Validation of parametrized forcing in NWP and climate models

Sean Milton, Ian Culverwell and David Cameron

Met Office, UK.

1 Introduction

The parametrization of subgrid-scale physical processes is a crucial component in the predictions of numerical weather prediction (NWP) and climate models. It is essential to try and evaluate the parametrized physical processes in atmospheric models in order to (i) identify and correct the deficiencies in the parametrized processes and (ii) understand the role played by subgrid-scale parametrized forcing in influencing the large scale flow, determining weather elements (precipitation, fog etc.) and in modulating feedback's in climate change prediction.

Diagnosing the sources of errors in complex numerical models of the atmosphere is a difficult process involving distinguishing between initial condition error and model error, and then further attributing model error to specific subgrid-scale processes or model numerics/resolution. Progress on these issues has always been hampered by the lack of observational data on the planetary scale to directly validate parametrizations. Field experiments and ongoing observation programmes (e.g. BSRN, ARM) can provide detailed information locally on some aspect of the physical processes such as the land surface scheme (Betts et. al., 1993) or surface radiation balance (Chevalier and Morcrette, 2000). The other source of global data is from satellites.The ERBE and ISCCP data have been used extensively in the climate and NWP communities to validate clouds and radiation (Kiehl et al. 1998). Precipitation in the tropics is directly linked to the diabatic forcing in the column and precipitation products now available from the tropical rainfall measuring mission (TRMM) provide a valuable resource to validate tropical precipitation on synoptic timescales.

One can also make progress in *inferring* possible errors in the parametrizations from examining model tendencies and diagnosing their influence on the large-scale flow. The concept of balance between time-mean diabatic and adiabatic tendencies has been used with some success to diagnose errors in parametrizations from NWP models (Klinker and Sardesmukh, 1992). In addition, sensitivity studies with parametrizations are useful in evaluating whether model performance depends heavily on aspects of parametrizations that have a high degree of uncertainty. This will highlight the need for better observations in a given area but ad-hoc tuning of the parametrizations on this basis should be avoided.

This paper provides a review of a number of diagnostic/validation techniques that have been developed by the NWP and climate research communities to help identify errors/uncertainties in parametrized processes. Examples of their application in the Met Office Unified Model are given in the following sections. The atmospheric component of the Unified Model is used for both NWP and Climate modelling. The NWP atmospheric component of the Unified Model is used for both NWP and Climate modelling. The NWP component currently consists of a global version (60km, 30L) and a regional model for the UK (12km, 38L). The climate model has a resolution of 250km and 30L. Where possible the same parametrizations are used across all configurations. This provides a rigorous testbed for model parametrizations across many spatial and temporal scales.

In section 2 examples are shown of using initial tendency balance diagnostics to infer errors in time-mean momentum and thermal forcing. Section 3 extends the use of the balance diagnostics to include potential vorticity balance in the extratropics. This is used in conjunction with other diagnostic tools such as lagrangian trajectories to examine the role of parametrizations in forcing errors in extratropical cyclones. Section 4 considers the use of observations from the ARM site and direct satellite radiances to diagnose potential errors in diabatic heating and clouds in the tropics. Discussion and conclusions are given in section 5.

2 Balance Diagnostics

Direct observations of diabatic heating, moisture forcing or momentum forcing are not available on the planetary scale. In general these quantities can only be diagnosed indirectly using a budget approach, where the forcing is determined as a residual in some balance equation. There is a long history of using both observational data and global analyses generated from NWP data assimilation to estimate the observed diabatic forcing on the atmosphere (see review by Holopainen, 1987). Examples include determining the diabatic heating as a residual in the thermodynamic equation (Hoskins et. al., 1989), estimating frictional torque from the atmospheric angular momentum balance (Swinbank, 1985), or estimating diabatic generation and mechanical dissipation terms in the energy cycle (Arpe et. al, 1986). In theory one can then use these budget estimates to compare with those produced directly from the sub-gridscale parametrizations in a global NWP or climate model (Fortelius, 1995).

Klinker and Sardesmukh (1987) developed this idea by using the ECMWF NWP system to output the dynamical and parametrized tendencies after 1-timestep of model integration. The initial dynamical tendency averaged over a large number of forecasts is equivalent to using the analyses to determine the diabatic forcing as a residual, with the advantage that the calculations are consistent with the model's numerics. On the monthly/seasonal timescale the dynamical and parametrized tendencies should balance each other (allowing for a monthly trend), and the degree of imbalance (the budget residual) points to possible error sources. The advantage of the technique are that it emphasises local errors, removing the complication interactions between local error growth and that due to remote forcing (e.g. errors in rossby wave propagation from tropics to extratropics). In addition the feedbacks amongst parametrizations (and dynamics) are minimised which should allow a clearer attribution of errors to particular parametrizations.

Klinker and Sardesmukh (1992) used the technique to diagnose errors in the momentum balance of the ECMWF model. This technique has been used routinely at the Met Office since 1993 to help diagnose errors in parametrized sub-gridscale forcing and some examples of its use are given below.

2.1 Momentum Balance - Deficiencies in parametrized orographic forcing.

The day 3 forecasts of zonal wind during December 1993 showed the following systematic errors

- westerly biases in the extratropical troposphere,
- easterly biases in the tropical BL and upper troposphere.
- Easterly bias in the stratosphere in the northern hemisphere

A series of 1-timestep integrations were run from the four analyses available every day for December 1993 and the parametrized and dynamical tendencies calculated. The monthly-mean momentum budget for zonal wind (Fig.1) shows that the parametrized and dynamical tendencies do largely balance each other. There are large dissipation rates in the boundary layer (10-20 m/s/day) due to turbulent mixing and in the northern

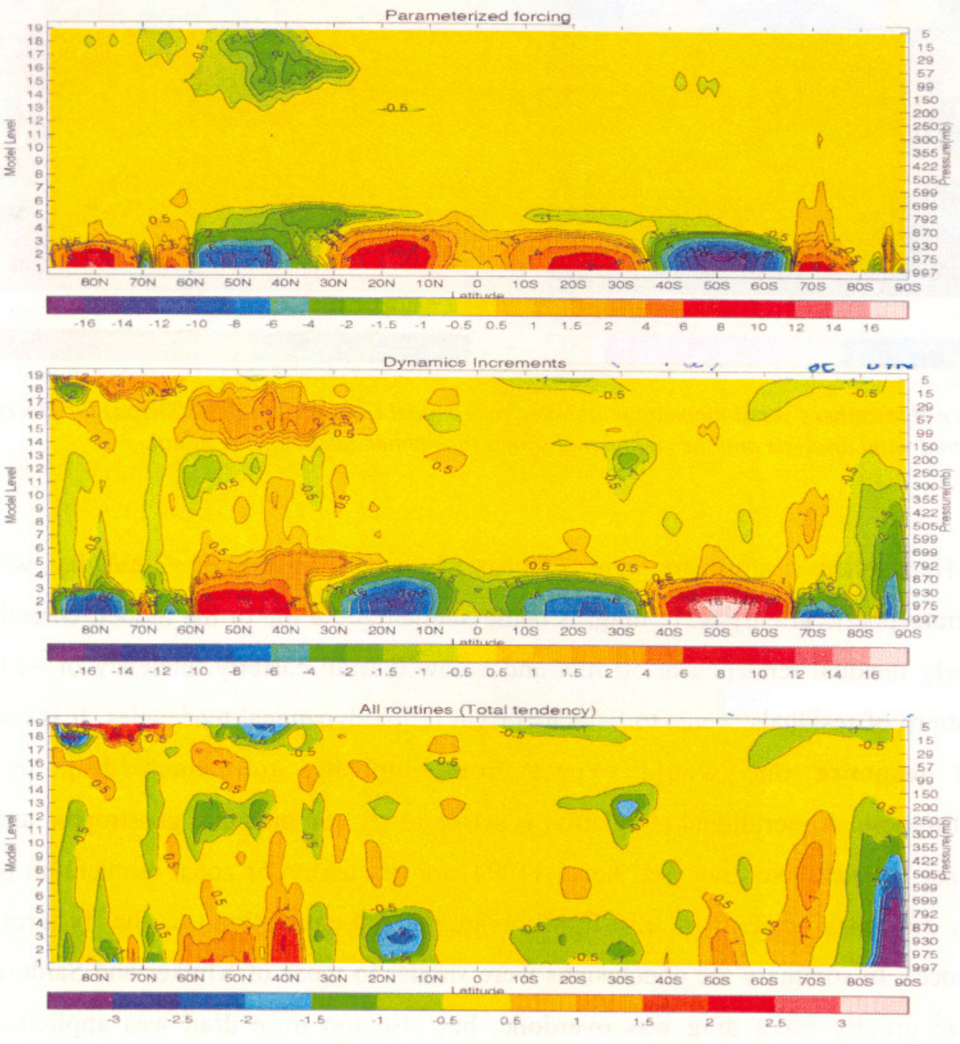


Figure 1 - Zonally averaged zonal momentum balance for December 1993 calculated from 1-timestep model integrations.

hemisphere stratosphere (~2m/s/day) due to the gravity wave drag parametrization. The lack of balance between parametrized and dynamical terms (Fig 1 - bottom panel) bears a resemblance to many of the forecast errors at day 3 (not shown) and shows a number of interesting features. The westerly/easterly residuals of between 1 and 2 m/s/day in the extratropical/tropical boundary layer are perhaps indicative of too little parametrized drag. Care must be taken in interpreting these momentum budget residuals as errors in the dynamical momentum tendencies may arise through the either the effect of erroneous diabatic heating on the evolution of the wind field, or, errors in the initial analysis. Examination of the geographical distribution of tendencies/residuals in the boundary layer of the northern hemisphere strongly suggests that the major imbalance is associated with the orography, with westerly tendencies of ~10 m/s/day over the Rockies, Alps, and Asian mountain complexes (Fig. 2).

