind. Doubling times are considerably longer in the tropics than in the extratropics. The doubling times calculated here are longer than the doubling times of 500 hPa geopotential height found by Simmons and Hollingsworth (2002), but this is due to the use of temperature and wind on η levels rather than geopotential height: the latter is a vertically integrated metric that effectively measures more of the barotropic component of the error. The barotropic error modes are known to grow more rapidly than baroclinic error components (Errico, 2000).

5.2 Observation Error Impacts

If two sources of error are independent, then the total error variance of the combined errors is simply the sum of the variances of both types of error. Because the observation errors that are deliberately applied to the synthetic observations are not correlated with the background or model errors, the error variance of the analysis state due to the ingestion of observation errors may be calculated by subtracting the total analysis error variance of the case with no applied observation error from the analysis error variance of the corresponding case that includes added observation errors. This result is also valid for the impact of observation errors on the forecast error variance if errors introduced into the forecast by independent observation errors also remain uncorrelated during the forward integration process; this will generally only hold as long as the error dynamics are linear.

Figure 10 shows the ratio of the error variance of the Control to the error variance of the NE cases for both pairs of experimental cases for the two NRs. In the absence of model error,

if the growth rate of the error variance stemming from observation errors is the same as the growth rate of the error variance arising from other sources of error, this ratio should remain constant during the forecast integration. The Twin cases, having no model error, only show a nearly constant ratio of error variances for the temperature field in the Southern Hemisphere and for both zonal wind and temperature in the tropics. For other variables and regions, the ratios of error variances for the Twind cases approach one as the forecast time increases.

It is expected that only a fraction of the errors present at the analysis time will project onto growing modes, while the rest will project onto damped modes. If the fraction of observation errors that project onto growing modes is small, then the ratio of error variances will approach one during the initial forward integration. Damping is expected to occur rapidly over the first day or two of the forecast, reflected in the initial decrease in the ratio of variances noted for zonal wind in the extratropics. However, the ratio continues to decline through the entire five-day forecast period. This indicates that the timescale for the damping of initial errors is at least several days, and/or that the error growth behaves nonlinearly.

When model error is present, as in the Control and NE cases, it is not expected that the ratio of error variances in the two cases will remain constant during forward integration due to the growth of model error. Error variances due to model error will have a very different growth rate with multiple timescales compared to the growth rate of errors stemming from initial condition error. Because model error is defined to be zero at the analysis time (although the background error and thus analysis error are likely to be correlated with systematic model error), an increase in model error variance with time will result in the ratio of variances of the Control and NE cases tending towards unity.

6 Discussion

When a numerical weather prediction model with data assimilation is cycling in a stable regime, where the errors of the analysis field are not persistently increasing or decreasing, there must be a balance between the growth of errors between cycle times and the 'work' done by ingestion of observations to improve the analysis compared to the corresponding background state (Daley and Menard, 1993). The error growth between cycles depends on the behavior of model error and initial condition errors, with the initial condition errors themselves influenced by observation error, inaccuracies in the DAS process, and background errors accumulated during previous model cycles. Some of the initial condition errors project

onto growing modes, while a large fraction projects onto damped modes. Model error has a complex growth rate as it consists of error processes with different time scales, including fast physical processes such as convection and numerical filtering, as well as slower dynamical and physical processes.

In the identical twin experiments, the fields of analysis error and analysis increment illustrate the regions of the globe that are constrained by assimilation of observations, as well as those that are not. The tropics are found to have the greatest variance and mean analysis error in the identical twin cases, with the least error over the continental United States and western Eurasia.

Due to imperfections in the data assimilation process and representativeness errors inherently present in the synthetic observations, errors will be introduced into the Twin NE fields even when initializing from a 'perfect' background field. Even in cases where an ingested observation locally improves the analysis state compared to the background, the resulting analysis state may be dynamically or physically unstable such that the forecast model will then amplify the error. For example, ingestion of temperature observations at only select vertical levels may result in undesirable vertical instability of the column, with subsequent convective processes acting to increase the forecast error. This type of behavior is noted in both Twin cases for temperatures in the tropics and summer convective areas, where convective correction of initial condition instability results in an apparent temperature bias of the background and analysis fields, although there is no model error and minimal observation biases.

The impact of initial condition errors on forecast accuracy is explored in these experiments. The DENSE case can be considered a "best observing network" scenario, where the initial condition error is minimized within the current data assimilation framework. A true "perfect initial condition" state does not actually exist because the ECMWF NR cannot be completely or perfectly represented using the GEOS-5 model due to differences in model grids and representation of parametrizations and variables. Initial attempts to approximate the ECMWF NR state directly on the GEOS-5 grid met with technical difficulties such that the resulting initial state would likely be no more accurate than the analysis states of the DENSE case. Instead, the DENSE case was chosen as a suitable alternative to study the role of initial condition errors.

Rapid growth of model error is clearly illustrated in the DENSE case during the first 24 hours of forward integration. This rapid error growth is not seen in the Twin cases that

have initial condition error of similar magnitude in the extratropics, implying that this error growth is most likely due to model error. The analysis errors in the DENSE case are very low due to the strong constraint of the global rawinsonde network, and thus the initial state is very close to the ECMWF NR state but in considerable disagreement with the preferred climatology of the GEOS-5 model. In the Control and NE cases, the relative sparseness of the observational network allows the analysis state to retain more of the GEOS-5 systematic model error with respect to the ECMWF NR, so that much less adjustment towards the GEOS-5 preferred climatology occurs during the early forecast period.

While it is desirable to attempt to quantify exactly the growth of model error and initial condition errors in conjunction with a simple theoretical framework of error growth, there are limitations to the practical application of such simplified theory to the experimental results. An attempt was made to fit the error growth rates to the simple model of exponential initial condition error growth with constant model error growth, as in Leith (1978). However, the fit of the data to theory was poor and did not yield useful estimates of model error growth rates.

Instead, some general statements quantifying model error growth in relation to initial condition error growth can be made. The results show that for the GMAO OSSE system, model error is dominant over initial condition error for medium range forecasts. The very rapid error growth observed during the first 24-48 hours of forward integration in the DENSE case is caused by model error, rather than growth of small initial condition errors. This sets a strict limit on the possible improvement in forecast skill to be achieved by reduction of initial condition errors, with the best possible improvement in forecast lead times being on the order of 24-36 hours for medium-range forecasts. The Twin cases have improved forecast skill of approximately two days in the extratropics and more than three days in the tropics compared with the Control, indicating that more significant gains in forecast skill are possible with improvements to the forward integrating model. It should be noted that the nature of model error growth in this case is actually a comparison of the ECMWF forecast model (circa 2005) with the GEOS-5 forecast model, and not a true comparison of the GEOS-5 model with the real world.

