THE DIAGNOSIS OF SYSTEMATIC ERRORS IN NUMERICAL WEATHER PREDICTION MODELS

Ernst Klinker and Prashant D. Sardeshmukh
European Centre for Medium-range Weather Forecasts
Reading, United Kingdom.

ABSTRACT

A new approach to the diagnosis of systematic errors in NWP models is developed and applied to the ECMWF model. This is based on examining the initial tendency of a large number of operational forecasts. Local budget residuals for each of the model's degrees of freedom (each model variable at each gridpoint) are then identified simply as the average initial tendency of these forecasts.

In principle, the partitioning of these precisely determined budget residuals into local data analysis and modelling errors poses a difficult problem. Some progress is possible by assuming that there are systematic data analysis errors only in the divergent component of the flow and that the errors in the other analysed variables are uncorrelated. Ideas as to how one might proceed in case these assumptions are not justified are also presented. In the majority of the examples considered here, however, the patterns of the residuals are already highly suggestive of errors in the parametrized forcing, and are consistent with the acknowledged uncertainties in the parametrization schemes.

The calculations based on data for January 1987 suggest that (i) the vertical diffusion of horizontal momentum is too strong above the boundary layer, (ii) the orographically induced gravity-wave drag is too strong in the lower stratosphere, (iii) there are errors in the boundary layer formulation over the oceans and near mountains over land (which are probably compounded by an error in the specification of the orographic height), (iv) the radiative cooling is too weak in clear skies, (v) the parametrized heating is too weak in the tropical convective regions at 700 mb (which contributes significantly to the tropical spin-up problem), and (vi) the humidity analyses are unreliable in middle latitudes.

Budget studies such as these cannot, of course, in themselves prove these statements. However they do point to a local problem and suggest some likely answers at little cost. We are sufficiently encouraged by these results to recommend that all existing general circulation models be benchmarked by using them first to assimilate a standard set of global observations (say a subset of FGGE) and then to ascertain the degree to which they violate the requirements of balance by performing similar calculations.

1. INTRODUCTION

Medium range weather prediction is now sufficiently advanced that continued progress is increasingly dependent upon improvements in the details of model formulation and the initial state. The major forecasting centres can still differ dramatically in their prognoses of synoptic development in individual cases, but these are usually traced back to differences in current data availability and use. Many features of the average or systematic errors at these centres are surprisingly similar, and are catalogued elsewhere in this volume. Prominent among these are a poleward and upward shift of the subtropical jet streams, a loss of both steady and transient eddy kinetic energy in middle latitudes, a generation of excessively strong easterlies in the tropical upper troposphere, and a weakening of the upper branch of the Hadley circulation. Extensive experimentation at ECMWF has shown such errors to be relatively insensitive to further increases in model resolution. The fact that essentially similar errors are also obtained in the extended forecast range, when the details (and errors) of the estimated initial state are presumably no longer important, suggests that the problem lies with the modelled mechanical and thermal forcing. At ECMWF, sensitivity has indeed been shown to changes in the specification of orography, the inclusion of gravity-wave drag, and the modelling of shallow convection, and some improvement in medium range forecast skill has been reported. The overall characteristics of the systematic errors mentioned above have, however, not changed, and no clear guidance as to their cause or remedy has as yet emerged.

The interpretation of systematic error maps is hampered by the fact that they show an integrated response to errors in remote as well as local forcing. In themselves, the maps only tell us that something is wrong in the forcing, not what or where, unless the patterns are in some way reminiscent of the response forced by mountains or diabatic heating in simple models. Thus Wallace et al. (1983) were able to point to erroneous mountain forcing which eventually led to the introduction of the 'envelope' orography, and Sardeshmukh and Hoskins (1988) to the major impact of the erroneous upper-level convective outflow over the tropical western Pacific ocean. But these are only pointers. The general problem of deducing forcing errors from some integrated measure of the forecast error is a difficult one involving the calculation of a model's adjoint, and has not yet been attempted in the literature.

Systematic errors at shorter forecast ranges are more easily related to local problems, but are increasingly contaminated by observational and data assimilation errors. Some aspects of the error, such as that in the globally averaged precipitation rate, exhibit a fluctuation of large amplitude at short forecast ranges before settling down to a rate of increase more consistent with a gradual approach to the climate error. This 'spin-up' problem arises from a mismatch between the estimate of the initial state provided by the data-assimilation system and that actually desired by the model. In its most general sense, 'initialization' is a process of adjusting the initial state so as to eliminate the spin-up when inserted into the model, and is perhaps best handled within the framework of a continuous four-dimensional data assimilation system. It should however be remembered that the problem can arise just as well from supplying a perfect initial state to an imperfect model as vice versa.

With these considerations in mind, we shall examine here the systematic errors of the operational ECMWF forecasts at the shortest possible forecast range, i.e., after the first time step! These will provide an enormous advantage of isolating local data or model problems, and will be completely equivalent to performing local budget calculations for each of the model's prognostic variables. The budget residual for any variable will be identified simply as the initial tendency for that variable averaged over a large ensemble of forecasts. For perfect data and a perfect model, this tendency would be close to zero. The separate contributions to the budget from the adiabatic and diabatic tendencies will also have been evaluated at the full model resolution, and will be completely consistent with the model's numerical and physical parametrization schemes. The mean adiabatic tendency with its sign reversed will thus represent an estimate of the diabatic forcing as given by the requirement of a balanced budget. This will be compared with the diabatic forcing actually generated by the model's parametrization schemes, i.e., the computed mean diabatic tendency, to obtain a better view of the nature of their difference, the budget residual.

In essence, our aim is to determine precisely the extent to which observations assimilated with a state-of-the-art data assimilation system fail to satisfy the equations of a state-of-the-art general circulation model. The model in question is a 19-level, primitive-equation, global spectral model with triangular truncation at wavenumber 106 (T106) that was operational at ECMWF in January 1987. It uses a hybrid vertical coordinate that gradually changes from a sigma coordinate at the surface to a pressure coordinate above the tropopause. The lower boundary is specified by an 'envelope' orography. Among the physical processes parametrized are large-scale condensation and precipitation, cumulus convection, radiative transfer, and the vertical eddy fluxes of heat, momentum, and moisture. A small amount of horizontal diffusion is included to control noise at the limit of the model's resolution.

The data assimilation is run as an intermittent analysis-initialization-forecast cycle in which a six-hour forecast from the previous initialized analysis is used to provide a first guess for the next analysis. This first guess is updated using observations from all sources that become available within three hours of the analysis time. After preliminary screening and quality control of the data, the apparent corrections to the first guess (defined as observation minus first guess) are analysed using a box-type, multivariate, optimum interpolation scheme (Hollingsworth, 1986) that allows for both first guess and observational errors. The observational errors are assumed to be spatially uncorrelated, and each observation type is ascribed a different expected error. The expected structure of the first guess error is specified with empirically modelled correlation functions that are separable in the horizontal and vertical coordinates. The current 'best' estimate of the correction to the first guess obtained through such an analysis (the so-called 'analysis increment') is thus consistent with both the current observations and the expected observational and first guess errors.

The mass and wind fields are rendered sufficiently out of balance as a result of this process to make some form of initialization necessary; otherwise, an unacceptably high level of gravity-wave noise is generated in the forecast. The method used at ECMWF is diabatic nonlinear normal mode initialization (Wergen, 1986) applied to the five gravest vertical modes. Only the gravity modes are initialized; the Rossby modes and the diurnal and semidiurnal tides are left unchanged. This finally yields a product which is in a suitable form for starting a numerical forecast.

One expects, of course, that errors of observation, analysis, initialization, and model formulation will all contribute in some measure to our precisely determined budget residuals. In line with the assumption that observational errors are mostly uncorrelated, we will assume their role in the large scale budgets to be negligible. As described above the analysis algorithm is designed to eliminate as much of the spatially correlated first-guess error as possible. However, the divergent component of the first guess wind error is not accounted for on the scale of the analysis box; indeed the system does not respond adequately to divergent wind information on larger scales either (Daley, 1985). It is true that different observing systems can give conflicting indications as to what this component of the flow ought to be, and the analysed values are probably still within the range of observational error (Hollingsworth et al., 1988). Nevertheless, the final initialized analysis of the divergent flow is affected significantly by the model's balance through the first guess fields and through the initialization process. Therefore, to the extent that it is associated directly with diabatic forcing, errors in it will reflect errors in the model's forcing. As discussed below the effect of this will be to introduce erroneous adiabatic tendencies that will cancel some, but not all, of the budget residual that would result from the erroneous forcing alone if data were available in sufficient quantities to constrain the divergent flow strongly. One should stress that the observations must be making some impact; otherwise one would not be seeing any residuals at all. One must also not ignore the fact that reasonably accurate medium range forecasts are generated at ECMWF with such observations as are available.

Thus, although the operational ECMWF analyses are probably reliable enough to yield useful estimates of the atmosphere's diabatic forcing as inferred from the adiabatic terms (see for example, Hoskins and Sardeshmukh, 1987; Holopainen, this volume), it is a different matter altogether to use them to deduce errors in the assimilating model's diabatic forcing. A more careful approach is required, the starting point of which is to determine the budget residuals as consistently and accurately as possible. The basic elements of a formalism for partitioning the residuals into analysis, initialization, and model errors will be developed below. We note in passing that any reservations one may have about such an approach would apply equally, if not more, to almost all large-scale budget studies that have been reported in the past, including those conducted with data from field experiments such as GATE. One could equally view those as showing the results of one time-step forecasts made from often crude analyses, paying little heed to aspects of the atmosphere's dynamical balance in which one has confidence, such as its tendency to evolve along some 'slow' manifold. In the case of field experiments such as GATE one has, of course, superior data on a denser observational network. However, no observational data set is ever likely to be so accurate and so complete as not to require at least some analysis and synthesis in order to enable one to compute the adiabatic tendencies, and a discussion of the budget residuals one has introduced through this process will always be necessary. At an operational forecasting centre such as ECMWF such errors will, in any case, be important in their own right.

The calculations presented here were performed using all the initialized analyses (31 x 4 = 124) available for the month of January 1987. Most of the budgets were also evaluated with the first-guess fields for those analyses, in order to get a clearer view of the different changes made to the adiabatic and diabatic terms by the data assimilation increments. In addition, the precise contribution of the (probably suspect) humidity analyses to the heat budget was isolated by evaluating that budget with the analysed and first guess values separately, using the initialized analyses for all other variables.

The zonal mean momentum budget is presented in section 2. The separate contributions from the parametrized vertical diffusion and orographic gravity-wave drag terms are assessed followed by a brief discussion of their possible impact on the analysed mean meridional circulation. The low level momentum budget is discussed in some detail in section 3, as large residuals are seen near mountains as well as over the oceans. The heat budget is presented in sections 4 and 5, concentrating on the major imbalances obtained at 700 mb over the principal areas of tropical convection and also in middle latitudes. Some concluding remarks are made in section 6.

2. THE ZONAL MEAN BUDGET OF ZONAL MOMENTUM

The time-averaged zonal mean momentum equation may be written

$$\overline{\left(\frac{\partial[u]}{\partial t}\right)} = \overline{\left(\frac{\partial[u]}{\partial t}\right)}_{\text{adiabatic}} + \overline{\left(\frac{\partial[u]}{\partial t}\right)}_{\text{diabatic}} \qquad \ldots(1)$$

where square brackets denote a zonal mean and an overbar a time mean. The adiabatic term includes contributions from the zonally averaged steady and transient eddy fluxes of zonal momentum as well as from the advection and Coriolis terms involving the mean meridional circulation. The diabatic term includes all explicitly modelled and parametrized physical effects, such as the vertical diffusion of horizontal momentum and orographically induced gravity wave drag. Their sum (apart from a very small contribution arising from the actual tendency of the zonal flow averaged over the month) represents the mean budget residual. The individual terms in (1) are all evaluated on model coordinate surfaces at the full model resolution. However for ease of viewing and to emphasize the larger scale features we shall show horizontally smoothed versions of these interpolated to standard pressure levels. Also as discussed above we shall show the adiabatic terms with their sign reversed, so that they may be interpreted directly as the 'inferred' physical forcing (inferred, that is, from the adiabatic terms). The explicitly modelled and parametrized physical forcing will be referred to as the 'model-generated' forcing.

Figure 1 displays the budget for $\overline{[u]}$ as a function of latitude and pressure. For convenience the distribution of $\overline{[u]}$ itself is given in the top panel (Fig 1a). As expected the (minus) adiabatic term shown in Fig 1b implies a need for strong drag in the planetary boundary layer, on time scales of a day or less. More surprisingly, it also shows a requirement for damping in the upper troposphere, especially near the jet stream maxima. The somewhat smoother model-generated forcing (Fig 1c) appears at first sight to be largely consistent with this; however, inspection of the budget residual (Fig 1c minus Fig 1b) in Fig 1d actually reveals quite significant differences, both in the boundary layer and in the free atmosphere above. Interpreted purely as reflecting errors in the parametrized forcing, Fig 1d suggests that the drag on the low-level flow is not strong enough and that on the upper tropospheric and lower stratospheric flow in middle latitudes is too strong. The detailed structure in the tropical upper troposphere and lower stratosphere is also apparently linked to similar features in the model generated forcing. The residuals are generally smaller in the southern hemisphere, and follow the pattern of the adiabatic term closely above the boundary layer.

