Meteorologische Zeitschrift, Vol. **16**, No. 6, 1-3 (December 2007)

Article

The Estimation of Analysis Error Characteristics Using an Observation System Simulation Experiment

RONALD M. ERRICO[1,2,*], RUNHUA YANG[3,2], MICHIKO MASUTANI[4], JOHN S. WOOLLEN[4]

[1] Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County
[2] Global Modeling and Assimilation Office, Goddard Space Flight Center
[3] Science Systems and Applications Incorporated
[4] National Centers for Environmental Prediction

(Manuscript received March 6, 2007; in revised form June 30, 2007; accepted July 7, 2007)

Abstract

Observation system simulation experiments (OSSEs) have been performed at the National Centers for Environmental Prediction primarily for the purpose of evaluating the forecast improvement potential of proposed new observation instruments. The simulations have been validated primarily by comparing results from corresponding data denial experiments in both simulated and real data assimilation contexts. Additional validation is presented here using comparisons of some statistics of analysis increments determined from a baseline simulation using the entire suite of observations utilized during a reanalysis for February 1993. By exploiting the availability of a data set representing “truth” in the simulations, the background and analysis errors produced for the baseline simulation are computed. Several statistics of these errors are then determined, including time means and variances as functions of location or spherical harmonic wave number, vertical correlations, Kalman gains, and balance as measured by projections onto normal modes. Use of an OSSE in this way is one of the few means of estimating analysis error characteristics.

Although these simulation experiments are among the best calibrated ones existing, the additional validation here indicates that some unrealism remains present. With this caveat, several interesting characteristics of analysis error have been revealed. Among these are that: longitudinal variations of error variances in the Northern Hemisphere have a similar range as latitudinal variations; corresponding background and analysis error variances are very similar in most regions so that the Kalman gains are generally small with the notable exception of regions and times well observed by rawindsondes; correlation lengths (both vertical and horizontal) are very similar for background and analysis errors; error variances at horizontal scales shorter than those corresponding to approximately spherical harmonic wavenumber 70 are as large as corresponding variances of the fields themselves; approximately 30% of the error energy resides in gravitational normal modes; and the variance spectrum of the wind error is approximately white for scales below its peak near wavenumber 20. Although limited by remaining imperfections in this OSSE, these results reveal ways in which future experiments should be examined.

Zusammenfassung

Am National Center für Environmental Prediction (NCEP) wurden Beobachtungssystem - Simulations-Experimente (observation system simulation experiments - OSSEs) durchgeführt, mit dem primären Ziel, potentielle Verbesserungen in der Vorhersage auf Grund neuer Beobachtungen zu evaluieren. Diese Simulationen wurden in erster Linie dadurch evaluiert, dass Ergebnisse von Experimenten mit einander verglichen werden, in denen in simulierten und operationellen Datenassimilationsumgebungen Daten entweder verwendet oder nicht verwendet werden. Hier wird eine weitere Evaluierung vorgestellt, in welcher Vergleiche von Statistiken von Analyseinkrementen verwendet werden, die aus einer Referenzsimulation bestimmt wurden, die sämtliche Beobachtungen verwendet, die für eine Reanalyse des Februar 1993 zur Verfügung standen. Unter Verwendung der ”Wahrheit”in den Simulationen werden Hintergrund- (”background”) und Analyse-Fehler für die Referenzsimulation berechnet. Eine Reihe von Statistiken dieser Fehler werden in Folge bestimmt, darunter zeitliche Mittel und Varianzen als Funktionen des Orts oder der Wellenzahl (basierend auf sphärischen Basisfunktionen), vertikale Korrelationen, Kalman Gewichte, und Balance (beschrieben durch Projektionen auf Normalmoden). Die Verwendung von OSSE Experimenten ist eine der wenigen Methoden, auf Grund derer die Eigenschaften von Analysefehlern abgeschätzt werden können.

Diese Simulationsexperimente gehören zu den am besten kalibrierten Experimenten, die derzeit verf—”ugbar sind. Dennoch zeigt diese zusätzliche Validierung, dass unrealistische Züge vorhanden sind. Unter Berücksichtigung dieser werden hier einige interessante Eigenschaften von Analysefehlern aufgedeckt. Dazu gehören: Variationen in Nord-Südrichtung der Fehlervarianzen haben auf der Nordhemisphäre eine ähnliche Streubreite wie die Variationen in West-Ostrichtung. Die sich entsprechenden Backgound und Analysefehler-varianzen sind in den meisten Gebieten sehr ähnlich, sodass die Kalmangewichte im allgemeinen sehr klein sind: eine bemerkenswerte Ausnahme besteht in jenen Gebieten und zu jenen Zeiten, welche durch Radiosonden gut beobachtet sind. Längenskalen für vertikale und horizontale Korrelationen sind für Background und Analysefehler sehr ähnlich. Fehlervarianzen auf horizontalen Skalen, die kleiner sind als ihre Entsprechung zu Wellenzahl 70, sind so gross wie die entsprechenden Varianzen der Felder selbst. Etwa 30% der Energie der Fehler wird von den Gravitations-Normalmoden getragen. Das Varianzspektrum des Windfehlers is nahezu weiss auf Skalen unterhalb des Maximums nahe Wellenzahl 20.

0941-2948/2007/00XX $ 2.25

DOI: 10.1127/0941-2948/2007/00XX

1 Introduction

Atmospheric analyses are used for many purposes, including numerical weather prediction, climate analysis and observation instrument validation. For all of these purposes, knowledge of the analysis error statistics would be useful for proper interpretation of any results. Applications that explicitly depend on analysis uncertainty, such as ensemble forecasting or data assimilation, particularly require estimates of analysis error statistics.

Since operationally produced analyses are intentionally designed to be the best available estimates of an atmospheric state and all observations are imperfect, it is very difficult to directly estimate realizations of analysis error. Several indirect methods are available, however, although each has its own deficiency. Fits between observations that have been used and the analysis produced from them are routinely examined, but since this is an incestuous examination, it can only provide consistency measures. Data assimilation systems based on the Kalman filter, including 3–dimensional variational (3DVAR) schemes, formally include algorithms for determining both a best–estimate of the atmospheric state and an estimate of the error covariance of that state. The latter, however, is very dependent on a priori assumptions and computational approximations. Observations that are independent of the analysis can be used as validation, but since most of the best available observations presumably have already been incorporated into the analysis, few additional quality observations exist, except perhaps within small, possibly unrepresentative regions where an intense observation experiment has been conducted. Comparison of distinct analyses produced using different techniques, assimilation models, and different but overlapping observational data sets may yield insight regarding errors, although assumptions about how close each is to truth must be made. Analysis systems are also routinely validated by scoring forecasts subsequently produced using them as initial conditions. This validation technique may be considered as an error amplifier, since some errors in sensitive regions will grow rapidly. Not all errors grow however, and some forecast errors are due to model formulation rather than initial condition error. Placing too much confidence in any of these useful techniques can lead to gross misinterpretations (e.g., as revealed in ERRICO AND LANGLAND, 1999).

One technique for indirectly estimating statistics of analysis error that has received little attention is the use of observation system simulation experiments (OSSEs). Typically, the motivation for such experiments has been to estimate potential impacts of changes (additions or deletions) to the present global observing network. OSSEs can also be applied, however, to help answer other questions that are both fundamental and critical. The advantage of OSSEs over other kinds of investigations of similar questions is that, since the "truth" is known exactly, the corresponding analysis or subsequent (simulated) forecast errors can be measured exactly. Their primary weakness is that the equivalence of their results to what would be obtained in the context of analyzing the real atmosphere, even when using an identical data assimilation system, will depend on how well the real, but often unknown, observation, representativeness, and model errors have been characterized in the OSSEs. If OSSEs are appropriately motivated, adequately calibrated, and carefully interpreted, their advantages greatly surpass these weaknesses.

2 OSSE design

The OSSEs examined in this study were produced at the United States National Centers for Environmental Prediction (NCEP, MASUTANI ET AL. 2006). The assimilation system is a September 2005 version of the NCEP 3DVAR, also known as spectral statistical interpolation (SSI; PARRISH AND DERBER 1992). The technique is sequential, producing an analysis every 6 hours. The assumed background error statistics used by the SSI are the same as used operationally; no retuning of those statistics was performed within the OSSE framework. The assimilation model is a February 2005 version of the NCEP spectral model at T170L42 resolution. It is the same model that was used to produce subsequent forecasts operationally at that time.