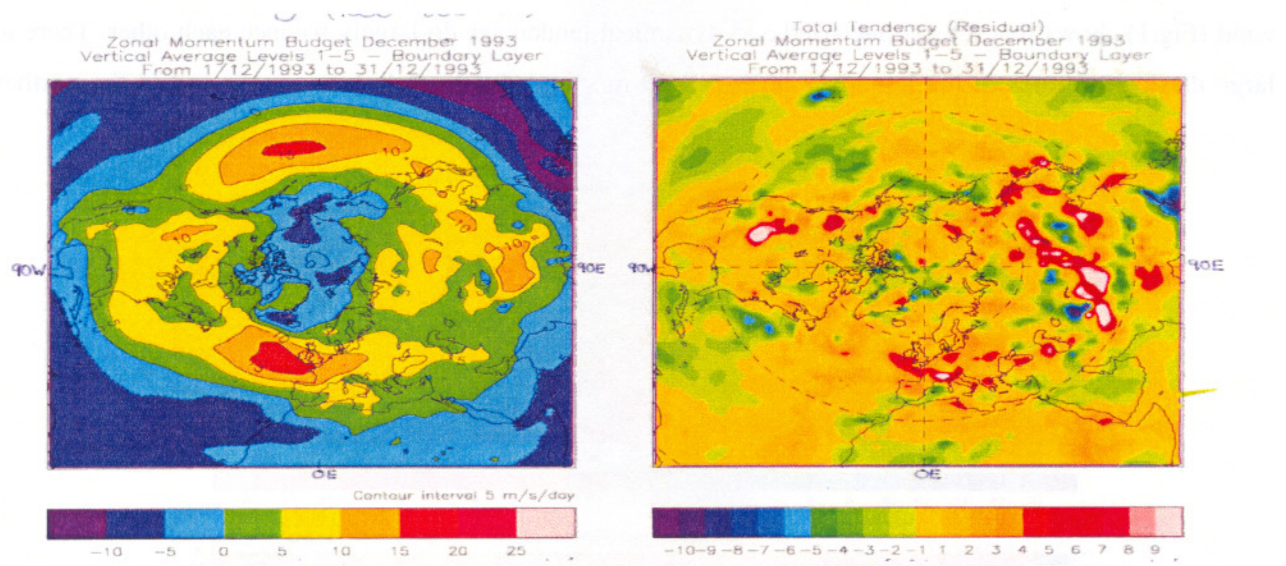


Figure 2 - Northern Hemisphere zonal momentum balance over model levels 1-5 (1000-800Hpa) for December 1993. Left panel shows the zonal wind analysis and the right panel shows the momentum budget residual.

In the Northern Hemisphere stratosphere there is an easterly budget residual of ~2 m/s/day, which suggests that too much stratospheric gravity wave drag, is being applied at the top of the model. Directly below this there is a westerly residual tendency at 50 hPa and below that an easterly tendency of ~1.0 m/s/day at 200hPa. This pattern of residuals seems to have its origin in the dynamical tendencies. It is consistent with the dynamical response one would expect from imposing too much drag in the upper troposphere/stratosphere. A meridional circulation is set up to try and maintain geostrophic and hydrostatic balance in the face of excessive drag (see James (1994) for discussion of zonal momentum balance). The westerly/easterly dynamical tendencies arise from this meridional circulation (via the -fv term in the zonal momentum balance). In summary the conclusions were similar to those of Klinker and Sardesmukh (1992) that stratospheric gravity wave drag was overdone, but also too little drag was applied in the lower

In conjunction with this diagnostic work two new parametrizations were developed which greatly reduced the systematic errors in both NWP (Milton and Wilson, 1996) and climate versions (Gregory et. al, 1998) of the Unified Model. A new gravity wave drag scheme including low level wave breaking mechanisms and a parametrization of form drag using effective roughness concepts. Following the introduction of these changes to orographic drag in January 1996 the budget residual for the following winter were markedly reduced and systematic errors and verification scores improved.

2.2 Thermal balance - impacts of increased resolution

The increments from the parametrizations and dynamics have also proved very useful when testing changes to the model formulation. This is true for both modifications to the parametrizations, where the change in forcing may be well understood but the impacts on the large scale flow or other parametrizations are generally not known, and also for changes to resolution, numerics or data assimilation which may impact on parametrization performance.

An example of this was the change to vertical and horizontal resolution of the global Unified Model in 1998. The main impact of changing the vertical resolution from 19 to 30 levels was a reduction the upper tropospheric cold bias and the moist bias in the model. To investigate the source of the improvements in temperature we examined the thermal balance of both HIRES and CONTROL over the first 24 hours of the forecasts from a winter trial of the increased resolution. The parametrized diabatic heating (Fig.3) in the CONTROL shows warming in the tropics and in the extratropical stormtracks, which is approximately balanced by the cooling from ascent from the dynamics. The largest imbalance between parametrized and dynamical heating rates is at the tropopause. The parametrized cooling exceeds the warming from the dynamics and the total temperature tendency in CONTROL bears a strong resemblance to the day 3 and day 5 systematic errors in temperature in the Unified Model (not shown).

The difference in the total temperature tendency between HIRES and CONTROL is the exact opposite of the CONTROL total tendency (Fig. 3). The tropopause cooling tendency is reduced in HIRES, as is the warming tendency in the tropical mid-troposphere and at 60N and 60S. This in agreement with reduced mean errors in temperature in the HIRES forecasts. The differences (HIRES-CONTROL) in the parametrized and dynamical heating are complex, but generally show that the largest contribution to the improved thermal balance at the tropopause comes from changes in the parametrized diabatic heating. A small increase in the dynamical warming at the tropopause in the extratropics also contributes to the improved thermal balance. Most of these changes at the tropopause are due to the increased vertical resolution. A similar impact was seen on increasing vertical resolution in the climate version of the Unified Model (Pope et. al., 2000). Elsewhere in the troposphere the changes to the parametrized and dynamical heating rates tend to oppose each other.

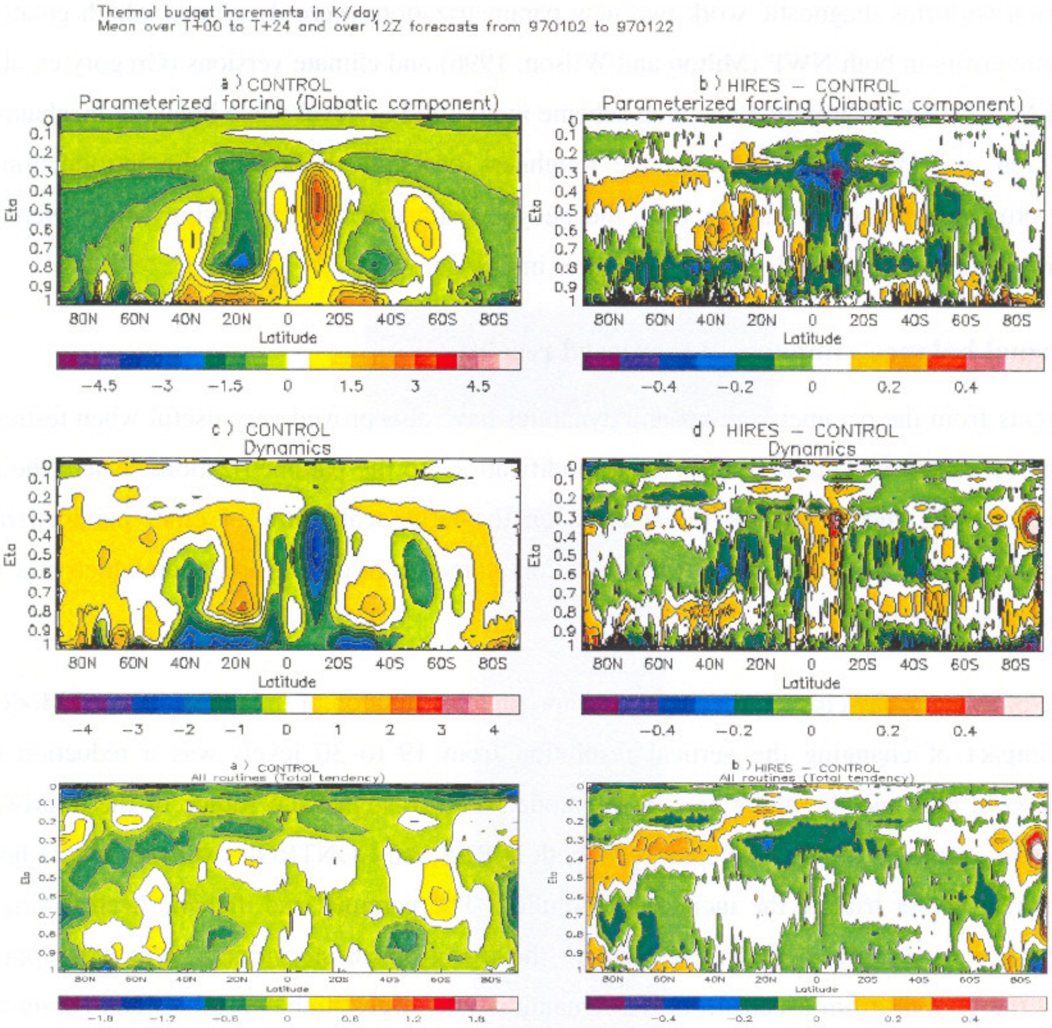


Figure 3 - T+24 Thermal budget for January 1997 showing the effect of increasing model resolution (see text for details). Left panels show the parametrized tendencies, dynamical tendencies and the budget residual for the CONTROL resolution. The right hand panels show HIRES-CONTROL difference for the same quantities. Units in K/day.