The contribution of the parametrized vertical diffusion and the gravity wave drag terms to the model-generated forcing shown in Fig 1c is given in Figs 2a and 2b respectively. The contribution of the former is clearly overwhelming.

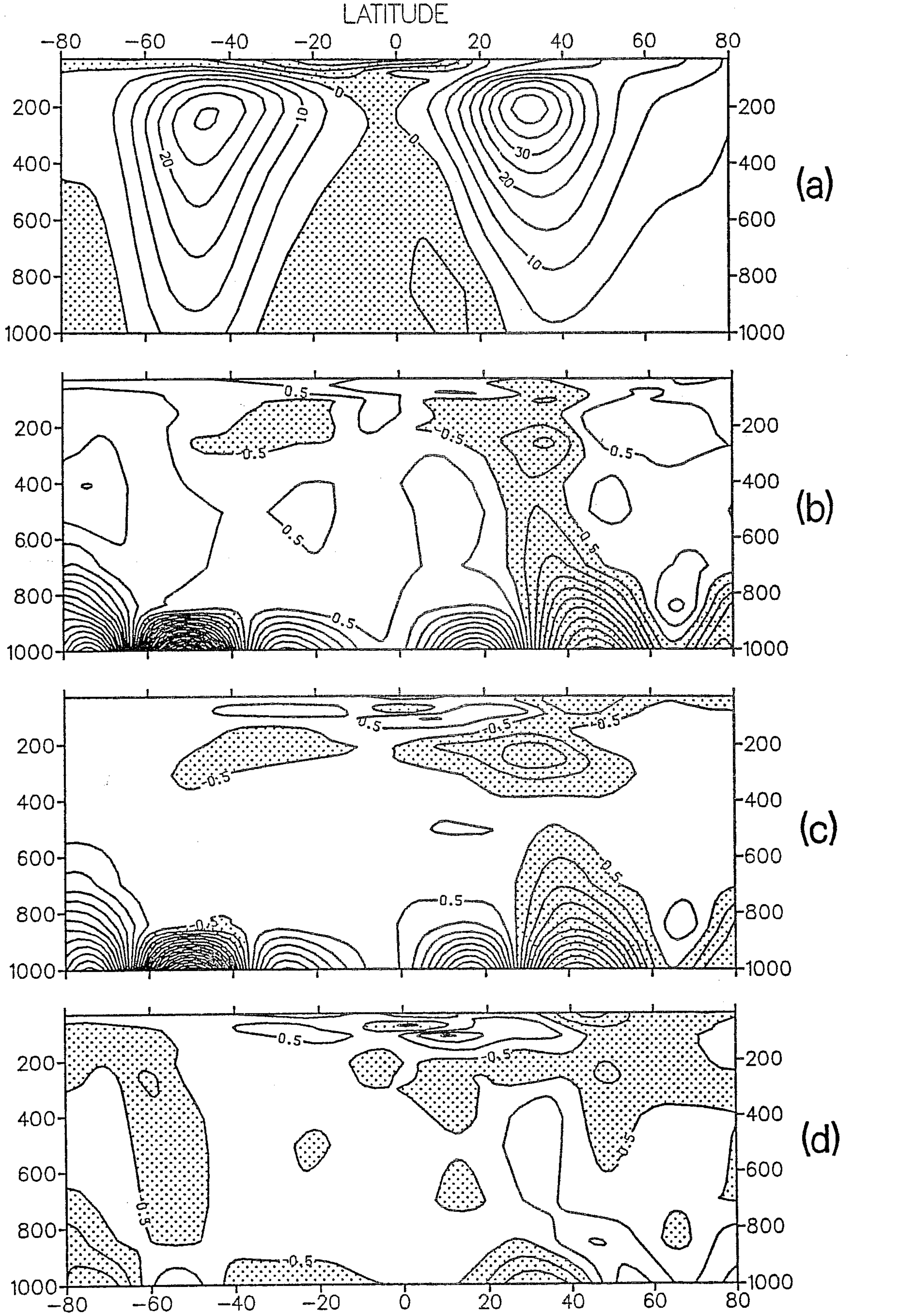


Fig. 1 The budget for the zonal mean zonal momentum for January 1987. (a) $[\bar{u}]$ in m/s. Easterlies are indicated by shading (b) the negative of the mean adiabatic tendency (c) the mean diabatic tendency, and (d) the budget residual ((c) minus (b)). Contours drawn are ±0.5, ±1.5, ±2.5, ... m/s/day and negative values are indicated by shading.

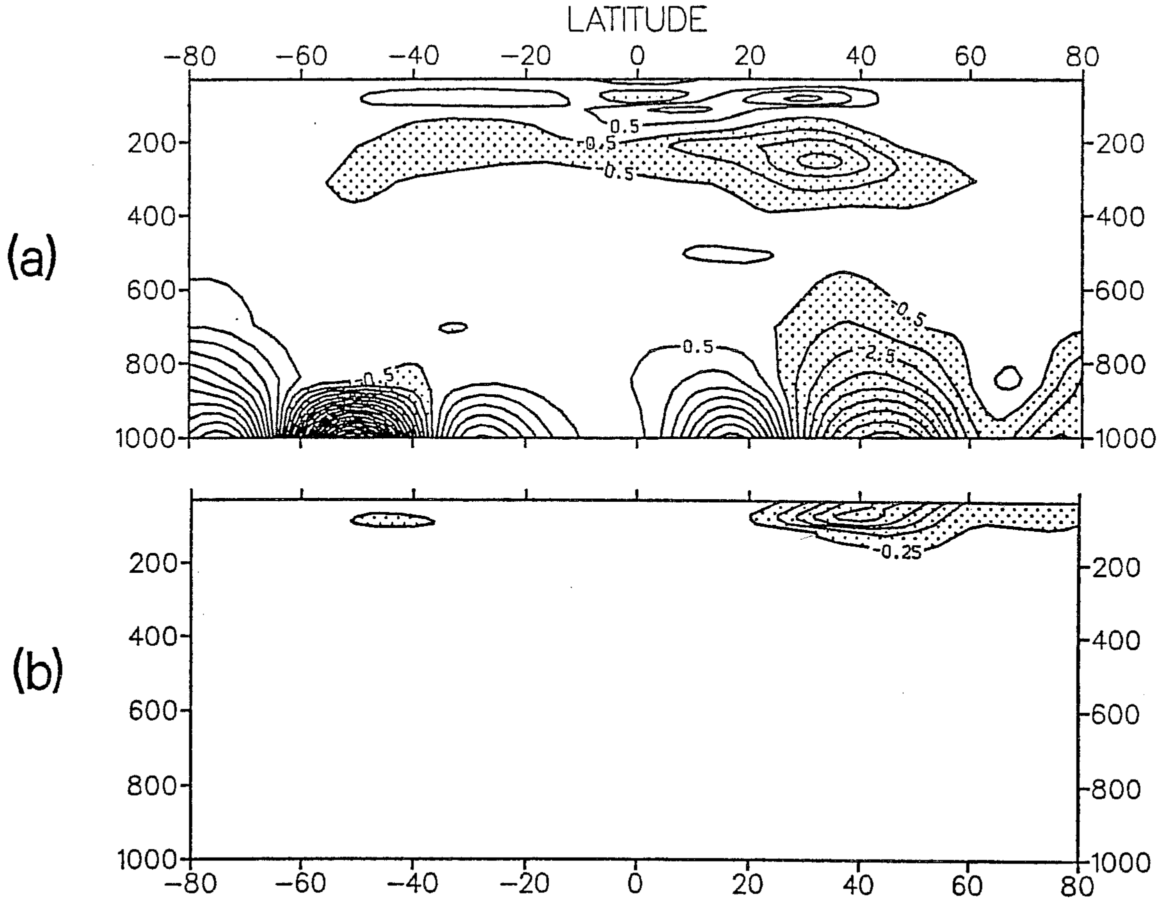


Fig. 2 The contribution to the mean diabatic tendency in Fig. 1c of (a) the vertical diffusion of horizontal momentum and (b) the orographically induced gravity wave drag. Note that the contour interval in (b) is half that in (a).

Gravity wave drag as modelled at ECMWF is apparently a minor player in the general circulation of the atmosphere. Its main impact is in the middle latitude lower stratosphere, and even there it is inconsistent with the balance requirement in the momentum budget. Note also that consistent with a tendency for excessive smoothing of the vertical profile of the zonal flow, the plot of the budget residual shows easterly acceleration at the jet maximum and westerly acceleration above and below. Thus the vertical diffusion as parameterized also appears to be too strong.

Further justification for these statements comes from examining the horizontal distribution of the model forcing and the budget residual at 70 mb (Figs 3a and 3b). Easterly tendencies associated with gravity wave drag are seen clearly over the Rockies and the Tibetan Plateau. The smoothing tendency of the vertical diffusion scheme is manifested as westerly acceleration in the jet stream regions of the subtropical western Pacific and Atlantic oceans. Both of these features are seriously at odds with the requirement from the

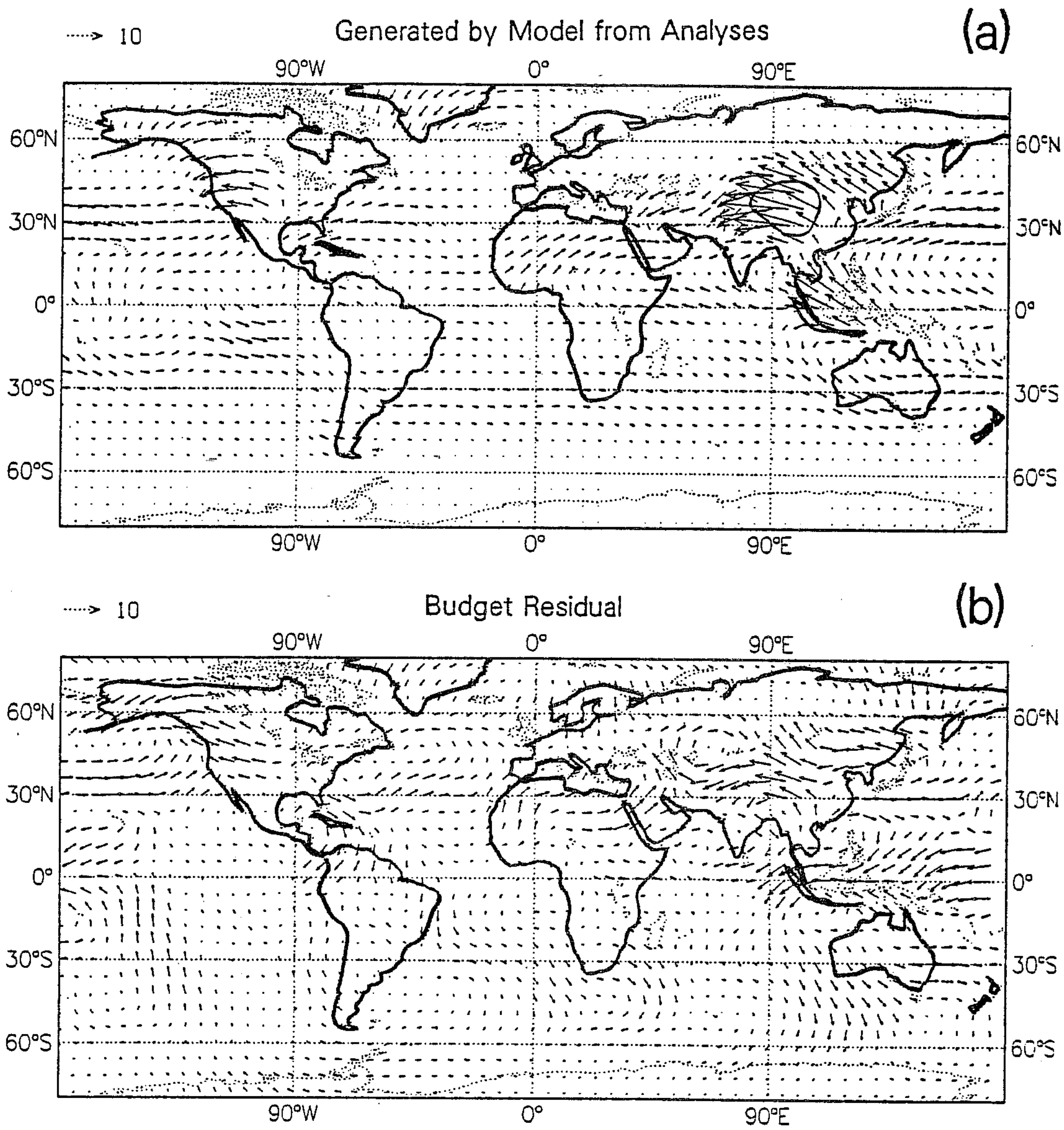


Fig. 3 (a) the mean diabatic tendency of horizontal momentum at 70 mb, and (b) the momentum budget residual at 70 mb. The arrow on the top left represents 10 m/s/day.

adiabatic term, with the result that the map of the budget residual closely resembles that of the model generated forcing. Inspection of the maps at the jet level and below confirms the points already made. And finally, the six hour forecasts when verified against actual observations (and not just analyses) show errors that are consistent with these errors in the forcing.