7 Acknowledgments

The OSSE was conducted with the assistance of Ricardo Todling, Meta Sienkiewicz, King-Sheng Tai, and Joseph Stassi at the GMAO. The ECMWF Nature Run was provided by Erik Andersson through arrangements made by Michiko Masutani. Support for this project was encouraged by Michele Rienecker and provided by GMAO core funding.

8 Appendix

A basic sense of how initial condition and model error act together to produce forecast error is revealed by considering a simple one-component model of linear dynamic growth with an added external forcing. Consider an error x governed by

$$\frac{dx}{dt} = \lambda x + f(t) \tag{2}$$

where t is time, λ is a growth rate due to linearized error dynamics and $f(t)$ is the external forcing. The latter represents model error due to imperfect formulation, independent of the forecast error itself. The general solution to (2) is

$$x(t) = e^{\lambda t}x(0) + \int_0^t e^{\lambda(t-t')}f(t')dt' \tag{3}$$

First, consider the case when $f(t)$ is time independent with $\lambda \neq 0$. Then

$$x(t) = e^{\lambda t}x(0) + \frac{f}{\lambda}(e^{\lambda t} - 1) \tag{4}$$

The first term describes the result of initial condition error and the second of model error. Both have an exponential component, indicating that as model error creates forecast error, the same dynamics apply, yielding the same growth rate. For short forecast times ($t << 1/\lambda$), however, the forecast error growth due to model error is dominated instead by the addition of model error rather than dynamic growth of previously added model error. The effect of model error is only seen after the error has had sufficient time to act, unlike the effects of initial condition error that may be seen immediately. At longer forecast times the dynamic growth of earlier added model error dominates the effect of such error still being added and, at such times, it is difficult to distinguish the presence of initial condition and model error based on the temporal behavior of the error alone.

If a set of independent cases are considered for which the means of the initial condition and model errors are zero and the variances are $V(0)$ and F, respectively, with no correlation

between the two types of error, then the variance of the forecast error as a function of time, computed by averaging over all the cases, is

$$V(t) = e^{2\lambda_i t}V(0) + (e^{\lambda_t} - 1)^2\frac{F}{\lambda^2} \tag{5}$$

If the forcing is instead stochastic white noise, with F_s indicating its variance, then

$$V_s(t) = e^{2\lambda t}V(0) + (e^{2\lambda t} - 1)\frac{F_s}{2\lambda^2} \tag{6}$$

The very similar forms of (5-6) reveal that similar comments about dynamics acting on previously applied model error in the case of time-independent forcing apply. Since white noise and temporally constant forings can be considered somewhat as extremes, their similar results provide confidence of the generality of the results for other likely types of forcing.

REFERENCES

Buizza, R. 2010. Horizontal resolution impact on short- and long-range forecast error. *Quart. J. Roy. Meteor. Soc.* **136**, 1020–1035.

Compo, G., Whitaker, J., Sardeshmukh, P., Matsui, N., Allan, R., Yin, X., Jr., B. G., Vose, R., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R., Grant, A., Groisman, P., Jones, P., Kruk, M., Kruger, A., Marshall, G., Maugeri, M., Mok, H., Nordli, Ø., Ross, T., Trigo, R., Wang, X., Woodruff, S. and Worleyu, S. 2011. The twentieth century reanalysis project. *Quart. J. Roy. Meteor. Soc.* **137**, 1–28.

Dalcher, A. and Kalnay, E. 1987. Error growth and predictability in operational ECMWF forecasts. *Tellus* **39**, 474–491.

Daley, R. and Menard, R. 1993. Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.* **121**, 1554–1565.

Errico, R., Ehrendorfer, M. and Raeder, K. 2001. The spectra of singular values in a regional model. *Tellus* **53A**, 317–332.

Errico, R. M. 2000. The dynamical balance of singular vectors in a primitive equation model. **126A**, 1601–1618.

Errico, R. M., Yang, R., Privé, N., Tai, K.-S., Todling, R., Sienkiewicz, M. and Guo, J. 2013. Validation of version one of the Observing System Simulation Experiments at the Global Modeling and Assimilation Office. *Quart. J. Roy. Meteor. Soc.* . doi: 10.1002/qj.2027, in press.

Errico, R. and Privé, N. 2013. An estimate of some analysis error statistics using the GMAO observing system simulation framework. *Quart. J. Roy. Meteor. Soc.* . in press.

Eyre, J. and Hilton, F. 2013. Sensitivity of analysis error covariance to the mis-specification of background error covariance. *Quart. J. Roy. Meteor. Soc.* **139**, 524–533. doi:10.1002/qj.1979.

Kleist, D., Parrish, D., Derber, J., Treadon, R., Wu, W.-S. and Lord, S. 2009. Introduction of the GSI into the NCEP global data assimilation system. *Weather and Forecasting* **24**, 1691–1705.

Leith, C. 1978. Objective methods for weather prediction. *Ann. Rev. Fluid Mech.* **10**, 107–128.

Lorenz, E. 1982. Atmospheric predictability experiments with a large numerical model. *Tellus* **34**, 505–513.

Orrell, D., Smith, L., Barkmeijer, J. and Palmer, T. 2001. Model error in weather forecasting. *Nonlinear Processes in Geophysics* **8**, 357–371.

Privé, N., Errico, R. and Tai, K.-S. 2013a. The influence of observation errors on analysis error and forecast skill investigated with an observing system simulation experiment. *J. Geophys. Res.* . in press.

Privé, N., Errico, R. and Tai, K.-S. 2013b. Validation of forecast skill of the Global modeling and Assimilation Office observing system simulation experiment. *Quart. J. Roy. Meteor. Soc.* . doi: 10.1002/qj.2029, in press.

Rienecker, M., Suarez, M., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R., Gelaro, R., Stajner, I. and Nielsen, J. 2008. The GEOS-5 data assimilation system - documentation of versions 5.0.1, 5.1.0 and 5.2.0, *Technical Report 27*, NASA.

Schubert, S. and Chang, Y. 1996. An objective method for inferring sources of model error. *Mon. Wea. Rev.* **124**, 325–340.

Simmons, A. and Hollingsworth, A. 2002. Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.* **128**, 647–677.

Untch, A., Simmons, A., Hortal, M. and et al, C. J.: 1999, Increased stratospheric resolution in the ECMWF forecasting system, *Proceedings of the SODA Workshop on Chemical Data Assimilation*, KNMI, pp. 45–52. De Bilt, the Netherlands.

