Diagnosis of systematic analysis increments by using normal modes

N. Žagar,a* J. Tribbia,b J. L. Anderson,b K. Raederb and D. T. Kleistc†

aUniversity of Ljubljana, Ljubljana, Slovenia
bNational Center for Atmospheric Research‡, Boulder, Colorado, USA
‡National Centers for Environmental Prediction, Camp Springs, Maryland, USA

*Correspondence to: Nedjeljka Žagar, Department of Physics, Faculty of Mathematics and Physics, Jadranska 19, SI-1000 Ljubljana, Slovenia. E-mail: nedjeljka.zagar@gmail.com
†The contribution of Kleist was prepared as part of their official duties as a United States Federal Government employee.
‡The National Center for Atmospheric Research is sponsored by the National Science Foundation.

This paper applies the normal-mode functions for the three-dimensional diagnosis of systematic analysis increments in the operational systems of ECMWF and NCEP, in the NCEP/NCAR reanalyses and in the ensemble data assimilation system DART/CAM which is developed at NCAR. Non-zero systematic increments are interpreted as the analysis system bias. The main region of tropospheric biases in all systems is the Tropics; most of the large-scale tropical bias resides in the unbalanced (inertio-gravity) motion with the eastward-propagating component being dominant in some datasets. The magnitudes of tropospheric wind-field biases in July 2007 were in the range between 1 m s\(^{-1}\) and 2 m s\(^{-1}\), illustrating the importance of diagnosing analysis systems in the Tropics where magnitudes of the large-scale variability are of the same order. The systematic increments integrated over the whole models’ atmosphere appear predominantly balanced; this balance is associated with biases in the zonally averaged balanced state and, in case of ECMWF and NCEP, with the biases at the model top levels. Spectra of analysis increments averaged over one month show that, on average, 20% to 45% of total energy in increment fields belongs to the inertio-gravity motions. Copyright c⃝ 2010 Royal Meteorological Society

Key Words: Hough functions; bias; Rossby waves; inertio-gravity waves

Received 31 May 2009; Revised 7 September 2009; Accepted 30 September 2009; Published online in Wiley InterScience 19 January 2010


1. Introduction

Eigensolutions of the linearized primitive equations, known as normal modes, have been used for almost forty years. As Dickinson and Williamson (1972) pointed out, normal modes are useful tool for addressing the problem of initialization of numerical weather prediction (NWP) models, for the identification of model modes which have significance for climate simulations and for comparison of amplitudes of model modes with those observed in the real atmosphere. The first of the listed applications has been extensively pursued through the development of the nonlinear normal-mode initialization (NNMI) (Errico, 1997).

Global horizontal structures of normal modes, known as Hough functions, have been used to identify the large-scale structure of some of the leading balanced (quasi-rotational or Rossby type) modes (e.g. Hirooka, 2000; Madden, 2007). However the main application of normal modes with respect to observations has been in the Tropics; here,
2. The datasets and methodology

2.1. Diagnosis of biases in terms of normal modes

The set of normal modes used in this study was derived by Kasahara and Puri (1981). The attractive feature of their derivation is that the three-dimensional modes are orthogonal; therefore, the energy in each mode can be quantified. The application of this function set was recently revived by Žagar et al. (2009a). The same analysis datasets are used in the present paper together with their background fields; on the output, the structure of analysis increments is represented in the space spanned by a zonal wavenumber, a meridional mode and a vertical eigenstructure.

In the discrete formulation, the three-dimensional global data vector \( \mathbf{U} = (u, v, g^{-1}P) \) is represented by the following finite series:

\[
\begin{bmatrix}
  u \left( \lambda, \theta, \sigma \right) \\
  v \left( \lambda, \theta, \sigma \right) \\
  g^{-1}P \left( \lambda, \theta, \sigma \right)
\end{bmatrix}
= \sum_{m=1}^{N_m} \sum_{n=1}^{N_n} \sum_{k=1}^{N_k} \sum_{p=1}^{N_p} \chi_{m,n,k,p} S_m \mathbf{H}^{m,n,k,p} (\sigma) e^{i2\pi} \Pi_m(\sigma).
\]  

Equation (1) describes the global motions, described by the wind components \( u \) and \( v \) and the mass field, represented by a modified geopotential variable \( P = \Phi + RT \ln(p) \), in terms of inertia-gravity and Rossby waves on the sphere.

The definition of the \( P \) variable involves the geopotential, \( \Phi \), the gas constant \( R \), the surface pressure field \( p_s \) and the globally averaged temperature on horizontal model levels, \( T_e \).

The four indices in (1), \( k, n, m, p \), constitute the four-component modal index \( v = (k, m, n, p) \) and they represent the zonal wavenumber, the meridional mode index, the vertical mode and the wave type, respectively. The longitude, the latitude and the vertical model level indices are \( \lambda, \theta \), and \( \sigma (\sigma = p/p_s) \), respectively. The truncation indices \( N_k \) and \( N_p \) correspond, respectively, to the number of waves along a latitude circle and a maximal number of meridional modes. The vertical eigenfunction is denoted by \( \Pi_m(\sigma) \) and the number of vertical eigenmodes is \( N_m \).