Finally, we consider the individual contributions to the parametrized diabatic heating (Fig. 4). Most of the signal at the tropopause comes from changes in the radiative heating, with a reduced cooling in the extratropics in HIRES. This comes from reduced LW cooling associated with reductions in high cloud amount, but also has a small contribution from the LW clear sky component, reflecting the beneficial reductions in upper tropospheric humidities. The signal of reduced high cloud formation can be seen in the large-scale condensation which shows a reduction in latent heating in HIRES. The increased radiative cooling in mid-troposphere is due to a reduction in the downward LW radiation from the upper level cloud.

In the tropics both LW and SW contribute to the radiative signal but tend to have opposing effects (not shown). The increased cooling at 300hPa and 10S comes from a reduction in the downward LW radiation from the reduced high cloud cover at 100-200hPa. In the deep tropics and in the boundary layer there are reductions in the convective heating in HIRES compared to CONTROL, but increased warming in the subtropics in the mid-troposphere. Finally, there is an increase in the diabatic heating in the stormtracks in HIRES, which has its origins in increased latent heating due to formation of large-scale condensation. Independent tests have shown that this signal is mostly due to increasing the horizontal resolution. This increased diabatic heating is beneficial to our cold temperature biases observed in the NH stormtracks in the CONTROL version of the UM.

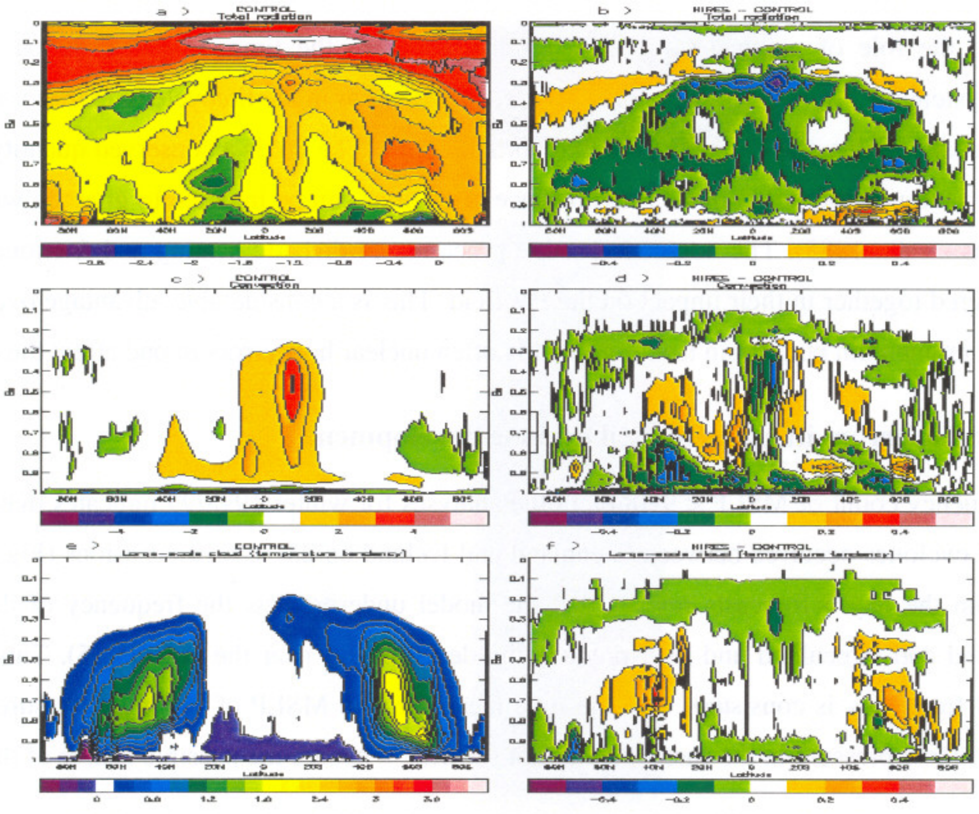


Figure 4 - as figure 3 but for individual parametrized tendencies, total radiation, convective heating and LS condensation.

This example demonstrates the complexity in changes to the parametrized and dynamical forcing and balance between these components from changes to the model resolution. The model tendencies are an essential diagnostic in unravelling the various complex changes and interactions between model physics and dynamics. Further diagnosis of the tropical thermal balance is discussed in section 5.

2.3 Other studies

The most common use of the tendency diagnostics has been in diagnosing error sources in monthly-mean budgets. However, a wealth of information is contained in the time and space variations of parametrized forcing and model errors and a number of studies have attempted to exploit this. Klinker and Ferranti (1999) used statistical techniques (e.g. SVD) to determine the time and space relationships between model forcing and errors in the large-scale flow in an attempt to derive the regime dependent errors in parametrized forcing. Schubert and Chang (1996) used a multiple regression to find the linear transformations to a model's forcing terms which produce the smallest short range forecast error in a least squares sense. They were able to partition the systematic and non-systematic component of error into contributions from various parametrized and dynamical forcing terms. They used this technique to study errors in the moisture budget over the U.S. In the following sections examples of using the time and space variation of parametrized forcing to diagnose parametrization errors are given.

3 Diagnosing Parametrized Forcing in Extratropical Cyclones - PV balance

In addition to the role of parametrizations in forcing errors in the time-mean circulation, we are also interested in systematic errors in model variability, and in particular in the lifecycle of extratropical cyclones. A useful diagnostic in this respect is the potential vorticity (PV). The PV is a conserved quantity for inviscid adiabatic flow, and with proper boundary and balance conditions can be inverted to give the wind and mass fields (Hoskins et. al., 1985). The non-conservative processes of diabatic heating and frictional dissipation can be considered together in their impact on the PV field. This is a considerable advantage over diagnosing the thermal or momentum budgets in isolation, as it is often unclear how errors in one affect the other.

3.1 Systematic errors in extratropical cyclone development

A simple frequency count of MSLP < 990hPa in the analyses for winter 1998/99 indicates that the majority of the deepest cyclones occurred between Greenland and Iceland on 25-30% of occasions (Fig.5). A similar diagnostic from the day 5 forecasts reveals that the model underpredicts the frequency of deep cyclones between Iceland and Greenland and over-predicts the development near the UK (Fig.5). This error in the climatological frequency is consistent with the time mean error in MSLP of too high pressure in the West Atlantic and too low pressure in the east Atlantic (not shown). What are the possible causes of this *systematic*

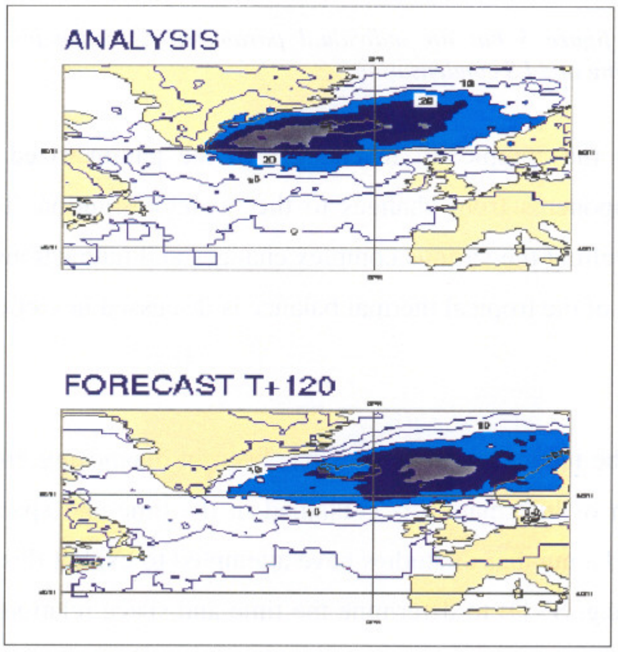


Figure 5 - Frequency of occurrence (%) of mean sea level pressures less than 990 hPa in analysis (top) and day 5 forecasts (Bottom)

error? Direct systematic effects on cyclone development could include insufficient resolution, errors in frictional effects (e.g. orographic forcing), or errors in the parametrized diabatic heating. Other error sources could be deficiencies in the scale interactions with the planetary waves. To try and uncover the role of parametrizations in cyclone development we have selected a single forecast that displays these characteristic errors in development.

3.2 Case study of rapid deepening - 18-20 November 1998

The case involves the rapid deepening of a surface cyclone moving from the sub-tropics to the extra-tropics (Fig 6). At 12UTC on 18/11/98 the analysis shows a cyclone of 998 hPa at 40N and 60W which deepens to 971hPa in the next 24 hours and 961 hPa 48 hours later when it is located between Greenland and Iceland. The operational day 3 forecast (initialised at 12Z 15/11/98) has a good representation of the cyclone on the 18/11/98 but completely fails to represent the explosive deepening occurring in the next 48 hours. The day 5 forecast central pressure on the 20/11/98 was only 989 hPa compared to 961 hPa in the analysis. On many occasions one would probably attribute such an error in development to errors in the initial conditions (Rabier et. al., 1996), but given a possible link to systematic deficiencies in extratropical cyclone behaviour we have diagnosed the role of the various forcing terms in this development.