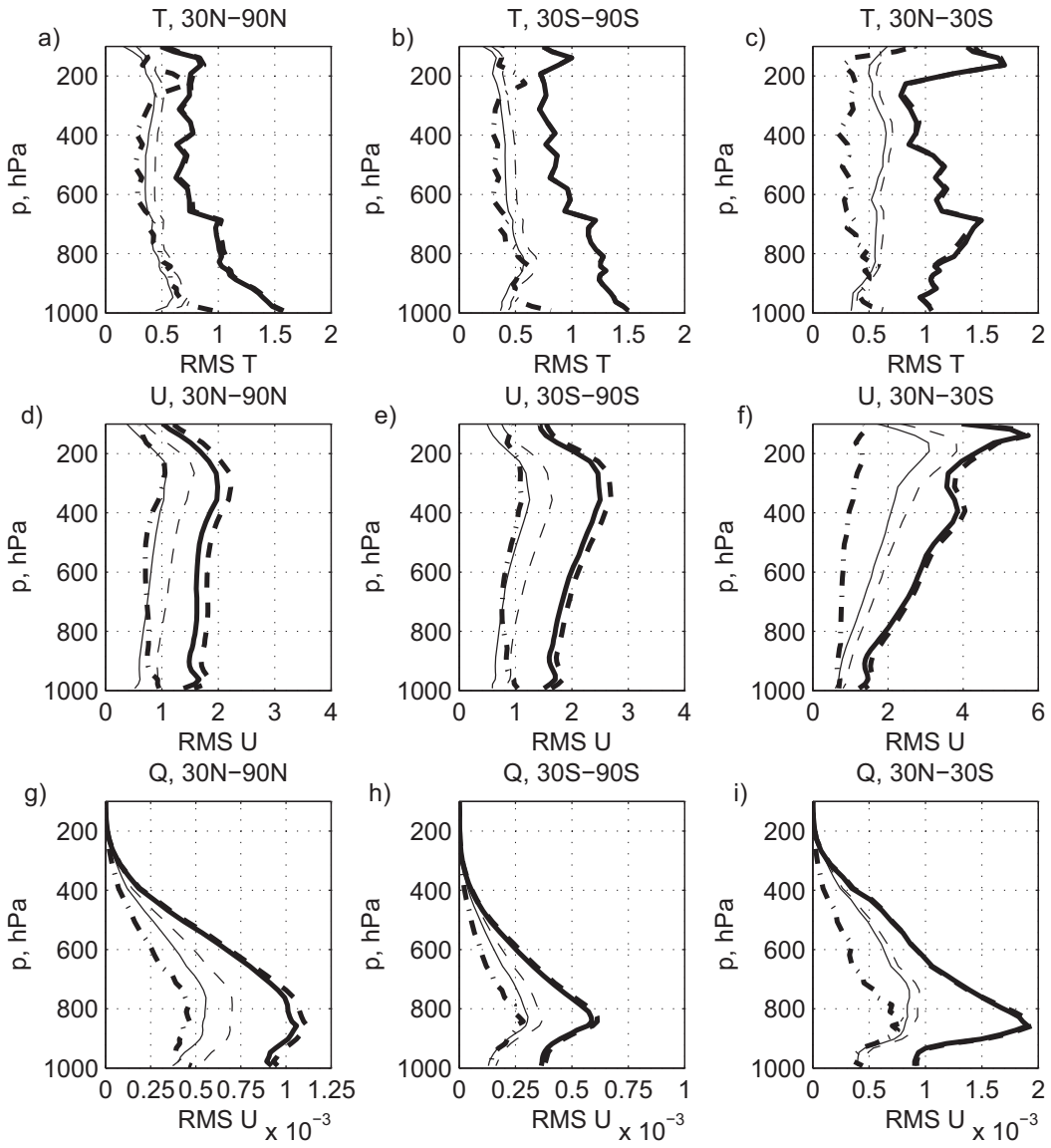

Figure 1. Global analysis RMSE for July 2005. Left column, 30N-90N; center column, 30S-90S; right column, 30S-90S. Top row, temperature (K); center row, zonal wind ($m\,s^{-1}$); bottom row, specific humidity (kg/kg). Solid heavy line, NE case; heavy dashed line, Control case; heavy dash-dot line, DENSE case; thin solid line, Twin NE case; thin dashed line, Twin Control case.