The atmospheric simulation that replaces nature in these experiments (denoted as "truth" or as the "nature run") is provided by a 1993 version of the forecast model then used at the European Centre for Medium–Range Weather Forecasts (ECMWF). Its resolution is T213L31. It defines heights of the surface topography using "envelope orography" (TIBALDI 1986), tending to yield higher topography than that defined for the NCEP system. This ECMWF model is the same that was used for the ERA–15 reanalysis (GIBSON ET AL. 1997) but at higher resolution. The data set is actually a 31 day forecast started at 06 UTC 5 February 1993. The NCEP model uses the same sea surface temperature as used for the nature run, except interpolated to the different resolution grid. By using different models for the nature run and assimilation system, a form of forecast model error is introduced in the assimilation system, although how this may be similar or different than forecast model error in the real assimilation system is unknown.

The set of simulated observations is in one–to–one correspondence with the set of real observations actually used by an NCEP assimilation applied to all observations archived during February 1993. Only those real

*Corresponding author's address: Ronald M. Errico, Global Modeling and Assimilation Office, NASA/GSFC Code 610.1, Greenbelt, MD 20771, USA. e–mail: rerrico@gmao.gsfc.nasa.gov

observations that passed both data selection and quality control procedures are used to define the corresponding simulated ones. This set therefore changes each analysis time; e.g. as cloud–screening algorithms select only particular radiance observations. The simulated observations are colocated with the corresponding real ones but are assumed to have no gross errors that require further quality control beyond the selection process applied to the real observations. Thus the utilized observation data count and distribution in time and space are identical to those for the real assimilation.

Note that, with regard to the nature run, the simulated radiances may unrealistically be in cloudy regions rather than in the cloud free regions where the real observations were. Analogously, simulated cloud track wind observations may be in cloud free regions of the the nature–run fields. If whether observations are in cloud free or clear regions profoundly affects their impacts, e.g., because it determines whether they are in "active" regions prone to forecast error growth, then this unrealism may detrimentally affect the OSSE. Currently, the impact of this unrealism is unknown.

Observations include rawindsondes, surface and aircraft reports, cloud track winds, and satellite observed radiances (simulated High–Resolution Infrared Sounder and Microwave Sounding Unit on NOAA–11 and NOAA–12 satellites). The daily numbers of observations actually used averaged approximately 230,000 for conventional and cloud track wind data and 255,000 for satellite radiances. These are the same observations used for the same period of the NCEP/NCAR reanalysis (KALNAY ET AL. 1996).

Observational errors are produced by adding random numbers drawn from a truncated–normal distribution having a zero mean and a variance equal to the corresponding real instrument plus representativeness error variances specified in the data assimilation system. The truncation simply removes random errors with absolute values greater than 3 standard deviations from the mean since these would be more akin to the gross errors detected for observations subsequently discarded by quality control algorithm. Simulated radiances are produced from the nature–run fields using a different radiative transfer model than the one used by the assimilation scheme. Observations not at nature–run grid points are produced by the same spatial interpolation scheme as used by the assimilation system. In the horizontal, this is bilinear, using the four surrounding grid points. Thus portions of representativeness error have potentially been added in two ways for some observations, since error is inserted by both adding random numbers and by using these forward and interpolation models (since they differ or act on a different grid compared with the ones employed by the assimilation system).

Note that although error biases or correlations have not been explicitly added, unknown biases and correlations have been introduced because, for example, the assimilation consistently uses a different radiative transfer model than the one used to simulate the observations. Although many real radiance observations have large biases, most (but not all) of such bias is removed by bias correction algorithms within the assimilation scheme. It currently remains unclear how different the simulated error biases or correlations are compared to those of such real "corrected" errors.

3 Examination method

The analysis errors are simply defined as the differences between the analyzed fields produced by the OSSE and the corresponding truth fields. These differences are computed by first interpolating both fields to a common grid. Horizontally, this is a Gaussian grid with 432 longitudes and 216 latitudes that allows exact spherical harmonic transforms for T213 resolution. Vertically, there are 31 levels above the surface. First, the fields on the truth grid are interpolated to the common grid by transforming those fields on their hybrid level surfaces to T213 spherical harmonic (hence denoted as spectral) coefficients and then projecting them onto the new grid. This projection is done to reduce data storage issues since the Gaussian grid designed for alias–free quadratic calculations on which the data has been originally provided is not needed for these computations. The analysis data that are provided as spectral coefficients defined on 42 sigma–coordinate surfaces in addition to the surface, are projected onto the common horizontal grid but their own vertical grid. Then, the projected analysis data are vertically interpolated to the same pressure–level values that apply to the truth data. That interpolation is linear in logarithm of pressure for all fields except specific humidity, for which the same interpolation is instead applied to the relative humidity. For extrapolations that may be required when analysis data surfaces do not exist both above and below a level to which the fields are to be interpolated, such as near the surface, a zero vertical gradient is assumed (but for q, in terms of relative rather than specific humidity). Several tests were performed to ensure that interpolation itself did not create significant differences at any horizontal scale. Background errors with respect to the OSSE truth fields are computed analogous to the analysis errors.

The fields that have been examined are temperature T, specific humidity q, eastward wind u, northward wind v, vertical component of vorticity ζ, and horizontal wind divergence δ. The latter two are computed with respect to the hybrid or σ–level surfaces being considered horizontal. Although the common grid uses a hybrid vertical coordinate, particular levels will be referred to by a value of $\sigma = p/p_s$ where p is the pressure computed on that surface when the surface pressure $p_s = 10^5$Pa. Using this denotation, the only levels examined have been $\sigma = 1., 0.9, 0.8, \ldots, 0.1$. Results suggest that the climatologies of the NCEP and ECMWF models are very different in the stratosphere, where few observations exist to constrain the analysis, so while differences at $\sigma = 0.05$ and 0.01 have also been determined, they are not discussed in what follows.

Analysis increments themselves are examined on the analysis grid, rather than on the common grid, since both analysis and background fields are defined on that grid. The differences between those fields are computed on their sigma–coordinate surfaces, with no vertical interpolation to a common pressure level. By not using the former common grid, no additional differences due to interpolation are introduced.

Spectra are computed from spherical harmonic decomposition of gridded fields defined on the hybrid–coordinate surfaces. For levels near the surface, the spectra of temperature in particular therefore reflect topographic effects on the fields to some degree, since those surfaces effectively follow the topography. For fields on the analysis grid which use a pure sigma coordinate, a topographic effect is felt even into the stratosphere. On the hybrid–coordinate common and nature grids, however, the effect lessons with height and is negligible in the stratosphere, where the grid surfaces become constant pressure surfaces.

Spectra of T are those of simple variances. Wind spectra on the other hand are in terms of separate power spectra for kinetic energies of rotational (RKE) and divergent (DKE) wind components. These all are presented in terms of 2–dimensional wavenumber n (the degree of the Legendre polynomial of the spherical harmonics considered). Thus, for each n both the variance and power spectra are computed as sums over all Legendre polynomials for orders m, $0 \le |m| \le n$. The larger the value of n, the greater the number of independent values summed, and therefore spectral statistics for larger n are computed using a larger sample of independent values. Since the 3–week period used to determine all the OSSE statistics is rather short, spectra presented for very small values of n may not be statistically representative of what results may be for a longer or different analysis period, but those for larger n should be more reliable. For this reason, results for $n < 4$ will not be discussed.

The fractional variance reduction for any field, termed the "Kalman gain" here, is measured as

$$K = \frac{s_b^2 - s_a^2}{s_b^2}, \tag{1}$$

where s^2 is the error variance for either the background or analysis as indicated by an appropriate subscript. For the scalar case of analyzing a single variable with a single observation, this is equivalent to the Kalman gain, which explains the generalized use of the term. Here the variances are first computed either for each grid point or for each spectral component after removing the time mean values. Then the Kalman gain at each grid point or spectral component is computed. Finally, for presentation, these individual Kalman gains are averaged zonally or over spectral components with the same value of n. What appears in those presentations is therefore an average of individually computed Kalman gain rather than the Kalman gain computed from averaged or total variances. Also note that this is not the true Kalman gain since spatial and field correlations are not considered.

Dynamic balances are measured by first projecting fields onto standard normal modes (DALEY 1991). The modes' vertical structures are determined using a spatially independent reference temperature of 270K and the hydrostatic formula used by the NCEP model (Other reasonable choices of either this temperature or the hydrostatic formula yielded negligible differences of the measures to be presented). Horizontal components of the normal mode are determined as spectrally truncated (at T98) Hough functions as in ERRICO (1987). The projections are then measured in terms of their energy, that can be considered equivalent to a weighted sum of variances of the corresponding T, u, v and p_s fields (ERRICO 2000).