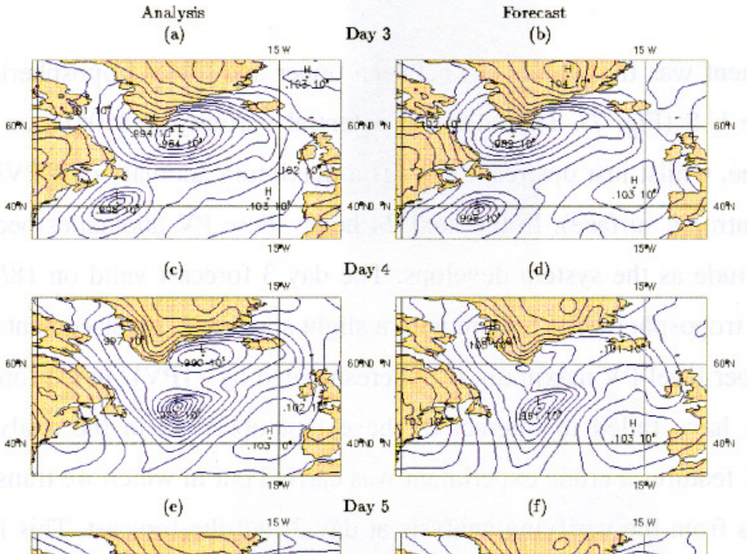


Figure 6 - Mean sea level pressure for analyses (left column) and corresponding day 3,4, and 5 forecasts (right column) for the period 12 UTC 18 Nov - 20 Nov 1998.

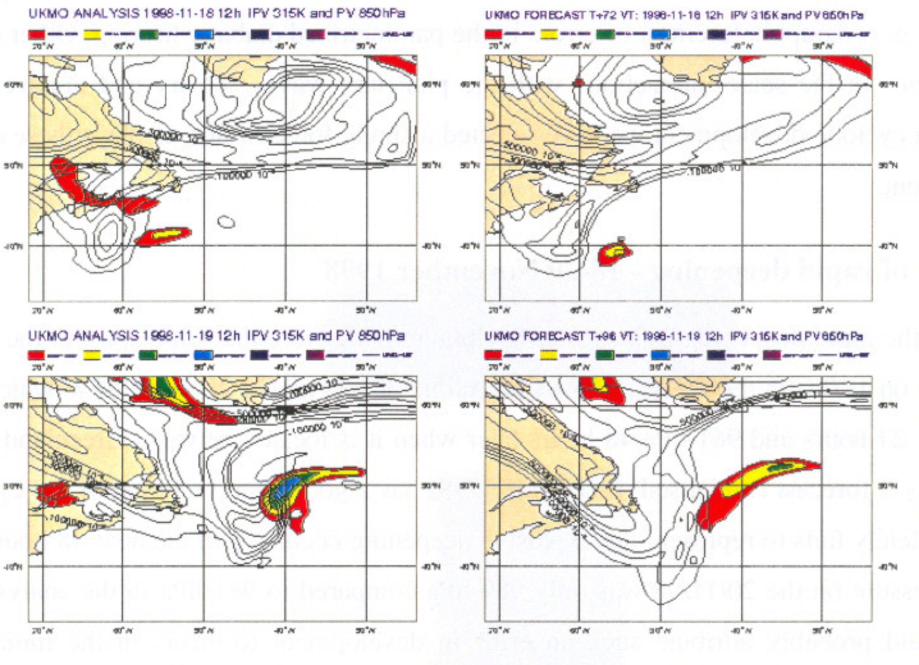


Figure 7 Potential vorticity on the 315K isentropic surface (contours) and at 850hPa (colours) for analyses (left panels) and corresponding day 3 and 4 forecasts (right panels) for the period 18-19 November 1998. Contours are every 1.0 PVU. For the 850hPa PV the first colour (red) represents 0-1 PVU and subsequent colours are in 0.5 PVU steps.

A key to the development was the interaction between upper and lower tropospheric PV maxima over the eastern seaboard of the U.S (Fig. 7). The analysed potential vorticity at 850hPa on 18/11/98 is coincident with the surface cyclone, whilst just upstream there is a maximum anomaly of 4 PVU[1] located in the lower stratosphere (315K isentropic surface). In the next 24 hours these PV anomalies become phase locked and both increase in magnitude as the system develops. The day 3 forecast valid on 18/11/98 has a reasonable prediction of the lower tropospheric PV feature with a slight southward displacement relative to the analysis. The intensity of the upper level PV maximum is underestimated by ~1PVU in the forecast. One day later the forecast PV anomalies have failed to interact to the extent observed in the analyses. To determine the importance of these PV features a crude experiment was carried out in which we transplanted the upper level winds and temperatures from the verifying analysis at day 3 into the forecast. This led to a deeper forecast low of 967 hPa at day 5 (not shown), confirming the importance of the upper level PV feature in the development. A further experiment replacing the lower tropospheric winds and temperatures with analysed values at day 3 had little impact.

3.3 Lagrangian trajectory diagnostics

To examine the time-evolution of the upper level PV features we applied a lagrangian trajectory calculation to forecasts and analyses (Methven, 1997). A cluster of particles were released from the region surrounding

[1] PVU= potential vorticity unit where 1 PVU=10-6 kg/m2/s2.

the upper level PV feature on 18/11/98 and the 6 hourly analysed wind field used to calculate back trajectories to 15/11/98 when the forecast in question was initialised. Following Wernli and Davies (1997) we used a number of criteria to isolate trajectories of interest. Figure 8 shows the back trajectories of particles released at 400hPa with initial amplitude > 1 PVU. Many of the the particles arriving at the target area on the 18/11/98 originate from a location in the East Pacific on 15/11/98 in which there was a significant upper level PV feature (>8 PVU). The time history of the trajectories suggests they originate in the lower stratosphere and descend to 400hPa as they cross the U.S. A set of forward trajectories were then released from 250hPa at this location in the East Pacific on 15/11/98, and those reaching the target area near Newfoundland on 18/11/98 selected for study. The evolution of PV for these trajectories shows the analysis varying in time but maintaining PV values between 4 and 8 PVU (Fig 8). In contrast the forecast trajectories initialised from the same positions maintain high PV values (5-7 PVU) for the first two days of the forecast then show a rapid decline between day 2 and 3.

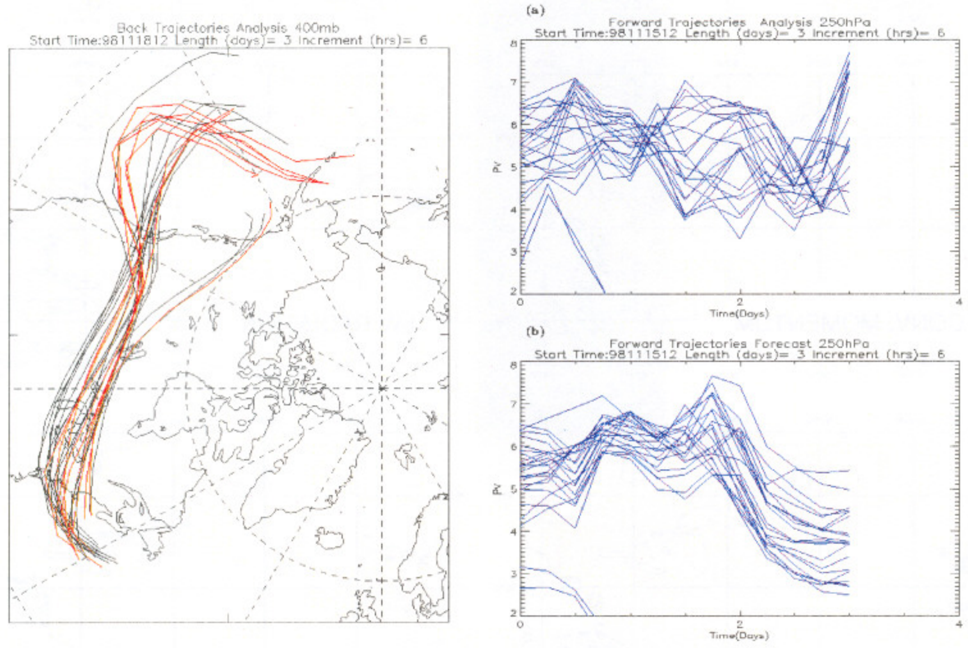


Figure 8 - Left panel - 72 hour analysed back trajectories for parcels started at 400hpa from an analysis on 18/11/98 surrounding the PV feature in figure 7. The trajectories shown were selected to have PV > 1PVU at their starting positions. Right panels - 72 hour forward trajectories for forecasts and analyses starting from 15/11/98 at a location in the East Pacific suggested by the back trajectories and at a height of 250hPa. Trajectories selected are those that end up in target area in Newfoundland 3 days later.