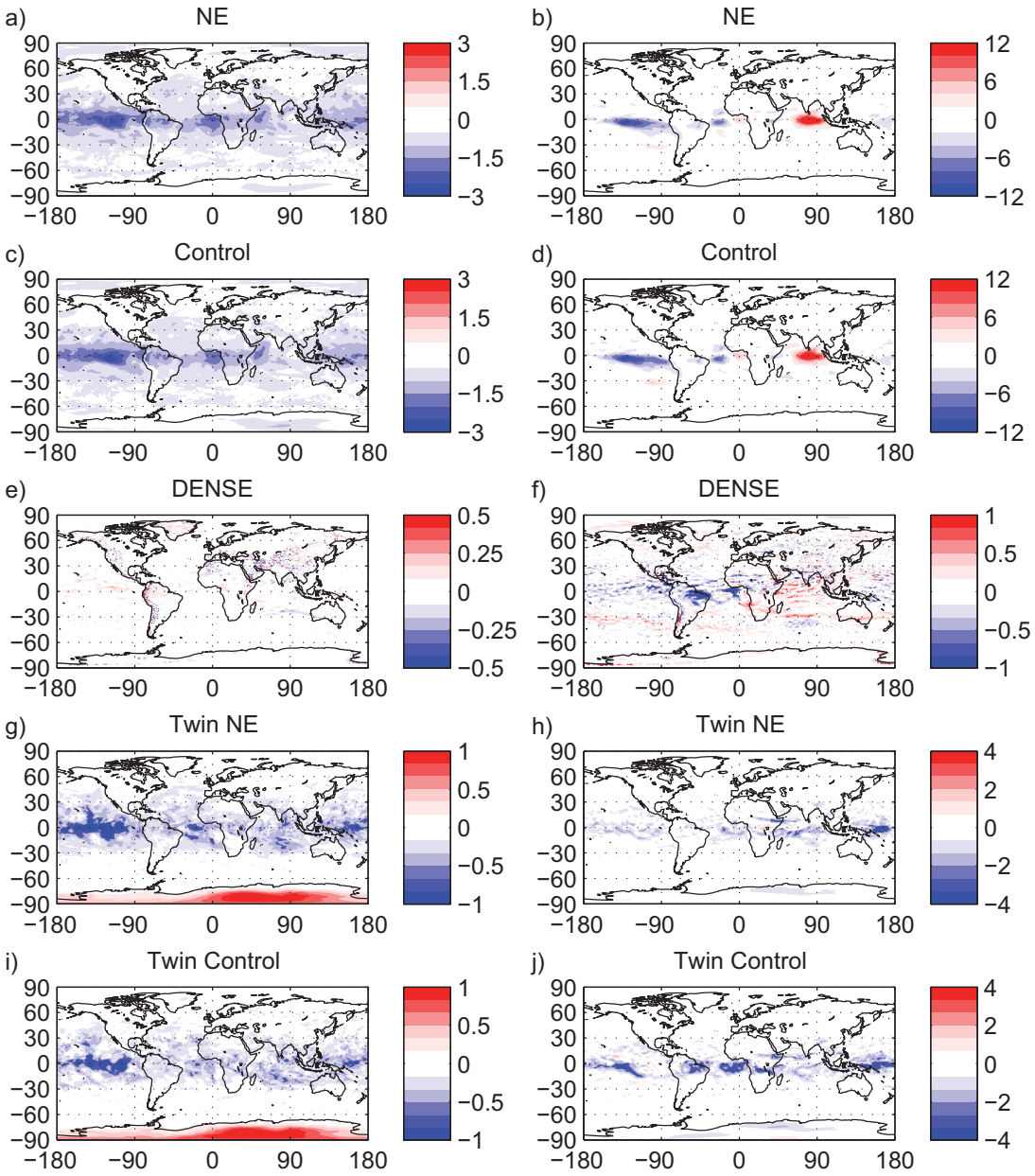

Figure 2. Monthly mean analysis error for the experimental cases in July 2005. Left column, temperature at 500 hPa, K. Right column, zonal wind at 250 hPa, $m\,s^{-1}$. Note varying contour intervals between panels.

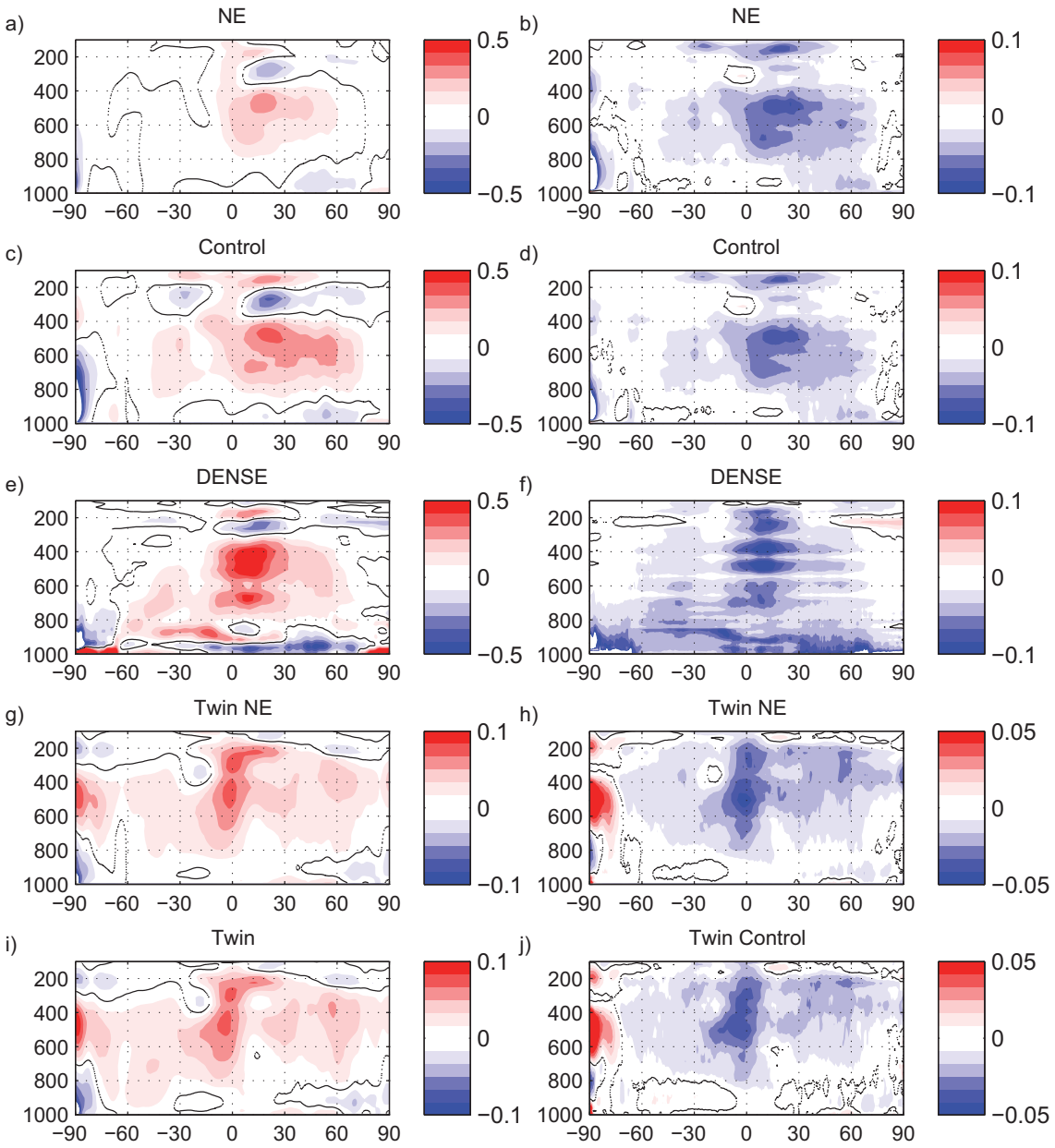

Figure 3. Zonal mean time mean analysis increment (left column) and $|A| - |B|$ (right column) for temperature (K), July 2005. Negative values of $|A| - |B|$ indicate an improvement of the analysis field compared to the background. Note different contour range for Twin and ECMWF NR cases.