The projection energies are summed over all modes for a given type (rotational or gravitational) and vertical mode index k. The value of $k-1$ is the number of nodes for the vertical structure of wind or geopotential for that set of modes. Thus, $k = 1$ refers to the set of external modes that are almost barotropic throughout the troposphere. Each vertical mode is characterized by an equivalent depth H_k. At small horizontal scales where the Coriolis force becomes insignificant, atmospheric gravity waves having the kth vertical structure will propagate with the same horizontal phase speed as shallow water waves in an ocean with a mean depth H_k. Thus H_k denotes a dynamical, rather than spatial, property of the mode.

4 OSSE validation

The primary validation of the NCEP OSSEs has been performed using data denial comparisons (MASUTANI ET AL. 2006). In these, selected observation types are removed from both the OSSE analysis and corresponding real analysis. The means and standard deviations of the differences between analyses produced with and without those observations are computed and corresponding results for the OSSE and real systems are compared. Of course, the correspondence is imperfect, especially regarding mean differences. Yet, in most respects the agreement is sufficiently adequate to declare the OSSE "realistic."

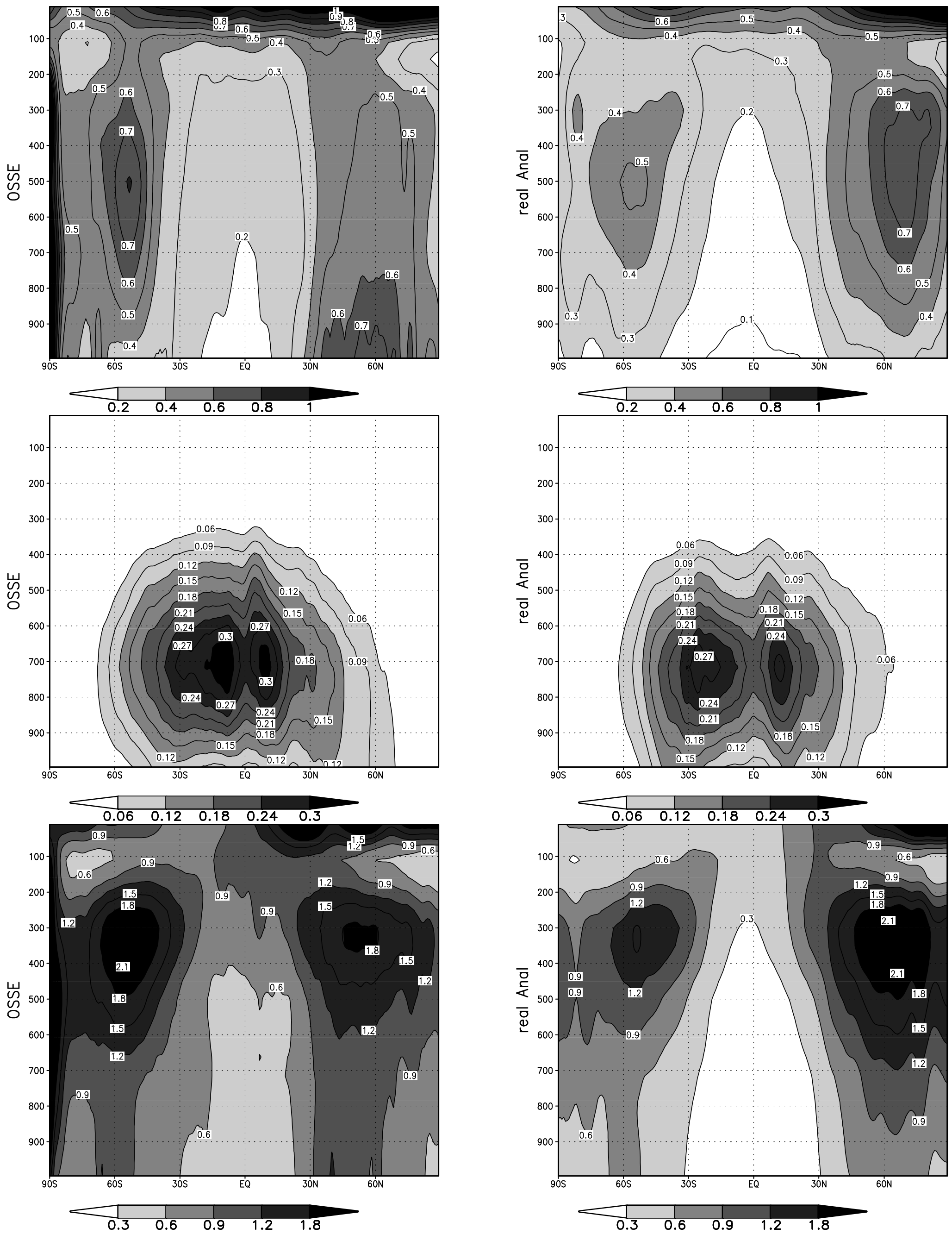

Figure 1: The square roots of zonal means of the temporal variances of analysis increments determined from the OSSE (left) and real analysis (right) for T (top), q (middle), and u (bottom). Units are K, g/kg, and ms^{-1}.

Here, additional validation is performed by comparing means and standard deviations of analysis increments produced by the OSSE with corresponding statistics produced by the real analysis. The square roots of zonal means of the temporal variances of analysis increments appear in Fig. 1 for both the OSSE and real analysis.

In the real analysis, the tropospheric statistics for T exhibit two local maxima associated with the two hemispheres' subtropical jets, with the northern hemisphere value being 30% larger than the southern hemisphere one. In the OSSE, there is a much larger local maximum near the South Pole. The maximum associated with the southern jet in the OSSE is more like that for the real analysis northern jet. The tropospheric maximum in the northern hemisphere for the OSSE occurs near the surface rather than near jet level.

For the q field, both the OSSE and real analysis show double maxima in the lower troposphere. In the northern hemisphere, the maxima in the two analyses are almost co–located, with the OSSE value only 10% larger than in the real analysis. In the southern hemisphere the OSSE

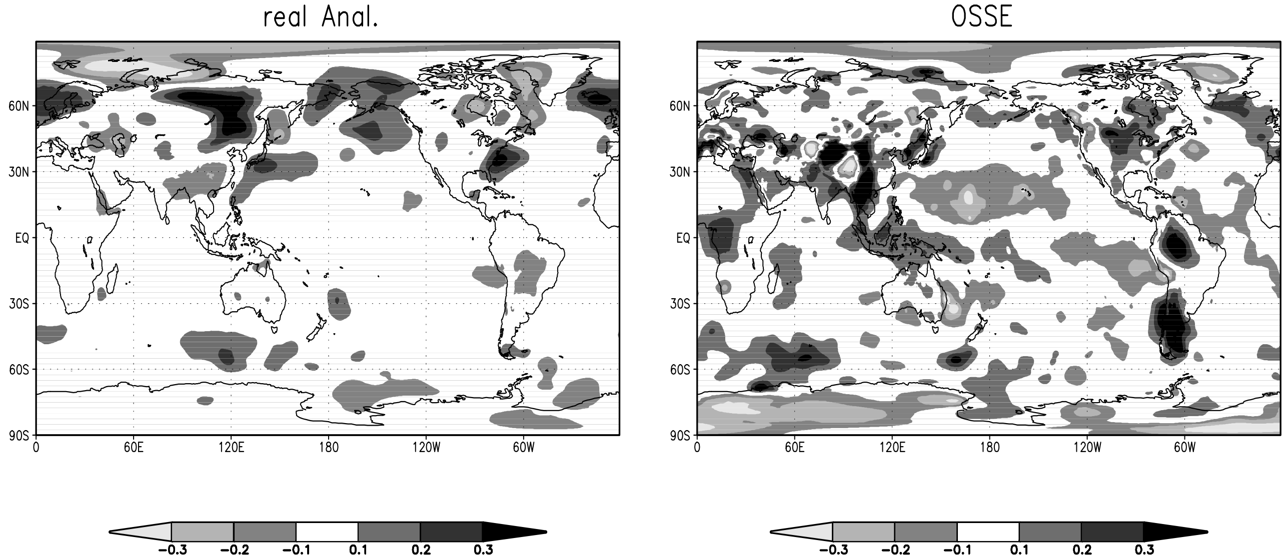

Figure 2: The time mean analysis increments of T on the $\sigma = 0.7$ surface for the OSSE (right) and real analysis (left). Units are K.

maximum is 18% greater and displaced 17° equatorward compared to the real case. This is the field for which the agreement between the two sets of increment statistics appears best.

