The Diagnosis of Mechanical Dissipation in the Atmosphere from Large-Scale Balance Requirements

ERNST KLINKER
European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire, United Kingdom

PRASHANT D. SARDESHMUKH
Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, Colorado

(Manuscript received 28 July 1989, in final form 17 June 1991)

ABSTRACT

The momentum budget for January 1987 is evaluated with global observations analyzed at the European Centre for Medium-Range Weather Forecasts (ECMWF). The dissipation term is diagnosed from the budget as a balance requirement, that is, as that required to balance the sum of the advection, Coriolis, pressure gradient, and local tendency terms. This is then compared with the parameterized subgrid-scale effects in the ECMWF model's momentum equation, with a view of identifying possible errors in those parameterizations.

The balance requirement does not support the high parameterized values of orographically induced gravity-wave drag in the lower stratosphere. A deeper analysis also does not suggest a major role for turbulent vertical transports above the boundary layer. On the other hand, our budget does indicate that more effort be spent on a better representation of the potential enstrophy cascade associated with Rossby wave breaking in the upper troposphere. These statements are qualified by the errors in the balance requirement itself. The extent to which this is a problem is discussed.

A distinctive feature of these calculations is their internal consistency; that is, all the terms in the budget are evaluated as in the version of the ECMWF model used for assimilating the observations. This offers several advantages, not the least of which is that it makes our budget residuals identical to the systematic initial tendency errors of the operational weather forecasts, thus facilitating their computation and routine monitoring. As such, our calculations explain a large fraction of the systematic short-range forecast errors and, because of their local character, provide clues as to the possible sources of those errors. Experiments with and without gravity-wave drag are described to illustrate its large contribution during this period to the southerly wind error of the operational weather forecasts at 70 mb over western North America.

1. Introduction

The immediate motivation for this study arose from an attempt to rationalize the diagnosis of systematic errors in the operational ECMWF forecasts. Extensive experimentation had shown these errors to be relatively insensitive to further increases in horizontal resolution (Tibaldi et al. 1987), and the fact that essentially similar errors were also obtained in the extended forecast range (beyond day 10) suggested that the difficulty lay not so much with the initial state as with the modeled mechanical and thermal forcing. The problem was to estimate the relative importance of the two—indeed, if possible to isolate the dominant physical process being misrepresented.

Maps of the systematic error at the medium and extended forecast ranges (climate "drift") are generally unhelpful in this regard, because they show an integrated response to remote as well as local model errors. In themselves, such maps only tell us that something is wrong, not what or where, unless they are reminiscent of the patterns associated with localized orographic or thermal forcing in simple models. Thus, Wallace et al. (1983) were able to point to erroneous mountain forcing that led to the concept of an "envelope" orography, and Sardeshmukh and Hoskins (1988) to the global impact of erroneous forecasts of the upper-level convective outflow over the tropical western Pacific Ocean. But these are only pointers. The general problem of unscrambling individual model errors from forecast errors is a difficult one involving the calculation of a model's adjoint, and has not yet been attempted in the literature. In lieu of this, expensive and sometimes ad hoc experimentation is often undertaken to identify aspects of the physical parameterizations to which the forecasts are most sensitive. In the absence of prior hypotheses this is a hit-and-miss approach at best, and can lead to unphysical tuning.

Systematic errors at the shorter forecast ranges of, say, one day are more easily related to local problems,
because the influence radius is expected to be smaller, but are now contaminated with observational and data assimilation errors. In addition, one has to contend with spinup errors, which are often thought of as "physical" initialization errors. Although such a nomenclature implies that these errors are caused by erroneous structures in the estimated initial state, it is good to remember that they can arise just as well from supplying a perfect initial state to an imperfect model as vice versa. In the latter scenario the effect is analogous to the adjustment produced in a GCM on different time scales when a model parameter is suddenly and arbitrarily changed in the middle of a long integration.

Such complications notwithstanding, the advantage offered by bringing the local sources of error into sharper focus is potentially enormous, and we have taken this approach to its logical limit here by concentrating upon systematic initial tendency errors. At this extreme there is a further simplification that the errors in the model's degrees of freedom become (almost) decoupled; for example, an error in the radiation scheme cannot give an initial tendency error in the vorticity equation.

These points are well illustrated in Fig. 1, which gives the systematic forecast errors for January 1987 of the zonal-mean zonal wind and temperature at different forecast ranges. At day 10 (Figs. 1a,b) is seen the familiar problem of the intensification and poleward shift of the midlatitude jets, as well as a strong easterly error in the tropical upper troposphere. The temperature error is dominated by a strong warming in the tropical upper troposphere. The meridional gradients in this field are roughly consistent with the vertical gradients in the wind errors through the thermal-wind relation, even quite near the equator. This makes it difficult to disentangle the effects of erroneous mechanical and thermal forcing in the two panels. Given also that the time scale for the meridional dispersion of Rossby waves is of order 5 to 10 days on the globe, one cannot be sure if tropical or middle-latitude sources of error are playing a dominant role. The errors at day 1 (Figs. 1c,d) are smaller in magnitude, but exhibit a broadly similar structure. In view of the shorter time scale, one could take this as evidence that the middle-latitude errors at day 10, in the Northern Hemisphere at least, are probably of midlatitude origin. One still cannot

![Fig. 1. Zonal-mean cross sections of the systematic error of the zonal wind (left) and temperature (right) forecasts for January 1987 at day 10 (a, b) and day 1 (c, d). The units are meters per second for the wind and degrees kelvin for the temperature. Negative values are indicated by shading. The zero contour is suppressed for clarity. The initial tendency errors of these fields, expressed in units of per day, are shown in (e, f) in an identical format. Note that the ordinate is model level; their approximate location in pressure is also indicated.](image)
distinguish between mechanical and thermal modeling errors, of course, because the time scale for geostrophic adjustment is even shorter, of the order of a few hours. Another way of stating this is that even if one were able to isolate the error in the potential vorticity forcing, one would not be able to invert it to obtain its mechanical and thermal components because of the lack of additional information such as thermal-wind balance linking them.

Given their overall similarity with the day-10 errors, one could also conclude from Figs. 1c,d that the systematic errors in the initial state are either small or their memory is mostly lost by day 1, at least as far as these variables are concerned. The real surprise, however, is in Figs. 1e and 1f, which show the systematic initial tendency errors, scaled as one-day forecast errors to make them directly comparable with Figs. 1c,d. These are definitely affected by systematic errors in the initial state, yet their degree of similarity with the top two sets of panels is considerable. There is no reason to expect forecast error patterns to be similar at different forecast ranges, of course. Nonetheless, that they are here suggests that a substantial portion of the systematic errors at not only the medium but even much shorter forecast ranges is associated with model errors, not errors in the initial state. Furthermore, there are apparently substantial errors in the parameterized mechanical as well as thermal forcing terms.