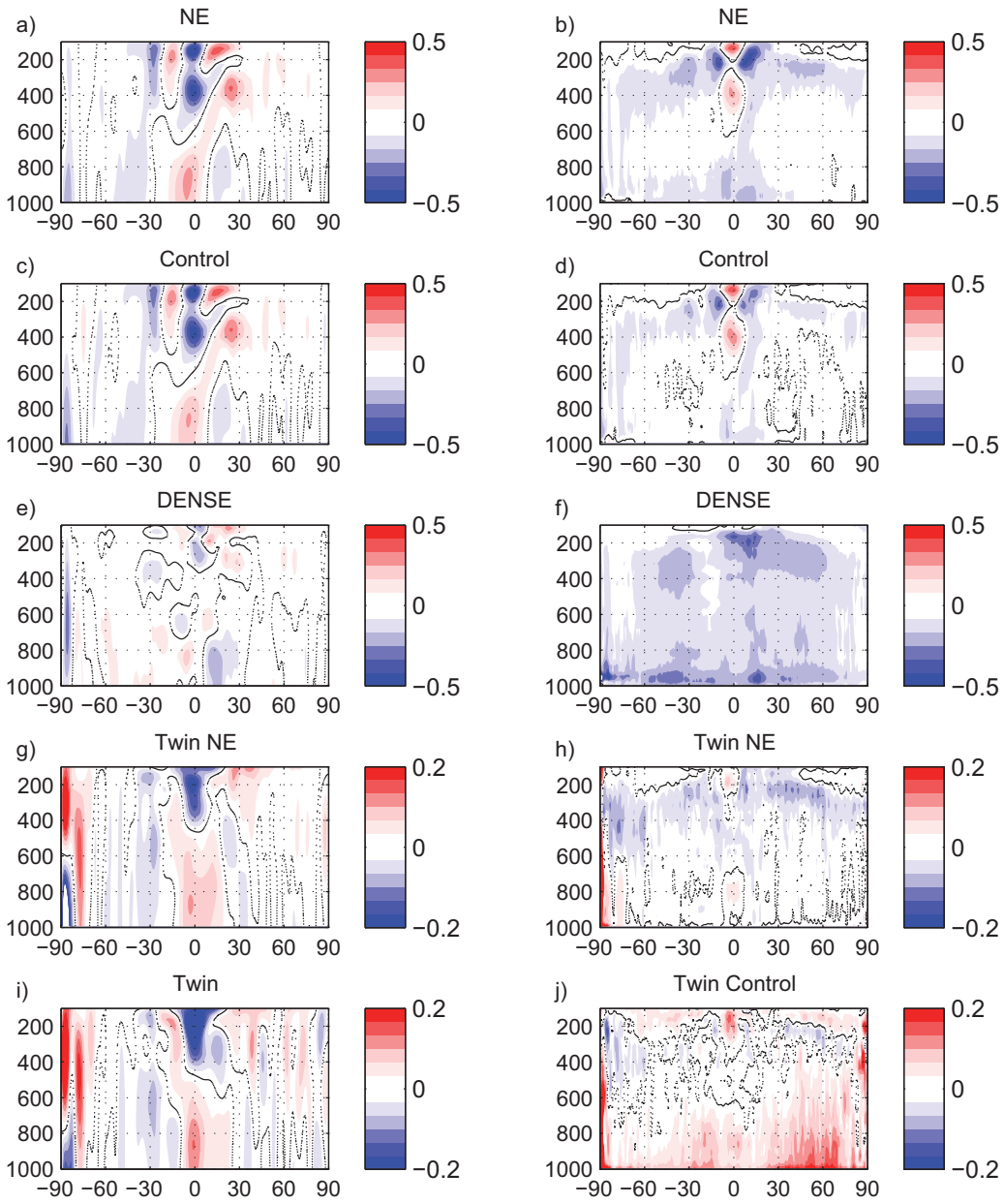

Figure 4. Zonal mean time mean analysis increment (left column) and $|A| - |B|$ (right column) for zonal wind ($m\,s^{-1}$), July 2005. Negative values of $|A| - |B|$ indicate an improvement of the analysis field compared to the background. Note different contour range for Twin and ECMWF NR cases.

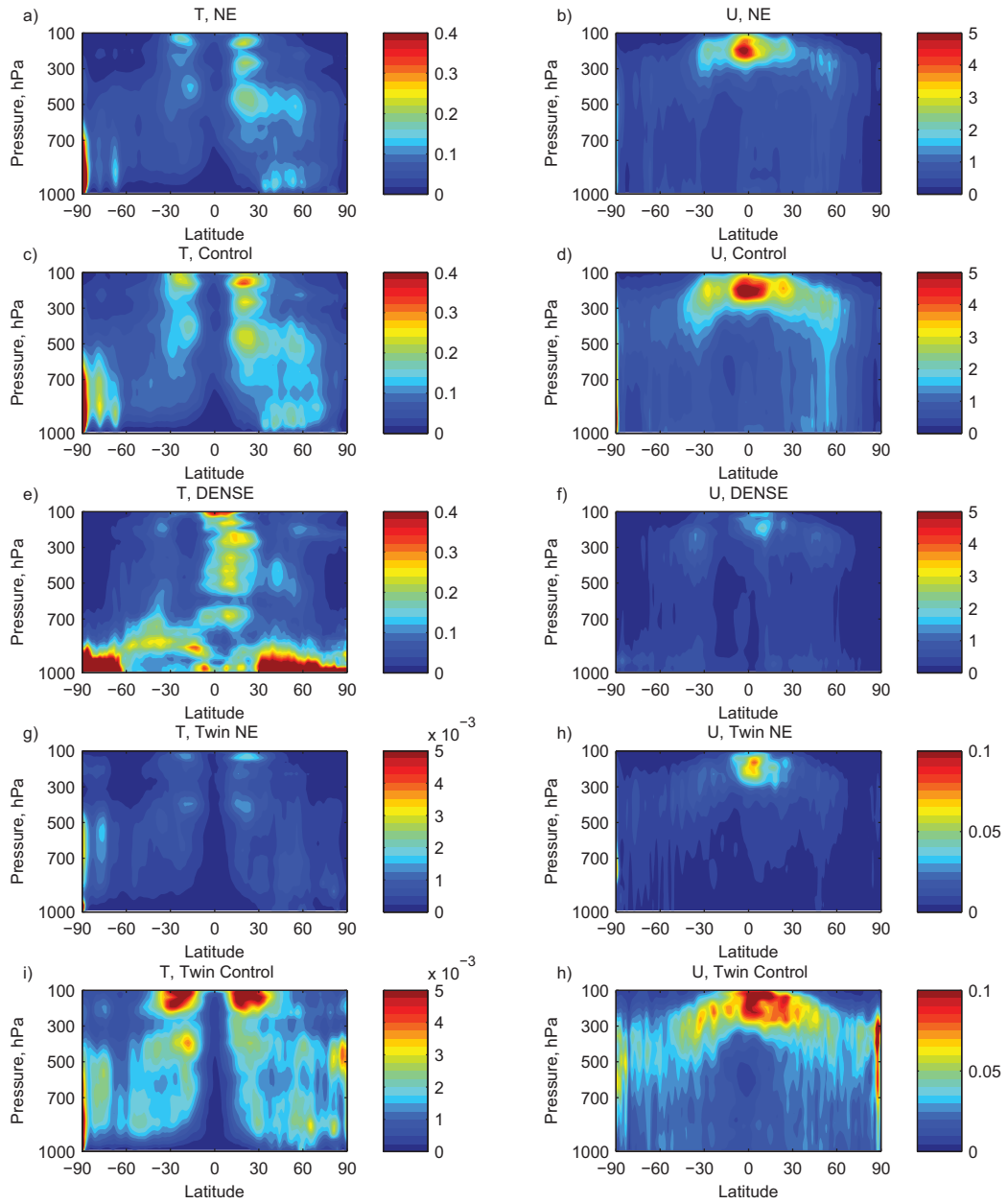

Figure 5. Zonal mean variance in time of the analysis increment for temperature (K^2), left, and zonal wind ($m\,s^{-1}$), July 2005.