The errors in the model forcing are probably even greater than is apparent from Fig 1d, because the data assimilation generates compensating errors in $\overline{[v]}$ in the absence of sufficient data to constrain the analysis of the divergent flow. To see this consider a simplified diagnostic equation for the streamfunction ψ of the mean meridional circulation under Boussinesq, quasi-geostrophic scaling:

$$f_o^2 \, \psi_{zz} + N^2 \, \psi_{yy} = - f_o \, \overline{[u^*v^*]}_{yz} - (g/\theta_o) \, \overline{[v^* \theta^*]}_{yy}$$

$$+ f_o \, \overline{[M]}_z + (g/\theta_o) \, \overline{[Q]}_y \qquad (2)$$

Here ψ is such that $\overline{[v]} = - \psi_z$ and $\overline{[w]} = \psi_y$. N is the Brunt-Väisälä frequency and θ_o a standard value of the potential temperature θ. The asterisks denote deviations from a zonal mean, and $\overline{[M]}$ and $\overline{[Q]}$ refer to the zonally averaged momentum and heat sources and sinks. For meridional and vertical length scales $L \sim 1000$ km, $H \sim 4$ km, values of

$\overline{[u^*v^*]} \sim 50 \ m^2 s^2$, $\overline{[v^* \theta^*]} \sim 40 \ ms^{-1} K$, $\overline{[M]} \sim 5 \ ms^{-1}$ day, and $\overline{[Q]} \sim 4K \ day^{-1}$

will all provide comparable forcing for ψ in middle latitudes. The observed typical values of

$\overline{[u^*v^*]} \sim 50 \ m^2 s^2$, $\overline{[v^* \theta^*]} \sim 25 \ ms^{-1} K$, $\overline{[M]} \sim 2 \ ms^{-1}$ day, and $\overline{[Q]} \sim 1K \ day^{-1}$
(from Fig 1c)

show the predominant role of the poleward fluxes of heat and momentum in the dynamics of the Ferrel cell. The question here is about the role of $\overline{[M]}$. If the values in Fig 1c are representative they imply a significant forcing of ψ in a sense opposite to that of the fluxes which has hitherto largely been ignored. The use of diabatic initialization at ECMWF ensures that the initialized ψ is roughly consistent with an equation such as (2); indeed it is consistent with a higher-order balance (see Leith, 1980). An error $\delta \overline{[M]}$ in [M] therefore implies an error $\delta\psi$ in ψ given approximately by

$$f_o^2 \, (\delta \, \psi)_{zz} + N^2 \, (\delta \, \psi)_{yy} = f_o \, (\delta \, \overline{[M]})_z \, .$$

The two terms on the left hand side are in the ratio

$$\frac{N^2 \, (\delta \, \psi)_{yy}}{f_o^2 \, (\delta \, \psi)_{zz}} = \frac{N^2 \, H^2}{f_o^2 \, L^2} = B$$

For $N \sim 1.1 \times 10^{-2}\ s^{-1}$, $f_o \sim 10^{-4}\ s^{-1}$, this gives $B \sim 0.2$. Thus

$$f_o^2\ (\delta\ \psi)_{zz} \sim f_o\ (\delta\ \overline{[M]})_z/(1+B)\ ,$$

which, apart from a barotropic error in $\overline{[M]}$, gives

$$\delta\ f_o\ \overline{[v]}\ \sim -\ \delta\ \overline{[M]}/(1+B).$$

Thus in the momentum budget

$$[u]_t \simeq -\ \overline{[u^*v^*]}_y + f_o\ \overline{[v]} + \overline{[M]}\ ,$$

the residual $[u]_R$ arising mainly from errors in these three terms is given by

$$\begin{aligned}[u]_R &\simeq -\ \delta\ \overline{[u^*v^*]}_y + \delta\ f_o\ \overline{[v]} + \delta\ \overline{[M]}\ ,\\ &= -\ \delta\ \overline{[u^*v^*]}_y + \left(\frac{B}{1+B}\right)\ \delta\ \overline{[M]}\ .\end{aligned}$$

The net contribution to the residual from a baroclinic error in $\overline{[M]}$ is reduced by a factor $B/(1+B)$ which for $B \sim 0.2$ is only 0.16. In effect $\delta\ \overline{[M]}$ induces an erroneous secondary circulation which, given that (2) is elliptic, tends to spread the tendencies from $\delta\ \overline{[M]}$ in the vertical. In the limit of vanishingly small B this spreading would be totally efficient; one would then see only the barotropic part of $\delta\ \overline{[M]}$ in the residual. However, B is not quite zero, not even for very large scale motion in middle latitudes, and so a carefully evaluated budget such as ours yields useful information about the dominant errors in the baroclinic part of $\overline{[M]}$ as well.

The evidence thus points to a need for reducing the vertical diffusion of horizontal momentum above the boundary layer, and by an amount considerably greater than a naive interpretation of Fig 1d would suggest. A revised version of the diffusion scheme giving practically zero tendencies above the boundary layer has subsequently been used operationally at ECMWF since January 1988 (Miller and Viterbo, personal communication). This has led to some improvement in forecast skill and a reduction in the systematic error; in particular the loss of eddy kinetic energy in the forecast has been reduced.

This is consistent with the fact that the removal of any process which internally redistributes horizontal momentum is tantamount to removing a net dissipative mechanism for the domain integrated kinetic energy.

The zonal mean momentum budget for January 1988 (not shown) confirms that the revised formulation of the vertical diffusion has greatly alleviated the upper tropospheric problem apparent in Fig 1d. However, the opposite sign of the budget residual now obtained at the jet level suggests that the change may have been too drastic. The character of the residual elsewhere is unchanged. Interestingly, the adiabatic term in January 1988 is similar to Fig 1b, apart of course from a significant reduction in amplitude at the jet level as expected from the argument given above. The requirement of a damping from the parametrized effects as indicated by these terms is evidently not entirely spurious, although what form it should take is not clear. Perhaps an adjustment of the horizontal diffusion, to make it more consistent with the downgradient potential vorticity fluxes associated with the enstrophy cascade to unresolved scales, is in order. Note that the meridional structure evident in Fig 1d below the jet level along the sharply inclined isentropic surfaces is not inconsistent with this!

3. THE LOCAL BALANCE OF HORIZONTAL MOMENTUM AT LOW LEVELS

As with equation (1) the local budget equation for horizontal momentum may be written

$$\overline{\left(\frac{\partial \underset{\sim}{v}}{\partial t}\right)} = \overline{\left(\frac{\partial \underset{\sim}{v}}{\partial t}\right)}_{\text{adiabatic}} + \overline{\left(\frac{\partial \underset{\sim}{v}}{\partial t}\right)}_{\text{diabatic}} \qquad (3)$$

with the individual terms interpreted in a similar manner. The adiabatic term includes the advection, Coriolis, and pressure gradient contributions (plus a term involving the logarithm of surface pressure), with all the physical effects grouped together in the diabatic term. As before we shall interpret the adiabatic term with its sign reversed as the 'inferred' diabatic forcing, and the diabatic term itself as the 'model generated' forcing.

To avoid complications arising from the interpolation of boundary layer values to standard pressure levels we shall discuss this budget as evaluated on model levels. We shall also concentrate on the flow averaged over the depth of the boundary layer (i.e. model levels 15 to 19), recognizing that the wind

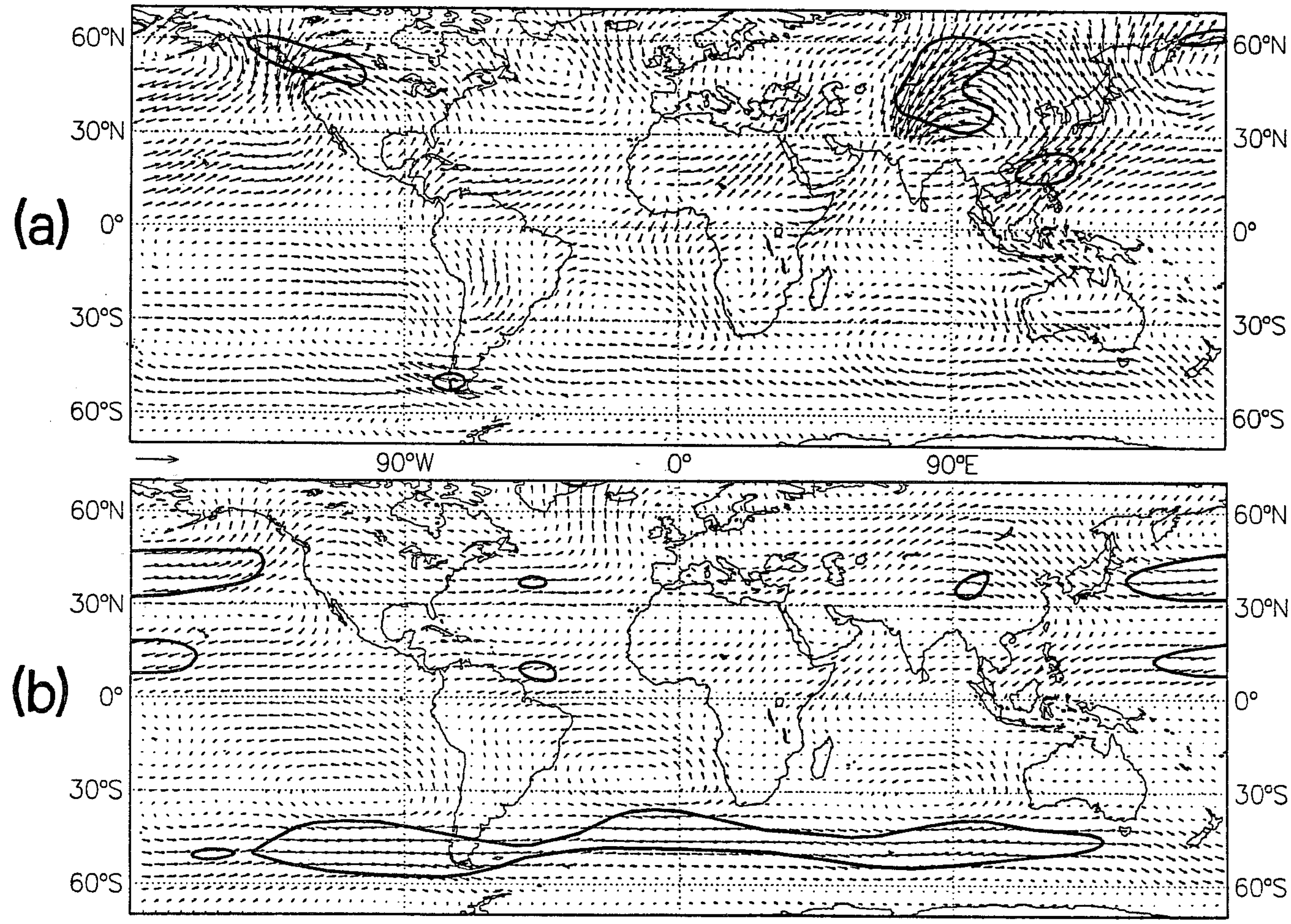


Fig. 4 (a) the negative of the mean adiabatic tendency of the horizontal flow averaged over the depth of the boundary layer, and (b) the mean horizontal flow averaged over the same layer. The arrow shown between the panels denotes 20 m/s/day in (a) and 40 m/s in (b). The contour shown denotes 10 m/s/day in (a) and 10 m/s in (b).

observations are probably not reliable enough to resolve the detailed boundary layer structure and that the pressure observations by themselves only constrain the analysis of the geostrophic component of the flow.