Three values of \( p \) correspond to the eastward-propagating IG modes (EIG), the westward-propagating IG modes (WIG) and the balanced motion (ROT). The symbol \( \mathbf{H}^{m,n,k,p}_{\sigma,\lambda,\theta} \) in (1) stands for the meridional Hough structure functions while the symbol \( S_m \) denotes the \( 3 \times 3 \) diagonal matrix with elements \( \{ (gH_{eq}) / \sigma \}^{1/2}, \{ (gH_{eq}) / \sigma \}^{1/2}, (H_{eq}) \} \) which removes dimensions from the input data vector after the vertical projection. The constant \( H_{eq} \), the 'equivalent depth', is a property of each vertical mode and it couples the horizontal and vertical motions. For further details on the derivation of NMFs the reader is referred to Žagar et al. (2009a) and Kasahara and Puri (1981).

The forward expansion is performed in two steps. The vertical projection is performed first followed by the normalization involving the \( S_m \) matrix. Afterwards, the Hough expansion coefficients \( \chi_{m,n} \) are computed according to the following (2), where the superscript \( * \) denotes complex
conjugation.

\[ \chi_v = \frac{1}{2 \pi} \int_0^{2\pi} \int_0^1 (u_m, v_m, l^{-1} P_m)^T (H_m)^* e^{-i k \mu} \, dl \, d\mu. \]  

(2)

Here, \( \mu = \sin(\theta) \). The zonal expansion is performed by using the fast Fourier transform while the Gaussian quadrature approximates the integration in the meridional direction. The coefficients \( \chi_v \) are non-dimensional. Multiplying them by \( (gH_m)^{1/2} \) gives the total energy in mode \( v \), \( E_v = gH_m(m \chi_v (\chi_v)^*) \), in units \( m^2 s^{-2} \) or \( J kg^{-1} \).

Of our special interest is the Kelvin mode, \( v = (k, 0, m, 1) \), the most energetic IG mode with substantial effects on the vertical momentum transport in the Tropics; Zagar et al. (2009b) show how their vertical propagation can be diagnosed in the normal-mode space. In the normal-mode representation developed by Kasahara (1976, 1978) which we follow here, the MRG mode has been included as the \( n = 0 \) ROT mode, \( v = (k, 0, m, 3) \). On the other hand, the MRG mode contains significant divergence; it is most often studied in the form of a special solution of the shallow-water equations on the equatorial \( \beta \)-plane, in addition to other IG and equatorially trapped balanced modes (e.g. Wheeler and Kiladis, 1999). We shall thus always refer to IG+MRG modes when both are included and to IG modes when EIG and WIG are summed together. ROT will refer to all \( p = 3 \) modes without \( v = (0, 0, m, 3) \) mode which is a constant needed to have the expansion complete (Kasahara, 1978).

The expansion (1) is applied separately to the first-guess and analysis fields every six hours during July 2007, for a total of \( N_{\text{smpl}} = 124 \) samples. The difference field between the analysis and first-guess fields in the modal space is denoted by \( \Delta \chi_v \), and it constitutes the main input information for our diagnosis. The time-averaged analysis increment in mode \( v \) is

\[ \Delta \chi_v = \frac{1}{N_{\text{smpl}}} \sum_{t=1}^{N_{\text{smpl}}} \left[ \chi_v^{\text{an}}(t) - \chi_v^{\text{bg}}(t) \right]. \]  

(3)

The superscripts 'an' and 'bg' denote analysis and background fields, respectively. Equation (3) formally defines analysis system bias and the majority of our results will be presented in terms of the \( \Delta \chi_v \) variable at various model levels or summed across a single modal index.

The averaged increment field transformed from the modal space back to the physical space should ideally correspond to the averaged analysis increments in the model space; the degree of agreement depends on the accuracy of (1). The accuracy of the expansion (1) is discussed in detail in Zagar et al. (2009a) as a function of the truncation parameters \( N_k \) and \( N_m \); these are tuned so that (1) represents as much of the input data variance as possible at each model level.

Another quantity presented is the time-averaged energy difference between the analysis and background fields in the modal space defined as

\[ \Delta E_v = \sum_x gH_m \left[ \chi_v^{\text{an}} \left( \chi_v^{\text{an}} \right)^* - \chi_v^{\text{bg}} \left( \chi_v^{\text{bg}} \right)^* \right]. \]  

(4)

where \( x \) denotes any of \( (k, n, m) \) and \( (\ldots) \) represents the time averaging. The \( \Delta E_v \) represents the average tendency of the assimilation system to place or subtract the energy to/from a particular mode.

The energy distribution in increment fields at each analysis step is obtained as

\[ E_v^n(t) = gH_m \left[ \chi_v^{\text{an}}(t) - \chi_v^{\text{bg}}(t) \right]. \]  

(5)

and we shall discuss the time-averaged spectra of this quantity. Derived spectra should ideally correspond to spectra which would be obtained if the difference between analysis and background was taken in the physical space at each time step and the average of \( N_{\text{smpl}} \) increment fields transformed from physical to the modal space.