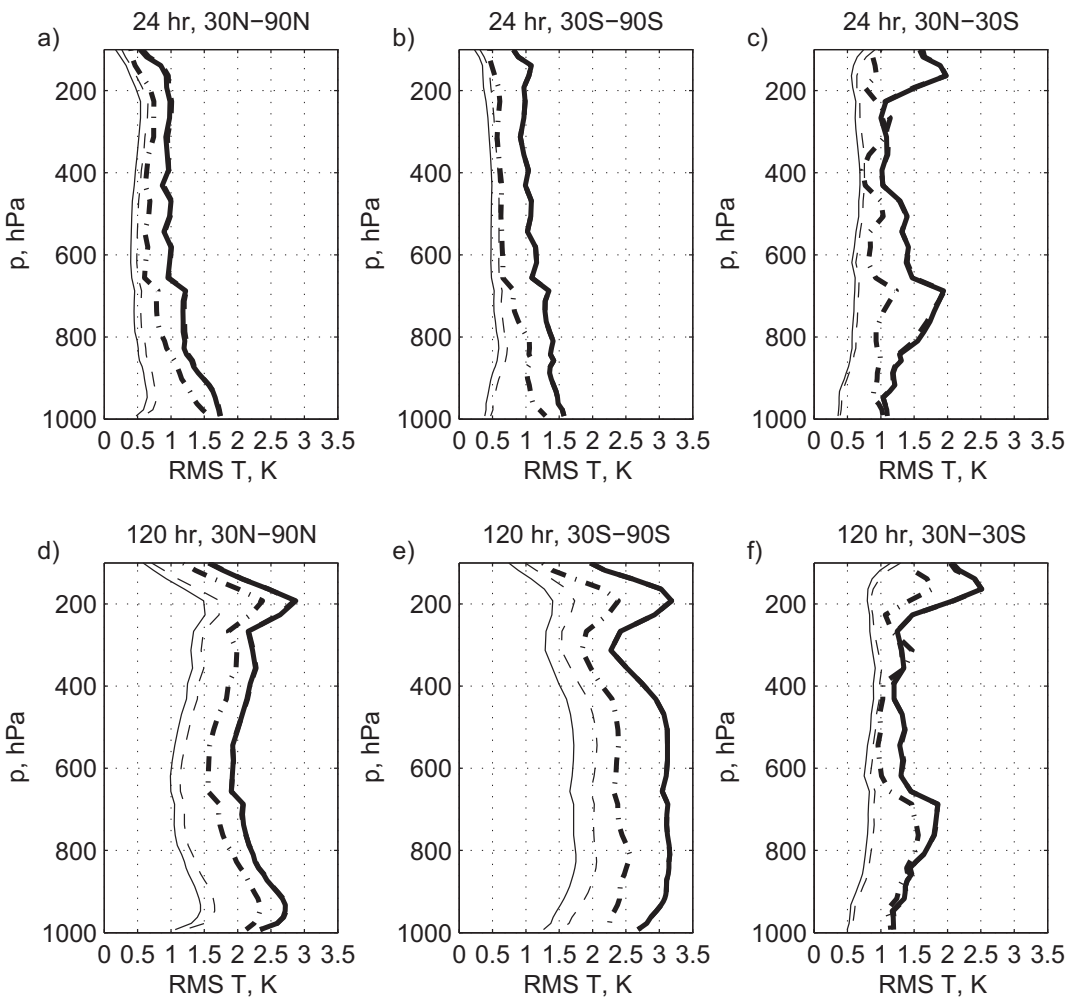

Figure 6. Forecast RMSE for temperature, July 2005. Left column, 30N-90N; center column, 30S-90S; right column, 30S-30N. Top row, 24 hour forecast; bottom row, 120 hour forecast. Solid heavy line, NE case; heavy dashed line, Control case; heavy dash-dot line, DENSE case; thing solid line, Twin NE case; thin dashed line, Twin Control case.