The u increment statistics reveal the greatest differences between the two systems' results. Both systems show local maxima associated with the subtropical jets, but the dominant results are in opposite hemispheres. In many locations, values differ by factors of 1.5 or greater.

It would of course be preferable if the agreement between the two analysis increment statistics were much better. It is reasonable to expect that discrepancies as large as those obtained for these statistics (e.g., differences as large as 30%) may be associated with undesirably large discrepancies in analysis error statistics. The present results must therefore be considered a benchmark for future examinations rather than definitive in any sense.

As an example of corresponding time means of analysis increments, values for T on the $\sigma = 0.7$ surface appear in Fig. 2. Magnitudes for the two systems are similar. The OSSE suggests biases associated with most of the world's high mountain ranges, including the Andes, Himalayas, Alps, Antarctica, and Rocky Mountains. In contrast, the real analysis has no such indication and much less bias outside the Northern Hemisphere extra–tropics in general. This bias may be partly due to the different topography employed in the ECMWF and NCEP models.

Quantitative discrepancies between corresponding increment biases should be expected. In the real analysis system, such biases occur primarily due to systematic errors in observation instruments and operators (forward models) as well as in the NCEP model climatology. In the OSSE, while the simulated instrument errors are unbiased, the forward models used to generate the observations may be. Moreover, whereas in the real analysis, the critical bias is discrepancy between the NCEP and real atmosphere climatologies, in the OSSE it is between those of the NCEP and ECMWF models. There is no reason for these two discrepancies to be correlated with each other. That the biases are not much smaller in the OSSE suggests that the climatologies of the ECMWF and NCEP models are not more alike than either is with regard to the real atmosphere, which is of general concern for OSSEs (i.e., the worry that the models are more like each other than either is like the real atmosphere, and therefore that the OSSE is more like a perfect model, than real atmosphere, experiment).

5 Results

5.1 Analysis error variances

The square roots of zonally averaged temporal variances of analysis errors appear in Fig. 3. In general, values are two or more times larger than the corresponding statistics of the analysis increments in Fig. 1. This relationship can be expected when the assumed error variances (i.e., those specified in the data assimilation algorithm) for most observation types are larger than the assumed variances of the background errors. The actual OSSE–derived background error statistics corresponding to Fig. 3 are slightly larger than those of the analysis, as will be shown in the next section.

The structure of the field for T shown in Fig. 3 is very unlike that for the analysis increments. The statistic is largest near both poles near the surface, with a maximum of 5K. Throughout much of the tropical and Northern

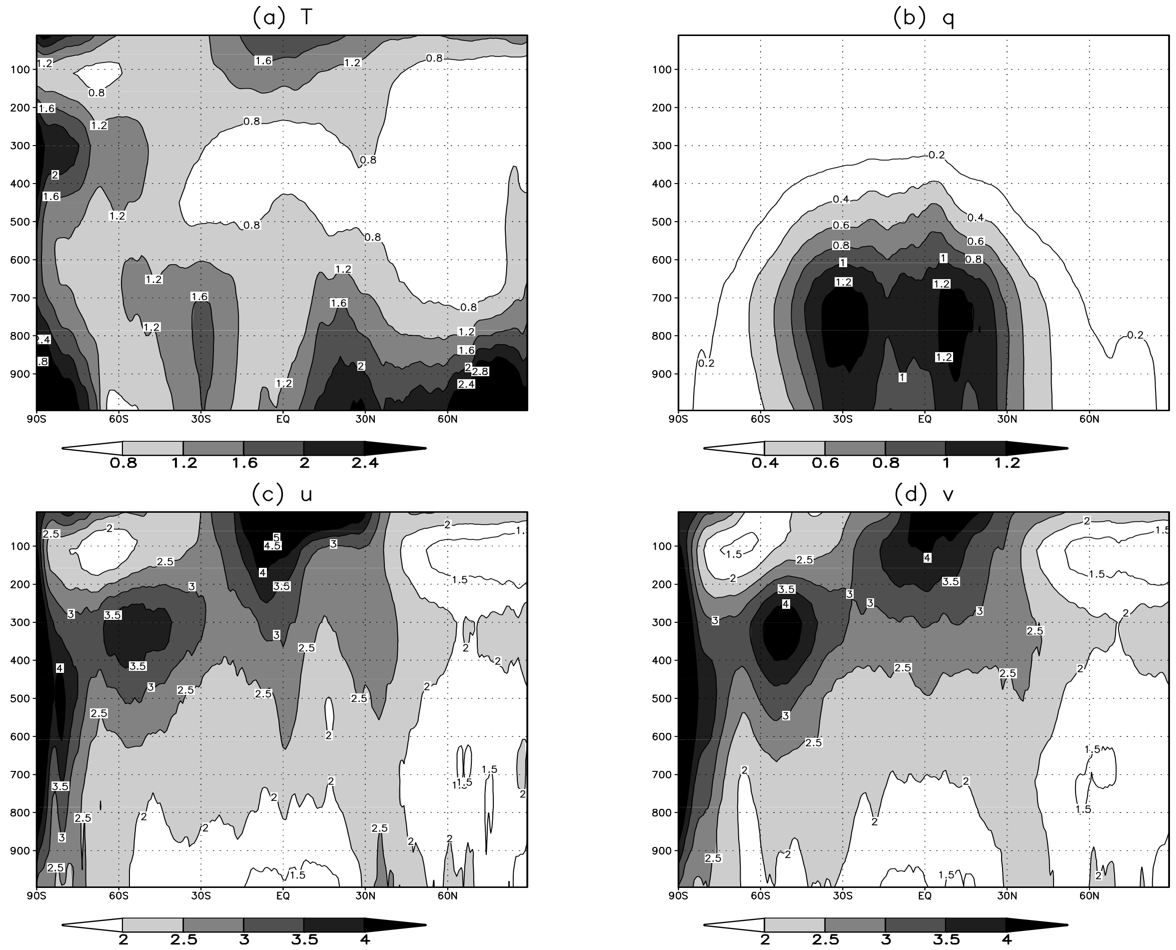

Figure 3: The square roots of zonally averaged temporal variances of analysis errors for (a) T (units K), (b) q (g/kg), (c) u (ms^{-1}), and (d) v (ms^{-1}).

Hemisphere troposphere, however, values are less than 1.5K, with a minimum near $\sigma = 0.4$.

The statistic for q in Fig. 3 has a structure similar to that of the analysis increments. Specifically, the values are largest near $\sigma = 0.7$, with equal peaks near 30°S and 10°N. The largest values shown are 1.3 g/kg, with values poleward of 45° below 0.8 g/kg at all levels.

The statistics for both wind components (Fig. 3) are very similar to each other, except for the large values for the u field near the equator above $\sigma = 0.1$. There are notable local maxima over Antarctica at $\sigma = 0.5$ and the southern hemisphere subtropical jet at $\sigma = 0.3$ (10 ms^{-1} and 4.5 ms^{-1}, respectively). Values in the extratropical Northern Hemisphere are less than 3 ms^{-1} at all levels.

As a representative horizontal plot of analysis error standard deviation, values for u on the $\sigma = .5$ surface are presented in Fig. 4. This reveals that the standard deviation over the North Pacific Ocean is as large (up to 6 ms^{-1}) as in the extratropical Southern Hemisphere. Maximum values over the North Atlantic Ocean

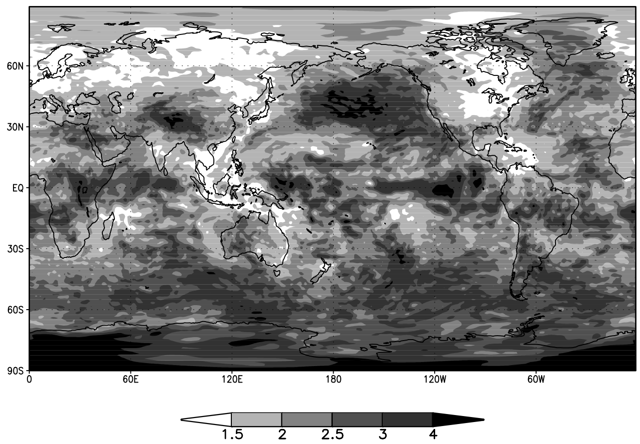

Figure 4: The analysis error standard deviation for u on the $\sigma = .5$ surface. Units are ms^{-1}.

are about half this. Values over North America and Eurasia are less than 2 ms^{-1}. Thus, there is significant longitudinal dependence and a contrast between more