A snapshot of increment fields is provided in Figure 1 which also illustrates the accuracy of the expansion (1). We choose to present the ECMWF model level 10 (about 1.3 hPa), at which significant increments were present at all times. First of all, the comparison of Figures 1(a) and (b), which present the difference between input analysis and background fields and the result of inversion of \( \Delta \chi_v \) respectively, verifies that the projection (1) accounts for nearly all variance in the input data and, at the same time, that the projection acts as a smoother and filter of the smallest scales. The other two panels illustrate the main advantage of using normal-mode functions by showing the wind and geopotential increments split into contributions from balanced (ROT, Figure 1(c)) and unbalanced (IG+MRG, Figure 1(d)) motions. It can be seen that inertia-gravity motions at 1200 UTC on 20 July 2007 constituted a significant portion of increment fields at this level, especially in high latitudes. A majority of unbalanced flow is contributed by the WIG modes, especially over Antarctica (not shown). Since our projection represents the input data well, we can systematically perform all statistics in the modal space and derive the spectra (5) and other quantities of interest directly from \( \chi_v(t) \) coefficients. This provides a different view of the analysis system balance as compared to previous studies which presented spectra in terms of divergence and vorticity fields.

2.2. Analysis datasets

Our primary dataset is the DART/CAM system; we present its main characteristics in the following subsection and after that we briefly describe the other three datasets and provide references for further details.

2.2.1. The ensemble analysis system DART/CAM

DART is the assimilation system based on the ensemble adjustment Kalman filter (EAKF) of Anderson (2001, 2003). The assimilation method is one of the ensemble square-root filter (Tippett et al., 2003) methods to approximate the solution to the Kalman filter equations. The goal is to account for the flow-dependency of the forecast errors in the representation of the background-error covariance matrix to improve over the present variational methods (e.g. Lorenc, 2003; Kalnay et al., 2007).

The forecast CAM model, the atmospheric component of the Community Climate System Model (CCSM; Collins et al., 2006a) is described in Collins et al. (2006b). We use...
Figure 1. Analysis increments at 1200 UTC on 20 July 2007, at model level 10 (∼1.3 hPa) of the ECMWF system: (a) difference between the input analysis and background fields in physical space, (b) inversion of $\chi^a_{\nu} - \chi^b_{\nu}$, (c) is as (b), but only ROT modes, and (d) is as (b), but IG+MRG modes. The contouring interval for the geopotential variable is 30 m, starting from ±30 m. Positive values are shaded and negative values are contoured.

an improved version of the model and concentrate on the large scales and time-averaged properties; for this purpose the model skill in short-range forecasts is comparable to that exhibited by NWP models (Tribbia and Baumhefner, 2004). The spectral version of the model with T85 truncation and 26 vertical hybrid levels was applied between the surface and the 3.7 hPa level. The time step for the Eulerian integration was 10 min. The atmospheric model is coupled to a land-surface model and monthly mean sea-surface temperature fields produced at NCEP.

The initial conditions for the ensemble are taken from the various CCSM simulations for July performed within the AMIP (Atmospheric Model Inter-comparison Project). Initially large ensemble spread quickly reduces and after the first 36 h of the assimilation the variations of the spread are small. Since the analysis increments for the first few analysis steps are large, the first eight steps are removed from the statistics so that $N_{\text{sample}} = 116$ for this system. The ensemble contains 80 members, requiring the application of both the localization and inflation of the covariances as discussed in Anderson (2007). In the presented experiments, the covariance inflation was implemented as time constant and spatially varying inflation applied to the analysis ensemble. The covariance localization following Gaspari and Cohn (1999) is implemented as a three-dimensional ellipsoid. The covariance localization increases imbalance (Kepert, 2009); as DART/CAM is not used for forecasts on an operational basis, the balance issue has not yet been properly addressed. This is a possible reason for increased levels of the IG energy in both analysis and background fields.

No observation is assimilated above 100 hPa and the vertical correlation functions spread the observed information from below this level so that the analysis increment and therefore also the analysis bias as defined by Eq. (3) are zero above model level 7 (around 70 hPa). The forecast model also has a significant damping at these levels making the stratospheric solution additionally unreliable.

DART/CAM makes a limited use of available observations; it assimilates radiosondes, aircraft measurements and satellite cloud motion wind vectors. In addition to not using satellite radiances, moisture observations and surface pressure measurements are not assimilated. The observation network is therefore sparse compared to that used by operational NCEP and ECMWF systems, especially over the oceans, in the Southern Hemisphere (SH) and in the Tropics. In spite of the lack of observations compared to the operational systems, the large-scale structures of DART analyses compare well with those shown by the operational systems (Žagar et al., 2009a,b).

2.2.2. Other datasets

Operational analyses of ECMWF and NCEP provide initial conditions for their medium-range weather forecasts as
well as the lateral boundary conditions for numerous mesoscale NWP applications worldwide. Both models solve the primitive equations spectrally on vertical hybrid levels. The number of levels is 91 and 64 with the model top full level at 0.01 hPa (about 80 km) and 0.32 hPa (about 60 km) in the ECMWF and NCEP systems, respectively. Their respective horizontal truncations are T799 and T382 corresponding to around 25 km and 50 km horizontal resolution. Both systems use the variational approach to obtain the solution for the analysis increments assimilating millions of satellite measurements in addition to the conventional observations.

The ECMWF system employs the 4D-Var scheme with the 12 h window in spectral space (Andersson et al., 2004), whereas at NCEP the statistical interpolation method in model grid-point space and 3D-Var are used (Wu et al., 2002).