2. A different approach to budget studies

A moment’s reflection shows that the initial tendency errors shown in Figs. 1e,f are in fact identical to the budget residuals one would have obtained in the zonal-mean budgets of zonal wind and temperature if one had evaluated the individual terms exactly as in the model, using all the gridded analyses available for January 1987 as data. This would in fact be true for all of the model’s degrees of freedom (vorticity, divergence, temperature, moisture, and surface pressure at each point) and linear combinations of them. This suggests a very easy and accurate way of evaluating observational budgets of direct interest to modelers, consistent with the state-of-the-art techniques of calculating the individual terms, without having to know the details of the numerical differencing and parameterization schemes. One simply runs a one-time-step forecast from each global analysis (performed four times a day at ECMWF) available in a month and determines the average initial tendency, minus the easily calculated monthly trend, as the mean budget residual for that month. The contribution from any individual parameterization scheme can also be extracted when these runs are made; alternatively, one may repeat these inexpensive forecasts with that particular parameterization switched off. (We shall describe the relevant features of the ECMWF data assimilation/forecasting system as we go along. The interested reader may consult papers by Hollingsworth et al. 1985, Shaw et al. 1987, Tiedtke et al. 1988, and Simmons et al. 1989 for further details.)

In this paper we shall examine the structure of such budget residuals and try to arrive at some conclusion about the accuracy of the parameterized terms. For ease of discussion we will define the balance requirement as the sum of all the other terms in the budget, that is, the sum of all the adiabatic terms in the tendency equation with the sign reversed. We shall then interpret the budget residual as the difference between the parameterized terms and this balance requirement. As previously indicated, the balance requirement will be determined exactly consistent with the model’s numerics as (minus) the average initial tendency of a parallel series of one-time-step forecasts made with all the physical parameterizations switched off.

The balance requirement will of course itself be subject to analysis errors, particularly as it will contain important contributions from unobservable or poorly observed aspects of the flow such as the horizontally divergent wind and the associated vertical motion. This is a problem with all budget studies, even those conducted with data from field experiments. Unless the observations are sufficiently abundant and accurate to determine it by themselves, a discussion of how the divergent flow is estimated will always be necessary. At an operational forecasting center such as ECMWF, it is significantly affected by the diabatic initialization (Wergen 1987), a process of adjusting the analyzed initial state until the projection of the initial tendency on the fast gravity modes is small. We need not go into details here, but what this amounts to for the divergent flow is essentially consistent with (it is, in fact, better than) solving the quasigeostrophic omega equation with inhomogeneous terms involving the parameterized diabatic heating and frictional damping. Thus, the initialized analyses we use to evaluate our balance requirement are affected by precisely the parameterized terms whose errors we wish to isolate. As we shall see, their main effect is to introduce compensating errors in the balance requirement with the result that the balance requirement tends to “forgive” errors in the parameterizations; that is, the budget residual is smaller than it would be otherwise. This is a particularly large effect in the heat budget in the tropics. However, this compensation cannot be complete, because the erroneous parameterized forcing also projects onto slow Rossby modes that are not initialized, that is, not adjusted so as to give a prescribed initial tendency. Since we are mainly interested in the erroneous forcing of these slow modes in any case, this is precisely the signal we wish to isolate. This can be done provided the terms are evaluated consistently using the initialized analyses, as described above.

There will be two other sources of errors in our balance requirement, associated with observational and data analysis errors. Their effect should be relatively
small on the systematic, large-scale residuals examined here for the basic reasons that (i) observational errors, being mostly uncorrelated in space, are not likely to give rise to large-scale residuals, and (ii) analysis errors are unlikely to contribute to systematic residuals, because the analysis scheme, given a reasonably accurate first guess (a 6-hour model forecast at ECMWF) and the statistical structure of the first-guess errors, is designed to eliminate the spatially correlated first-guess error in a statistical (for our purposes, time mean) sense. These statements are less true for the heat budget, in which spatially coherent errors in satellite-derived temperatures can have a significant impact not only on the balance requirement, but also on the output from the convection scheme. Large uncertainties in the humidity analyses further compound the study of that budget. This means that even when one is able to settle that the budget residual is associated with an error in the parameterized heating, it becomes important to determine whether that error arises from errors in the inputs to the convection scheme or in the scheme itself. For this reason we shall discuss the heat budget elsewhere and concentrate here on the momentum budget.

There is one final point to consider before we end this section. At most operational forecasting centers a model forecast provides a first guess for the analyzed fields. At ECMWF this first guess is updated using all available observations made within 3 hours of the analysis time. If no observations were available, the first guess would be retained as the “analysis.” We should therefore stress that the only reason we do get large residuals in our budgets is that there is an input from real observations. Furthermore, that only those observations are admitted to the analysis scheme that do not stray too far from the first guess, and yet budget residuals are obtained, suggests that there is information in those observations, and in the resultant analyses, which is not known to the model. It is often argued that budget studies conducted with initialized analyses only tell us about the balances in the assimilating model and nothing about those in the “real” atmosphere. As discussed earlier, this is only partially true. It would become more true if the comprehensive continuous four-dimensional data assimilation techniques currently being developed at several forecasting centers were to be implemented. The main thrust of these techniques would be to adjust the initial state until the trajectory of a forecast made from that state minimizes the distance from observations over, say, a 12-h period, while keeping within the bounds of observational error. This would almost be equivalent to initializing the Rossby as well as gravity modes, that is, assuming that all the initial tendency errors seen in Figs. 1e, f are associated with errors in the initial state, which we find questionable given their similarity with 10-day and 1-day forecast errors (Figs. 1a,b, and 1c,d). However, even in this scenario a very wrong model would lead to analyses that would deviate from observations toward the limits of observational error, and one would avoid this only by relaxing if not eliminating the constraint on the forecast trajectory, approximating the situation we have at present.

3. The zonal mean budget of zonal momentum

As discussed previously, the budget equation may be viewed in the form

\[ \text{Residual} = \text{parameterized terms} - \text{balance requirement} \]  

where the balance requirement includes contributions from the zonally averaged steady and transient eddy fluxes of zonal momentum, the advection and Coriolis terms involving the mean meridional circulation, as well as the small monthly trend of the zonal-mean zonal flow. The parameterized terms include the horizontal and vertical diffusion terms plus gravity-wave drag. The terms in (1) are evaluated on 19 unequally spaced model coordinate surfaces in the vertical at the full model resolution of T106. However, for ease of viewing and also to emphasize the larger-scale structures we shall show horizontally smoothed versions of these truncated at T21 in the manner of Sardeshmukh and Hoskins (1984).