Figure 4a gives the inferred forcing of this depth averaged flow. For comparison the depth averaged flow itself is given in Fig. 4b. As expected the adiabatic terms indicate the need for a drag acting on the time scale of 1-2 days. It is encouraging that the analyses are reliable enough to suggest a need for stronger drag over land (particularly near the mountains) than over the oceans. The pattern of the parametrized drag is broadly similar to this and is therefore not shown. We show instead its difference from Fig. 4a, i.e. the budget residual, in Fig. 5a (note the magnified scale). Wherever these arrows are in the same (opposite) sense to those in Fig. 4a, the parametrized drag is stronger (weaker) than that dictated by balance requirements. Over large areas of the tropics equatorward of about 30 degrees latitude, the arrows are generally opposite; however, the situation elsewhere is less

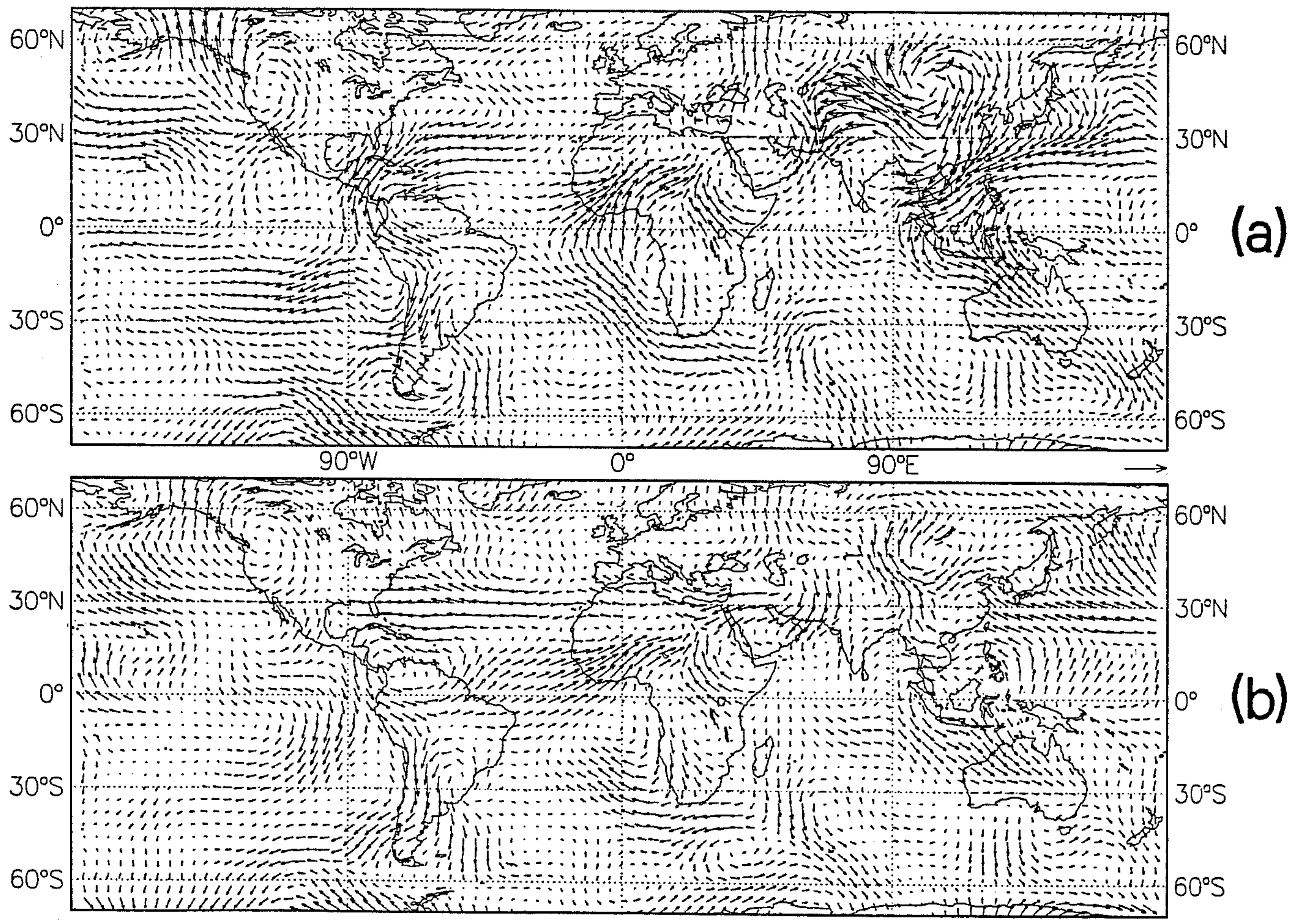


Fig. 5 (a) The mean momentum budget residual averaged over the depth of the boundary layer, and (b) the six-hour forecast errors averaged over the same layer, and also expressed in m/s/day. The arrow shown denotes 5 m/s/day in both panels.

clear-cut. In particular the indications from the Pacific and Atlantic storm track regions are in apparent conflict, as are those from the Rockies and the Tibetan Plateau.

Given the variations in data density and quality as well as the difficulty in determining the ageostrophic component of the flow, Fig. 5a is clearly difficult to interpret as it stands. Another complicating factor to bear in mind is the highly nonlinear stability dependence of the parametrized drag.

Figure 5b shows that the systematic errors of the 6-hour forecasts, verified against the initialized analyses, are entirely consistent with the initial tendency errors depicted in Fig. 5a. In regions of large errors there is a noticeable Coriolis turning of the winds to the right (left) in the northern (southern) hemisphere. This has implications for the determination of the ageostrophic motion in the subsequent analysis.

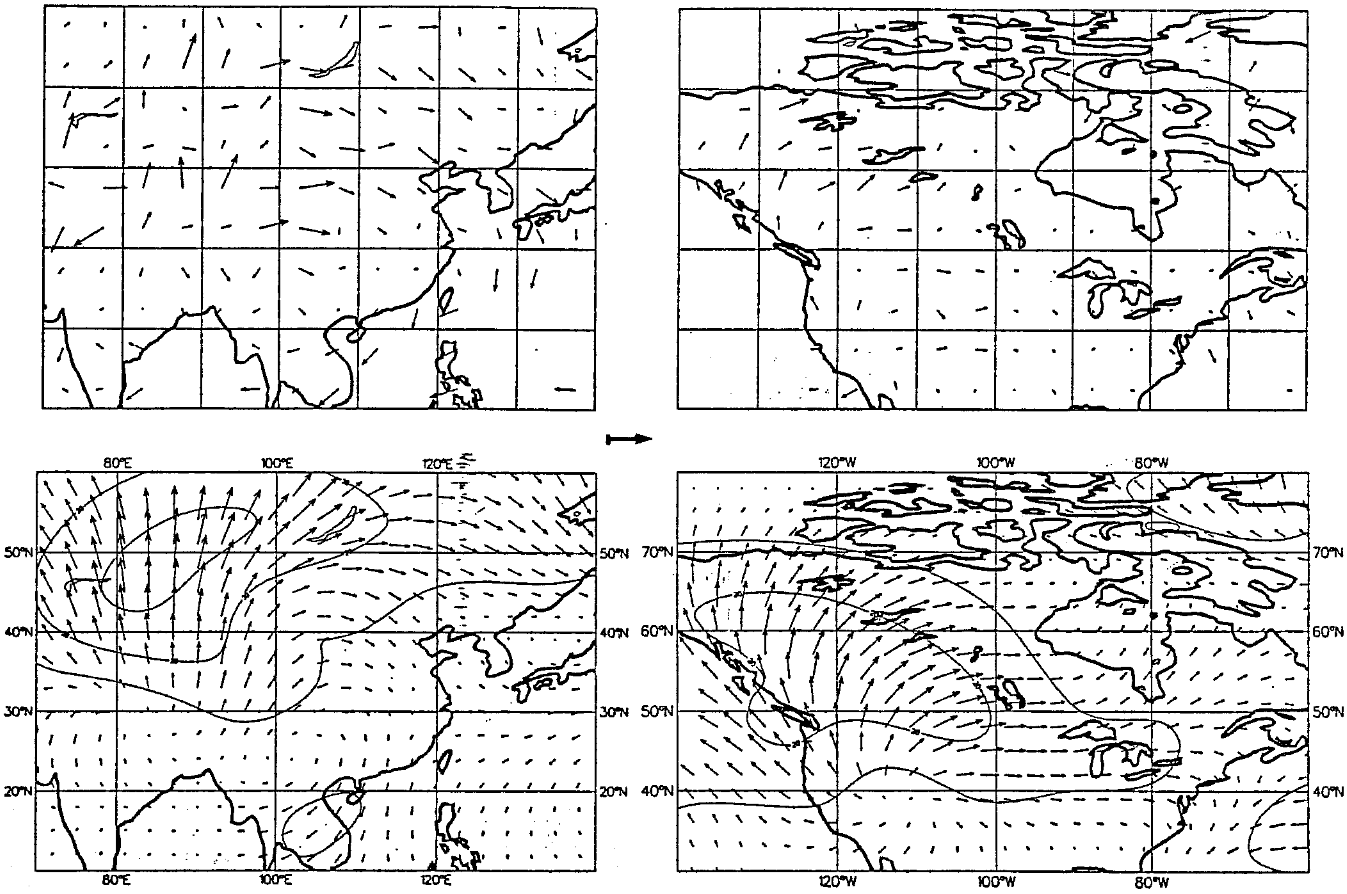


Fig. 6 The top panels show the 6-hour 10 metre wind forecast error verified against actual observations grouped in 5°x5° boxes. The mean initial tendency error at model level 19 is given at the bottom. The arrow in the middle of the figure represents 6 m/s in the top panels and 40 m/s/day in the bottom.

Note that the demarcation line between easterly and westerly drag in Fig. 4a lies slightly poleward of the latitude separating the tropical easterlies from the midlatitude westerlies (see Fig. 4b). The parametrized drag (not shown) follows the pattern of the low level winds rather more closely, with the result that the residual in Fig. 5a shows easterly forcing errors around 30°N and 35°S. This signal is large enough to appear even in the zonal mean budget (see Fig. 1). The problem probably arises from an erroneously determined ageostrophic flow, which contributes significantly to the values in Fig. 4a. However, in view of the inability of the observations to constrain the analysis of this component in the shallow boundary layer, it appears likely that the error originates in the model itself.

When considering the budget residual near mountains, an additional complication with the adiabatic terms is the large cancellation between the horizontal gradients of geopotential and the logarithm of surface pressure when evaluated on sharply inclined coordinate surfaces:

$$\left(\frac{\partial \underset{\sim}{v}}{\partial t}\right)_{adiabatic} \sim -\hat{k} \times (f+\xi)\, \underset{\sim}{v} - \underset{\sim}{\nabla}_\sigma \left(\phi + \frac{|\underset{\sim}{v}|^2}{2}\right) - R\,T\, \underset{\sim}{\nabla}_\sigma\, \ell n p_s \qquad (4)$$

If the slope of the orography is such as to imply a drop in surface pressure of 100 mb in 1000 km, the tendency from the last term is of order 1000 m/s/day ! The net pressure gradient is a very small difference between this and the second term on the right hand side of (4), so a small error in the analysis of the surface pressure or low level temperature could give rise to spuriously large adiabatic tendencies. However, the process of initialization should take care of this, particularly since the error would then be felt through the bulk of the troposphere.

The six hour forecasts when verified against actual observations over land show systematic errors that are consistent with these initial tendency errors. Figure 6 gives close-ups of these over East Asia and North America. Note that the large values of order 20 m/s/day in the plots of the residuals are confined to model levels 18 and 19 in the boundary layer in these regions; the values above the boundary layer are much smaller.

We have argued that a large mismatch between the surface pressure and geopotential fields is unlikely; but is it possible that the orography itself has been wrongly specified? The 'envelope' part of the orography may essentially be viewed as a parametrization of subgrid scale orographic effects, and it is of interest to determine how an error in this would manifest in our budget. Defining reference surface pressure and air temperature distributions $\tilde{p}_s$ (x,y) and $\tilde{T}_s$ (x,y) such that

$$\underset{\sim}{\nabla}_\sigma\, \phi_s + R\, \tilde{T}_s\, \underset{\sim}{\nabla}_\sigma\, \ell n \tilde{p}_s = 0$$

and assuming that the deviations p_s' and T_s' of the actual values from these are small, we obtain at a sigma level infinitesimally above the surface:

$$\underset{\sim}{\nabla}_\sigma\, \phi + R\,T\, \underset{\sim}{\nabla}_\sigma\, \ell n p_s \simeq -\frac{1}{\tilde{\rho}_s}\, \underset{\sim}{\nabla}_\sigma\, p_s' - (\rho_s'/\tilde{\rho}_s)\, R\, \tilde{T}_s\, \underset{\sim}{\nabla}_\sigma\, \ell n \tilde{p}_s$$

$$\simeq -\frac{1}{\tilde{\rho}_s}\, \underset{\sim}{\nabla}_\sigma\, p_s' - \left(\frac{p_s'}{\tilde{p}_s} - \frac{T_s'}{\tilde{T}_s}\right) R\, \tilde{T}_s\, \underset{\sim}{\nabla}_\sigma\, \ell n \tilde{p}_s$$

If the specification of the envelope is significantly in error, $\delta(R\tilde{T}_s \underset{\sim}{\nabla}_\sigma \ln\tilde{p}_s)$ could be as large as 500 m/s/day. The terms in the brackets would ordinarily be expected to be small (of order 1/100) and tend to cancel, but perhaps not in the very stable boundary layers that occur in the cold ridges near mountains ($p'_s > 0$, $T'_s < 0$); so a contribution to the residual in Fig 6 arising purely from an excessively high mountain is possible. The problem is probably aggravated in the boundary layer by the erroneously analysed values of p'_s and T'_s generated mainly by an erroneous first guess there.