The effective spatial resolution of the analysis systems, i.e. the resolution of the increment fields, is the full resolution of the first guess and driving model in the case of NCEP (T382) and the resolution of the final inner loop of the minimization in the case of ECMWF (T255). The ECMWF 4D-Var cost function includes a control term that applies a digital filter initialization (Gauthier and Thépaut, 2001) as a weak constraint. The operational NCEP assimilation system includes the tangent-linear normal-mode constraint (TLNMC), which is applied directly to the analysis increment during every iteration of inner loop during minimization within their 3D-Var (Kleist et al., 2009a,b).

For a better comparison with the DART/CAM system and because we are interested primarily in the large-scale flows for which the normal-mode expansion can be trusted best, both datasets are interpolated to the regular Gaussian grid used for the outputs of the CAM model, the N64 grid made of 256×128 points in the longitudinal and latitudinal directions, respectively. Furthermore, before subjecting the data to the normal-mode expansion, the vertical interpolation from the hybrid to the corresponding sigma (σ) levels is performed.

Compared to the above two systems, the NCEP/NCAR reanalysis system (Kalnay et al., 1996) has a lower resolution and is based on an older version of the NCEP assimilation system which did not assimilate satellite radiances but retrievals. This dataset has been the subject of numerous studies and many of its features are well documented. Here we utilize the analysis and first-guess fields available from the NCMR archive on a T47 regular Gaussian grid and 28 σ levels in the range between 0.995 and 0.0027. Thus the horizontal grid includes 192×94 points in longitudinal and latitudinal directions, respectively, with spacing around 1.9°.

Given that details of the average tropical circulation appeared very similar in operational analyses of NCEP and ECMWF (Žagar et al., 2009a,b) on the one hand, and that the CAM model is the atmospheric component of the NCAR’s climate modelling system often verified by comparisons against reanalyses on the other hand, the proposed application of the normal modes can be a useful tool for tuning climate models on short time-scales, especially with respect to the impact of physical parametrizations in the Tropics. The model biases arise during the early days of integration; in the case of CAM, a comparison with DART/CAM analyses using normal-mode functions may offer an attractive way of dealing with the model biases. An effort in this direction is undertaken by carrying out the data assimilation experiment in which the assimilating model is perfect. Since DART/CAM assimilates only conventional observations that can be considered largely unbiased and the observation operator is simple, the results of a perfect-model experiment simply demonstrate that biases found in the referent DART/CAM experiment are likely to be due to model biases, which appear in the background. Another rationale of this experiment is to verify that we are studying the same problem in DART/CAM as in the operational systems and not some properties of the still relatively unexplored ensemble Kalman filter approach (e.g. imbalances) i.e. biases introduced by the DART data assimilation system itself.

3. Results

Now we present the normal-mode diagnosis of analysis increments. Their characteristics depend on the observation coverage and the analysis scheme, especially the specification of the background-error covariances. Given the differences among the four systems, in particular the difference in the model depth, a quantitative comparison of the results is not possible. We simply demonstrate the usefulness of the approach on different systems and, where possible, qualitatively discuss their differences as captured by the normal modes.

3.1. Average energy spectra of analysis increments

The energy of analysis increments is presented in Figure 2 as one-dimensional spectra for the RO+MRG and IG modes. The energy of increment fields (E) is averaged over all samples for each mode and then summed over all meridional and vertical modes for each motion type, i.e.

\[ \sum_{m=1}^{N_m} \sum_{n=0}^{N_n} E_{m,n} \]

Figure 2 thus quantifies the degree of dynamical balance of increment fields in four systems. It can be seen that the amplitudes of increments are smallest in the DART/CAM system as it assimilates fewest observations. The percentage of energy contributed by the IG part is smallest in the NCEP analyses over all zonal wavenumbers and the slope of the spectra become steeper for k > 30. This is highly likely a direct consequence of the implementation of the TLNMC which is an effective filter of the IG contribution to the increment fields. Relative contributions of IG energy vary from 20% (NCEP) to 40% (DART/CAM and ECMWF) and 45% (NCEP/NCAR). Energy percentages in IG and ROT motions provided in the figure are relative to the total energy of increment fields. However, excluding k = 0 from the summation changes percentages little; it reduces the balanced contribution between 1% (ECMWF and NCEP) and 5% (NCEP/NCAR), i.e. it increases relative contribution of EIG or Wig motions up to 1% at most.

These magnitudes can be compared with the IG contribution to the full fields (Žagar et al., 2009a); in this case the IG energy makes up about 9%, 12%, 15% and 45% of the wave (k > 0) energy in the NCEP, DART/CAM, ECMWF and NCEP/NCAR datasets, respectively. Therefore, except for NCEP/NCAR, the increment fields contain about 2–3 times larger energy percentages in IG wave motions than...
Figure 2. Average spectra of the energy of analysis increments in July 2007 as a function of the zonal wavenumber. The three curves shown in the key display balanced modes including mixed Rossby-gravity mode (ROT+MRG), and eastward- and westward-propagating inertia-gravity modes (EIG and WIG, respectively). (a) DART/CAM, (b) ECMWF, (c) NCEP/NCAR and (d) NCEP datasets. Summation is performed over all vertical modes (m) and all meridional modes (n) for each analysis system. %ages are relative to the total energy of increment fields. Dashed lines are labelled with spectral slopes.

the full wave fields. It should also be pointed out that 40% of the increment energy in IG motions in the DART/CAM system effectively means more imbalance in midlatitudes than the same number in the ECMWF system as the tropical analysis increments in DART/CAM are relatively small and all increments are contained within the troposphere.