The budget is displayed in Fig. 2. For convenience, the distribution of the zonal-mean zonal wind \( u \) itself is given in Fig. 2a. The balance requirement for the frictional terms is given in Fig. 2b. As expected, it implies the need for the boundary-layer flow to be strongly damped, on time scales of a day or less. More surprisingly, it also shows a requirement for damping in the upper troposphere near the jet stream maxima. [A cruder version of this picture is shown in Palmer et al. (1985) as the budget residual, which they use to support their case for gravity-wave drag.] The parameterized terms are shown in Fig. 2c, and appear at first sight to be largely consistent with the balance requirement; however, inspection of the budget residual (Fig. 2d) reveals quite significant differences. Interpreted purely as reflecting errors in the parameterized terms, Fig. 2d suggests that the drag 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 Fig. 2c. There is a hint of an imbalance in the boundary layer, but this will become clearer in later sections. The residuals are generally smaller in the Southern Hemisphere, and follow the structure of the balance requirement above the boundary layer.

The contribution of the vertical diffusion and gravity-wave drag terms to the total in Fig. 2c is given in Figs. 3a and 3b, respectively. (Note the smaller contour interval in Fig. 3b.) The parameterized horizontal biharmonic diffusion that is included in the ECMWF
Fig. 2. The zonal-mean budget of zonal momentum for January 1987. (a) $\bar{u}$ in meters per second. Easterlies are indicated by shading. (b) The balance requirement for the frictional terms, (c) the parameterized frictional terms, and (d) the budget residual, (c) minus (b). The contours drawn are ±0.5, ±1.5, ±2.5, ⋯ m s$^{-1}$/day and negative values are indicated by shading.
model to control numerical noise at the limit of the model’s resolution is very small on these large scales and is therefore not shown. Clearly, much of the structure in Fig. 2c is associated with the turbulent vertical fluxes of horizontal momentum, which are parameterized at ECMWF as a vertical diffusion. The contribution from gravity-wave drag is much smaller; its main impact is in the midlatitude lower stratosphere, and even there it is inconsistent with the balance requirement. 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 term also appears to be too strong.

Further justification for these statements comes from examining the horizontal distribution of the parameterized terms (Fig. 4a) and the budget residual (Fig. 4b) at 70 mb, a level at which the parameterized gravity-wave drag is especially strong. 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 in this plot as westerly acceleration in the subtropical jet stream regions of the western Pacific and Atlantic oceans. Both of these features are seriously at odds with the balance requirement (not shown) with the result that the plot of the budget residual closely resembles that of the parameterized terms. Inspection of similar maps at the jet level and below (not shown) confirms the points already made. And finally, the 6-hour ECMWF forecasts, when verified against actual observations (and not just analyses), show errors that are consistent with these errors in the forcing.

The errors in these parameterized terms are probably much larger than is apparent from Fig. 2d because, as discussed above, the diabatic initialization scheme generates compensating errors in [v] (and thence in the balance requirement) in the absence of sufficient data to constrain the analysis of the divergent flow. To see this more clearly, consider a simplified diagnostic equation for the streamfunction $\Phi$ of the mean meridional circulation under Boussinesq, quasigeostrophic scaling:

$$f_0^2 \Phi_{zz} + N^2 \Phi_{yy} = -f_0 [u^* v^*]_{yy}$$

$$- (g/\Theta_0) [v^* \theta^*]_{yy} + f_0 [M]_z + (g/\Theta_0) [Q]_y.$$  (2)

Here $\Phi$ is such that $[v] = -\Phi_z$ and $[w] = \Phi_y$; $N$ is the Brunt–Väisälä frequency, and $\Theta_0$ a standard value of
the potential temperature $\theta$. The asterisks denote deviations from the zonal mean, and $[M]$ and $[Q]$ refer to the zonally averaged momentum and heat sources and sinks. For meridional and vertical length scales $L \approx 1000$ km, $H \approx 4$ km, values of

$$[u^*v^*] \approx 50$ m$^2$ s$^{-2}$, $[v^*\theta^*] \approx 40$ m$^{-1}$ K,

$$[M] \approx 5$ m$^{-1}$ s/day, $[Q] \approx 4$ K day$^{-1}$,

will all provide comparable forcing for $\Phi$ in middle latitudes. The observed typical values of

$$[u^*v^*] \approx 50$ m$^2$ s$^{-2}$, $[v^*\theta^*] \approx 25$ m$^{-1}$ K,

$$[M] \approx 2$ m$^{-1}$ s/day, $[Q] \approx 1$ K day$^{-1}$,

show the predominant role of the poleward fluxes of momentum and heat in the dynamics of the Ferrel cell. The question here is about the role of $[M]$, that is, whether the values of $\approx 2$ m$^{-1}$ s/day given in Fig. 2c are representative. If they are, they would imply a significant damping of the Ferrel cell, which has hitherto largely been ignored in general circulation studies. The use of diabatic initialization at ECMWF ensures that the initialized values of $\Phi$ are roughly consistent with an equation such as (2); indeed, they are consistent with a higher-order balance (see Leith 1980). An error $\delta[M]$ in $[M]$ therefore implies an error $\delta\Phi$ in $\Phi$ given approximately by

$$f_0^2(\delta\Phi)_{zz} + N^2(\delta\Phi)_{yy} = f_0(\delta[M]).$$

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

$$N^2(\delta\Phi)_{yy} / f_0^2(\delta\Phi)_{zz} \approx (N^2H^2/f_0^2L^2) \approx B.$$
For $N \approx 1.1 \times 10^{-2} \text{ s}^{-1}$, $f_0 \approx 10^{-4} \text{ s}^{-1}$, this gives $B \approx 0.2$. Thus,
$$f_0^2 (\delta \Phi)_{x, 2} \approx f_0 (\delta [M])_x / (1 + B),$$
which, apart from a barotropic error in $[M]$, gives
$$\delta [u]_v \approx -\delta [M] / (1 + B).$$
Thus, in the approximate momentum budget
$$[u] \approx -[u^* v^*]_v + f_0 [v] + [M],$$
the residual $[u]_{x, R}$ arising mainly from the errors in these three terms is given by
$$[u]_{x, R} \approx -\delta [u^* v^*]_v + \delta f_0 [v] + \delta [M]$$
$$\quad \approx -\delta [u^* v^*]_v + (B(1 + B)) \delta [M].$$

The net contribution to the residual from a baroclinic error in $[M]$ is reduced by a factor $B / (1 + B)$, which for $B \approx 0.2$ is only 0.16. In effect $\delta [M]$ induces an erroneous secondary circulation, which given that (2) is elliptic, tends to spread the tendencies from $\delta [M]$ 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 $\delta [M]$ in the budget 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 $[M]$ as well. Essentially the same arguments apply with minor modifications to the local vorticity budget.

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. 2d would suggest. (The naive interpretation works better for gravity-wave drag in Figs. 2d and 4b because the high values of stratospheric stability imply that $B$ is larger.) A revised version of the diffusion scheme giving practically zero tendencies above the boundary layer has subsequently been implemented in the operational ECMWF model by Miller and Viterbo (personal communication), who also arrived at this conclusion by a different approach. This has led to some improvement in forecast skill in the shorter forecast range and a reduction in the systematic error. In addition the loss of eddy kinetic energy in the forecasts has been reduced. This is consistent with the fact that the removal of any process that internally redistributes horizontal momentum is tantamount to removing a net dissipative mechanism on the domain-integrated kinetic energy.