Thus in spite of inevitable errors in the observations and in their assimilation, there is evidence of serious deficiencies in the boundary layer formulation both over the oceans and land. It is difficult to make more precise statements based purely on such time-averaged statistics, given that the parametrized boundary layer fluxes have a highly nonlinear dependence on the flow Richardson number and hence on the prevailing synoptic situation. We have concentrated here on the time mean budgets in order to minimize the difficulty in estimating the local tendencies of momentum from the analyses. However, in view of the large residuals obtained in this as well as other budgets, our strategy appears in retrospect to have been overly cautious. A very considerable amount of useful information is contained in the behaviour of the budget terms with time, which will be extracted and discussed elsewhere.

In any event, regardless of whether one examines the budget terms in space or in time, one is in general going to face the problem of determining if the residuals arise from errors in the adiabatic or the diabatic terms. Furthermore, even in cases where one knew the output of a parametrization scheme to be at fault, one would still have to decide if the problem lay with the inputs to that scheme or with the scheme itself. In the following we give some ideas as to how one might proceed. The basic idea is to exploit the fact that the adiabatic and diabatic terms have a different functional dependence on the same inputs. Thus a valuable clue is provided by the fact of the budget residual following, more or less, the pattern of one term or the other. This thinking was implicit in our interpretation of Fig 3 for example. What follows is simply a formalization of this concept.

Let us suppose for illustration that the budget equation for a prognostic variable η consists of three terms

$$\eta_R = \eta_a + \eta_b + \eta_c$$

$$= (\eta_{a\ true} + \delta\eta_a) + (\eta_{b\ true} + \delta\eta_b) + (\eta_{c\ true} + \delta\eta_c)$$

$$= \delta\eta_a + \delta\eta_b + \delta\eta_c \tag{5}$$

$$= \eta_a \left(\frac{\delta\eta_a}{\eta_a}\right) + \eta_b \left(\frac{\delta\eta_b}{\eta_b}\right) + \eta_c \left(\frac{\delta\eta_c}{\eta_c}\right)$$

where the δ's denote errors as before and η_R refers to the budget residual. Assuming that there are no errors in at least one term, say η_c (so that $\delta\eta_c=0$) we attempt to explain the structure of η_R in space or time with the regression equation

$$\eta_R \sim \tilde{\eta}_R = \eta_a\alpha_a + \eta_b\alpha_b \tag{6}$$

The regression coefficients α_a and α_b are alternatively interpreted as probable fractional errors in η_a and η_b respectively. An optimal 'least-squares' solution is generated by minimising the penalty

$$J(\alpha_a, \alpha_b) = \overline{(\eta_R - \tilde{\eta}_R)\ (\eta_R - \tilde{\eta}_R)} \tag{7}$$

with respect to α_a and α_b. Here the overbar refers to a spatial or temporal average. This gives two equations in the two unknowns α_a and α_b

$$\begin{aligned} \overline{\eta_a\eta_a}\ \alpha_a + \overline{\eta_a\eta_b}\ \alpha_b &= \overline{\eta_R\eta_a}\ , \\ \overline{\eta_a\eta_b}\ \alpha_a + \overline{\eta_b\eta_b}\ \alpha_b &= \overline{\eta_R\eta_b}\ , \end{aligned} \tag{8}$$

whose solution may be expressed in the form

$$\begin{aligned} \alpha_a &= \left(\frac{\overline{\eta_R\eta_R}}{\overline{\eta_a\eta_a}}\right)^{\frac{1}{2}} \left\{\frac{\varepsilon_{Ra} - \varepsilon_{Rb}\ \varepsilon_{ab}}{1 - \varepsilon_{ab}^2}\right\}\ , \\ \alpha_b &= \left(\frac{\overline{\eta_R\eta_R}}{\overline{\eta_b\eta_b}}\right)^{\frac{1}{2}} \left\{\frac{\varepsilon_{Rb} - \varepsilon_{Ra}\ \varepsilon_{ab}}{1 - \varepsilon_{ab}^2}\right\}\ , \end{aligned} \tag{9}$$

where ε_{Ra}, ε_{Rb}, and ε_{ab} are the correlations between η_R and η_a, η_R and η_b, and η_a and η_b respectively. The fractional variance of η_R explained by this solution is

$$\varepsilon = \frac{\overline{\tilde{\eta}_R \tilde{\eta}_R}}{\overline{\eta_R \eta_R}} = \varepsilon^2_{Ra} \left\{ \frac{1 + \tilde{\tilde{\varepsilon}}^2_{ab} - 2\tilde{\tilde{\varepsilon}}_{ab} \, \varepsilon_{ab}}{1 + \varepsilon^2_{ab} - 2\varepsilon_{ab} \, \varepsilon_{ab}} \right\} , \tag{10}$$

where $\tilde{\tilde{\varepsilon}}_{ab} = (\varepsilon_{rb}/\varepsilon_{ra})$. This provides a valuable a posteriori check on the assumptions made in (6). If ε is not close to 1 the split (6) is not useful and one would have to repeat the analysis with another regrouping of terms on the right hand side. Also if the residual η_R is a very small difference between η_a and η_b then $\varepsilon_{ab} \sim -1$ and (10) reduces to

$$\varepsilon \sim \varepsilon^2_{Ra} \tag{11}$$

Note that in this limit the two equations in (8) are not quite linearly independent and the expressions in (9) approach an indeterminate form, making their evaluation sensitive to sampling errors.

If ε is close to 1 one knows that one is on the right track and the α's determined from (9) give some indication of the relative contributions of the terms in (5) to the error budget. Suppose now that a major contributor turns out to be a 'model-generated' term, say η_b, which is of the form

$$\eta_b = \gamma q$$

where q is a model variable representing the 'input to the parametrization scheme' and γ a model parameter representing the 'scheme itself'. In this case (6) becomes

$$\eta_R \sim \tilde{\eta}_R = \eta_a \, \alpha_a + \eta_b \, \alpha_q + \eta_b \, \alpha_\gamma ,$$

and minimising the penalty J with respect to α_a, α_q and α_γ gives three equations instead of two as in (8), but two of which are now identical. Thus one cannot determine α_q and α_γ separately, only their sum $\alpha_b = \alpha_q + \alpha_\gamma$ from (9). In effect this states that one cannot distinguish between an error in γ and an error in q in η_b.

The way out of this dilemma is either to consider another equation in which γ or q appears, or, if a q also appears in η_d, to rewrite (6) as

$$\eta_R \sim \tilde{\eta}_R = (\eta_a - \eta_{aq})\, \alpha_a{}' + (\eta_{aq} + \eta_b)\, \alpha_q + \eta_b\, \alpha_\gamma$$

and repeat the analysis (6) - (10).

Thus one is able to progress beyond a simple interpretation of the budget residuals in a quantitatively useful manner. This technique will be developed further in Klinker, Hollingsworth, and Sardeshmukh (1988). Preliminary calculations have shown promise and are generally in the sense of confirming the conclusions drawn above. The simple approach outlined here can readily be adapted to a variety of other situations, such as for extracting information about possible errors in the surface stress formulation by considering the momentum budget integrated through the depth of the atmosphere.

4. THE HEAT BUDGET

Figure 7 gives zonal mean cross-sections of the diabatic heating

(a) as inferred from the time-averaged adiabatic terms determined using the initialized analyses, and

(b) as generated by the model's parametrization schemes using those same initialized analyses as input.

The units are K/day. The inferred pattern shows, as expected, positive values in the tropics associated with organized cumulus convection with a maximum in the middle to upper troposphere. The negative values in the subtropics are associated mainly with radiative cooling in the descending branch of the Hadley cell. This being the northern winter, the significant asymmetry between the northern and southern hemispheres is also as expected. The positive values at 500 mb near 40 N are associated with latent heat release in the Pacific and Atlantic storm track regions. Further aloft as well as poleward lie regions of cooling, determined here mainly as requiring to balance the upward and poleward transport of heat by the baroclinic eddies.

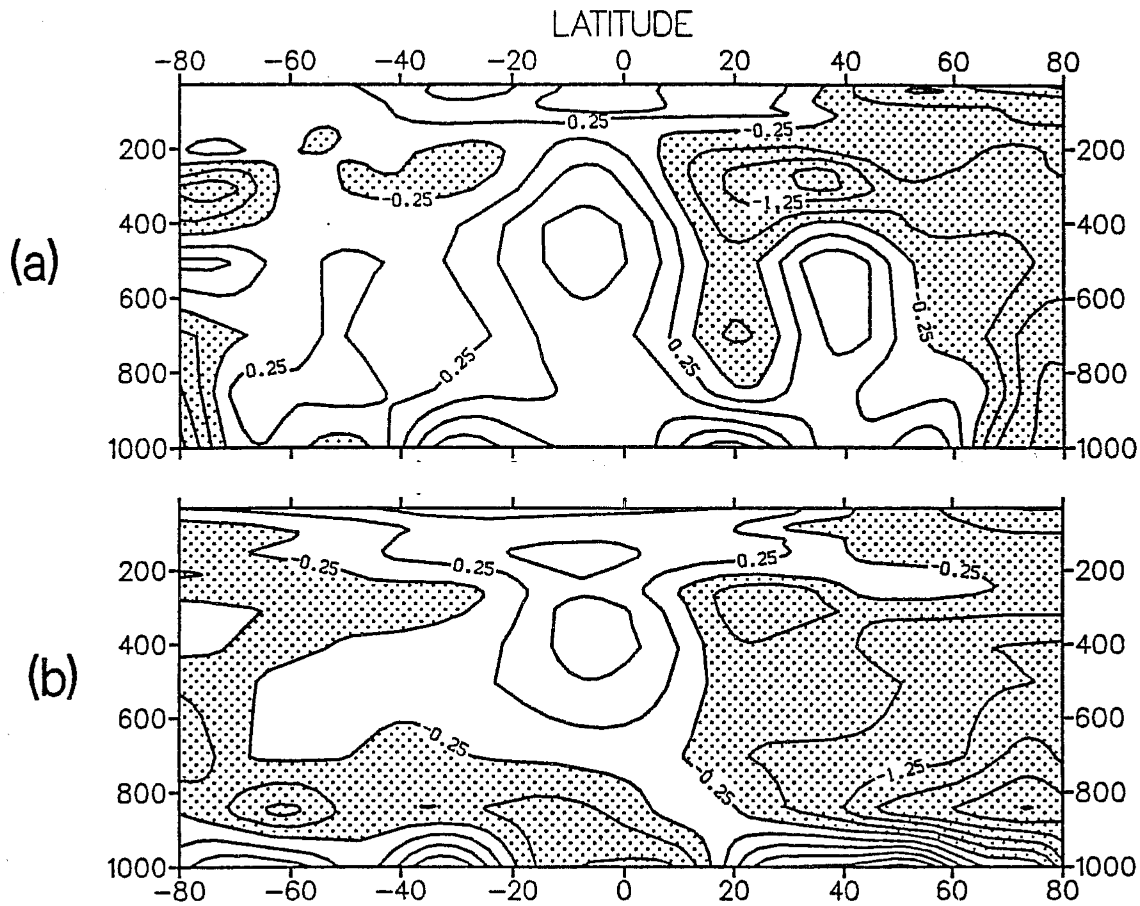


Fig. 7 The zonal mean temperature budget, evaluated using the initialized analyses. (a) the negative of the mean adiabatic tendency (b) the mean diabatic tendency. The contours drawn are ±0.25, ±0.75, ±1.25, ... K/day and negative values are indicated by shading.