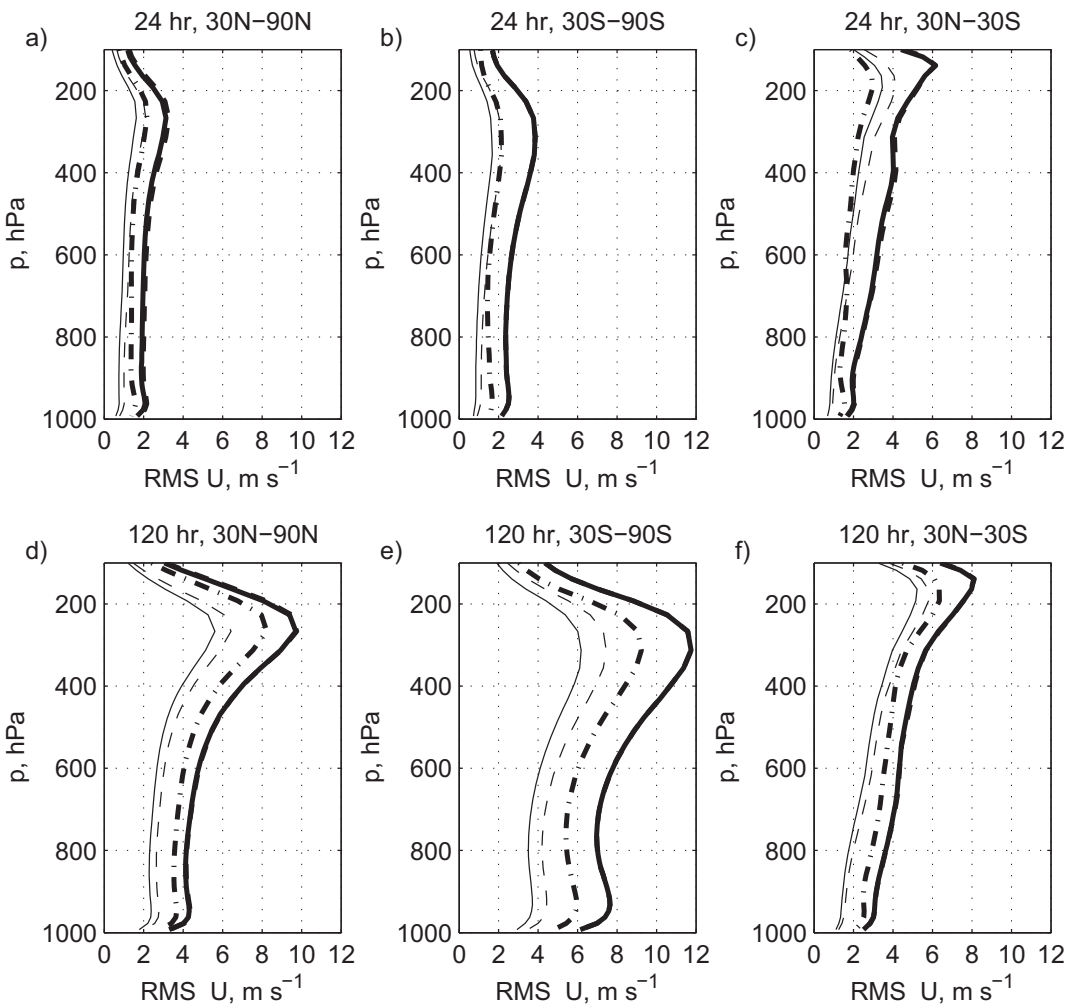

Figure 7. As for Figure 6, but for zonal wind u, $m\,s^{-1}$.

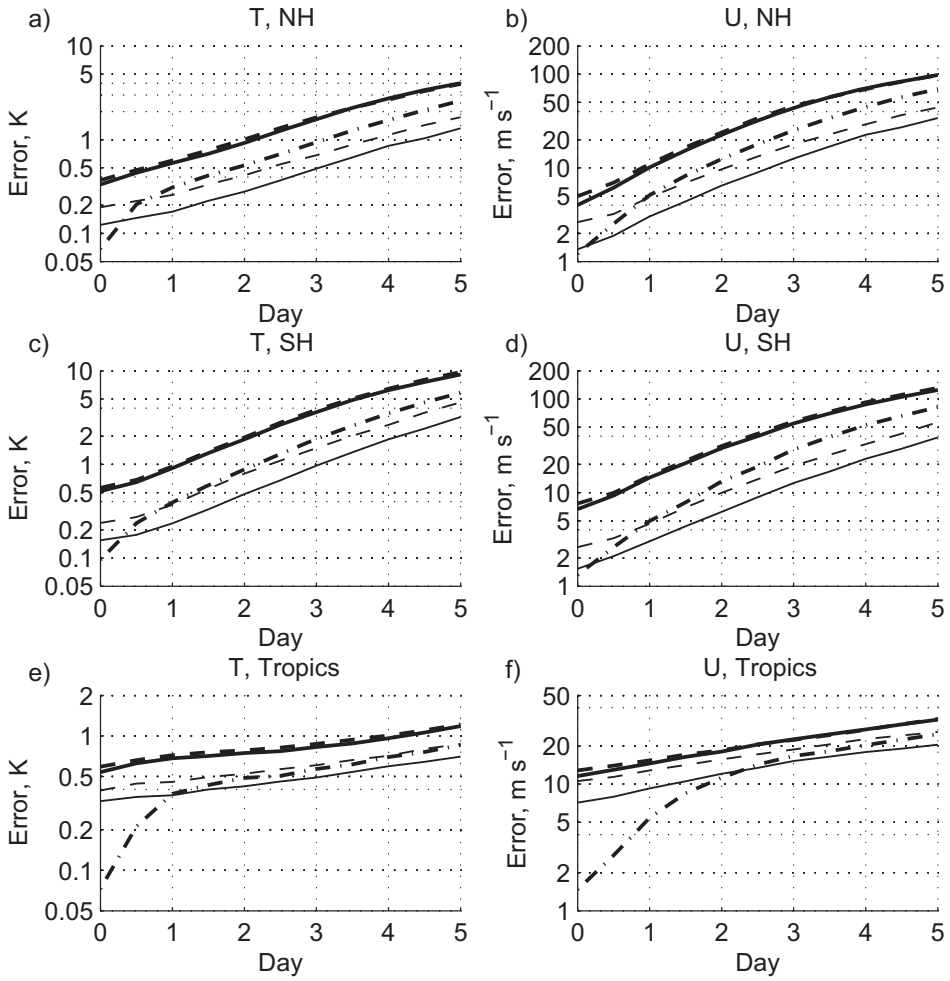

Figure 8. Global error variance as a function of forecast time, July 2005. Left column, 500 hPa temperature (K^2); right column, 250 hPa zonal wind (m^2, s^{-2}). Top row, 30N-90N; center row, 30S-90S; bottom row, 30N-30S. Solid heavy line, NE case; heavy dashed line, Control case; heavy dash-dot line, DENSE case; thing solid line, Twin NE case; thin dashed line, Twin Control case.