The vertical diffusion is modeled as the vertical derivative of a stress; the revision effectively sets this stress to zero above the boundary layer. Note that the surface value of the stress (which is a parameterization for the turbulent surface momentum flux and is related to the low-level stability and vertical wind shear through Monin–Obukhov similarity theory) is not affected directly by this change. Therefore, one might naively expect the vertically averaged part of the flow not to be affected to first order in the medium-range forecasts. However, the change in the vertically averaged momentum flux by the mean meridional motion, $\delta ([u] [v]) \approx [u] \delta [v] \approx 5 \text{ m}^2 \text{ s}^{-2}$, is not small at the jet latitude, and is in the correct sense of reducing the systematic error in Fig. 1a. Also, the increased level of eddy kinetic energy affects a change in the flow through different momentum fluxes, plus the subsequent change in the low-level flow affects the surface stress. It is the importance of precisely such second- and third-order effects that makes it difficult to use simple theory to diagnose model errors from forecast or climate errors.

The tendencies associated with gravity-wave drag are also modeled as the vertical gradient of a stress. The surface value of this stress, which is proportional to the subgrid-scale momentum flux associated with gravity waves excited by subgrid-scale orography, is parameterized at ECMWF in terms of the low-level flow, low-level stability, and subgrid-scale orographic variance. It is not immediately apparent from Fig. 4 if the problem at 70 mb arises from the modeling of the surface stress itself (which predicts the overall level of gravity-wave activity in a grid column) or just the gradient of the stress (which predicts where those waves are going to break) at 70 mb. In the next section, we shall attempt to isolate the errors in the surface stress by considering the momentum budget integrated over the depth of the atmosphere.

4. The vertically integrated momentum budget

The results will be presented in a format similar to that of the previous section. The balance requirement will now include the local advection, Coriolis, and pressure gradient terms. (Note that in a terrain-following vertical coordinate system the last includes a term involving the gradient of the logarithm of surface pressure.) The parameterized terms will include the horizontal and vertical diffusion and gravity-wave drag terms as before. Because the tendencies of the horizontal wind $v$ from both the vertical diffusion and gravity-wave drag terms are of the form

$$v_i = g \delta r / \delta p,$$

where $r$ is the parameterized upward momentum flux, we shall show mass-weighted vertical integrals rather than averages; thus, we will be looking at the distribution of the surface stress instead of the surface stress divided by surface pressure.

The calculations were repeated for January 1988 and 1989 and the results in Figs. 5 and 6 show averages over 1987, 1988, and 1989. Figure 5a shows the balance requirement for the drag on the vertically integrated horizontal flow. Note that this should be viewed as an estimate of the surface stress only if all of the subgrid-
Fig. 5. The vertically integrated momentum budget averaged over January 1987, 1988, and 1989. (a) The balance requirement for the frictional terms, (b) the parameterized frictional terms, (c) the budget residual, and (d) the contribution of gravity-wave drag in (b). The arrow shown denotes 0.4 N m$^{-2}$.

scale terms to be parameterized are of the form (3); horizontal or quasi-horizontal mixing processes in the atmosphere cannot be represented in this form. With this reservation, it is encouraging to see that the arrows are generally opposite in sense to the surface flow over most of the hemisphere. This should be seen not only as a general validation of (3) but also of the quality of current upper-air operational analyses. The vertically integrated parameterized terms are given in Fig. 5b. Apart from the very small horizontal diffusion term this picture shows the distribution of the modeled surface stress. The generally opposite sense of these arrows to that of the low-level flow is therefore to be expected.

The agreement between the parameterized stress and the balance requirement is generally good, but not over the Rockies and the Tibetan Plateau. This is highlighted in the plot of the budget residual (Fig. 5c). Note that the discrepancy does not reflect a small difference between large numbers; the balance requirement and the parameterizations are seriously at odds with one another in these regions. The residual is of the order of 0.2 Pa, and both its pattern and magnitude correspond to that of the parameterized gravity-wave stress (Fig. 5d). The problem of Fig. 4 is thus seen even in this
vertically integrated budget. However, this does not appear to be the whole story, even near mountains. The contribution to these integrals from different layers of the atmosphere will be considered next. Note that these are now affected by the errors in the divergent flow, which one would expect to cancel out, to first order in Rossby number, in the vertically integrated budget. The contribution to Fig. 5c from the top 14 layers of the atmosphere, corresponding roughly to the free atmosphere above the boundary layer (see Fig. 2), is shown in Fig. 6a. Most of the problem in Fig. 5c is clearly related to the budget in this portion of the atmosphere. The contribution from the boundary layer (layers 15 to 19), shown in Fig. 6b in a magnified form, is much smaller, although we shall see that it is important for the evolution of the low-level flow. Gravity-wave drag is negligibly small in this layer (not shown, but see Fig. 3). Given also that the stress associated with vertical diffusion is much smaller at the top of the boundary layer than at the bottom, the figure suggests that the errors in its surface value are relatively small. This is consistent with the revision made by Miller and
Viterbo in January 1988 to the diffusion scheme only above the boundary layer.

The balance requirement and parameterized terms for the top 14 layers are shown in Figs. 6c and 6d, in a format identical to that of Figs. 5a and 5b. Figure 6d essentially shows the distribution of the modeled stress at the top of the boundary layer. The contribution of gravity-wave drag to this (not shown) is almost identical to that given in Fig. 5d. Thus, most of the structure over the mountains is associated with gravity-wave drag. The balance requirement, however, has a totally different structure: no damping is needed over the Tibetan Plateau and even the sense of the forcing appears to be wrong over Canada. Thus, the residual over Canada is even larger than the contribution from gravity-wave drag alone, and cannot be understood without allowing for independent errors in the balance requirement as well.

One possible candidate is the orography. The envelope part of the orography is essentially a partial parameterization of subgrid-scale orographic effects, with an implicit assumption that these amount to increasing the blocking effect of the mountains. Since we have not explicitly included it with our parameterized terms, an error in the choice of the best envelope should appear in our balance requirement, but exactly how is not clear. It has been suggested that it will affect the data analysis, which is performed on constant pressure surfaces through the provision of first-guess fields at a slightly wrong pressure level. It is hard to imagine this being a major source of error, but a quantitative assessment of its importance remains obscure. However, if the first-guess error structure functions specified in the analysis scheme are modeled on actually perceived first-guess errors, as they are over North America, this effect should already have been accounted for to some extent. We should stress again that so long as the first-guess error is small enough that the basic linearity assumption in the analysis scheme is valid, and so long as its statistical structure is specified accurately, its actual magnitude and structure are irrelevant for the time-mean analyzed values that dominate our budgets.