To explain why the relative contribution of the IG energy in the increment fields is 2–3 times larger than in full analyses is not easy. In particular, the difference is larger for the ECMWF system (which uses 4D-Var) than the NCEP system based on 3D-Var. Smaller scales of the IG increments relative to ROT increments in ECMWF than in NCEP beyond $k = 15$ are most probably due to the strong impact of TLNMC in the NCEP system which operates on the entire analysis increment and removes a substantial portion of the increment that projects onto IG waves. Overall, a comparison of Figure 2 with analysis spectra presented in Žagar et al. (2009a, Figure 8) poses several questions such as whether the models on large scales are too balanced, whether the partitioning of the energy of increment fields is overinfluenced by the imposed balance operators, and whether the models’ diabatic fields are inconsistent with observations. Note that Figure 2 is not directly comparable with similar plots showing the divergence and vorticity spectra, as IG motion contains both vorticity and the divergence. In particular, at large scales the equatorial Kelvin and MRG waves can be vorticity dominated; for example, the $k = 1$ Kelvin wave (KW) with phase speed $15 \text{ m s}^{-1}$ contains four times more vorticity than divergence (Hendon, 1986).

Figure 2 does not include the zonally averaged state ($k = 0$) which contains between 12% (DART/CAM and NCEP/NCAR) and 28% (ECMWF) of the total (all $k$) energy of increment fields. The contribution of $k = 0$ in the NCEP system is about 17%. In all four systems, the relative portion of ROT energy in $k = 0$ is larger than in $k > 0$ but there is a notable difference between NCEP/NCAR, which has the ROT energy percentage in $k = 0$ increment fields about 94%, and other systems where the difference is about 5%. Overall, the NCEP/NCAR reanalyses have a similar distribution of energy in their increment fields as in their full
analysis fields. Furthermore, this is the only system showing a larger energy content in the EIG increments than in the increments associated with the WIG motions (Figure 2(c)). Both properties of NCEP/NCAR may be related to the absence of the initialization and the assimilation of retrievals in the reanalysis system as speculated in Zagar et al. (2009a), but we do not know exact causes. On the other hand, analysis spectra for all four datasets (Figure 8 in Zagar et al., 2009a) indicate about 2–3% more wave energy in the EIG than in the WIG motions relative to the total wave energy; the increased level of EIG wave energy is associated with the KW.

3.2. Systematic increments in physical space

First we present monthly averaged analysis increments on a single horizontal level in the upper troposphere close to 200 hPa (Figure 3).

Figure 3 presents the total bias which is obtained by the normal-mode expansion of analysis and first-guess fields, statistics in the modal space and finally inversion to the model space of \( \Delta \chi_\nu \). The increment fields are commonly averaged in the model space; thus, Figure 3 can easily be verified, as illustrated in Figure 1 for a single time instant. A similar comparison is performed for averaged increments to find out that the large-scale structure of the bias is captured well whereas some small-scale structures are filtered out (not shown). The main advantage of the modal analysis is that Figure 3 can be split into parts associated with different types of motion. The part of the total bias associated with the IG and MRG modes is shown in Figure 4. Its difference from Figure 3 depicts the part of the bias represented by the ROT modes. Comparing these two figures for DART/CAM suggests that at this model level a large part of the bias is represented by the IG+MRG modes both in the Tropics and in midlatitudes. The upper-tropospheric zonal wind bias over the SH midlatitude oceans reaches 2 m s\(^{-1}\). Largest wind biases are found north of the Equator in the Pacific, the Caribbean area and Sahel Africa. Associated with the cyclonic wind bias in the Gulf of Mexico, there is a minimum of the geopotential bias centred over Florida (Figure 3(a)). The negative \( P \) bias in the upper troposphere is spread over Central America and US. Besides the balanced wind bias off the east coast of US, biases seen as easterlies in the Indian and Atlantic oceans and anticyclonic circulation around the Himalayas and continental Europe are balanced. On the contrary, the wind bias in the southern Atlantic is represented by the IG modes (Figure 4(a)); the part of the IG bias close to southern Africa belongs to the EIG motions (not shown).

Among the four systems, ECMWF has smallest tropospheric biases and they are concentrated in the Tropics, especially the wind biases (Figure 3(b)). Their magnitudes are largest over the northern Indian Ocean, but the maximum amplitudes are still well below 2 m s\(^{-1}\) (maximal zonal and meridional winds in Figure 3(b) are 1.75 m s\(^{-1}\) and 1.67 m s\(^{-1}\), respectively). The bias stays mainly in the IG modes (Figure 4(b)). A minor part of the total IG bias comes averaged from the WIG modes; most of it is accounted for by the EIG motions (not shown). The negative geopotential bias coupled with the easterly wind bias over the Indian Ocean is suggestive of the KW. This is confirmed in Figure 5 which shows the bias represented by the Kelvin mode. Besides the signal in the Indian Ocean which has amplitude up to about \(-1\) m s\(^{-1}\), an equally strong signal of opposite sign can be seen over equatorial South America.