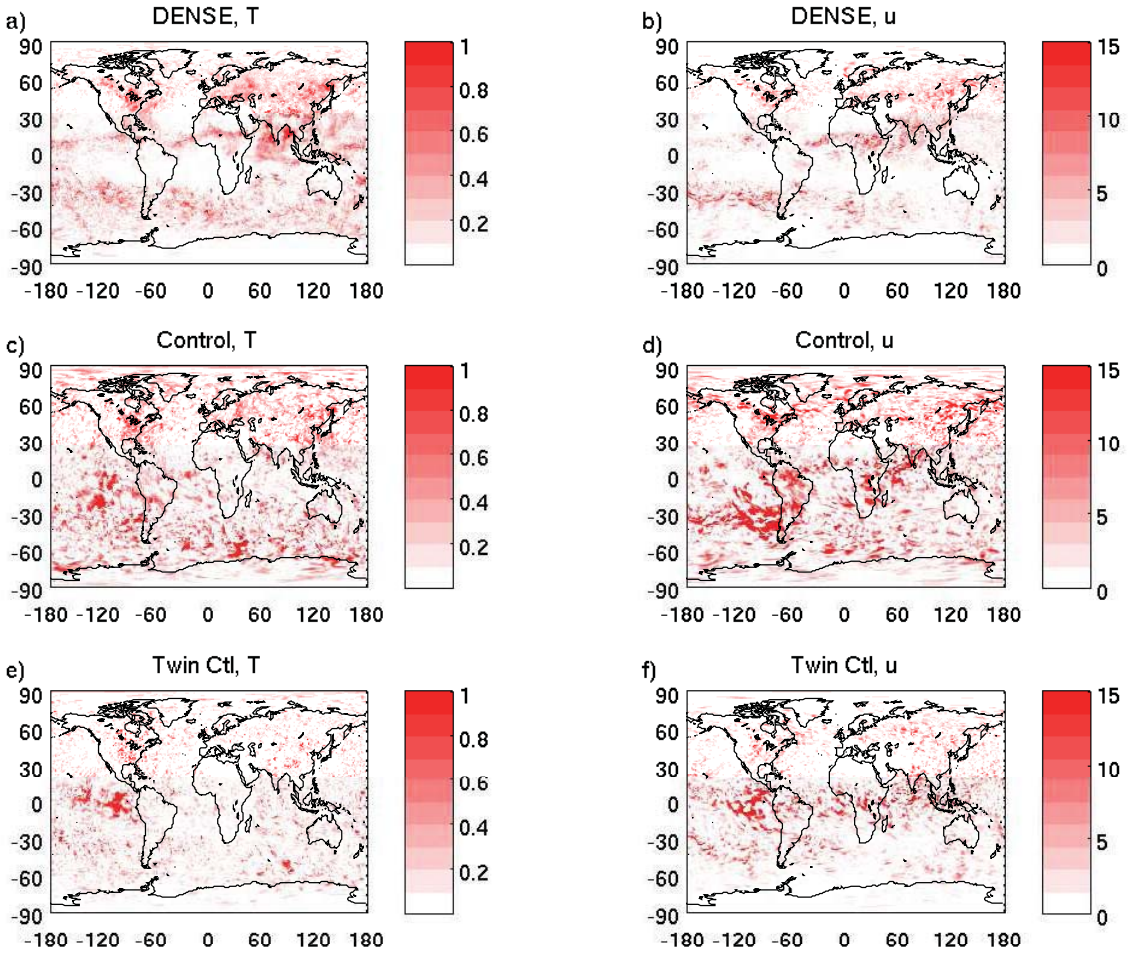

Figure 9. July 2005 monthly mean error variance at the 12 hour forecast minus the analysis error variance. Left column, temperature variance (contour interval 0.1 K^2) on model surface nearest 500 hPa; right column, zonal wind variance (contour interval 1.5 $m^2\ s^{-2}$) on model surface nearest 250 hPa. Top, DENSE case; center, Control case; bottom, Twin Control case.

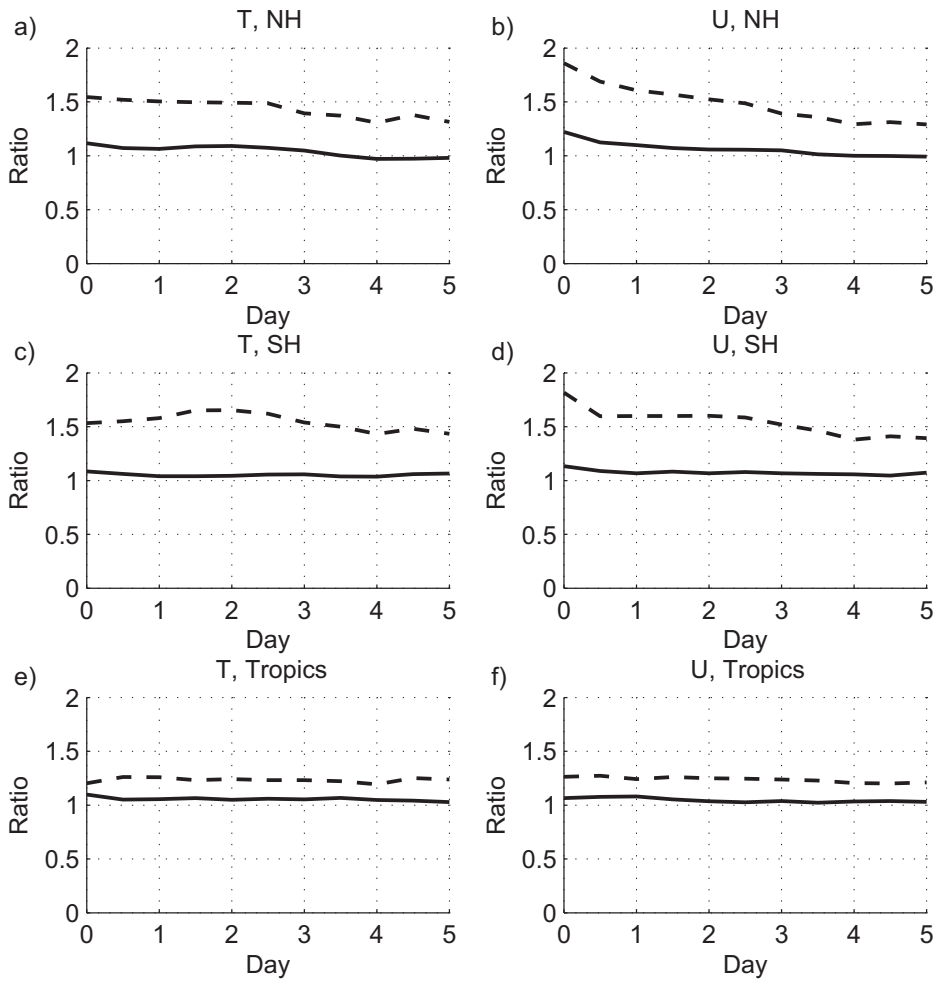

Figure 10. Ratio of error variance as a function of forecast time between the Control and NE cases (solid line) and between the Twin Control and Twin NE cases (dashed line), July 2005. Left column, ratio for global mean temperature at 500 hPa; right column, ratio for global mean zonal wind at 250 hPa.

Table 1. Description of all OSSE cases included in this manuscript. NR options are the ECMWF-generated NR or the identical twin GEOS-5 NR. "Added Obs Err σ" refers to whether or not synthetic observation error was explicitly applied to synthetic observations. "Obs Network" refers to the choice of synthetic observation network, with "operational" the data types used operationally, and "RAOBs" a hypothetical global network of rawinsondes only.

	NR	Added Obs Err	Obs Network
Control	ECMWF	yes	operational 2005
NE	ECMWF	no	operational 2005
DENSE	ECMWF	no	RAOBs
Twin Ctl	GEOS-5	yes	operational 2011
Twin NE	GEOS-5	no	operational 2011

Table 2. Doubling times in days for RMSE of temperature and zonal wind, calculated using a fit to forecast error from 48 to 96 hours.

	Temperature			Zonal Wind		
	NH	SH	Tropics	NH	SH	Tropics
Control	2.8	2.3	10.7	2.6	2.6	7.7
NE	2.5	2.3	10.7	2.4	2.6	6.9
DENSE	2.5	2.0	7.7	2.2	2.0	4.8
Twin Ctl	2.8	2.3	9.2	2.5	2.3	7.7
Twin NE	2.5	2.0	8.2	2.2	2.1	6.9