If a substantial portion of the flow is over the orography, rather than around it, the choice of a too high or too low mountain would affect the surface values of the vertical velocity and thus the vertically integrated divergence field. This issue is perhaps best examined through the mass budget

$$\frac{\partial \ln p_s}{\partial t} = -\mathbf{v} \cdot \nabla \ln p_s - \vec{D}$$  \hspace{1cm} (4)

where the caret refers to a vertical average. The elements of this budget are displayed in Fig. 7. Figures 7a and 7b give the distribution of the vertically averaged flow and the logarithm of surface pressure, respectively. A substantial portion of the flow is indeed over the orography and implies large-scale ascent upstream and descent downstream of the mountains. The divergence–convergence couplets seen in these regions in Fig. 7c are consistent with this.

The contribution to these from the envelope part of the orography (not shown) has a similar structure. The initialization process ensures a fine balance between the terms on the right-hand side of (4), as is evident from the small residual of order $0.05 \times 10^{-6} \text{ s}^{-1}$ in Fig. 7d. The implied $\sim 3\%$ error in the vertically integrated divergence resulting from incomplete initialization is, however, too small to explain the error in the balance requirement shown in Figs. 5a or 6c. This budget, of course, does not tell us anything about errors in either the surface pressure or the vertically averaged divergence, but only the extent to which the two are made consistent with mass balance. The point of Fig. 7 is really to emphasize the horizontal structure of the divergence fields associated with mountain forcing. With such a structure one expects anticyclonic initial vorticity tendencies upstream and cyclonic tendencies downstream of the mountains, and therefore northerly flow tendencies along the mountain ridges. The model was rerun with and without envelope orography to confirm this. Figure 8 shows the initial tendency difference for the depth-averaged (model layers 1-14) horizontal flow. This is consistent with the qualitative argument given above and shows a structure that is entirely different from the structure of the budget residual or the balance requirement in the vertically integrated momentum budget.

A second candidate is the analyzed momentum fluxes. However, these are prominent in the budget over the oceanic storm-track regions, not over western North America or East Asia. Also, the upper-air network is relatively dense over these land areas. Note that a vertically averaged budget residual of order $\sim 3 \text{ m s}^{-1}/\text{day}$ implied by Fig. 5c would imply a major error in the magnitude of the fluxes if attributed entirely to them. This seems unlikely, especially in view of the fact that a major contribution to transient eddy terms in local momentum budgets is associated with self-correlation terms (Hoskins et al. 1983). And finally, it is difficult to relate the horizontal structure of the residual to a presumed error in the fluxes.

An additional investigation was undertaken to isolate the effect of the parameterized gravity-wave drag itself on the balance requirement by running a series of data assimilation and initialization experiments with and without gravity-wave drag for a relatively short period of two days. The changes induced on the balance requirement by the drag were generally compensatory, as expected from the arguments of the previous section, but not in a vertical average, where they were opposite, leading to a budget residual larger than the drag itself. One clearly needs to exercise caution in applying the arguments of section 3 too rigorously.

We conclude that, over land, a major contributor to the residual in our vertically integrated budget (Fig.
5c) is gravity-wave drag. The situation is very different over the oceans, especially over the Atlantic, where the dissipative terms are very small (Fig. 6d). This plus the boundary layer being relatively well balanced (cf. 5c, 6b) implies from (3) that the errors in the parameterized surface stress are relatively small and do not
strands, a partial solution might be to make the horizontal diffusion stronger as well as more scale selective.

5. The boundary-layer budget

Although we have indicated that the stresses given by the gravity-wave drag scheme may be too large, this may reflect an error in the inputs to the scheme rather than in the scheme itself. In other words, it is important to establish that the first time step is not grossly unrepresentative of later time steps. The fact that we have evaluated our budgets with initialized data more or less guarantees this. Nevertheless, we have verified (Arpe, personal communication) that the output from the scheme does indeed maintain magnitudes of $\sim 0.2$ Pa, shown in Fig. 5d throughout the 10-day forecast period.

The stress at the surface is parameterized as (see Miller and Palmer 1986)

$$\tau_{GWD} = -K \rho \nabla \sigma_{h} \mathbf{v},$$

(5)

where $K$ is a proportionality constant ($\sim \pi/\lambda$ for a sinusoidal mountain of wavelength $\lambda$), $\sigma_{h}$ a (directionally dependent) measure of the subgrid-scale orographic variance, $N$ a measure of the low-level stability, and $\mathbf{v}$ the horizontal flow averaged over the lowest three model layers. Because of this strong dependence on low-level quantities it is instructive to take another look at the boundary-layer budget, even though its contribution to the vertically integrated budget is relatively small. Figure 9a shows the budget residual in units of $\text{m s}^{-1}/\text{day averaged}$ over the boundary layer. The 6-hour forecast errors, expressed in the same units, are given in Fig. 9b. The considerable degree of correspondence between the two panels suggests that the first time step is indeed representative of later time steps. (We emphasize again that one should not expect an exact one-to-one correspondence between the panels in any case.) The significant thing about Fig. 9b is, however, that we are able to perceive the forecast errors at all. The arrows in this figure with their sense reversed show precisely the correction that the analyses make to the first-guess fields; this would have been zero if the observations were having no impact. Given that the analyses fit the data to within observational error, one can be reasonably confident that the input $\mathbf{v}$ of order $7 \text{ m s}^{-1}$ in the gravity-wave scheme is accurate to within about 10% in the monthly mean. This said, the boundary-layer budget is difficult to interpret as it stands. Note that Fig. 9a shows a decidedly stronger divergent component than Fig. 9b. One might suspect this to be a result of incomplete initialization: only the five gravest vertical normal modes are initialized at ECMWF [the rough equivalent in section 3 is solving (2) with five levels in the vertical]. However, the higher uninitialized modes have very low frequencies, low enough that their effect would certainly be apparent 6 h later in Fig. 9b. A likelier explanation is that the

---

**Fig. 8.** Initial momentum tendency integrated over the free atmosphere, model layers 1–14. The figure shows the difference between a model run with envelope orography and mean orography. The arrow denotes 0.4 N m$^{-2}$. 

---

Contribute to the residual in Fig. 5c. Therefore, we have to consider an error in the balance requirement that arises from analysis errors. The major contribution to the balance requirement in Fig. 5a is associated with eddy momentum fluxes; the contribution from the divergent flow is much smaller. Therefore, if all the terms to be parameterized are of the form (3), Fig. 5c must reflect an error in the fluxes. If so, the error in the fluxes would have to be of a magnitude similar to that of the fluxes themselves. This seems unlikely, even after allowing for considerable uncertainty in their estimates over the relatively data-sparse eastern Atlantic Ocean.

The daily analyses cannot be claimed to be so poor in this region. If they were, the day 2–3 ECMWF forecasts farther downstream in their influence zone over western Europe would hardly be better than persistence forecasts.

