Conservation of Mass in Three Dimensions in Global Analyses

KEVIN E. TRENBERTH, JAMES W. HURRELL, AND AMY SOLOMON*

National Center for Atmospheric Research,† Boulder, Colorado

(Manuscript received 7 March 1994, in final form 16 May 1994)

ABSTRACT

For a number of reasons, conservation of mass in the global analyses on pressure coordinates is violated, yet this constraint is required for budget studies of all kinds. The imbalances arise from postprocessing the variables onto pressure surfaces, problems of dealing with the lower boundary and substituting an artificial atmosphere below ground, and diurnal pressure tendencies associated with the semidiurnal tide and the timing and distribution of observations. Methods are described and illustrated for May 1988 for adjusting the monthly mean global European Centre for Medium-Range Weather Forecasts analyses in three dimensions on pressure surfaces so that the mass balance is achieved, but the problems are present in analyses on constant pressure surfaces from all centers. First, a correction is needed for the global mean vertical motion. Second, it is shown that a local adjustment to the horizontal divergent velocity field is needed for regions that are below ground on constant pressure surfaces and nearby. Third, a change in the lower-boundary condition is required to remove diurnal and tidal influences, and this produces a barotropic adjustment in the horizontal velocity field as well as an adjustment in the vertical motion field that compares favorably with the semidiurnal tide in the analyses as a function of height. Solution of a three-dimensional Poisson equation is required to provide a final adjustment that minimizes the changes in the three-dimensional flow field. A vertical coordinate change is required to facilitate the solution, and the equation solves for the adjustment in the three-dimensional velocity potential using spherical harmonic expansions and finite differences in the vertical so that a matrix inversion is required for each wavenumber. Rather than any universal single-correction technique, this four-step process proves to be necessary to produce reasonable results. Even if the corrections noted here are not implemented, the diagnostic results serve as a warning to users of the analyses of potential substantial problems for certain applications. The results also indicate how operational centers could desirably alter their postprocessing procedures to ensure that the velocity field archived on constant pressure surfaces in below-ground regions satisfies the constraint of conservation of mass.

1. Introduction

The most fundamental budget of the atmosphere is that for the mass of dry air, which should be conserved to a very good approximation. It is so basic, in fact, that it is frequently assumed and overlooked. Yet failure to satisfy the dry-air mass budget will have a major impact on any other budget being calculated. Considerable interest exists in establishing the sources and sinks of many substances or conserved quantities in the atmosphere and the way they are connected. The heat and momentum budgets are particular examples, but so too are any budgets involving trace gases, such as water vapor, carbon dioxide, ozone, and so on, whose concentrations in the atmosphere are typically expressed as a mixing ratio that gives the mass (or volume) of the substance relative to that of either the total or the dry air mass (or volume). Any errors in the budget of dry air will therefore be reflected strongly in the budgets of all these other quantities. Accordingly, it is not only important to recognize errors in the mass budget but also to correct for them in some way (Boer and Sargent 1985; Alexander and Schubert 1990). Typical errors in the vertically integrated heat budget due to mass imbalances, for example, easily exceed 100 W m^{-2} locally (Alexander and Schubert 1990). Moreover, as we will demonstrate here, it is important to determine the specific sources of errors and devise specific, rather than universal, solutions.

The mass budget, therefore, can and should be used to provide an indication of the size and nature of the errors arising from diagnostic studies of sources and sinks of trace gases, heat, and momentum. The most convenient and best means of deducing budgets of any of these quantities is from the operational global analyses from the major numerical weather prediction centers, such as the National Meteorological Center (NMC) and the European Centre for Medium-Range Weather Forecasts (ECMWF). However, the archived

* Additional affiliation: Center for Meteorology and Physical Oceanography, Massachusetts Institute of Technology, Cambridge, Massachusetts.

† The National Center for Atmospheric Research (NCAR) is sponsored by the National Science Foundation.

Corresponding author address: Dr. Kevin E. Trenberth, Climate Analysis Section, NCAR, P. O. Box 3000, Boulder, CO 80307-3000.

analyses from these centers do not conserve the mass of dry air (Trenberth 1991). We have identified three major sources for this lack of conservation.