3.4 PV budget analysis

To understand the decline in forecast PV we use the balance equation for PV (Hoskins et. al., 1985)

$$\frac{D(PV)}{Dt} = \frac{\zeta_a}{\rho}.\nabla\dot{\theta} + \frac{\nabla\theta}{\rho}.\nabla\times F$$

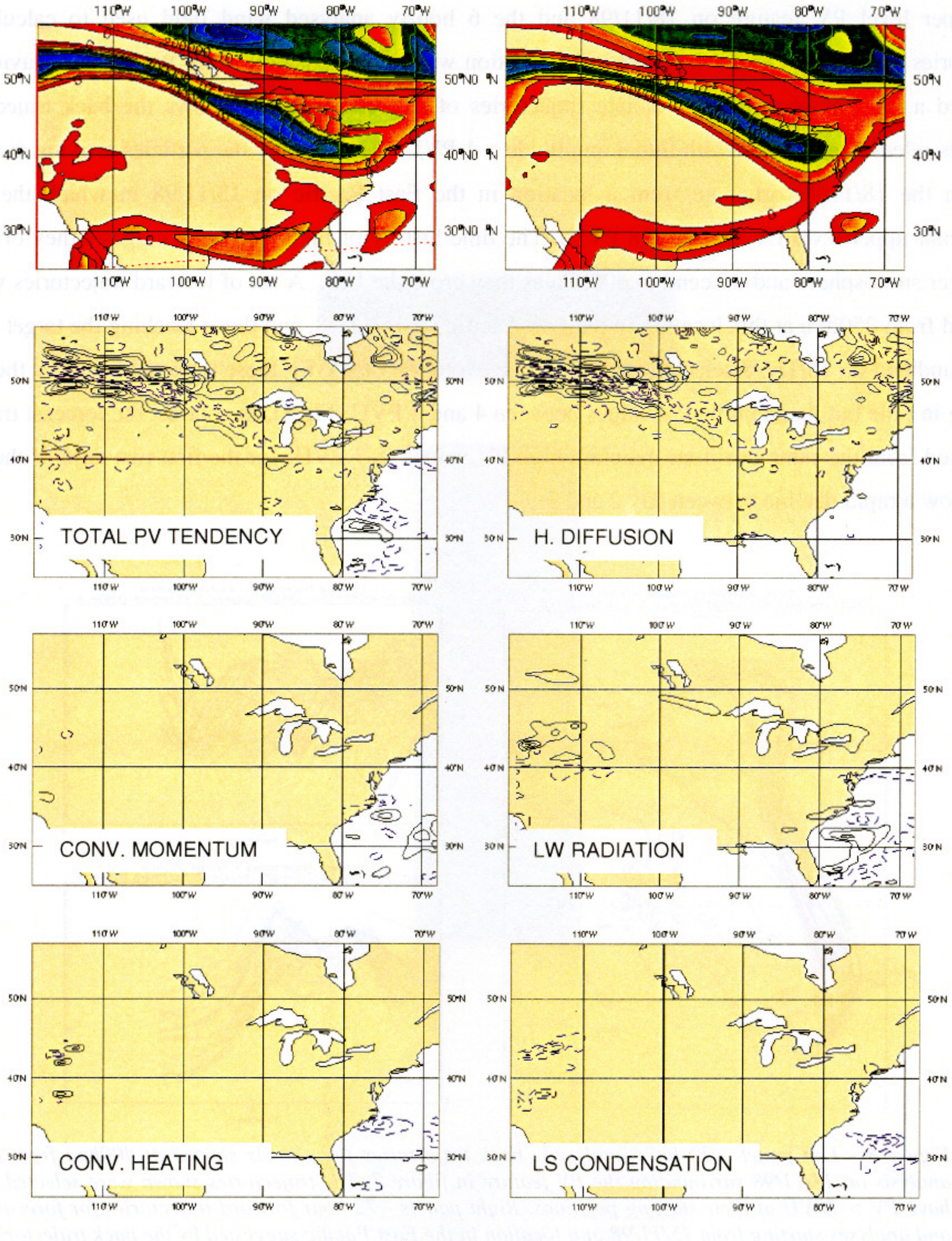


Figure 9 - The PV at 300hPa for analysis (top left) and corresponding 42 hour forecast both valid at 06UTC 17/11/98. Remaining panels show the contributions to the forecast PV budget from the various parametrized tendencies for the 36-42 hour forecast period. Negative PV tendencies shown as blue dashed contours.

The 6-hour model tendencies from the physical parametrizations can be used to calculate the contribution of the frictional (F) and diabatic heating ($\dot{\theta}$) terms to the time evolution of PV in the upper troposphere (~300hPa). Stoelinga (1996) discusses the role of physical processes in modifying PV in extratropical cyclones. For example, PV is generated by a positive gradient in diabatic heating in the direction of the absolute vorticity vector. For mid-tropospheric diabatic heating (e.g. due to latent heat release in a cyclone) this implies depletion of PV at upper levels $(\nabla\dot{\theta} < 0)$ and generation at lower levels $(\nabla\dot{\theta} < 0)$.

At the critical time when the forecast PV appeared to be declining over the central U.S. the largest contributor to this reduction appeared to be the horizontal diffusion in the model (Fig. 9). This smoothes the PV distribution where the gradients are largest. Other forcing terms such as the convective heating and radiative cooling play a role in the evolution of upper tropospheric PV associated with the cyclone moving northwards from the subtropics.

The model currently has ∇^4 diffusion with an e-folding time of ~1 hour. Kanamitsu and Saha (1995) also found the ∇^4 horizontal diffusion used in NCEP (formerly NMC) MRF model was excessive, causing a loss of rotational kinetic energy in the medium (n=11-40) and small (n=41-80) total wavenumbers. This case study was re-run with more scale selective ∇^6 diffusion with a ~5 hour e-folding time. The ∇^6 forecast produced a deeper low of 974 hPa at day 5, 11hPa deeper than with ∇^4 diffusion but still 12hPa too shallow compared to the model analyses (Fig. 10). The erosion of PV by the horizontal diffusion is greatly reduced resulting in larger PV values in the forecast (Fig. 11). One drawback is some numerical noise in the PV, which the high levels of diffusion are obviously controlling. A semi-implicit semi-Lagrangian version of the Unified Model is currently under development at the Met. Office (Cullen et. al., 1997). This numerical scheme should be more accurate and more stable than the current split-explicit integration scheme, and it is hoped that this will remove the need for high levels of explicit diffusion.

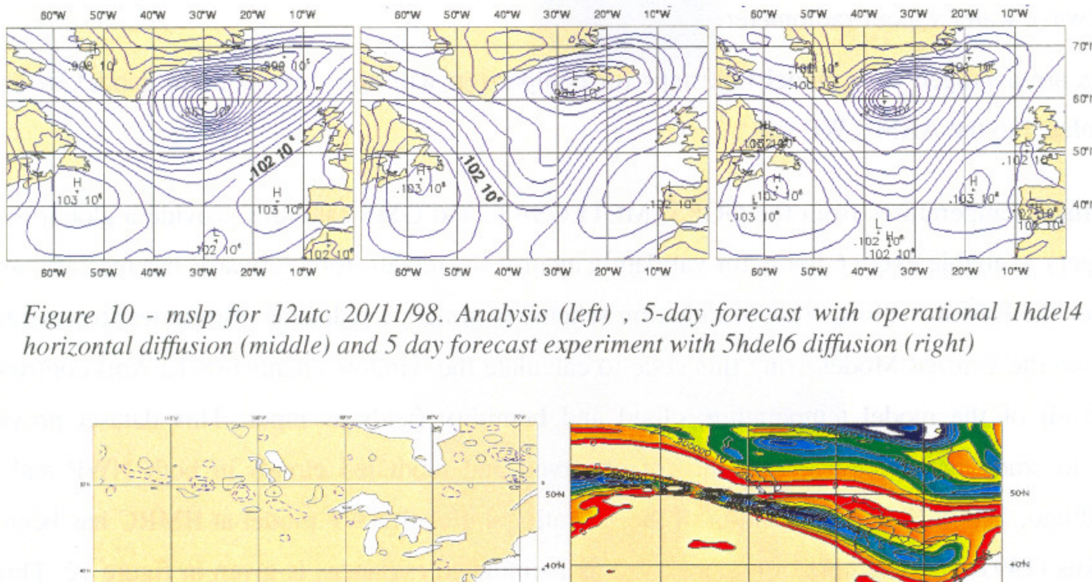


Figure 10 - mslp for 12utc 20/11/98. Analysis (left) , 5-day forecast with operational 1hdel4 horizontal diffusion (middle) and 5 day forecast experiment with 5hdel6 diffusion (right)

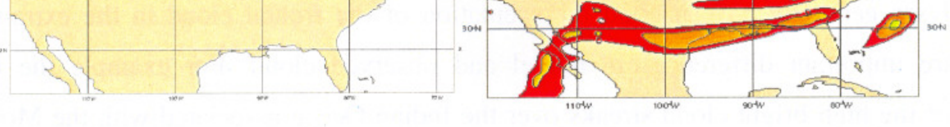


Figure 11 - The 36-42 hour PV tendency from horizontal diffusion (left) and the PV at 42 hours (right) from the experiment with 5hdel6 diffusion.

Of course this is only one case and in general errors in the initial conditions will dominate the error growth in these fast growing systems. However, it does demonstrate how the use of several diagnostic techniques combined with numerical experimentation can point to potential sources of errors in extratropical cyclones related to deficiencies in sub-gridscale physical processes.

4 Validation against Observations - ARM data and IR imagery

To diagnose the potential source(s) of tropical errors we have been using data from the Atmospheric Radiation Measurement (ARM) program's tropical West Pacific (TWP) site and IR brightness temperature data from geostationary satellites to validate the UM short-range forecasts and analyses for a test period during July 1998 (Culverwell and Milton, 2000). Similar comparisons of the climate version of the Unified Model against ARM are also underway.

4.1 Details of observational datasets

The ARM data consist of high frequency measurements from the following instruments active during July 1998

- Surface meteorology sensors
- Radiosondes
- Radiometers (surface SW and LW downwelling and upwelling components)
- Ceilometer and lidar (cloud backscatter profiles)
- Microwave water vapour radiometer.

These data are averaged over model timesteps for comparison with the corresponding model fields, which are interpolated to the ARM TWP site.