There is another possibility that the form (3) is inappropriate. This will be true if subgrid-scale horizontal fluxes play an important role. The likeliest regions for this to occur are the eastern halves of the oceanic storm tracks. As shown in many observational and theoretical studies (e.g., Hoskins et al. 1985), these are regions of intense Rossby wave breaking and hence of significant potential vorticity mixing. The horizontal diffusion used in many large-scale models describes this process only crudely; in particular, it does not incorporate the regional dependence one expects from such studies. Given that the main signature of Rossby wave breaking in the upper troposphere is the drawing out of stratospheric potential vorticity tongues into finer and finer
analysis system is unable to correct for an error in the forecast of the divergence implied by Fig. 9a even on these relatively large scales, and that this is too shallow to be handled subsequently by the initialization scheme. A significant portion of the agreement between Figs. 9a and 9b is therefore probably due to an error $\delta D$ in the analyzed boundary-layer convergence. This would imply an error $-f\delta D$ in the vorticity budget, which would contribute to the rotational part of Fig. 9a, and this extrapolated to 6 h would explain a significant portion of the rotational part of Fig. 9b. Note, however, that because such an error $\delta D$ is almost invisible to the assimilation system, it can ultimately arise only from model errors. Beyond this it is difficult to make more precise statements; near mountains it could well be related to errors in the envelope. Hollingsworth (personal communication) has demonstrated a striking dependence of the low-level forecast errors on the backing and veering of boundary-layer winds with height, which suggests difficulties with the stability dependence of the parameterized turbulent fluxes.

6. A numerical experiment

Further confirmation of our conclusion regarding gravity-wave drag comes from running a numerical forecast without it and verifying that this does indeed yield consistent results. For brevity we shall focus upon the lower stratosphere over western North America. Figure 10a gives the monthly mean budget residual at model level 4 (roughly 70 mb). As already discussed, this is very similar to and somewhat smaller than (ow-
Fig. 10. (a) The momentum budget residual at model level 4 (roughly 70 mb), (b) the contribution of gravity-wave drag to the budget, (c) the 6-hour operational forecast errors at model level 4, and (d) the contribution to this from gravity-wave drag, all expressed in units per day. See text for further explanation of (d).

...ing to the compensating effect in the balance requirement) the parameterized gravity-wave drag alone (Fig. 10b). Note that the tendencies from the drag scheme have a strong divergent component. Insofar as the evolution of the flow is governed by potential vorticity dynamics and the concept of balance between the mass and wind fields is valid, the relevance of such a divergent component that does not appear in the potential vorticity source is unclear. At any rate, one expects a significant Coriolis turning of the erroneous winds to result from this in the subsequent forecasts. (The effect is perhaps better understood in terms of the forced erroneous divergence–convergence couplet and the vorticity budget.) The systematic southerly wind errors of the 124 (=31 × 4) operational 6-hour forecasts in this month are consistent with this (Fig. 10c).

Instead of repeating each and every one of these 124 6-hour T106 forecasts without gravity-wave drag in order to isolate its impact, it was decided to run just two, one with and the other without, starting from the monthly mean analysis in each case. The difference of these after 6 h, shown in Fig. 10d, bears a striking re-
semblance to Fig. 10c over the Rockies. This is entirely consistent with our conclusion. It is interesting that one is able to reproduce much of the structure of the systematic error even in this simplified set of experiments; this is because \( \tau_{GW} \) as evaluated with monthly mean quantities happens to give a good estimate of its monthly mean value; that is, the nonlinearity in (5) is not very strong.

A final view of the impact of gravity-wave drag on the operational forecasts since its incorporation into the model in May 1986 is given in Fig. 11. This displays a time series of the bias of the 6-hour meridional wind forecasts at a radiosonde station located in the Rockies (47°N, 110°W). The change from a positive bias to a negative one after the introduction of gravity-wave drag is entirely consistent with the adjustment of the stratospheric winds in the 6-hour forecast shown in Fig. 10. A similar behavior is also obtained at other stations.

### 7. Discussion

Given the considerable body of evidence that the inclusion of orographically induced gravity-wave drag produces an improvement in GCM simulations (e.g., Boer et al. 1984; Palmer et al. 1986; Slingo and Pearson 1987; Miller et al. 1989; Iwasaki et al. 1989), our results from the observed momentum budget are surprising and so at least an attempt at a reconciliation seems necessary.

The three main influences of unresolved mountains upon the larger-scale flow are associated with (a) dynamical low-level barrier effects, (b) unresolved vertically propagating gravity waves, and (c) aerodynamic drag over irregular surfaces. The incorporation in general circulation models of envelope orography, gravity-wave drag, and enhanced low-level turbulent dissipation over irregular terrain represents attempts to account for these quite different physical processes. However, since they all act to reduce the westerly flow and therefore the excessive westerly problem in middle latitudes, care must be taken during model development to achieve the correct balance between them. There are evident pitfalls in tuning one representation to compensate for the deficiencies or absence of another.

All three effects are included in the version of the ECMWF model examined here. In our budget, the envelope orography effect is included in our “balance requirement” for the frictional and gravity-wave drag terms. If we assume that the balance requirement without the envelope orography contribution is basically correct, then the vertically integrated budget residual in Fig. 5c indicates that at least one, if not all, of the three effects is misrepresented in the model. Given the complexity of the subgrid-scale phenomena in question, one clearly has to be more circumspect in pointing to a particular parameterization scheme. Nevertheless, the facts that (i) gravity-wave drag makes a large contribution to the residual, (ii) the contribution from the envelope orography has a different spatial structure from that of the residual, and (iii) the boundary-layer budget is relatively well balanced, provide a hint that the parameterized values of gravity-wave drag may represent an overestimate.

Since the argument is basically about numbers, this is perhaps not altogether surprising. The behavior of internal waves excited by mountains is complex in the nonlinear regime and poses difficulties in parameterization that are well recognized by modelers. For Froude number of order one or greater, a condition quite easily met, there is a considerable enhancement of the low-level drag associated with wave breaking (Peltier and Clark 1979), which may or may not already be implicit in the concept of an envelope orography. It is not clear from saturation flux arguments (Lindzen 1981) that there should be additional levels of convective breaking in the troposphere or lower stratosphere. Shear breaking of the type incorporated into the ECMWF scheme can gain precedence over convective breaking, but only if the ambient flow is rapidly varying, in which case its existence cannot be characterized strictly in terms of a local Richardson number criterion as the WKB approximation becomes invalid (Pierrehumbert 1986). Even if such secondary breaking levels were to exist, it appears that the deposition of momentum would not in fact occur there but rather would be distributed throughout the vertical column between the breaking level and the mountain crest (Peltier and Clark 1986).