The amplitude of the mass-field biases in the NCEP/NCAR reanalyses is between those in the DART/CAM and ECMWF datasets (Figure 3(c)). The wind bias is largest over the Pacific, Australia and Indonesia. Only a small part of the bias is contributed by the ROT modes; most of it resides in the unbalanced motions (Figure 4(c)). There is a global nearly constant geopotential bias of about \(-10\) m which is in the MRG modes (not shown) and the rest of the geopotential bias is associated with IG motions, just like the wind biases (Figure 4(c)). In particular, the zonal wind part of the IG bias over the Pacific projects mainly onto the EIG modes (not shown). The total geopotential bias is overall negative in the upper troposphere and positive in the lower troposphere. The bias becomes again positive in the stratosphere over the Americas, east Asia, Australia and the Antarctic region; this positive mass-field bias is maximal at the model top level where the zonal wind bias is also maximal at somewhat over 3 m s\(^{-1}\) (not shown).

Finally, Figure 3(d) suggests that the largest positive geopotential bias in the NCEP analyses is located over the continental USA. Both balanced and unbalanced modes contribute to the bias without one motion type being dominant (Figure 4(d)). Compared to ECMWF, the NCEP dataset has wind biases also in the extratropics and a geopotential bias with twice the amplitudes. The three-dimensional structure of the bias associated with various modes can be significantly different in the two operational systems; for example, the bias associated with the KW in NCEP is smaller than in ECMWF and has a different spatial structure (not shown).

The solution of the perfect-model experiment equivalent to Figure 3(a) is shown in Figure 6. It shows that systematic increments disappeared except in the poorly observed SH midlatitudes; small increments left here are most likely due to sampling error. Given the observations and observation operators used, this figure suggests that the biases in the upper-tropospheric winds, seen in Figure 3(a), are model biases. This finding is consistent with too strong midlatitude westerlies in the CAM model reported in Hurrell et al. (2006).

Systematic analysis increments in both ECMWF and NCEP are largest at the top model levels and maximal in the SH extratropics. This is best seen in the ECMWF mesosphere, but the SH zonal wind biases dominate also in the upper stratosphere (figures not shown). The absolute bias maxima are at the top level (level 1, \( \sim0.32\) hPa) in the NCEP system and at model level 4 (\( \sim0.10\) hPa) in ECMWF; the latter applies an additional filter to the analysis increments over the topmost four levels.

In the Tropics, the zonal wind biases are smaller than in the SH midlatitudes, but they are still significant; for example, at level 4 (\( \sim0.1\) hPa) in the ECMWF system, the zonal wind bias has a maximal magnitude about 4 m s\(^{-1}\) and it is mainly associated with the ROT modes. The remaining zonal wind bias at the top ten model levels (\( \sim1\) hPa and above) projects onto the Kelvin modes. The tropical meridional wind bias has less than half the amplitude of the zonal wind bias and it is related to the IG modes (figures not shown). The physical interpretation of solutions at these levels is difficult; the vertical depth of both models requires gravity wave parametrizations in the middle atmosphere which are crude and these levels are treated as a ‘sponge layer’.

Copyright © 2010 Royal Meteorological Society
Figure 3. Systematic analysis increments in July 2007 at the model level closest to 200 hPa obtained by the inversion of $\Delta \chi$. (a) level 13 (~195 hPa) DART/CAM, (b) level 50 (~212 hPa) ECMWF, (c) level 11 (~210 hPa) NCEP/NCAR and (d) level 30 (~202 hPa) NCEP. The contouring interval for the $P$ variable is different in different panels: (a) 2 m starting from ±2 m, (b) 1 m starting from ±0.5 m, (c) 2 m starting from ±2 m, and (d) 1.5 m starting from ±1.5 m. Positive values are shaded and negative values are contoured.
Figure 4. As Figure 3, but showing the part of the bias due to inertio-gravity and mixed Rossby-gravity (IG+MRG) modes.
In spite of the differences among the four datasets, Figure 3 suggests that some features of the systematic analysis increment in the upper troposphere are shared among the systems. The most prominent is an area of the outflow in the ITCZ between 10°N and 20°N over the eastern Pacific. Other common features are westerlies over Sahel Africa and equatorial South America (except in DART/CAM) and easterlies over the Indian Ocean.

3.3. Energy distribution of systematic increments in modal space

Two-dimensional bias diagnosis in modal space is shown in Figure 7. We choose to present biases summed across all $k$ including $k = 0$. By showing the distribution as a function of vertical modes, we are not trying to ascribe physical meaning to particular vertical modes; the purpose is to show how various features of systematic increments discussed in the previous section appear in modal space. Magnitudes for different datasets should not be compared. For example, the single model level discussed in the previous section contains the largest biases in DART/CAM, but in modal space DART/CAM displays the smallest magnitudes.

The $(m, n)$ distribution for ROT modes in each system would look significantly different and have much smaller magnitudes if the zonally averaged state was not included. For example, the ROT distribution for ECMWF (Figure 7(d)) would stretch across several low vertical ($m = 1 - 4$) and meridional ($n = 2 - 5$) modes (not shown). With $k = 0$ included, the modal-space distribution is dominated by deficiencies at the top model levels in the SH. In Figure 1 we illustrated that at these levels analysis increments are large. Therefore, the maximum of energy is in $n = 2$, an even meridional mode which is asymmetrical (in height and the zonal wind field and with respect to the Equator). A similar explanation holds for the NCEP system (Figure 7(j)). A simple test performed by filtering only leading vertical ROT modes back to the model space recovers nearly all ROT bias in the upper stratosphere and the mesosphere. Related to the dominance of the top model levels, a majority of the ROT bias in the ECMWF and NCEP systems is located at the zonal wavenumbers $k = 1 - 3$ and it is positive, more so in the ECMWF system.