Diabatic heating in the tropics is balanced mainly by the adiabatic cooling of ascent. This is entirely consistent with (2), since the term involving the heating is by far the largest on the right hand side, and has to be balanced mainly by the second term on the left hand side given that the Burger number $B = (NH/fL)^2$ is now large. The pattern of diabatic heating inferred from the adiabatic terms thus depends crucially upon the initialized analysis of the vertical motion. In the absence of sufficient quantities of data in the tropics, this is influenced strongly by the model's parametrization of cumulus convection, both through the use of a model-generated first guess in the analysis algorithm and again through the use of a model-generated diabatic heating field in the initialization scheme. One might therefore have expected the pattern of the model generated heating in Fig 7b to be similar if not identical to that in Fig 7a. This is clearly not the case. At midtropospheric levels the tropical convective heating, the subtropical radiative cooling, and the midlatitude storm-track heating are all underestimated by the model. Elsewhere there are large discrepancies below 700 mb. The error of sign in the

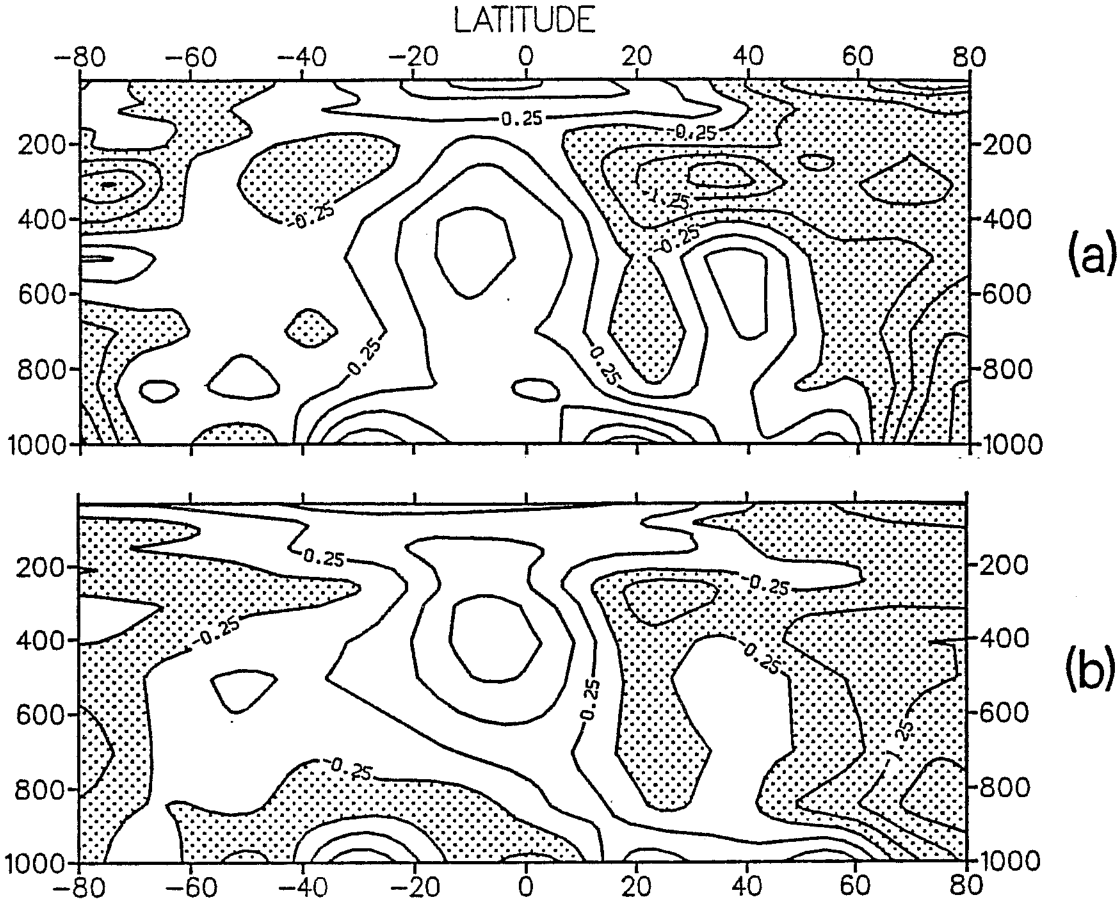


Fig. 8 As in Fig. 7 but evaluated using the first guess fields.

tropical lower troposphere is particularly striking. The boundary layer fluxes given by the parametrization schemes also appear to be too large in the subtropics and in middle latitudes. In fact it is only in the data sparse regions of the southern hemisphere above 700 mb that the differences can be said to be small.

Given their 'on'-'off' character, the model's parametrization schemes are sensitive to small errors in the input data structures implied by the analyses, and it might be argued that many of these discrepancies are of little or no consequence beyond the first few time steps of a forecast. Figure 8, which shows the inferred and model generated terms as calculated with the first guess fields, however demonstrates otherwise. As with Figs 4 and 5 the differences between the inferred patterns in Figs 7 and 8 are very minor (but look above 100 mb), and the model generated terms are also similar. The one major exception to this occurs in the subtropical and middle latitude boundary layer, where it seems that small differences between the surface temperature and that analysed at the lowest model level lead to erroneously large boundary layer fluxes in Fig 7b which quickly redress the imbalance.

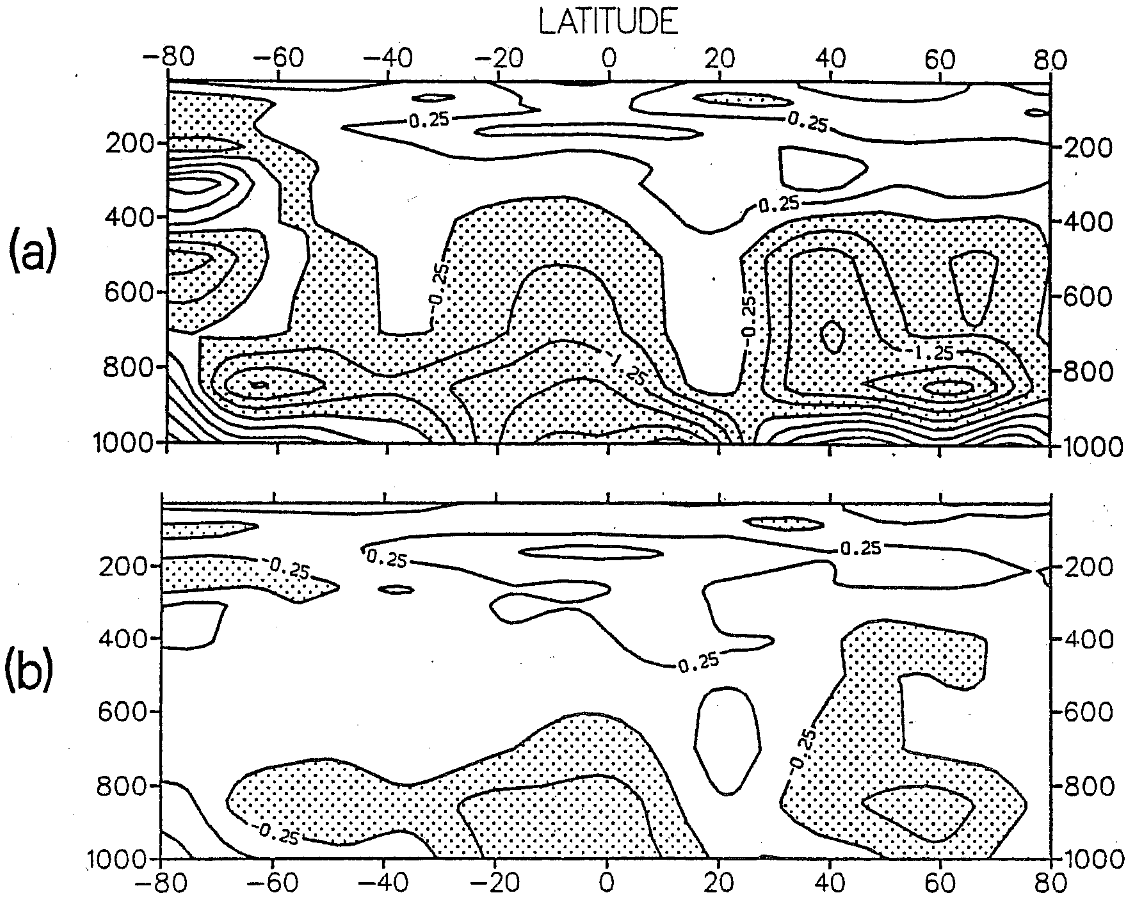


Fig. 9 From top: (a) the zonal mean temperature budget residual (Fig. 7b minus Fig. 7a) and (b) the six-hour forecast error, expressed in units of K/day. Contouring as in Fig. 7.

The budget residual calculated with the initialized analyses, Fig 7b minus Fig 7a, is shown in Fig 9a. The same quantity calculated with the first guess fields (not shown) is very similar except in the extratropical boundary layer as discussed above. Interpreted purely as reflecting errors in the model-generated heating, Fig 9 indicates that the tropical convective heating, the subtropical radiative cooling and the midlatitude storm track heating are all underestimated by the model. The positive residuals near the tropopause suggest, inter alia, that the vertical diffusion of temperature is too strong. The pattern of the systematic six-hour forecast error shown in Fig 9b is again largely consistent with these initial tendency errors.

These imbalances are not small, and would appear even more serious if the masking effect of the initialization were to be taken into account as in section 2. It is interesting to consider the implications of the principal balance in (2) for our heat budget residuals, given an error $\delta\,[Q]$ in $[Q]$ and an error $\delta\,[M]$ in $[M]$ and assuming no errors in the analysis of the momentum and heat fluxes. For B large in the tropics and small in middle latitudes we have approximately

$$\left(\frac{1}{B} + 1\right) N^2\ \delta\,[w]_y \simeq (g/\theta_o)\ \delta\,[Q]_y \tag{12}$$

in the tropics and

$$-\ (1 + B)\ f_o^2\ \delta\,[v]_z \simeq (g/\theta_o)\ \delta\,[Q]_y + f_o\ \delta\,[M]_z \tag{13}$$

in middle latitudes. Assuming that the main errors in the zonally averaged heat budget are associated with the terms

$$[\theta]_R \simeq -\ (\theta_o\ N^2/g)\ \delta\,[w] + \delta\,[Q] \tag{14}$$

and proceeding in an analogous manner to section 2 gives

$$[\theta]_R \sim \left(\frac{1}{1 + B}\right)\ \delta\,[Q]' + \delta\,[\tilde{Q}]\ \pm \frac{f_o\theta_o}{g}\ \left(\frac{B}{1 + B}\right) \frac{L}{H}\ \delta\,[M]' \tag{15}$$

in the tropics or in middle latitudes. Here $\delta\,[Q]'$ and $\delta\,[M]'$ refer to the meridionally and vertically varying parts of $\delta\,[Q]$ and $\delta\,[M]$ respectively and $\delta\,[\tilde{Q}]$ denotes the meridionally constant part of $\delta\,[Q]$. Thus in the absence of sufficient data to constrain the analysis of the divergent flow, the initialization scheme generates an erroneous divergent circulation associated with the erroneously parametrized thermal and momentum sources which affects our residuals. The term involving $\delta\,[M]'$ is small in the tropics because both $\delta\,[M]'$ and f_o are small. Given also that B is large, much of the detailed meridional structure in $\delta\,[Q]$ is masked there as well; the large meridional scales apparent in Fig 9a are consistent with this. There is less of a tendency to filter out the meridional structure of $\delta\,[Q]$ in middle latitudes; however the picture there is complicated by the large error in $[M]$. Assuming that the large momentum sink at the jet level displayed in Fig 2a is entirely spurious, the negative (positive) sign in front of the last term in (15) would obtain equatorward (poleward) of the latitude of the maximum momentum sink. This would be in the sense of masking the radiative cooling underestimate

(δ [Q] positive) in the subtropics and exaggerating it (to a lesser extent, because B is small) in high latitudes. Numerically, though, one expects these effects to be small.

The masking effect of the initialization on the zonal mean residuals can thus be estimated qualitatively. (Calculations enabling a more quantitative assessment and without the restrictive assumptions of quasi geostrophy are also feasible, and are being planned). The difficulty arises mainly from the wind observations not being sufficiently abundant to constrain by themselves the determination of the zonal mean divergent flow. This is less of a problem in the local budgets, at least on the large horizontal scales considered here. One should again stress that despite much of the model's balance being continually imposed on the analyses through the use of strict data-rejection criteria, a model-generated first guess, and the initialization process, such observations as are available give rise to significant residuals in our budgets. Admittedly one ends up looking more at the tip of the iceberg than the iceberg itself, but a careful evaluation of the residuals can nonetheless yield crucial information.

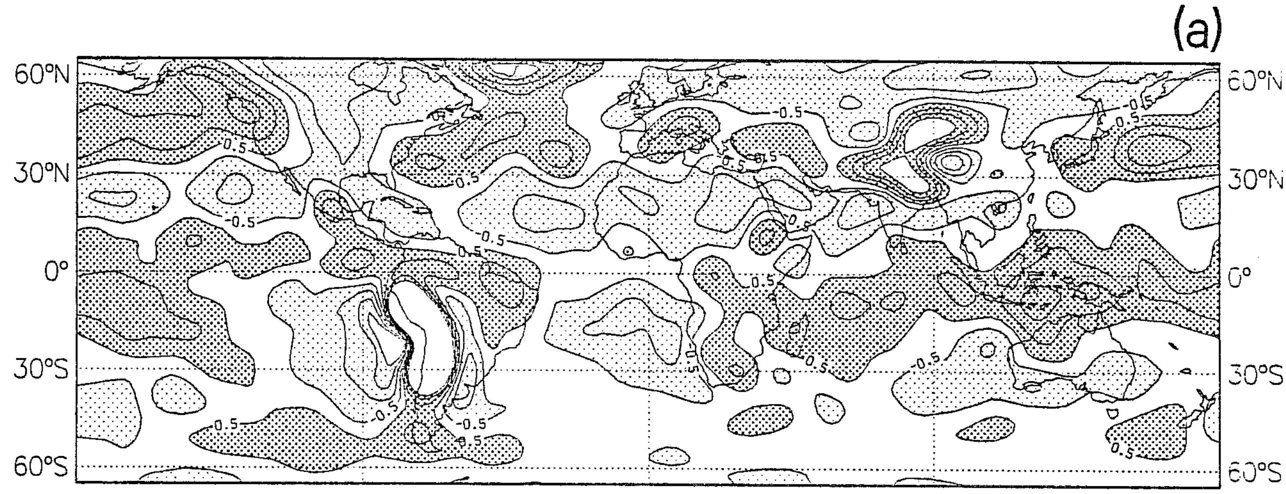
As discussed in section 1 the alternatives to using initialized analyses have their own difficulties. The reasons for initializing the analyses are excellent; indeed current operational wisdom among medium range forecasters is that initialization is necessary not so much to produce a better medium range forecast as to produce a better analysis, because the reduction of gravity wave noise in the short range forecast leads to a better first guess. One could of course think of dispensing with the model-generated first guess altogether but the alternatives of using, say, a six-hour persistence forecast from the previous analysis or some estimated 'climatology' as the first guess are even less attractive.