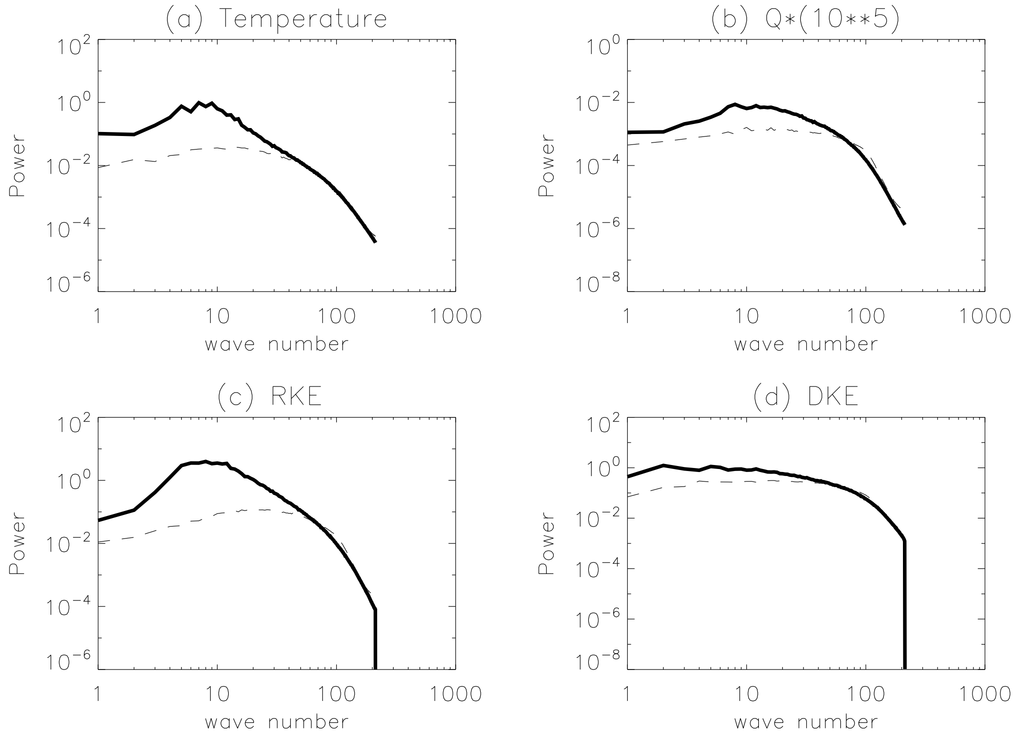

Figure 5: Variance spectra of (a) T and (b) q and power spectra for (c) RKE and (d) DKE, all on the $\sigma = 0.7$ surface. The time mean fields have been removed prior to the spectral analysis. The solid line indicates results for the ECMWF nature run; dashed indicates the analysis error. Ordinates have units (a) K^2, (b) $(g/kg)^2$, and (c,d) m^2s^{-2}.

poorly–observed ocean and well–observed land regions.

Selected horizontal power spectra of the ECMWF dataset and corresponding OSSE analysis error (with powers of the time mean fields removed) are presented in Fig. 5. These are for T, q, and wind (separated into rotational and divergent components) fields on the $\sigma = 0.7$ surface. The results are qualitatively representative of those at other levels in the troposphere above the planetary boundary layer.

First to note is that for all the fields, at some set of wavenumbers $n > n_c$, the variance (or energy) of the analysis error field is equal to or greater than that of the ECMWF simulated atmosphere. For all the fields, $n_c < 70$. For $n > n_c$, the error variance is 100% or more of the analysis field itself, although the analysis is performed at a resolution of T170.

If corresponding temporal variances of the analysis and ECMWF model are identical for some range of wavenumbers but the complex amplitudes of the two sets of spectral coefficients are temporally uncorrelated, then the variance of the analysis error will be twice that of the ECMWF analysis itself. This would be the approximate result, for example, if the moduli of the corresponding spherical harmonic amplitudes were similar constants but their phases had random differences between $\pm\pi$ drawn from a uniform distribution. It would also be the approximate result from computing variance by averaging over all days in a particular month over several years compared with variances of differences between randomly drawn days from different years within that dataset. In this case, the difference can be termed "saturated"in the sense that it results from uncorrelated information and can only be larger if the coefficients differenced are negatively correlated. There are only a few scales for which the analysis error is saturated in Fig. 5. Nonetheless, relative errors of 100% should be considered as very large.

The spectra of the analysis error variances peak at about twice the wavenumber of the peak for the corresponding ECMWF spectra. For most fields within the troposphere above the planetary boundary layer, this peak in error variance is near $n = 20$. The analysis error variances at any particular wavenumber for all fields and levels are not less than 1% of the corresponding ECMWF variances.

For spectra of 2–dimensional fields plotted in terms of a scalar 2–dimensional wavenumber with logarithmic ordinate and abscissa axes, white noise is indicated by a line with slope 1. This is the slope of the analysis error curve for the RKE power spectrum in the range $2 < n < 15$. For the other fields within this range of n, the slope is shallower, although still increasing with n. For a white–noise wind spectrum, strict geostrophy would imply a slope of -1 for the variance of geopotential. Note that this is unlike the variance of T (which with p_s determines geopotential) obtained here for large scales.

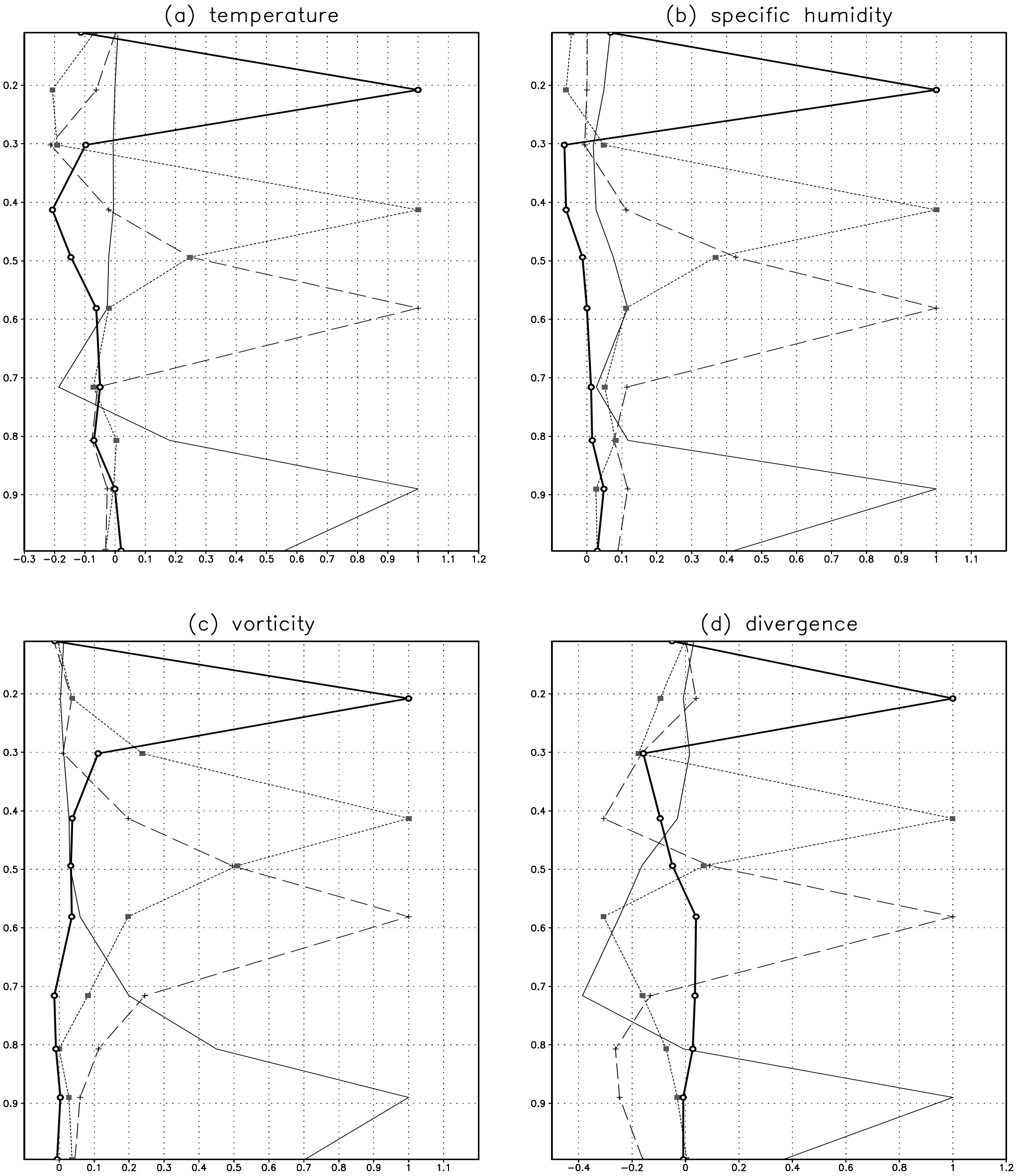

Figure 6: Vertical correlations determined from covariances averaged over the North Pacific Ocean for (a) T, (b) q, (c) ζ, and (d) δ. Separate lines are presented for the correlations between fields on the common grid surfaces near $\sigma = 0.2, 0.4, 0.6.0.9$ and all the other levels. Each line is to be identified by noting the level at which the correlation is 1, where the level is correlated with itself.