A better similarity between the DART/CAM and NCEP/NCAR systems on the one hand and the two operational systems on the other is related to the absence or relatively small stratospheric biases in the earlier cases. In the case of NCEP/NCAR and DART/CAM, negative biases in the barotropic mode at $n = 3$ as well as the negative bias centred at $m = 6$ are coming from the zonally averaged state. About half of the total ROT bias in DART/CAM is associated with the $n = 1$ mode. Its spatial distribution at the model level close to 200 hPa superposed on the mean-state bias is displayed in Figure 8(a) (to be compared with Figure 3(a)). It shows the zonal wavenumber 3 of the Rossby mode with maximal negative amplitudes in the Northern Hemisphere sub-tropics. The same mode in ECMWF and NCEP looks different, with the latter having the largest negative geopotential amplitude south of the Equator and the zonal wind component along the Equator, whereas the former is dominated by the southward wind across the Equator in the mean state (Figures 8(b, c)). Significant differences in the tropical wind-field bias among the systems are also illustrative of the lack of direct wind observations in this region and difficulties in tropical data assimilation (e.g. Žagar et al., 2004). The perfect-model experiment removes the bias from the DART/CAM system (not shown).

Figure 7 suggests that the energy of systematic increments in the balanced modes significantly exceeds that in the IG part when integrated over the whole atmosphere depth in each system. This may be expected, given that the bias in zonally averaged states as well as the middle atmosphere biases project mainly onto the balanced modes and that the amplitude of increment fields in the middle atmosphere is larger than in the troposphere (except in DART/CAM). However, this does not apply to the troposphere, as illustrated by showing the IG+MRG part of the total bias at a level in the upper troposphere (Figures 3, 4). Consequently, Figure 7 would look significantly different for ECMWF and NCEP if their top-model levels are not included in the projection, for example by taking the uppermost model level of the ECMWF system to be level 12 which is at about the same pressure (∼2.4 hPa) as top levels in the DART/CAM and NCEP/NCAR systems.

A qualitative similarity among the datasets is best seen in the negative bias in the Kelvin mode in the ECMWF, NCEP/NCAR and NCEP systems (Figure 7(e, h, k)). It suggests that all three assimilation systems on average act to suppress the KW motion. In particular, both ECMWF and NCEP pick the same vertical mode $m = 3$ which has zero crossings at the tropopause and at around 1 hPa (close to model level 9 in the ECMWF and level 2 in NCEP) and maximal amplitude at around 5 hPa. In both systems, a part of the increased bias above 1 hPa in the tropical zonal wind is associated with the Kelvin mode. In the NCEP/NCAR system, the KW bias is the largest negative IG bias and it resides in modes $m = 5 - 6$ which have lowest two zero crossings (from the surface) in the middle troposphere and around the tropopause level. This structure is typical of the standard KW representation in simple models, the so-called ‘first baroclinic mode’ of the tropical troposphere. In the DART/CAM dataset the tropical analysis increments are small because of few observations assimilated and the IG bias appears smaller than in the other systems. The $(k, n)$ distribution (not shown) shows that the lack of the KW bias in DART/CAM in Figure 7(b) is due to the opposite signs of this bias in $k = 2$ (positive) and $k = 1, 3$ (negative) waves. In NCEP/NCAR the negative bias comes from $k = 2$. Not considering the KW, the EIG bias in DART/CAM, ECMWF and NCEP is mainly located in $k = 0$ while NCEP/NCAR has a positive peak at $k = 2$ (not shown).

A similar discussion applies to WIG biases. Again, ECMWF and NCEP have most of the bias in the same vertical and meridional modes ($m = 2$ and $n = 0, 2, 4$) corresponding to the higher model levels. In DART/CAM, the WIG distribution appears similar to the EIG biases because most of the bias is associated with the zonally averaged state characterized by the same distribution in EIG and WIG modes, just like the full fields. Positive bias in both EIG and WIG modes in NCEP/NCAR has the maximum at $m = 3$, a vertical mode with its zero crossings below the tropopause (at around the 250 hPa level) and at model level 2, suggesting an opposite shape of the bias in the troposphere and the stratosphere. The corresponding zonal wave number is $k = 2$.

An important large-scale equatorial mode, the MRG wave, does not appear in Figure 7 because of its small amplitude relative to ROT modes. It does contain a part of the bias
4. Discussion and conclusions