In view of such uncertainties and also the additional complexity introduced by three-dimensional and rotation effects, it is difficult to form an impression of
the overall importance of gravity-wave drag in the general circulation from idealized theoretical considerations alone. Observational studies of this effect, on the other hand, yield widely varying estimates (Hoinink 1985), tend to be biased toward cases when one expects it to be large, and, given their local character, are again not revealing as to its overall role. There is, however, considerable indirect evidence, summarized clearly in Palmer et al. (1986). One especially persuasive argument, elaborated further in Miller et al. (1989), is based upon the link between the surface flow and the vertically integrated poleward momentum flux in GCM simulations without gravity-wave drag, and goes as follows: Low-resolution GCMs produce realistic surface westerlies in the Northern Hemisphere but too weak westerlies in the Southern Hemisphere. The situation there is much improved in the high-resolution runs, but now the Northern Hemisphere westerlies are too strong. At the same time the behavior of the momentum flux by the zonally asymmetric eddies is symmetric with respect to the hemispheres: it is realistic in the high- and too weak in the low-resolution runs. Assuming that the surface stress increases monotonically with the surface wind, one can interpret these results as reflecting a deficiency in the parameterized link between the surface stress and the surface flow (i.e., a linearized drag coefficient), but only in the Northern Hemisphere, and therefore probably connected with mountains. Thus, the weaker momentum flux in the low-resolution runs leads to weaker surface westerlies in the Southern Hemisphere but not in the Northern Hemisphere, where its effect is masked by a roughly compensating underestimate in the drag coefficient. This error is exposed only when the fluxes are modeled realistically in the high-resolution runs.

The essence of this argument is that if the vertically integrated momentum flux is right and the surface flow is wrong, there must be an error in the parameterization of the surface stress and/or the pressure torque associated with unresolved mountains. However, this ignores the fact that associated with the erroneous surface flow is an erroneous surface pressure distribution, and therefore an erroneous resolved mountain torque. The effect is significant enough that even the sign of the mountain torque can be wrong in high latitudes (Swinbank 1985). In any event, the poleward momentum flux in the high-resolution ECMWF runs, although more realistic than in the T21 runs, is actually too strong in both hemispheres (Arpe and Klinke 1986; Arpe 1988) and is consistent with the error in the zonal mean winds. It is also worth noting that the argument as formulated applies to the total poleward momentum fluxes \( [u,v] \), which have contributions not only from the zonally asymmetric eddies \( [u',v'] \), but also from the mean meridional circulation \( [u][v] \) and the unresolved eddies \( [u^*v^*] \). The last two may be small in an actual budget, but perhaps they cannot be ignored in an error budget, especially in cases when the errors in \( [u',v'] \) are relatively small. We may recall that we found the contribution of \( [u][v] \) to be quite useful in discussing the beneficial impact of the revision to the vertical diffusion scheme in section 3.

If our results are correct and gravity-wave drag is overestimated, why then does it prove beneficial in GCMs? It is possible that it compensates, at least in part, for another error in the parameterization of subgrid-scale momentum fluxes. Our calculations suggest an error in \( [u^*v^*] \) arising from a misrepresentation of the potential enstrophy cascade in storm-track regions that lie slightly poleward of the jets. It is interesting to note that Palmer et al. (1986, p. 1001), when describing the characteristics of the serious systematic errors in NWP models, point to their appearance as cyclones over the oceans that fail to fill adequately over land. Given that the maxima in observed transient eddy statistics occur well away from the eastern boundary between the oceans and land, this points to a deficiency of processes over the ocean. In our view, a likely candidate is an error in the representation of the barotropic decay of baroclinic eddies associated with Rossby wave breaking in the upper troposphere (Simmons and Hoskins 1980; Hoskins et al. 1985). The excessive surface westerlies over land are thus possibly related to excessive momentum fluxes associated with cyclones that should have been stopped farther upstream. If this error is not corrected, then introducing an additional stress over land that increases monotonically with the surface wind will reduce the surface westerlies and therefore the error in them.

An important factor contributing to the momentum flux error is the treatment of blocking. Blocking is associated with poleward (equatorward) excursions of low (high) potential vorticity air, and hence, with equatorward potential vorticity fluxes. A systematic underestimate of blocking activity implies a poleward error in the potential vorticity flux, and therefore an erroneous westerly forcing of the zonal-mean zonal flow. For equivalent barotropic structures \( ([v'/y] \sim -\partial[u'/y]/\partial y) \) this is consistent with an excessive poleward momentum flux error in the subtropics.

If Rossby wave breaking is indeed underestimated as we suggest, this would also imply an underestimate of the equatorward potential vorticity flux in the stormtrack regions and therefore an erroneous westerly momentum source slightly poleward of the zonal mean jets in both hemispheres. The midlatitude atmosphere would see primarily the barotropic part of this source; hence, the equivalent barotropic structure of the zonal wind errors. The relatively symmetric structure of the error between the hemispheres in Fig. 1a is consistent with this. Gravity-wave drag, although placed higher up (Fig. 3b), acts in a similar manner to compensate for these errors, but only in the Northern Hemisphere. Pierrehumbert (1986) has also suggested that it exerts
a spurious though beneficial effect on the systematic error in the westerlies aloft that should in reality be attributed to other physical causes.

Though our calculations indicate an overestimation of gravity-wave drag, errors in the balance requirement pose a problem for a quantitative assessment. More progress can be made by performing relatively expensive sets of data assimilation experiments for a longer period of time (more than one week) in which the choice of subgrid-scale orographic forcing can be made in such a way that a separation of the effect of envelope orography and gravity-wave drag on initial tendency errors and short-range forecast errors is possible.

8. Concluding remarks

Throughout this work we have wondered about how best to perform a budget study in order to isolate errors in the parameterized momentum and heat sources. We recognize that conclusions based on observational budgets alone must always be interpreted with caution, no matter how abundant and accurate the observations themselves, because meteorological budgets involve gradients of fields that ultimately can only be estimated. Some data analysis is always necessary; this always implies some smoothing (Hollingsworth 1986), and if the smoothing is excessive, an error in the gradients. The local momentum budget is especially tricky as it involves large canceling terms, one of which is in gradient form, and whose difference is intimately tied up with the estimate of the divergent flow. These difficulties are well appreciated in the pioneering studies of Holopainen and his collaborators (Holopainen et al. 1980, and references therein). All we can say is that we have done the best with the operationally available data, evaluated the terms consistently, and linked our residuals to forecast errors.

If a model has a climate drift, it must have model errors, and these must contribute to our budget residuals. The question is to what extent their presence is masked by data assimilation errors. One might think it better to evaluate the budget residual with uninitialized analyses. However, this would have a large component of no consequence associated with erroneously analyzed fast gravity modes. Such an error incidentally must also be present in many other budget studies reported in the literature, but is rarely discussed. Averaging in time will make it look smaller if the excitation of these spurious modes is random, but complete elimination can be achieved only by normal-mode initialization.

Another subtlety that detracts from the use of uninitialized analyses is that the analysis scheme (at ECMWF at any rate) does not adequately correct for an error in the first-guess divergence field, even on fairly large length scales (Daley 1985). Because the model-generated first-guess fields are equally if not more consistent with (2) as the initialized analyses, this means that a knowledge of the parameterized momentum sources $M$ is also implicit in the uninitialized divergence fields. Therefore, the balance requirement calculated with the uninitialized data has a similar structure to Fig. 2b. (See also Fig. 5 in Palmer et al. 1986.) In addition, it has a contribution from a systematic error in the correction that the objective analysis makes to the first-guess divergence that is difficult to estimate.