1) Mass imbalances arise from the methods of postprocessing the variables onto pressure surfaces. Because sophisticated four-dimensional data assimilation performs analyses of fields on model (sigma or hybrid) surfaces, the postprocessing performs interpolations from model coordinates to pressure coordinates with the result that the archived variables are really representative of individual levels rather than finite-sized layers. Mass imbalances result. Spurious residuals in the equation of continuity of up to 100% of the size of the horizontal divergence term, with largest errors in the Tropics, have been found in the free atmosphere (Trenberth 1991).

2) Representation of the earth's surface and below-ground regions is difficult. The traditional analyses on constant pressure surfaces are useful for many reasons and are familiar to the community, but they also suffer from extrapolation to the pressure surfaces that are below ground (in regions of high topography). Because variables are treated individually without consideration of any physical constraints in this process, gross violations of mass conservation result. (Incidentally, other expected constraints, like the geostrophic relation, are also violated below ground.) Of course, these are known artifices and the atmosphere below ground should be excluded from any analyses of the data. But this is easier said than done. On pressure surfaces, the region below ground can be tracked by proper consideration of surface pressure p_s, but inevitably there is an influence of the artificial atmosphere on the remainder—for instance, any time a gradient or divergence is computed.

3) The failure to properly resolve the diurnal cycle and, in particular, the semidiurnal solar tide results in the diurnal pressure tendencies being aliased onto apparent time-mean imbalances in mass. The problem is especially evident with once- or twice-daily data, while four-times daily data can capture most of the diurnal cycle in principle. However, because observations of the atmosphere above the surface are made once, or at most twice, daily and because of the distribution of the observations, additional local diurnal components are created in the analyses. The result is that a large-scale wavenumber 2 pattern with maxima in the Tropics occurs as a residual in the mass budget—the semidiurnal tide—but there are also many smaller-scale features in the residual (Trenberth 1991).

In addition, the vertical motion on pressure surfaces ω is not zero for the global mean in the analyses, as it must be to conserve the mass of the layer. This may be a manifestation of problem 1 but it has to be treated separately in correction procedures.

One solution to all these problems, that of using the model coordinates, also suffers from several problems (Trenberth et al. 1993; Trenberth 1995). Model levels consist, in simplest form, of a sigma (σ) terrain-following coordinate (e.g., at NMC) in which the lowest level corresponds to $p = p_s$, where p is pressure and $\sigma = p/p_s$. Hybrid levels (e.g., at ECMWF) consist of sigma near the surface but with a gradual transition to pressure above about 100 mb. One major issue with data on such levels is that the model surface often does not correspond with the earth's real surface. At ECMWF, an enhanced "envelope orography" is used that places the earth's surface at surface pressures up to 100 mb or more lower in value (Trenberth 1992). At NMC, a "silhouette orography" was used until 6 March 1991 when it was replaced by "mean mountains." Justification of the enhanced topographic heights is made on the grounds that the free atmosphere does not dynamically connect with air in valleys. In both cases the representation is spectral, so that the surface has extensive ripples over the ocean, arising from Gibbs phenomena.

Another important factor is that a change in resolution on model surfaces is not well defined, because the vertical coordinate also changes in the process. For instance, a reduction from a high resolution with 106 waves (T106 resolution) to a moderate resolution with 42 waves (T42), so as to reduce the size of datasets by a factor of about 6, is often desirable for climate purposes. This problem arises also if we wish to compare global analyses with climate model simulations at a different resolution. Even models at identical resolutions may have different surface topography and, therefore, different coordinate systems. Unfortunately, it is not a trivial problem to change resolution on model coordinates because the model coordinate itself changes so that horizontal and vertical gradients of variables become confounded. Horizontal smoothing on pressure surfaces is well posed and methods for truncating spherical harmonics at lower resolution are described in Trenberth and Solomon (1993) and Trenberth (1992). Similar truncation is not well posed with sigma or hybrid surfaces. To change resolution on sigma surfaces, it appears to be necessary to first interpolate to pressure, then truncate, and finally interpolate back to sigma. Ironically, the main advantage of working in model coordinates to avoid vertical interpolation errors is immediately lost. These aspects have been extensively explored and documented by Trenberth et al. (1993) and Trenberth (1995).