The brightness temperatures from the GOES, METEOSAT, and GMS satellites provide a global composite of IR imagery available every 6 hours for validation against the model. John Edwards has recently developed a version of the Edwards-Slingo (1996) radiation scheme to simulate radiances. Brightness temperatures are derived from the Unified Model using this code to calculate the window channel (8-12 Ãm) contribution to OLR with all of the model temperature, cloud and humidity fields as input. This dataset provides the capability to study the synoptic variability of observed and modelled clouds in both NWP and climate models (Slingo, 2000). Similar validation of the operational global NWP model at BMRC has been carried out by Rikus (1997). An example of the observed and simulated radiances is given in figure 12. This shows the model analyses generally have a good representation of the frontal cloud in the extra-tropics. In the tropics there are important differences in model and observed cloud. For example the model has no representation of the high bright cloud streaks over the Indian Ocean associated with the Monsoon outflow into the easterly jet flow.

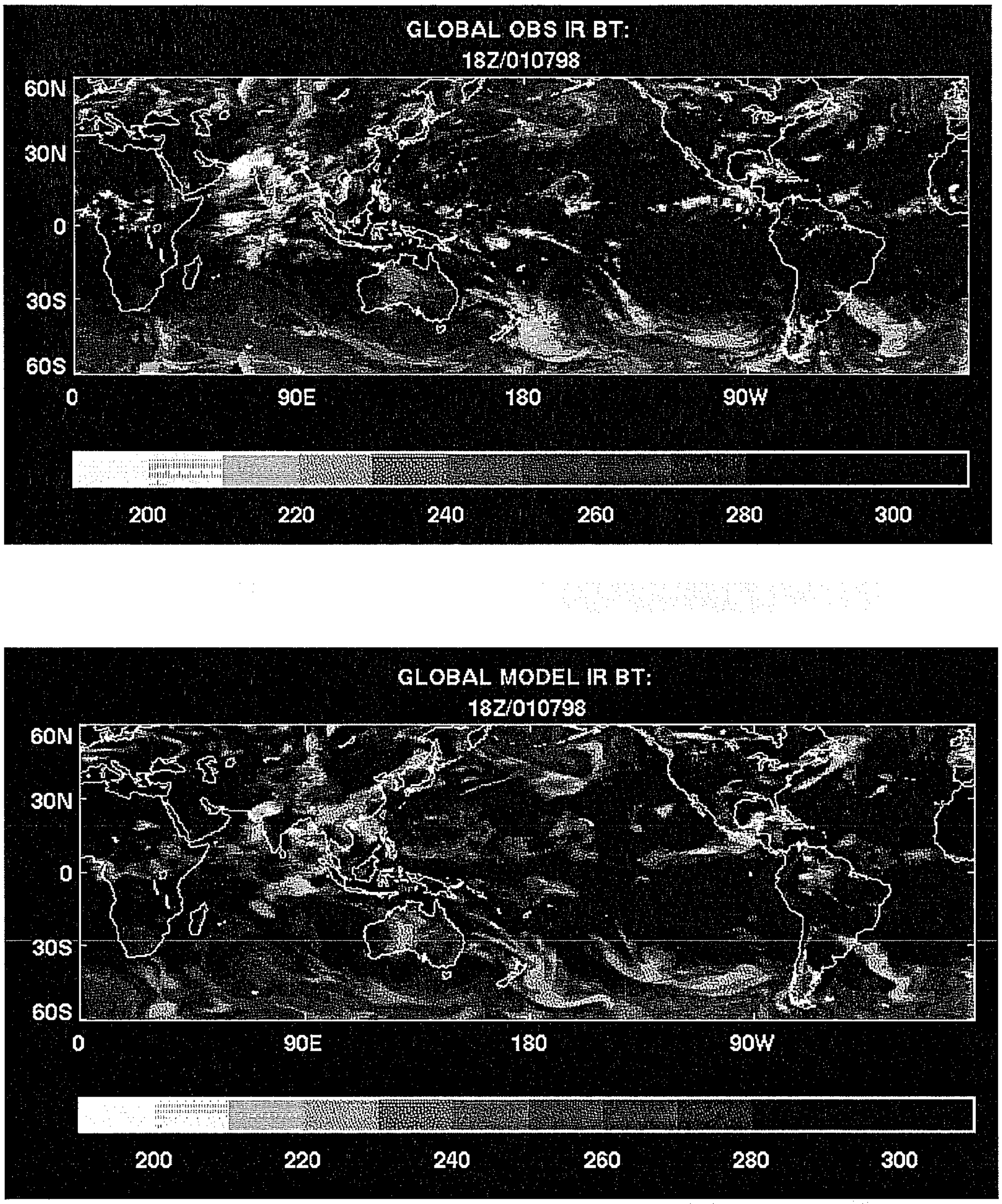
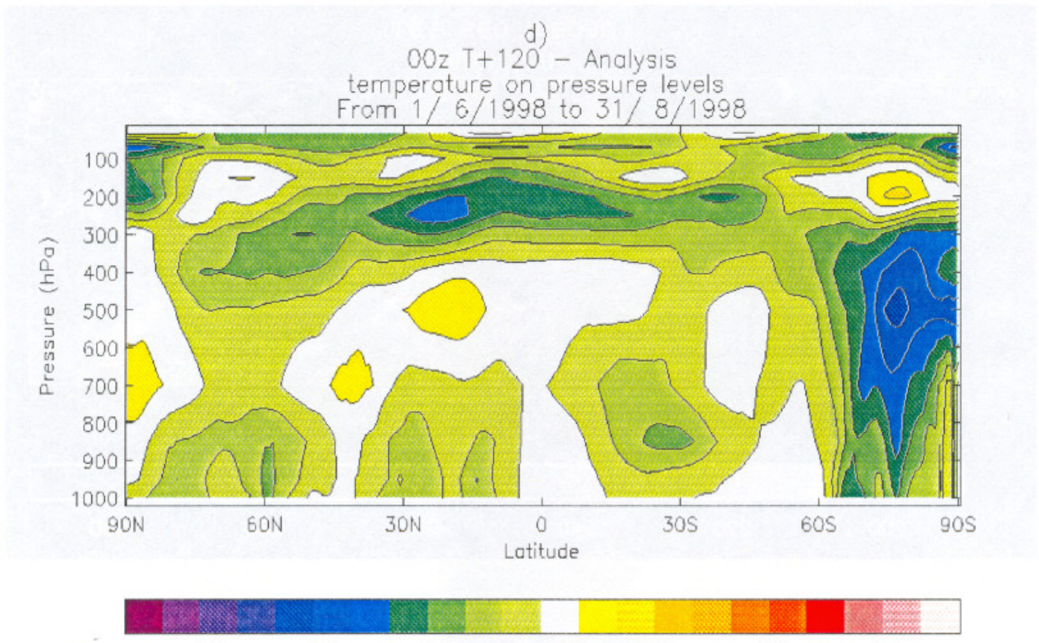


Figure 12 Observed and model (analysis) brightness temperature distributions for 18UTC 01/07/1998.

4.2 Tropical Systematic errors in the UM - Thermal balance

The global Unified Model has some robust tropical systematic errors that have proved difficult to reduce. These include a cold bias at the tropical tropopause and warm bias in the mid-troposphere (Fig.13), increased Hadley and Walker circulation's and a spin-up in tropical precipitation. The comparison of model and sonde temperatures at 00 UTC over the ARM site show both the model analyses and T+24 forecasts have the cold bias in the upper troposphere, warm bias in the mid-troposphere and cold bias in the boundary layer (Fig.14), the same errors that are apparent in zonally averaged temperatures for the whole tropical domain (Fig.13). The thermal balance for the T+00-T+06 forecast period for the whole of July 1998 (Fig.14) shows the budget residual has the same vertical structure as the temperature errors. The largest contributors in the balance are the convective heating and the radiative cooling, which are much larger than the residual tendency. It is



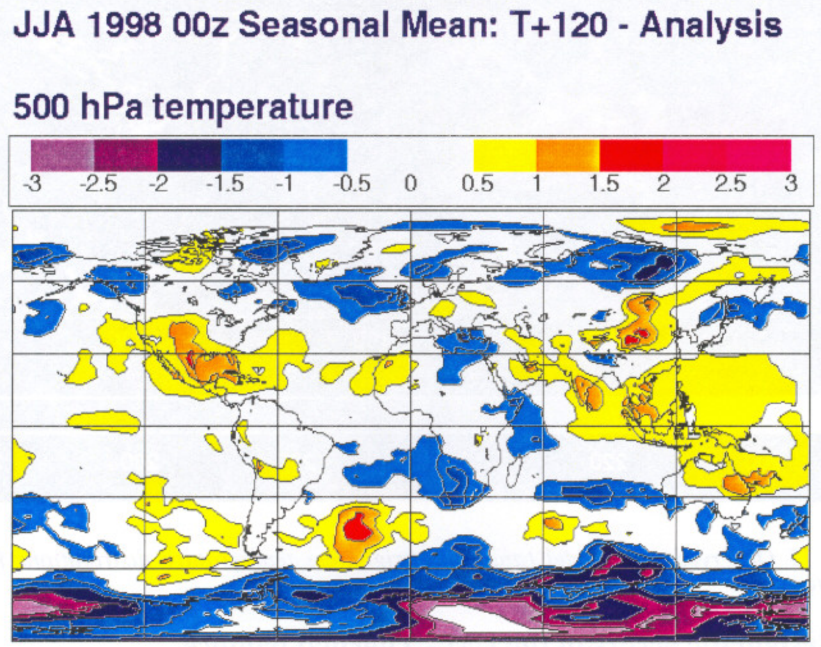


Figure 13 - Systematic errors in Met Office Unified Model day-3 forecasts during JJA 1998 for zonally averaged temperatures (top) and 500hPa temperatures (bottom)

difficult to infer which process is responsible for the errors. One possibility is that the convective heating profile is too shallow, with excessive heating in the mid-troposphere and too little in the tropopause region. Comparisons with earlier versions of the ECMWF model showed that convective heating was generally larger at the tropopause in the ECMWF model and temperature biases smaller (Klinker personal communication). Equilibrium studies with a single column version of the Unified Model initialised and compared with cloud resolving model (CRM) data also suggest that the convective heating profile is too