5. THE LOCAL HEAT BALANCE

Given the significant zonal variations in dynamical structures and perhaps more importantly in data density and quality, the zonal mean diagnostics presented above should essentially be viewed as providing a convenient summary and any conclusions based on them regarded as tentative until confirmed in the local budgets. Most of the heating imbalances discussed above are highlighted at 700 mb, so we shall concentrate on this level to save space. Figure 10 gives the inferred and model generated values as determined using the initialized analyses. In the tropics the inferred pattern closely resembles that of the outgoing longwave radiation for the same period (not shown). The pattern (Fig. 10a) elsewhere is also realistic especially in the northern hemisphere. (The large values seen over the Andes and the Tibetan Plateau result from a 'post-processing' problem of extrapolating the boundary layer values in these regions to levels below the ground, and should be ignored). One is of course less certain of magnitudes, but given the verifiable underestimate of the analysed boundary layer convergence in the tropics the likelihood is that the values there are, if anything, an underestimate. In view of this the even weaker model generated values (Fig. 10b) in the major convective areas give cause for concern. The radiative cooling given by the model is confirmed to be an underestimate almost everywhere, and so is an underestimate of the heating in the Atlantic and Pacific storm track regions.

The same fields as in Fig 10 but calculated using the first guess values are given in Fig 11. The inferred values are very similar to those in Fig 10a, consistent with the fact that the analysed values represent small modifications to an accurate first guess. However, the model generated values are now significantly larger, and in better agreement with the inferred fields, particularly over the Atlantic and Pacific storm track regions, but the radiative cooling values and the tropical convective heating maxima are still underestimated.

The model generated terms thus show an interesting sensitivity to apparently small analysis increments that the inferred terms do not. One candidate for this different behaviour is the humidity field, which plays a crucial role in the determination of the model generated but not the inferred terms. To explore this possibility, both these terms were recalculated using the initialized values for the wind and temperature but the first guess values for the moisture field. As expected, the inferred term was almost identical to Figs 10a and 11a. The model generated term, surprisingly, was also very similar to Fig 10b over most of the globe, but not in the northern hemisphere storm tracks, where it was much closer to the first guess term shown in Fig 11b. Thus in these regions the problem with the model generated heating in Fig 10 is seen to be a problem with the humidity analysis, the 'input to the scheme', rather than the 'scheme itself'.



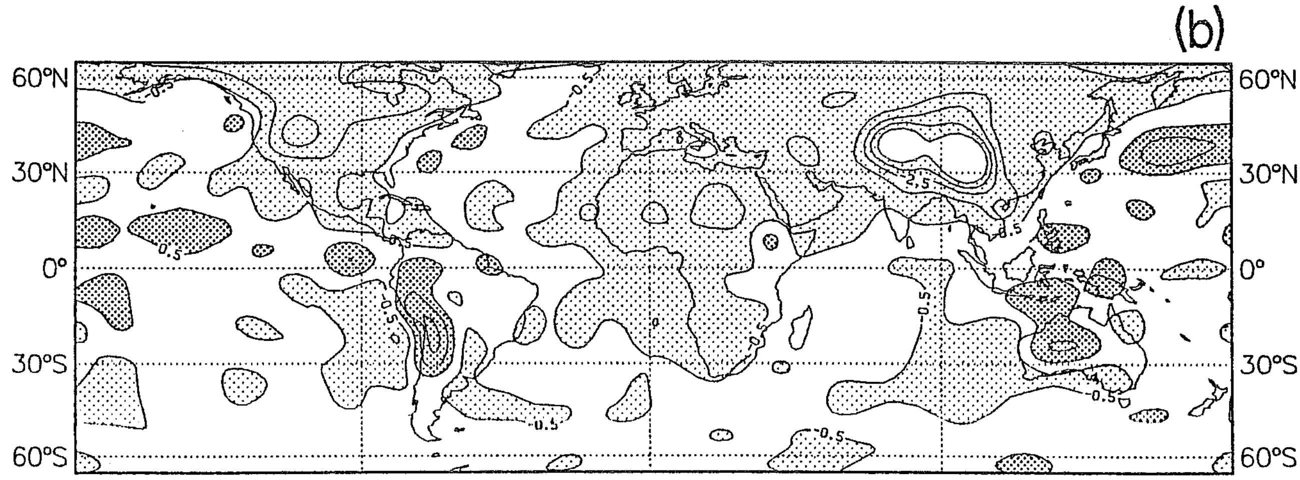


Fig. 10 The temperature budget at 700 mb. (a) the negative of the mean adiabatic tendency, (b) the mean diabatic tendency, both evaluated using the initialized analyses. Heavy shading represents heating, light shading represents cooling. Contours drawn are ±0.5, ±1.5, ±2.5, ±3.5 K/day.

In the tropics, the insensitivity of the model generated heating to the humidity analysis is also evident at other levels. Figure 12 gives two patterns of heating at 400mb, both calculated using the initialized values for wind and temperature but using the first guess moisture values in (a) and the initialized values in (b) in the manner described above. (Note that the field in Fig 12a has been determined using the first guess moisture values everywhere, not just at 400 mb). The inputs to the modified version of the Kuo convection scheme used at ECMWF are the vertically integrated moisture convergence and the thermodynamic stability. The difference between Figs 10b and 11b is thus associated with changes induced in the moisture field, the convergent flow, and the thermal structure by the analysis increments. From the similarity of Figs. 12a and 12b we have just ruled out the sensitivity to moisture. Furthermore, any important difference between the first guess and the analysed convergent flow would also imply an important difference between Figs 10a and 11a; but this is clearly small. The only remaining possibility is a crucial change in the thermal structure. The change shown in Fig 9b is

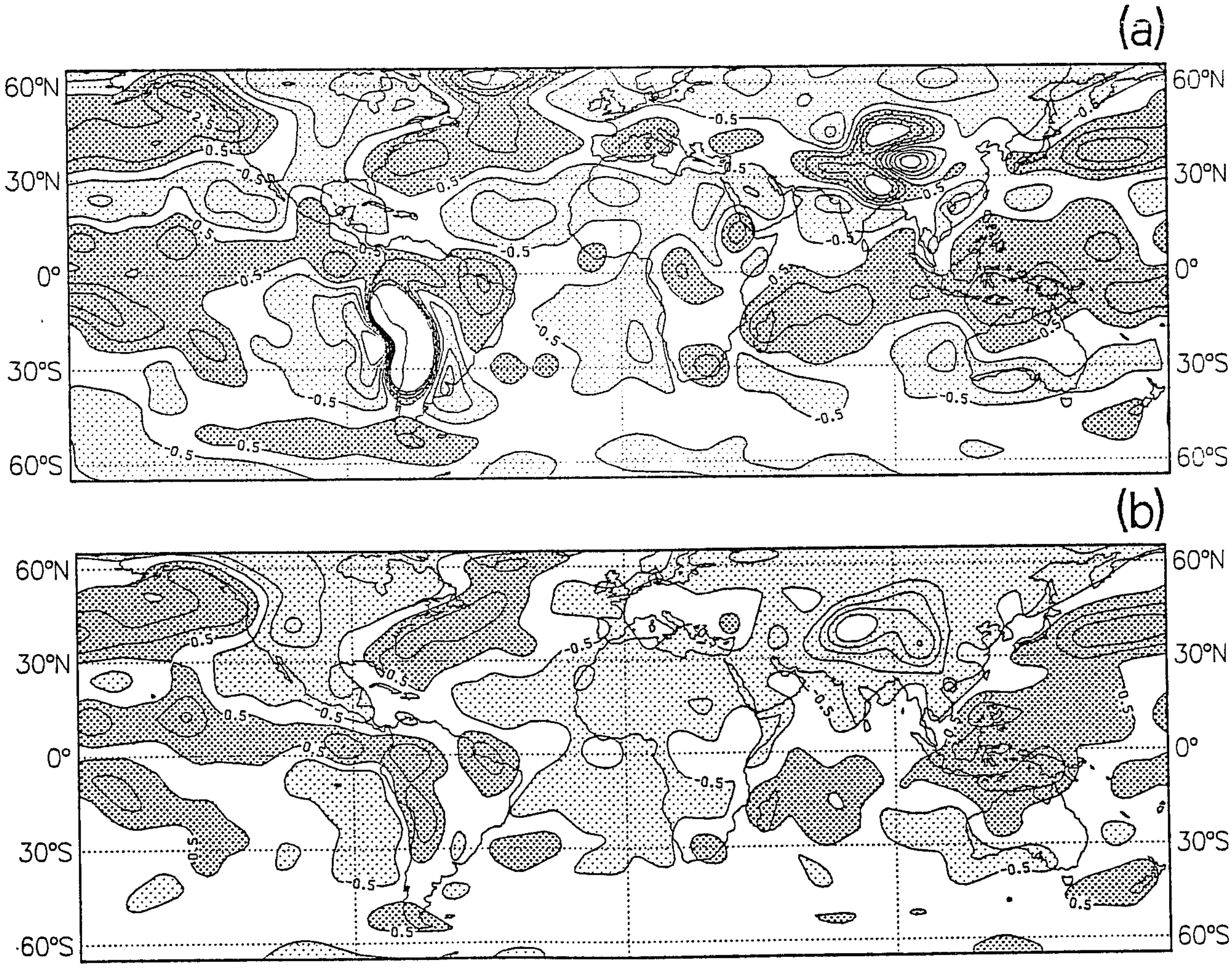


Fig. 11 As in Fig. 10 but evaluated using the first guess fields.

apparently large enough to affect significantly the calculation of the model generated heating (cf.Figs 10b and 11b) but not that of the inferred heating (Figs 10a and 11a). It is this differential response of the diabatic and adiabatic parts of the model to the same inputs that lies at the heart of the 'spin-up' problem (see Arpe, this volume, Fig 39). The fact that the chararacter of the spin-up has been found to be sensitive to the type of convection scheme used (Illari, 1987) is entirely consistent with this.

Given the sensitivity to the thermal structure, one now has to decide if the problem seen in Fig 10b arises from an erroneously specified temperature profile (the 'input to the scheme') or from a defective convection scheme that generates insufficient heating even with all the correct inputs (the 'scheme itself'). Routine comparisons of the first guess and analysed values with more than 1600 radiosonde ascents in the tropics confirm that the analysed profiles indeed fit the observations better than do the first guess profiles. Therefore in this case the likelihood is that the problem lies with the convection scheme itself. A correlative study of the type described in section 3 may help in diagnosing further the nature of this problem.