For $15 < n < 70$, the spectrum of rotational kinetic energy for the nature run is approximately proportional to n^{-3}, indicative of an inertial range of quasi–geostrophic turbulence (CHARNEY 1971). For $n > 100$, however, the dependence is more like n^{-8}. At much smaller scales than resolved here, where the turbulence becomes fully 3-dimensional rather than quasi–geostrophic, the dependence should be more like $n^{-5/3}$. This suggest that this particular ECMWF model may be over-damped. If true, not only is the variance at small scales grossly weak compared to the atmosphere, but the upward–scale transfer of error variance may be inhibited (LEITH AND KRAICHNAN 1972).

Vertical correlations of analysis error were also examined for four different geographical regions: the tropics, North America, North Pacific Ocean, and extratropical Southern Hemisphere. These were all computed from the vertical level covariances averaged for each region. Since only 12, approximately equally spaced, σ–levels of data were considered for this calculation, subtle changes in correlation could not be examined. At the level of detail available, correlations were very similar to those of the background errors; i.e., the analysis did not appear to appreciably alter the vertical scale of the errors compared with those of the background.

As a sample of the correlations examined, values averaged over the North Pacific Ocean (20°N–60°N and 150°E–130°W) for T, q, ζ, and δ appear in Fig. 6.

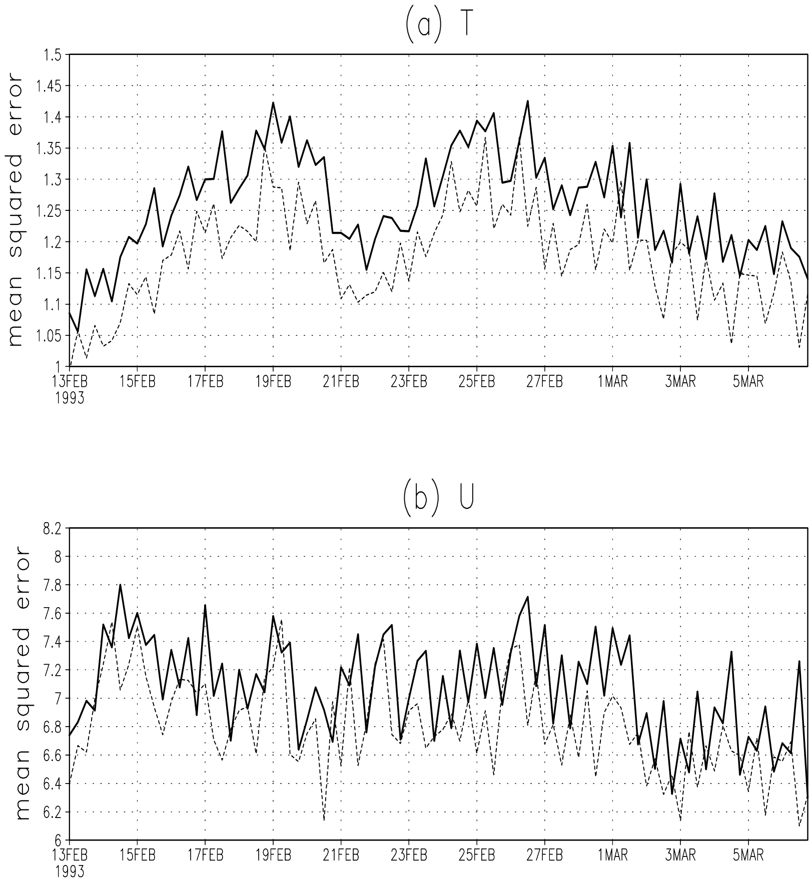

Figure 7: Time series of the global mean squared background (solid) and analysis (dashed) errors for (a) T and (b) u on the $\sigma = 0.5$ surface. Units are (a) K^2 for (b) and m^2s^{-2} and the abscissa is time starting from the beginning of the examination period, with 4 times per day plotted.

These show correlations of analysis errors between all σ–surfaces considered and each of the surfaces $\sigma = .2, .4, .6, .9$. Correlations for the u and v fields (not shown) are very similar to those for ζ.

The correlation between any error field at any levels σ_1 and σ_2 generally decreases with increasing $\Delta\sigma = |\sigma_1 - \sigma_2|$ except for the horizontal wind divergence and temperature which have significant ranges of negative correlations before asymptoting to zero at larger $\Delta\sigma$. The $\Delta\sigma$ corresponding to a correlation of 0.5, which defines a correlation "length" in terms of σ, generally decreases as correlations are considered between levels higher in the atmosphere, suggesting that correlation lengths prescribed in terms of height z or $\log p$ may be more constant with elevation. Vertical correlation lengths appear similar in the four regions examined except over North America, where the correlation lengths near the surface are 50% or less than those measured elsewhere for all the fields, and in the mid troposphere tropics for the rotational wind field, where the correlation length again is dramatically shorter. Further examination of these correlation functions, including their geographical inhomogeneity will await the availability of longer, future OSSE data sets.

5.2 Kalman gain

Time series of the global mean variances of both background and analysis errors for both T and u on the $\sigma = 0.5$ surface appear in Fig. 7. Values range between 1–1.3 K^2 and 6.1–7.8 m^2s^{-2}, respectively. For T at almost all times, the analysis value is less than the background value. The same is true for q (not shown), for which the two time–averaged variances are 0.29 and 0.26 g/kg, respectively. For both v (not shown) and u, there are many times at which the global–mean analysis error variance equals or exceeds that for the background.

Notable in Fig. 7 are the variations with a 12–hour period. This component of variability tends to have a minimum at hours 0 or 12 UTC for the analysis error variance and a maximum at those times for the background error variance. This 180 degree phase difference in the most rapid variations of the two variances results in much larger differences between them at 0 and 12 UTC compared with 6 and 18 UTC, especially in the case of u and v. This result is simply explained by the dearth of rawindsonde observations at 6 and 18 UTC.