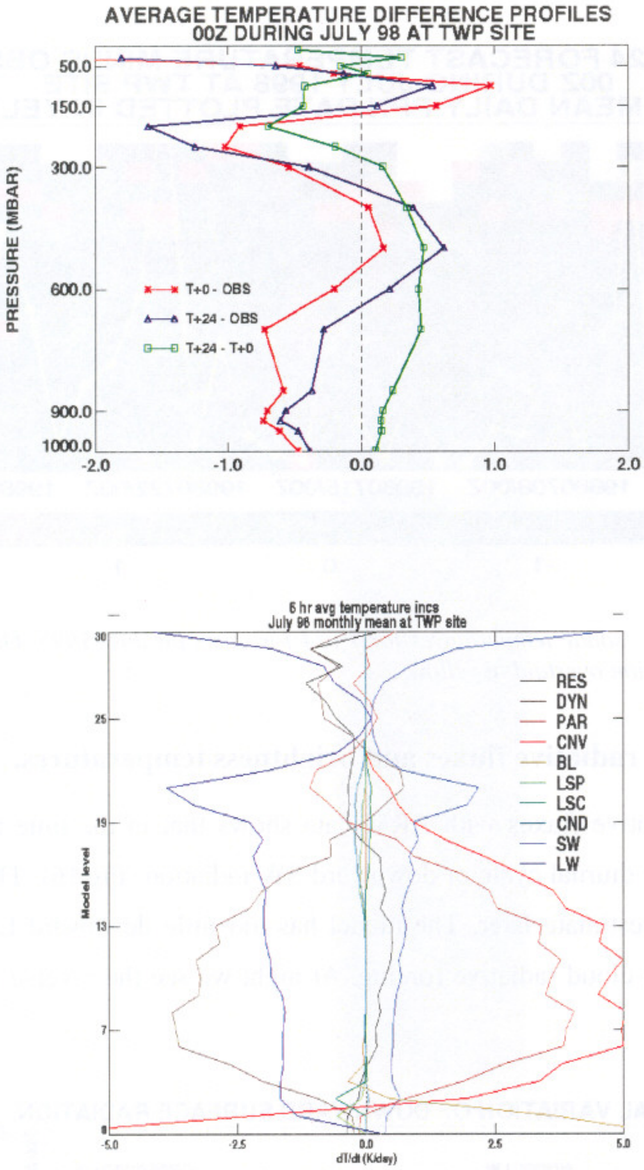


Figure 14 - Temperature errors in analyses and day 1 forecast verified against the 00UTC sonde from the ARM site (top). The Thermal balance in 00-06 hours of the forecast for July 98.

shallow (M. Gray, personal communication). Another possibility is that the LW radiative cooling is too large, perhaps through poor simulation of cloud and humidity at the tropopause. Examination of the time variation of the temperature biases confirms that the upper tropospheric cold bias is a regime dependent error and is largest in cases of strong convection (Fig.15). The time variation of parametrized convective heating and LW cooling over the whole month do not give any clear indication of error sources as the maximum LW cooling and maximum convective heating are coincident in time (not shown).

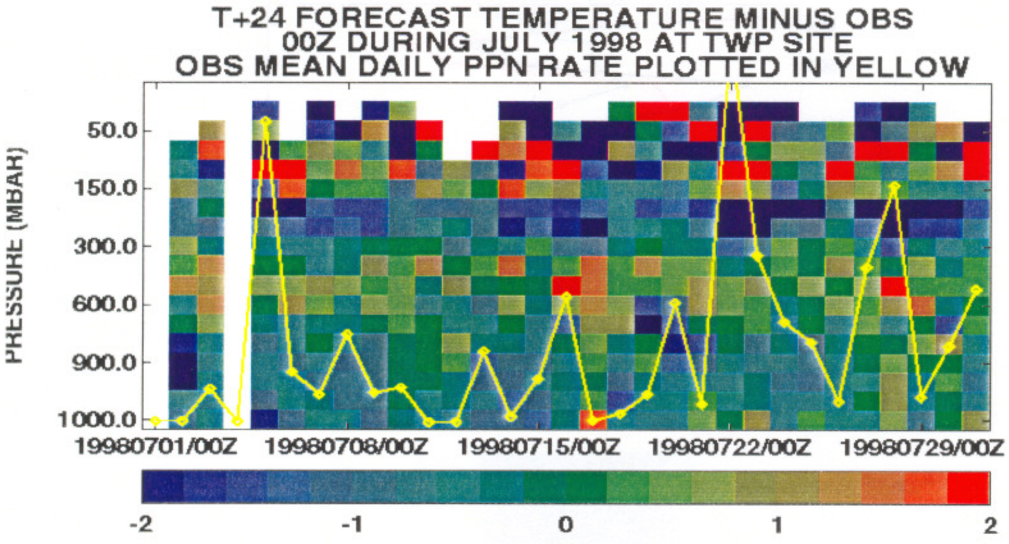


Figure 15 - Model - sonde temperatures for T+24 forecasts for July 1998. Observed daily averaged precipitation overlaid in yellow.

4.3 Validation of surface radiative fluxes and brightness temperatures.

Validation of the surface radiative fluxes with ARM data shows that in the time mean the model does a reasonable job of capturing the diurnal cycle of downward SW radiation (Fig.16). There is an underestimate early in the cycle and an overestimate later. The model has too little downward LW flux during daytime suggesting perhaps insufficient cloud radiative forcing. At night we see the reverse with an overestimate of

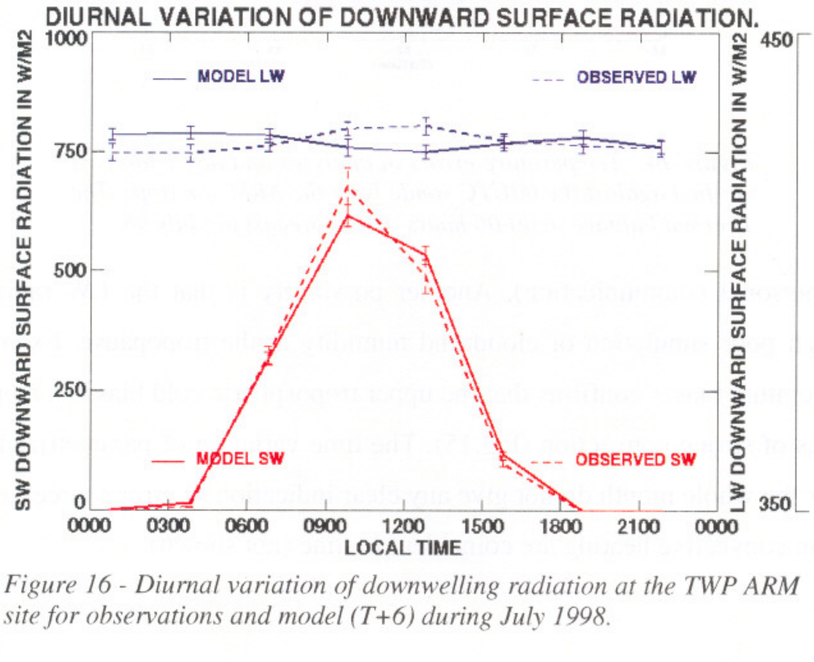


Figure 16 - Diurnal variation of downwelling radiation at the TWP ARM site for observations and model (T+6) during July 1998.

the downward LW flux.

The time-mean picture hides a lot of variability present in the model's surface flux errors. A time series of the 3-hourly averaged SW and LW flux errors (Model - ARM) shows a distinct regime dependence of the errors (Fig.17). During observed heavy precipitation events the model has insufficient cloud radiative forcing. The downward SW flux is overestimated by between 250 and 300 W/m^2 and the LW flux is underestimated by 10-25 W/m^2. The reverse occurs during observed periods of dry/clear weather, with too little downward SW and too much downward LW. The model has excessive cloud cover during these periods. This demonstrates the value of examining the radiative fluxes on synoptic timescales as the time-mean fluxes mask large