In our view, a better strategy in middle latitudes is to ignore the analyzed divergence field altogether and effectively let the initialization scheme regenerate it. This introduces errors of its own, but insofar as the large-scale rotational flow and mass fields are analyzed accurately, these are associated with the errors in the parameterized forcing itself, and can therefore be accounted for. A complicating factor is that the errors in the momentum and heat sources are now coupled, albeit only within a Rossby radius. This presents no difficulty in principle if one considers both the momentum and temperature budget residuals. In the illustrative example of section 3 a consistent set of equations for the unknowns $\delta \Phi$, $[\delta M]$, and $[\delta Q]$ is

$$
[u]_{i,R} = [\delta M] - f_0(\delta \Phi)_z \tag{6a}
$$

$$
[\theta]_{i,R} = [\delta Q] - (N^2/\Theta_0/g)(\delta \Phi)_\theta \tag{6b}
$$

$$
0 = L(\delta \Phi) - f_0(\delta M) - (g/\Theta_0)[\delta Q]_\theta \tag{6c}
$$

where $L$ is the operator on the left-hand side of (2). Equation (6c) is consistent with equations (6a) and (6b) only if

$$
f_0[u]_{i,R_z} = -(g/\Theta_0)[\theta]_{i,R\theta}; \tag{6d}
$$

that is, the momentum and heat budget residuals are in thermal-wind balance, as they must be if initialization is to project the initial tendencies onto the "slow" manifold. Equations (6a, b, c) provide a set of three equations for the three unknowns $[\delta M]$, $[\delta Q]$, and $\delta \Phi$. However, because of the consistency condition (6d), one can solve uniquely only for the vertical average of $[\delta M]$ and the horizontal average of $[\delta Q]$. Further progress is possible if one reformulates the problem as one of finding $[\delta M]$ and $[\delta Q]$, which minimizes some domain-integrated quadratic measure, say, $[\delta M]_z^2 + (g/f_0\Theta_0)^2 [\delta Q]_\theta^2$, subject to the constraint (6). Such a problem can be solved by standard singular value decomposition techniques, and by this means one can still demonstrate the existence of irreducible errors in the parameterized momentum and heat sources.

Note that we simplified the discussion in section 3 considerably by ignoring the contribution of $[\delta Q]$ in (6c). We believe this to be small in middle latitudes, as is evident from Fig. 1f and the statements following (2). In addition, it has been ascertained that much of the midlatitude error in Fig. 1f is associated with an error in the analyzed humidity field (arising from radiosondes rarely reporting 100% humidity), which is
very quickly redressed in the forecast. As the initialization scheme uses heating fields averaged over a trial 2-h forecast, a significant portion of \( \delta Q \) implied by Fig. 1 in middle latitudes does not enter (2) in any case.

Note that even partially initialized analyses have the same advantage over uninitialized analyses. Let us say for illustrative purposes that the data were initialized using \( Q \) but not \( M \) in the diabatic initialization scheme. In the example of section 3 this would imply solving for \( \Phi \) after ignoring \([M]\) on the right-hand side of (2), and the equivalent of (6c) becomes

\[
0 = L(\delta \Phi) + f_0[M]_z - f_0[\delta M]_z - (g/\Theta_0)[\delta Q]_y.
\]

(6c')

Consistency with (6a) and (6b) now requires that

\[
f_0[u]_{1,RZ} = -(g/\Theta_0)[\theta]_{1,RY} + f_0[M]_z.
\]

(6d')

The residuals in (6a) and (6b) are thus no longer quite in thermal wind balance, because of incomplete initialization. However, their departure from that balance is known; and therefore, the information contained in the modified set (6a, b, c') is the same. The basic assumption in this approach is that the atmosphere evolves along some slow manifold and that the true divergent flow is diagnostically related to the true momentum and heat fluxes and parameterized sources through some equation such as (2). If the error in the analyzed fluxes is small, the error in the divergence must be related to the errors in the sources. Thus, this is better than using completely uninitialized analyses, in which case one cannot use (6c) or (6c'), and \( \delta \Phi \) in (6a) and (6b) has to be estimated from the first-guess errors and the characteristics of the objective analysis system. Both of these are problematic for the reasons given previously.

We have illustrated these ideas here using quasi-geostrophic balance arguments. The principle remains the same even when a higher-order balance such as that imposed by nonlinear normal-mode initialization is considered; it is inherent in the concept of balanced motion. In this case the equivalent of (6) is complicated but one can still envisage an iterative approach to the problem of determining \( \delta M \) and \( \delta Q \) whereby one makes a guess based on the solution of (6), reinitializes with \( M \) and \( Q \) "corrected" by these values, and recalculates the budget residuals until they become negligibly small.

If one interprets the residuals as systematic tendency errors, one could of course also apply a set such as (6) at a later time in the forecasts, say day 3. However, one would then have to contend with the additional errors in the momentum and heat fluxes plus those in \( M \) and \( Q \) arising from the wrong inputs to the parameterization schemes. One could estimate these from the analyses themselves and, assuming that the response of the parameterization schemes to errors in the input is roughly linear, still be able to retrieve some useful information. However, the additional approximations involved would make this approach less attractive.

In summary, we have attempted here a more unified and consistent approach to large-scale budget studies than has been possible hitherto, via initial tendency calculations with a state-of-the-art general circulation model. Our main conclusion as regards the momentum budget is that the whole question of the importance of frictional processes above the boundary layer needs to be reconsidered. Similar calculations for other periods (not shown) have yielded essentially consistent results. The parameterization of gravity-wave drag in the operational ECMWF forecast model was revised in May 1989. Future work using this technique will concentrate upon thermodynamic errors and the errors in boundary layer. We believe that diagnostic work along these lines can be pursued fruitfully at most large-scale numerical modeling centers, bearing in mind of course the limitations inherent in a budget approach. The best one can do is to point to local problems and provide tentative answers, confirmation of which can come only from careful modeling and independent diagnostic work.

Acknowledgments. The major part of this research was completed when one of us (PDS) was a visiting scientist at ECMWF. The use of diagnostic software developed by their Research Department is gratefully acknowledged. We are indebted to A. Hollingsworth, whose proposals on diagnosing diabatic forcing errors from 6-h forecast errors stimulated us to develop the technique described here. Comments on an earlier version of the manuscript by A. Hollingsworth, B. J. Hoskins, W. L. Gates, M. J. Miller, T. N. Palmer, K. Puri, and A. J. Simmons were extremely helpful and their encouragement is much appreciated. Technical assistance from D. Dent proved invaluable. Figure 11 was generated by A. Radford.

REFERENCES


KLINKER AND SARDESHMKH 627