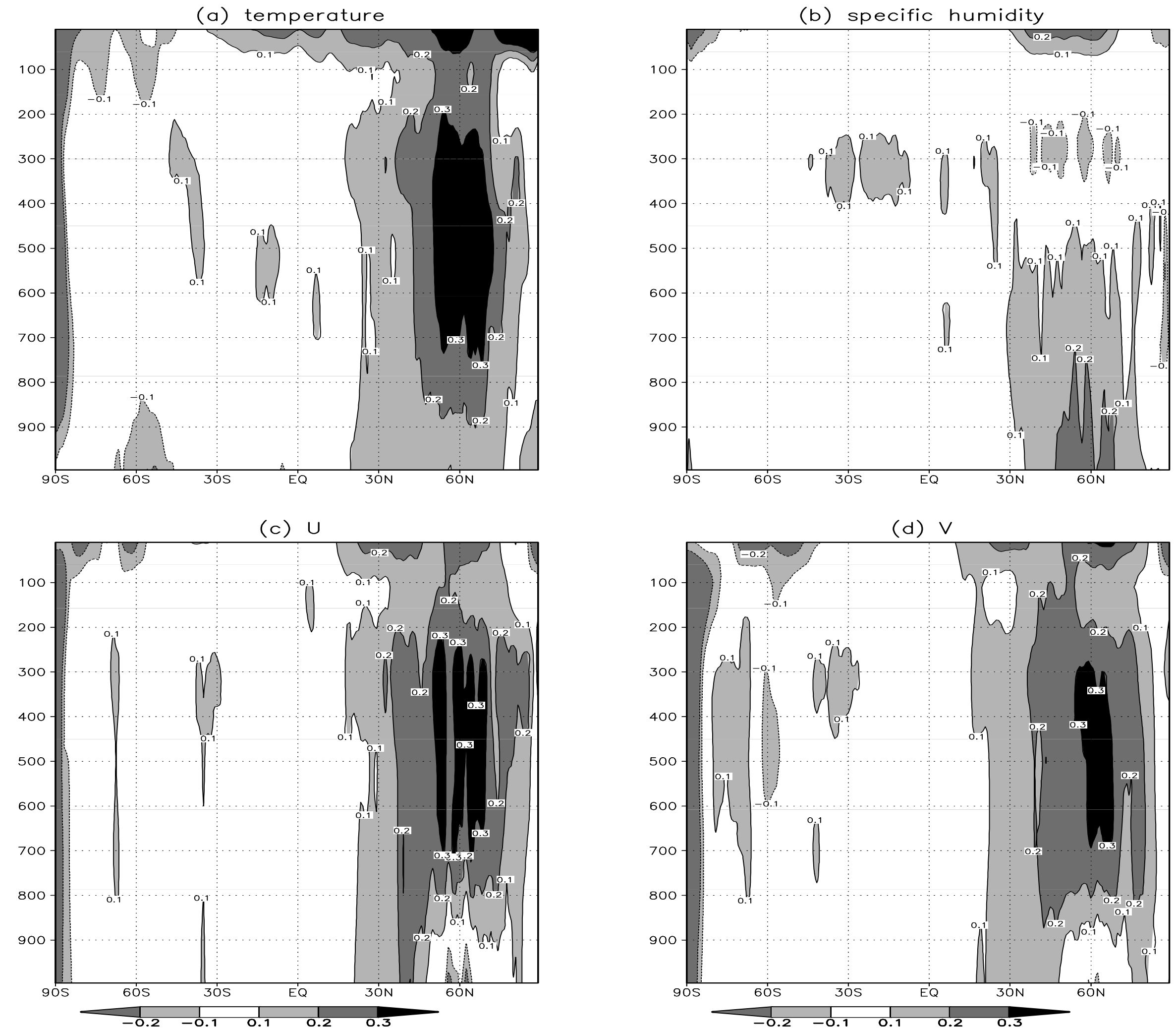

Figure 8: Zonal averages of Kalman gains for (a) T, (b) q, (c) u and (d) v computed for fields at 0 and 12 UTC only.

The analysis thus has more wind observations to reduce its errors at hours 0 and 12 UTC. For this reason, the subsequent background errors at 6 and 18 UTC also tend to be reduced compared with those at the other hours; i.e., the 6 and 18 UTC 6–hour forecasts begun from 0 and 12 UTC, respectively, are statistically better than those at 0 and 12 UTC begun from 18 and 6 UTC.

Zonal averages of Kalman gains for T, q, u and v appear in Fig. 8 where the variances are computed for only 0 and 12 UTC errors. Note that values are larger than 0.4 in the relatively well observed, Northern Hemisphere extratropics in the upper troposphere, except for the q field. In the same region except near the surface, the zonal mean K for the winds is negative. The same is true for q above $\sigma = 0.4$ in this same region, near the South Pole for winds and temperature, and for T and u near 60° S, with some mean $K < -0.2$. These negative values mean that the analysis has generally made the errors worse than the background errors in these locations. When errors at 6 and 18 UTC are also considered in determining the variances, the maximum zonal mean Kalman gain is reduced by a factor of 2 for all the fields and negative values appear in more locations. This reduction of Kalman gain has been explained in conjunction with Fig. 7.

Kalman gains have routinely been estimated within operational data assimilation systems. These are either produced as by–products of the analysis algorithm (e.g., from estimates of the Hessian matrix in variational procedures) or from comparison of differences between observations and both analysis and background information. Interpretations of these estimates as the true Kalman gain are clouded by the ubiquitous assumptions and tuning practices in the assimilation techniques. For example, a technique can be tuned to very closely fit the analysis to observations, inflating the Kalman gain estimated using those observations unless those observations and their representation (forward models) are themselves error free. Estimates produced using the Hessian are valid only if all the statistics and other ex-

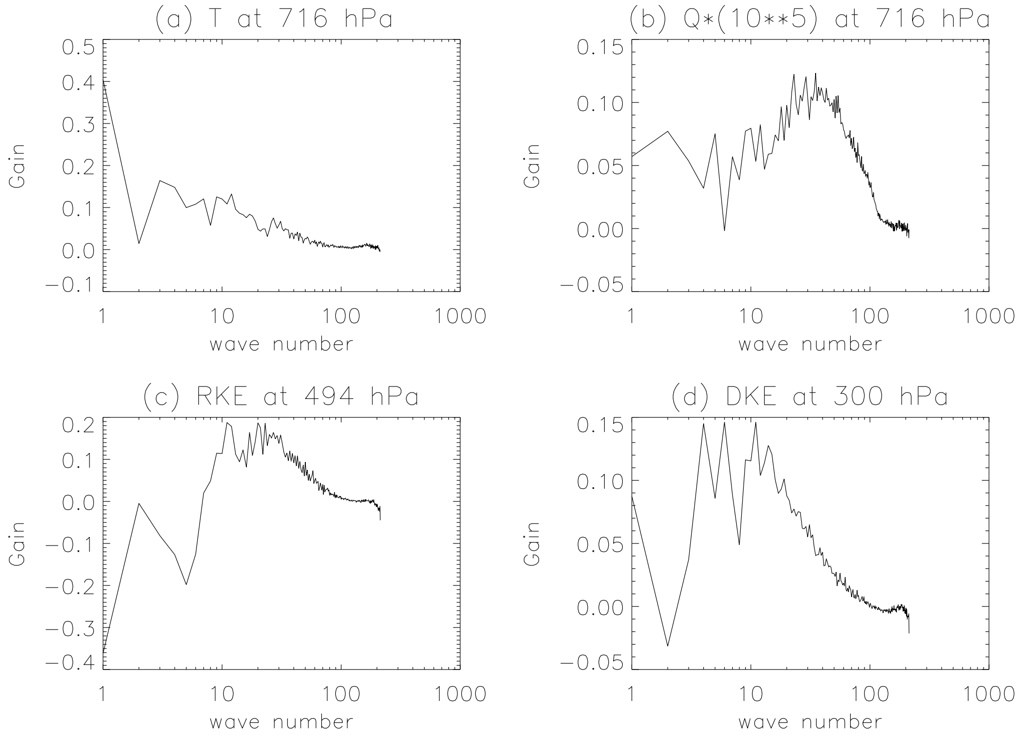

Figure 9: Zonal averages of Kalman gains for (a) T, (b) q, (c) u and (d) v computed for fields at 0 and 12 UTC only.

plicit or implicit assumptions or approximations employed by the assimilation technique are indeed accurate. With these caveats, we note that mean Kalman gains of near 0.4 have been informally reported for some current assimilation systems.

The Kalman gain of the variances of T and q (at $\sigma = 0.7$), power spectra of RKE (at $\sigma = 0.5$), and DKE (at $\sigma = 0.3$) appear in Fig. 9. Errors at only 0 and 12 UTC were considered for the computation. The σ–surfaces presented were chosen for their relevance to the particular fields, but the results described are typical of those at other σ–levels except near the surface and above the tropopause. For T, except for the statistically uncertain largest scales (i.e., those for very small n), the Kalman gain generally decreases until near $n = 100$ where the mean gain asymptotes near 0. For the other presented fields, the maximum K occurs near $n = 20$, but the gain near $n = 100$ has dropped to near 0. This indicates that the analysis is improving the q and wind fields most at length scales (computed as a wavelength divided by 2π) of approximately 300 km.

5.3 Dynamic balance

The partitioning of the "energy" of the analysis error between rotational (R) and gravitational modes (G) is presented in Table 1 as a function of selected vertical mode indexes k for a single date at 12 UTC. The corresponding H_k and fraction of error energy in gravitational components ($f_g = G/(R+G)$) are also presented. Note that although $H_9 = 49$m for the shallowest mode examined, this mode has only 3 zero crossings throughout the troposphere and therefore its small depth is no indication of the spatial distance between its nodes in the vertical. The modal partitioning was also performed for several other individual days and times of day, yielding results qualitatively similar to those in Table 1.

Table 1. The energy of the analysis error at one simulation time for gravitational (G) or rotational (R) normal modes and the fraction f_g of gravitational mode error energy for each of the first nine vertical modes k and corresponding equivalent depths H_k. Units are indicated.

k	1	2	3	4	5	6	7	8	9
H(m)	10943	4444	1538	628	311	175	109	71	49
G(J/kg)	.18	.16	.22	.32	.31	.32	.29	.28	.25
R(J/kg)	.82	.47	.38	.51	.52	.58	.56	.45	.33
f_g	.18	.25	.37	.39	.37	.36	.34	.38	.43

The smallest fraction of unbalanced error energy is for the external mode ($k = 1$). This reflects primarily the degree of balance between increments of the surface pressure gradient force, the gradient of the surface geopotential and the Coriolis force acting on the barotropic wind. This component of the error is primarily (linearly) balanced, but to a much less degree than the full analysis or nature fields for $k = 1$ for which $f < .02$.