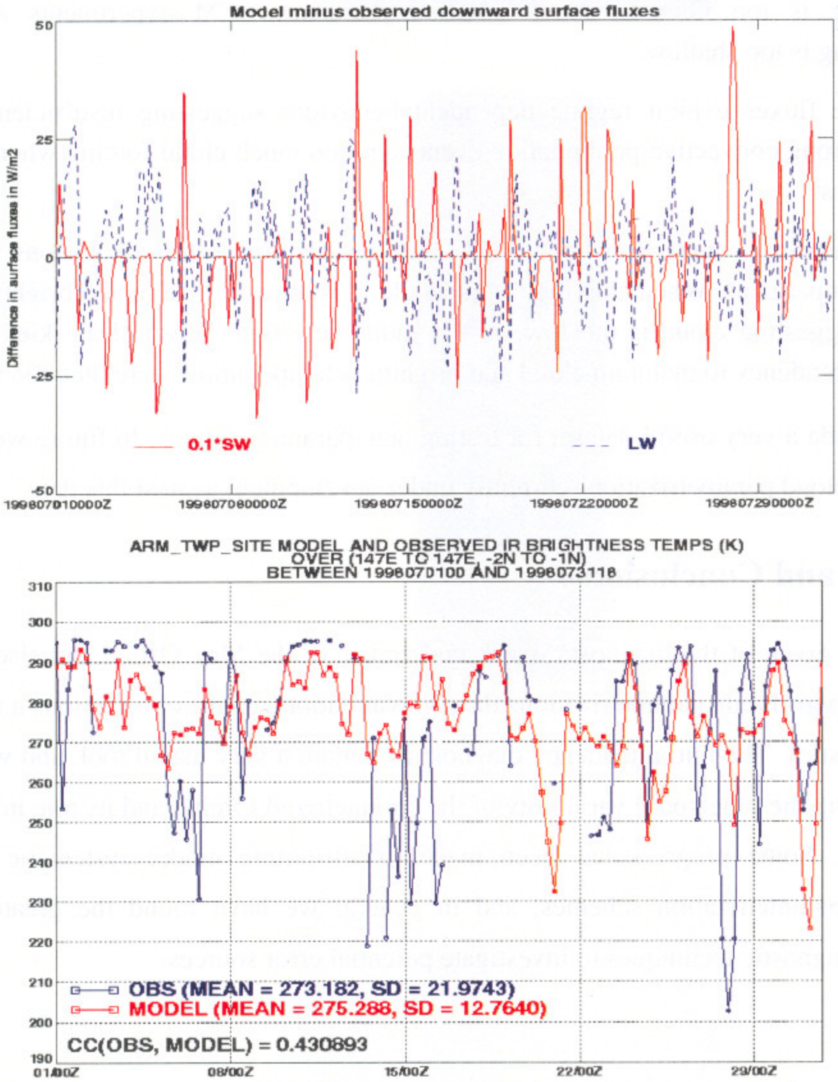


Figure 17 - Timeseries of difference between model (T+6) and observed downward surface radiative fluxes during July 1998 (top) and timeseries of observed and model simulated brightness temperatures at the ARM site every 6 hours during July 1998. (bottom).

compensating errors.

A comparison of the observed and simulated brightness temperatures at the ARM site reveals similar errors to the surface radiative fluxes (Fig.17). During the observed heavy precipitation events the model fails to capture the very low brightness temperatures observed by the satellite and indicative of cold high clouds.

This may suggest that the modelled cloud tops are too low, or cloud is not thick enough, or the modelled radiative properties of the cloud are in error. During clear-sky conditions the model invariably has too low brightness temperatures, showing that there is residual cloud. On average the model underestimates the variability in brightness temperatures.

In summary, comparisons with data at the ARM site and geostationary IR imagery show

- Cold bias at tropopause is episodic and largest during strong convective events
- Thermal balance suggests 2 possible scenarios (I) the convective heating profile is too shallow, (ii) LW radiative cooling is too intense. (cloud-radiation problem). SCM experiments also suggest that convective heating is too shallow.
- Surface radiative fluxes exhibit regime dependent behaviour suggesting insufficient cloud radiative forcing during strong convective precipitation events and too much cloud forcing when the observations suggest clear skies.
- Brightness temperature comparisons with GMS satellite reinforce the conclusions from the surface radiative flux comparisons. When cold high tropical cloud is observed the model brightness temperatures are too warm suggesting cloud is too low, or too radiatively thin. When clear skies are observed the model showed a tendency to maintain cloud and brightness temperatures were then too low.

The ARM sites provide a very useful dataset for testing new parametrizations. In future we hope to validate new convective and cloud parametrizations currently under development against this data.

5 Discussion and Conclusions.

Examples have been given of the diagnostic work undertaken at the Met. Office and elsewhere to try and diagnose errors in sub-gridscale parametrizations through validation against observations and use of a variety of large-scale diagnostics. The initial tendency diagnostics remain a very useful tool, and we hope to extend their use to considering the time-space variability of the parametrized forcing and its role in model errors in a more objective way. Some progress has been made in attributing model systematic errors to single components in the parametrization schemes, and in general we have found the greatest benefit when combining various diagnostic techniques to investigate potential error sources.

References

Arpe, K., Brankovic, C., Oriol, E., and Speth, P., 1986: Variability in time and space of energetics from a long series of atmospheric data produced by ECMWF. *Contrib. Atmos. Phys.*, **59**, 321-355.

Betts, A.K., Ball, J.H., and Beljaars, A.C.M., 1993: Comparison between the land surface response of the European Centre Model and the FIFE-1987 data. *Q.J.R. Meteorol. Soc*, **119**,975-1001.

Broad, A.S. and Oakes, A., 1996. Gravity wave momentum fluxes and surface pressure drags due to the Pyrenees: observations and numerical simulations. In *proceedings of the AMS Seventh conference on Mesoscale processes.*

Chevalier, F. and Morcrette, J.J., 2000: Comparison of Model Fluxes with Surface and Top-of-the Atmosphere Observations. *Mon. Wea. Rev.*, **128**, 3839-3852.

Culverwell, I.D. and Milton, S.F., 2000. An evaluation of the UKMO global NWP model by comparison with ARM data in the tropics. *Forecasting Research Technical Report*, No. 295 (Available from the Met. Office).

Cullen, M.J.P., Davies, T., Mawson, M.H., James, J.A, Coulter, S.C., and Malcolm, A., 1997. An overview of numerical methods for the next generation U.K. NWP and Climate model. In "*Numerical methods in Atmospheric and Oceanic Modelling : The Andre Robert Memorial Volume*", 425-444. NRC Research Press.

Edwards, J.M. and Slingo, A., 1996. Studies with a flexible new radiation code. I - Choosing a configuration for a large -scale model. *Q.J.R. Meteorol. Soc*, **122**, 689-719..

Fortelius, C., 1987. Inferring the diabatic heat and moisture forcing of the atmosphere from assimilated data. *J. Climate*, **8**, 224-239.

Gregory, D., Shutts, G.J., Mitchell, J.R., 1998. A new gravity wave drag scheme incorporating anisotropic orography and low level wave breaking: impact upon the climate of the U.K. Met. Office Unified Model. *Q.J.R. Meteorol. Soc.*, **124**, 463-493.

Holopainen, E., 1987: Recent estimates of diabatic forcing on the planetary scale: a review. In *proceedings of the ECMWF workshop on "Diabatic Forcing"*, 30 Nov - 2 Dec 1987.

Hoskins, B.J., McIntyre, M. and Robertson, A. 1985. On the use and significance of isentropic potential vorticity maps. *Q.J.R. Meteorol. Soc*, **122**, 689-719.

Hoskins, B.J., Hsu, H.H., James, I.N., Masutani, M., Sardeshmukh, P.D., and White, G.H., 1989: Diagnostics of the global atmospheric circulation- based on ECMWF analyses 1979-1989. World Climate Programme, World Meteorological Organization, WMO/TD- No. 326, 217 pp.

James, I.N., 1994. *Introduction to circulating atmospheres*. Cambridge atmospheric and space science series. 422pp.

Kanamitsu, M., and Saha, S., 1995. Spectral budget of the short-range forecast error of the NMC Medium-Range Forecast model. *Mon. Wea. Rev.*, **123**, 1834-1850.

Kiehl, J.T., Hack, J.J., Hurrel, J.W, 1998. The energy budget of the NCAR Community Climate Model: CCM3. *J. Climate*, **11**, 1151-1178.

Klinker, E., and Sardesmukh, P.D., 1987. The diagnosis systematic errors in numerical weather prediction models. In *proceedings of the ECMWF workshop on "Diabatic Forcing"*, 30 Nov - 2 Dec 1987.

Klinker, E., and Sardesmukh, P.D., 1992. The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sc.*, **49**, 608-627.

Klinker, E. and Ferranti, L, 1999. Use of tendency-balance to diagnose the performance of models and data assimilation systems. In *proceedings of the ECMWF seminar on "Diagnosis of models and data assimilation systems"*, 6-10 Sep 1999.

Methven, J., 1997. Offline trajectories: Calculation and Accuracy. *UGAMP Technical Report* 44, UGAMP.

Milton, S.F. and Wilson, C.A. 1996. The impact of parametrized subgrid-scale orographic forcing on systematic errors in a global NWP model. *Mon. Wea. Rev.*, **124**, 2023-2045

Pope, V.D., Jackson, D.R., Pamment, J.A., and Slingo A., 2000. The impact of increasing vertical resolution on the Hadley Centre climate model. To appear in *Climate Dynamics*.

Rabier, F., Klinker, E., Courtier, P., and Hollingsworth, A., 1996. Sensitivity of forecast errors to initial conditions. *Q.J.R. Meteorol. Soc*, **122**, 121-150.

Rikus, L, 1997. Application of a scheme for validating clouds in an operational global NWP model. *Mon. Wea. Rev.*, **125**, 1615-1637.

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

Slingo, A., Allan, R.P., and Ringer, M.A.,2000. Evaluation of climate model simulations of the radiation budget and clouds using satellite and other data. In proceedings WGNE workshop on "Model Systematic Errors", *BMRC Research report* No. 80, 205-211

Stoelinga, M. T., 1996. A potential vorticity based study of the role of diabatic heating and friction in a numerically simulated baroclinic cyclone. *Mon. Wea. Rev.*, **124**, 849-874.

Swinbank, R., 1985. The global atmospheric angular momentum balance inferred from analyses made during FGGE. *Q.J.R. Meteorol. Soc*, **111**, 977-992.

Wernli,H., and Davies, H.C., 1997. A lagrangian based analysis of extratropical cyclones. I- The method and some applications. *Q.J.R. Meteorol. Soc*, **123**, 467-489.