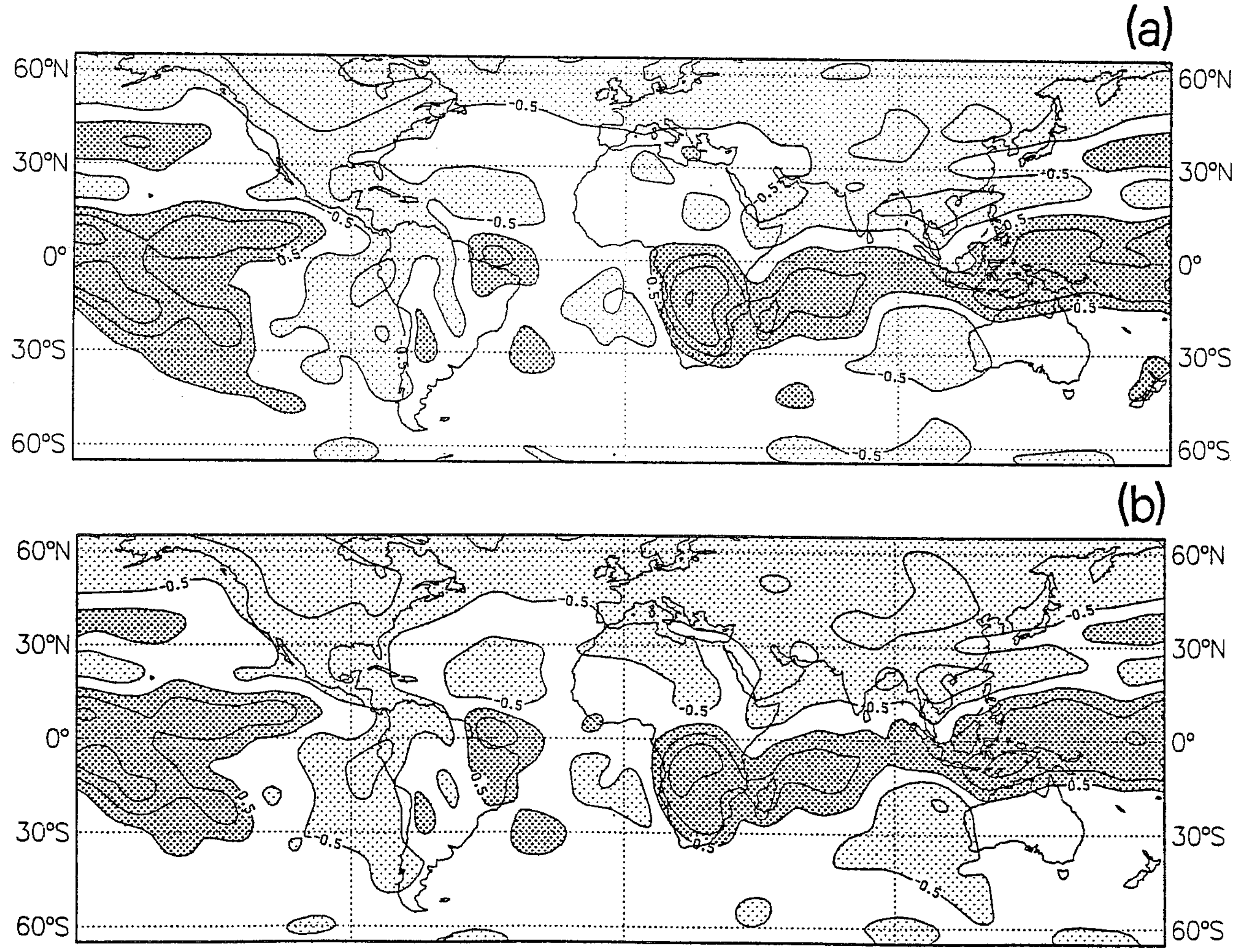


Fig. 12 The mean diabatic temperature tendency at 400 mb, evaluated using the initialized values for wind and temperature, but using the first guess values for moisture in (a) and the analysed values in (b). Contouring as in Fig. 10.

6. CONCLUDING REMARKS

In our view budget studies are usually conducted with one or more of the following aims in mind:

(i) to elucidate the role and relative importance of the individual terms and their behaviour in space and time;

(ii) to provide estimates of some term, whose direct determination is either difficult or prone to error, from a knowledge of the others as a 'residual'; or

(iii) in cases where all the terms can be evaluated, to draw implications about errors in some terms from a knowledge of the budget residual, or even to conclude that 'something is missing', or 'should be parametrized'.

All three approaches have their uses, and examples of all three abound in the literature. It is clear that the demands made upon the data and analysis quality must become increasingly more stringent depending on whether one chooses to go down paths (i), (ii), or (iii). We have concentrated here on the last of these, in the hope that the tremendous effort invested routinely in improving the quality of the analyses and the model at a major operational forecast centre allows one to address these issues in a qualitative if not quantitative sense.

Despite evaluating all the individual terms as consistently as possible, large residuals are obtained in the local as well as zonal mean momentum and heat budgets. We have tended to ascribe these systematic residuals on large scales to modelling errors for the basic reasons that: (a) observational errors, being mostly uncorrelated in space, are not likely to give rise to large scale residuals, (b) analysis errors are unlikely to contribute to systematic residuals because the analysis scheme, given a reasonably accurate first guess and the statistical structure of the first guess errors, is designed to eliminate the correlated first guess error in a statistical sense, and finally, (c) initialization errors, although large, act almost everywhere in the sense of only partially masking the effect of the diabatic forcing errors in the budgets. In effect we claim that the erroneous adiabatic tendencies associated with the erroneously determined divergent flow arise primarily from the use of the erroneous diabatic forcing itself in the initialization scheme. Thus the residuals that are finally obtained are probably indicative of even larger forcing errors than is apparent from a cursory examination.

That the model has forcing errors is not in doubt; otherwise it would have a perfect climate. That they contribute to our residuals is also not in doubt; the question is about the extent to which their effect is corrupted by the presence of data assimilation errors. In light of the above arguments and also considering that the systematic error tendency implied by the systematic 3- or even 10-day forecast errors (not shown) is comparable in magnitude to our residuals, the likelihood is that the contribution of the forcing errors is large on the large scale. The situation is probably quite different on smaller scales. In fact we intend to study the unsmoothed residuals in some selected areas with a view to extend the data monitoring capability of the ECMWF assimilation system, discussed by Hollingsworth et al. (1985), in a straight-forward manner.

With all these caveats in mind, there is evidence here to suggest that:

(1) the parametrized vertical diffusion of horizontal momentum is too strong in the free atmosphere above the planetary boundary layer;

(2) the parametrized orographically induced gravity wave drag is too strong in the lower stratosphere;

(3) the boundary layer formulation appears to be in error over the oceans and near mountains over land, where an erroneously specified terrain height probably compounds the problem;

(4) the radiative cooling is too weak in the clear-sky regions of the subtropics;

(5) the parametrized heating is too weak in the tropical lower troposphere; and

(6) the humidity analyses represent an underestimate in middle latitudes.

Some of these conclusions have found support in independent research carried out at ECMWF. In particular Miller and Viterbo (personal communication) arrived at conclusion (1) by a different route and have since demonstrated its impact on the medium range forecasts. The radiation error (4) was already well known (Morcrette, personal communication) but it is nonetheless encouraging to be able to see it in our heat budget.

Pending a more quantitative assessment of the impact of the error in the analysed divergence, statements (1) to (6) should be regarded as indicative and not conclusive. The regression approach proposed in section 3 may help with this requirement.

Another possibility is actually to solve a local 'omega-equation' (or 'normal mode initialization') version of equation (2) for the divergence error δD associated with the momentum and thermal forcing errors δM and δQ, assuming that there are no errors in the determination of the momentum and heat fluxes. This suggests an iterative procedure in which initial guesses for δM and δQ based on a simple interpretation of the budget residuals are modified until δM and δQ and the compensating δD thus determined are all consistent with the observed residuals.

A third possibility is to examine the potential vorticity budget. This has powerful advantages as well as disadvantages, but would mostly provide redundant information given our use of the initialized data if the iterative procedure described above turns out to be successful. Ignoring for simplicity the contribution of the twisting terms the expression for the potential vorticity $\underline{P}$ in pressure coordinates is

$$\underline{P} = - g \zeta \frac{\partial \theta}{\partial p} , \tag{16}$$

where ζ is the absolute vorticity. Hoskins et al. (1985) discuss at length the invertibility principle, i.e the problem of determining both ζ and θ from a knowledge of $\underline{P}$. This is obviously impossible from just one equation (16) so it is necessary to specify some additional information such as thermal wind balance linking ζ and θ. Now given a vorticity budget residual ζ_R and a heat budget residual θ_R, the potential vorticity budget residual $\underline{P}_R$ is approximately

$$\underline{P}_R \simeq - g \zeta_R \frac{\partial \theta}{\partial p} - g \zeta \frac{\partial \theta_R}{\partial p} . \tag{17}$$

In general ζ_R and θ_R will both have contributions from an erroneously determined vertical p-velocity ω:

$$\begin{aligned} \zeta_R &\simeq \bar{\bar{\zeta}}_R + \zeta \frac{\partial}{\partial p} (\delta\omega) , \\ \theta_R &\simeq \bar{\bar{\theta}}_R - \frac{\partial \theta}{\partial p} (\delta\omega) , \end{aligned} \tag{18}$$

which will tend to cancel in (17). The advantage of considering the potential vorticity budget is thus that the vertical velocity error is nearly subtracted out and $\underline{P}_R$ reflects mainly the 'true' diabatic forcing errors $\bar{\bar{\theta}}_R$ and $\bar{\bar{\zeta}}_R$. The disadvantage is that we cannot now disentangle these two types of error, because of the absence of any 'thermal-wind balance' linking them. One way out of this dilemma is to perform a regression analysis of (17) along the lines discussed in section 3. This may or may not prove useful depending upon the noise introduced by taking vertical derivatives and nonlinear products in (17). Also the terms on the right hand side will tend to be small near the equator in any case. Nevertheless the degree of cancellation obtained between the two will expose some of the importance of the vertical velocity error in (18). But again some cancellation is inevitable, given our use of the initialized data.

Thus although initialization introduces compensating adiabatic tendency errors in response to 'true' diabatic forcing errors, one could generate quantitative estimates of those errors. Whether this would give a more realistic view of the true importance of the forcing errors is, however, another matter. In this context it is interesting to recall that the direct impact of the initialization on the medium range forecasts is small. This would not occur unless the divergence changes induced by the current diabatic initialization scheme were dynamically almost inert, i.e not large enough to cause large and enduring changes in the forcing terms. Thus in some ways the residuals we see in our budgets with the initialized data are the residuals that matter most, because they expose those parts of the forcing errors that are not spent in generating a dynamically inert erroneous divergent circulation. They also suggest that the changes in the diabatic forcing would have to be substantial in order to have any impact on the forecasts.

Another potentially sensitive element is the humidity analysis and its impact on our heat budget. We have in fact confirmed that much of the budget residual in the storm track regions of middle latitudes can be ascribed to this error. Much less sensitivity is found in the tropics. This is of course not to suggest that the humidity analyses there are any more accurate; just that the major heating imbalances seen in these regions are not related directly to them. There is thus a case to be made for not attempting to control the tropical spin-up problem merely through modifications to the initial humidity field.

To conclude, there is enough 'real' information in the operational ECMWF analyses to give rise to large residuals in the local budgets of heat and momentum. Studying their structure and sensitivity to different inputs can yield very useful information about both analysis and model errors. The main strength of this approach derives from pointing to a local problem. Some outstanding problem areas have already been identified; their satisfactory solution will of course require further detailed study.

Acknowledgements

Numerous discussions with our colleagues in the ECMWF Research Department have clarified many of these issues in our minds. In particular we note the contributions of Drs. A. Hollingsworth and A.J. Simmons. Technical assistance from D. Dent is gratefully acknowledged. P. Lönnberg and A. Radford helped in verifying our results against observational data.

References

Arpe, K., 1988: Planetary scale diabatic forcing errors in the ECMWF model. This volume.

Daley, R., 1985: The analysis of synoptic scale divergence by a statistical interpolation procedure. Mon. Wea. Rev., 113, 1066-1079.

Hollingsworth, A., 1986: Objective analysis for Numerical Weather Prediction. In collection of papers presented at the WMO/IUGG NWP symposium, Tokyo 4-8 August 1986. T. Matsuno, ed. (Special volume of J.Met.Soc. Japan).

Hollingsworth, A., D.B. Shaw, P. Lonnberg, L. Illari, K. Arpe, and A.J. Simmons, 1985: Monitoring of observation quality by a data assimilation system. Mon. Wea. Rev., 114, 861-879.

Hollingsworth, A., J. Horn, and S. Uppala, 1988: Verification of FGGE assimilations of the tropical wind field: the effect of model and data bias. Submitted to Mon. Wea. Rev.

Holopainen, E., 1988: Recent estimates of diabatic forcing on the planetary scale: a review. This volume.

Hoskins, B.J., M.E. McIntyre and A.W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. Q.J.R. Meterorol. Soc., 111, 877-946.

Hoskins, B.J., and P.D. Sardeshmukh, 1987: A diagnostic study of the dynamics of the northern hemisphere winter of 1985-86. Q.J.R. Meteorol. Soc., 113, 759-778.

Illari, L., 1987: The 'spin-up' problem. ECMWF Technical Memorandum No. 137. Available upon request.

Klinker, E., A. Hollingsworth, and P.D. Sardeshmukh, 1988: In preparation.

Leith, C.E., 1980: Nonlinear normal mode initialization and quasi-geostrophic theory. J.Atmos.Sci., 37, 958-968.

Sardeshmukh, P.D., and B.J. Hoskins, 1988: The generation of global rotational flow by idealized tropical divergence. J. Atmos. Sci., 45, 1228-1251.

Wallace, J.M., S. Tibaldi and A.J. Simmons, 1983: Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. Q.J.R.Meteorol.Soc., 109, 683-717.

Wergen, W., 1986: Diabatic nonlinear normal mode initialisation for a spectral model with a hybrid vertical coordinate. ECMWF Technical Report No. 59. Available upon request.