We have applied the normal-mode functions to represent the time-averaged analysis increments, which serve as a proxy for analysis system bias, in terms of balanced and unbalanced contributions. Results show that the most significant tropospheric biases in the four studied systems are in the tropical inertio-gravity motions. The magnitudes of tropospheric wind-field biases associated with the large-scale IG motions in July 2007 were in the range 1–2 m s$^{-1}$. Within the Tropics, the horizontal distribution of the bias at any given level can be very different in different analysis systems, as illustrated by the Kelvin mode. This particular mode is of special interest because of its role in tropical variability and the global circulation. KWs coupled to convection have been identified from observations of convection such as the outgoing long-wave radiation (OLR) (e.g. Wheeler and Kiladis, 1999). These are observations of the mass field; as direct wind observations in the Tropics are sparse, the tropospheric wind field of the KW has previously been estimated by the regression of the observations of convection and reanalysis winds (e.g. Wheeler et al., 2000). Geographical regions with the largest climatological OLR variance for the KW coincide with the maxima of the KW bias found in the present study. For example, the upper-tropospheric maximum of the bias in the Indian Ocean in the ECMWF system (Figure 5(a)) coincides with the maximal climatological OLR variance for this wave in Wheeler et al. (2000). A secondary climatological OLR variance maximum from the same study coincides with the KW bias over equatorial South America. The magnitude of the bias is similar to the magnitude of the KW wind field obtained from the regression, illustrating the importance of diagnosing the analysis system biases in the Tropics where magnitudes of the large-scale variability can be comparable with the errors involved in the data assimilation process.

In Figure 9 we use once again the Kelvin mode in the ECMWF system to illustrate how the normal-mode expansion can be applied for the tracking of the analysis system behaviour. The time evolution of the energy in analysis increments shows that the positive energy increments are systematically placed at 12 UTC times, following a negative energy increment at 06 UTC times. An exception to this behaviour is a period in mid-July, during which large negative increments to the Kelvin mode resulted in an overall negative KW bias seen in Figure 7(e). A closer investigation of this particular event and its relation to the analysis and the forecast system uncertainties is currently being undertaken. Such strong episodic increments in a particular mode can occur for various reasons (e.g. a break of a satellite instrument with a significant bias correction applied in the assimilation) and they can mask the overall
Figure 7. The \((n,m)\) distribution of \(\Delta E\) for the (a–c) DART/CAM, (d–f) ECMWF, (g–i) NCEP/NCAR and (j–l) NCEP systems: (left) ROT+MRG modes, (middle) EIG modes, and (right) WIG modes. Units are m²s⁻² and the contour interval is different in different plots.
behaviour of the system, suggesting that longer periods should be used for the bias quantification.

Two operational systems, ECMWF and NCEP, share several properties concerning the bias distribution among various modes. Both analysis systems extend to the mesosphere and are characterized by significant biases at the model top levels which are most likely associated with inadequate parametrizations and artificial damping. In the extratropics, these biases project mostly onto the balanced motions and in July 2007 they reside predominantly in the
Figure 8. Systematic analysis increments due to $k = 0$, all modes, and $n = 1$ ROT mode at the model level closest to 200 hPa. (a) level 13 ($\sim 195$ hPa) DART/CAM, (b) level 30 ($\sim 212$ hPa) ECMWF, and (c) level 30 ($\sim 202$ hPa) NCEP. Positive values of the $P$ variable are shaded and negative values are contoured. The contour interval is 0.5 m starting from $\pm 0.5$ m.

Figure 9. Time evolution of energy in the two lowest EIG modes for increment variables in the ECMWF system. x-axis labels are at 1200 UTC.
winter hemisphere. In the Tropics, a significant portion of the model-top level biases projects onto the two most energetic IG modes, the Kelvin and the MRG mode. The former is associated with the zonal wind biases while the latter represents the cross-equatorial (meridional) wind bias. The amplitude of these biases should be taken into account when studying the structure of the equatorial motions at and above ~1 hPa based on analysis datasets.

Among the four studied systems, the recently developed DART/CAM ensemble assimilation system contains largest biases in the troposphere which could be expected since it is a research tool under development. A perfect-model assimilation experiment with this system demonstrated that the biases found are the model biases which appear in the background. The bias in the zonally averaged state in both balanced and IG modes in DART/CAM far exceeds biases in the wave (k ≠ 0) motions. In the case of the NCEP/NCAR reanalysis system, the vertical tropospheric structure of the main tropical biases follows the structure traditionally known as the first baroclinic mode.

We presented spectra of analysis increments averaged over one month as functions of the zonal wavenumber for balanced, EIG and WIG waves. On average, 20% to 45% of energy in increment fields belongs to the IG motion. Systematic increments (biases) appear predominantly balanced (e.g. over 90% in the case of ECMWF; Figure 9) which is associated with the mean-state bias and biases at the model top levels in the operational systems of ECMWF and NCEP. However, of most interest for operational NWP are tropospheric biases which are largest in the Tropics and dominated by unbalanced modes in all four datasets. This suggests that the representation of the model-error covariance matrix by the scaled B-matrix (e.g. Houtekamer et al., 2005), while useful on average, may not work well in the Tropics since the background-error covariance matrices are tuned to the midlatitude conditions. This fact is exhibited in the spectra of increment fields by nearly the same percentages of energy in the EIG and WIG motions, contrary to the energy distributions in the full fields which contain more energy in the EIG motions due to the KW.

In the ongoing study, we are applying the same methodology to the short-range forecast fields used to estimate the background-error characteristics for the ECMWF and DART/CAM systems. Results are expected to help us to better understand (im)balances contained in the B-matrix and their link with various properties of analysis increments diagnosed here.

Acknowledgements

The Hough function expansion used in this study was developed by Akira Kasahara; his interest and advice during the stay of N. Zagar at NCAR are gratefully acknowledged. We thank Carla Cardinali of ECMWF for discussions of the ECMWF assimilation system in the course of this work and to Tim Hoar and Nancy Collins of the DART team at NCAR for their support in using DART/CAM.

References