Internal mode increments exhibit a uniformly larger degree of relative imbalance (i.e., larger values of f_g)

than the external mode does. For $k \leq 9$, linearly balanced increments still dominate unbalanced ones. This is similar to the dominance of balanced components of the background error, which have similar values of f_g to those reported here. Since vertical modes contribute to the energy separately, their variances can be added. For the sums of R and G in table 1, the fraction of total unbalanced energy is 0.33. Since R and G are quadratic in mode amplitudes, the ratio of the amplitudes of the modes is more like $\sqrt{G/R}$. For $k = 9$ this value is large as 0.87. For the internal modes therefore, although the balanced components of the error are on the average larger than the unbalanced, they do not strongly dominate.

6 Conclusions

Validation of the NCEP OSSE by comparing its analysis increment variances with those of a real analysis system reveals enough similarity (e.g., overall magnitudes of values) that a cautious attempt at estimating some gross characteristics of analysis error can be made. Discrepancies between the two sets of variances are significant in many aspects, however, such as the shift in jet maxima wind increments from one hemisphere to the other. Since the behaviors of present data assimilation systems are quite complex and often nonintuitive (BAKER 2000) it remains unclear how improvements to the OSSE will alter any of the results that have been described here. Also, the steeper than inertial range slope of the nature–run power spectra of the wind suggests that the ECMWF model may be over–damped compared to the real atmosphere. This can diminish the dynamical cascade of error variance from more poorly observed small spatial scales to larger scales (LEITH AND KRAICHNAN 1972, LORENZ 1969) and thus affect the equilibrium Kalman gain (DALEY AND MENARD 1993). It is for these reasons and because this is the first reported use of an OSSE to characterize analysis error that the results presented here should be considered only a first cut of what can be examined using such data.

Background errors in the OSSE were examined in all the ways analysis error were. Few of these results were presented because publication space is limited and the analysis error statistics are more novel to present. Most background error statistics examined were qualitatively very similar to the corresponding analysis ones. In most regions, were the Kalman gain is small, even the quantitative results were very similar. Thus additional presentation of most background error statistics would be redundant.

Many data assimilation systems, including the one used for this study, use assumed background error variances that do not vary longitudinally. In contrast, variations of temporal variances of analysis error that have been determined here are almost as large along latitude 45°N as along the 90°W meridian. In particular, the large variances obtained over the mid North Pacific Ocean and the oceanic, extra–tropical Southern Hemisphere have similar values, presumably due to the lack of in–situ observations in both regions. The OSSE background error variances portray similar geographic variations. Indeed, the Kalman gains in the more poorly observed regions are generally small (i.e., < 0.1) so the values of the presented analysis error variances are quantitatively very similar to the background values. Thus an assumption of longitudinally independent error variances is a poor one in the Northern Hemisphere, with values instead ranging over a factor of 2 between data rich and poor regions.

Kalman gains are the greatest (> 0.4) at 0 and 12 UTC over regions well observed by rawindsondes. The gains at other times or over other regions are less than half of those values. This indicates that rawindsondes remain a critical, or even dominant, observation system in this OSSE. It is possible that the simulation of satellite observations or their errors may be lacking in some important way that diminishes their measured utility in the OSSE. Also, the satellite data in 1993 were somewhat poorer quality and less abundant than what are available today. The variation of gains, and the geographical dependence of the error variances, may therefore particularly change with an upgraded OSSE.

For both the total and rotational wind fields, power spectra of analysis error peak near $n = 20$. At larger scales the power spectra for rotational wind errors are white. Near $n > 70$ for most fields at vertical levels ($0.9 > \sigma > 0.1$), the power spectra of analysis error variance is approximately the same as that of the corresponding nature run field. Thus, although the OSSE has been performed at T170 resolution, for most of the higher wavenumbers there is little accurate information.

Since data assimilation is a blending of background fields having spatially correlated errors and observations having generally assumed less correlated errors, analysis errors may tend to have spatial correlations between those. Here, however, both vertical and horizontal correlation lengths of analysis and background error fields were generally similar. When they were measurably different, e.g., horizontally for the rotational wind field and vertically over some regions, the correlation lengths were just slightly shorter for the analysis errors. Also, both types of error had similar degrees of imbalances, as measured by projections of errors onto normal modes and measured in terms of an energy metric. Error energy in the rotational (approximately linearly balanced) modes is greater than in the gravitational modes at all the equivalent depths examined. The balance, however, is not overwhelming.

The assertion that any of these results characterize real analysis error rests on the quantitative validity of the OSSE. Here, these results reflect the total effect of the observation system on the assimilation errors. Any single observation type, whether existing or envisioned, would have a much smaller impact. OSSEs designed for quantitative instrument evaluation therefore have even more stringent validation requirements than do OSSEs for the purposes here. Although performing sufficiently validated OSSEs remains a difficult task, it is hard to obtain a better source of quantitative information for many data assimilation assessments, and therefore the community should be encouraged to commit to the great diligence required to conduct even more valuable OSSEs in the future. Characterization of analysis error should be one of the foci of any such cooperative endeavor.

Acknowledgements

This work was partially supported by the U.S. Navy Office of Naval Research grant number N00014–0310554, the U.S. National Science Foundation grant number ATM–0351938, and the Atmospheric Data Assimilation Development component of the NASA Modeling, Analysis and Prediction program grant MAP/04–0000–0080.

References

BAKER, N. L., 2000: Observation adjoint sensitivity and the adaptive observation–targeting problem. –Ph. D. Thesis, Naval Post–Graduate School, 265 pp.

CHARNEY, J. G., 1971: Geostrophic turbulence. –J. Atmos. Sci. **2**8, 1087–1095.

DALEY, R., 1991: Atmospheric Data Analysis, –Cambridge University Press, 420 pp.

DALEY, R., R. MÉNARD, 1993: Spectral characteristics of Kalman filter systems for atmospheric data assimilation. –Mon. Weather Rev. **1**21, 1554–1565.

ERRICO, R. M., 1987: A description of software for determination of normal modes of the NCAR Community Climate Model. –NCAR Technical Note, NCAR/TN–217+STR, 86 pp.

ERRICO, R. M., 2000: Interpretations of the total energy and rotational energy norms applied to determination of singular vectors. –Q. J. R. Meteorol. Soc. **126A**, 1581–1599.

ERRICO, R. M., R. LANGLAND, 1999: Notes on the appropriateness of "bred modes" for generating initial perturbations used in ensemble prediction. –Tellus **5**1A, 431–441.

GIBSON, J.K., P. KALLBERG, S. UPPALA, A. NOMURA, A. HERNANDEZ, E. SERRANO, 1997: ERA description. –In ECMWF ERA-15 Project Report Series, No. 1. European Centre for Medium-Range Weather forecasts, Shinfield, Reading, UK.

KALNAY, E., M. KANAMITSU, R. KISTLER, W. COLLINS, D. DEAVEN, L. GANDIN, M. IREDELL, S. SAHA, G. WHITE, J. WOOLLEN, Y. ZHU, A. LEETMAA, B. REYNOLDS, M. CHELLIAH, W. EBISUZAKI, W. HIGGINS, J. JANOWIAK, K.C. MO, C. ROPELEWSKI, J. WANG, R. JENNE, D. JOSEPH, 1996: The NCEP/NCAR 40–year reanalysis project. –Bull. Am. Meteor. Soc. **7**7, 437–471.

LEITH, C.E., R.H. KRAICHNAN, 1972: Predictability of turbulent flows. –J. Atmos. Sci. **2**9, 1041–1057.

LORENZ, E. N., 1969: The predictability of a flow which possesses many scales of motion. –Tellus **2**1, 289–307.

MASUTANI, M., J.S. WOOLLEN, S.J. LORD, T.J. KLEESPIES, G.D. EMMITT, H. SUN, S.A. WOOD, S. GRECO, J. TERRY, R. TREADON, K.A. CAMPANA, 2006: Observing system simulation experiments at NCEP. –National Centers for Environmental Prediction Office Note 451, 34 pp.

PARISH, D.F., J.C. DERBER, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. –Mon. Weather Rev. **120**, 1747–1763.

TIBALDI, S., 1986: Envelope orography and maintenance of the quasi-stationary circulation in the ECMWF model. Advances in Geophysics: Anomalous Atmospheric Flows and Blocking **29**, –Academic Press, 339–374.