In this paper, we address the question of how to best achieve a set of global analyses that satisfies the constraint of mass conservation in three dimensions for use in subsequent analyses. The framework is the traditional one of pressure coordinates, as these will continue to be widely used and there are difficulties attached to use of model levels (as noted above). We have devised methods for dealing with all three problems outlined above in three dimensions. Earlier, Trenberth (1991) showed how the vertical mean mass budget could be corrected for exactly and this paper is

an extension of that work. Previously, vertical mean corrections have been applied barotropically with a constant correction to the horizontal velocity throughout the atmosphere. We argue that this is appropriate, but here it is just one component of the suite of corrections. An alternative would be an adjustment such that the mass flow correction is constant with height. The mass flow differs from the velocity by weighting values with the density, so that the correction to velocity is inversely proportional to the pressure and increases with height. We have experimented with this and concluded that it is not desirable for the vertical mean adjustment, as will be discussed later.

Ehrendorfer et al. (1994) and Hantel (1986) have previously discussed the problem of imbalances in the equation of continuity, and they proposed variational modifications of the three-dimensional wind field that minimize the corrections in a least-squares sense. Rather than deal with the equation of continuity, however, Hantel considered only the divergence of the three-dimensional wind field. Ehrendorfer et al. advanced Hantel's method, but focused on local areas of the globe and individual rather than time-mean analyses. Because the variability from day to day is so large, it masks some of the problems that are more evident in time-mean fields, and Ehrendorfer et al. did not explicitly address problems 2 and 3. In contrast, our correction procedures are global, and we have found that the different problems do require different solutions and that it is essential to also adjust the state of the lower boundary. Application of a single method will either not address some of the problems or will impose an incorrect—or nonoptimal—solution. For instance, the imbalances arising from extrapolation below ground must be addressed first and should remain local in space to the region impacted by such extrapolation, whereas procedures to compensate for the diurnal cycle extend throughout the global atmosphere.

After a brief discussion of the datasets used, section 3 presents the mass conservation constraints, and section 4 documents the issues for dealing with these in discrete representations and how to correct for the mass imbalances. Application of these methods to the case for May 1988 is given in section 5, and the discussion is given in section 6.

2. Data

Global analyses are produced using a four-dimensional data assimilation system in which multivariate observed data are combined with the "first guess" using a statistically optimum scheme. The first guess is the best estimate of the current state of the atmosphere from previous analyses produced using a numerical weather prediction (NWP) model.

It must be emphasized that the operational analyses are performed under time constraints for weather forecasting purposes and not for climate purposes. Changes in the NWP model, data handling techniques, initialization, and so on, which are implemented to improve the weather forecasts, may disrupt the continuity of the analyses (Trenberth and Olson 1988, Trenberth 1992). Some aspects, such as detailed analyses of the conditions at the surface of the earth, may be of less importance for weather forecasting while of great importance for diagnostic studies. Analyses are not made at the true surface of the earth. In evaluating terms in the budget equations using the archived analyses, finite-difference approximations and other assumptions (such as interpolation procedures) must be made but may not be compatible with the NWP model or the way the analyses were produced.

In this study we have used the ECMWF World Meteorological Organization (WMO) archive (Trenberth 1992), which consists of initialized data at seven levels twice daily at 0000 UTC and 1200 UTC. The seven levels are p = 1000, 850, 700, 500, 300, 200, and 100 mb, but our methods can readily be extended to any other set of levels. The case for May 1988 is used to follow on from the work reported in Trenberth (1991). Other archives of ECMWF analyses are available at NCAR (Trenberth 1992), but these contain uninitialized analyses that have much greater mass imbalances and much more noise in the fields (Trenberth 1991). Although the conclusions of this study are specific to the ECMWF datasets interpolated onto pressure surfaces, the issues are relevant to all global atmospheric datasets.

The WMO ECMWF archive is available only on a 2.5° grid. However, the gridpoint values were put onto the grid at T106 resolution, resulting in considerable aliasing of the data (Trenberth and Solomon 1993). This is especially a problem for fields, such as divergence, that have a peak in the spatial global spectrum at wave 40–70 (Trenberth and Solomon 1993), and thus, a lot of structure cannot be resolved by the 2.5° grid. In practice this leads to noise in the velocity components at high latitudes within 10° of the poles. This problem is further compounded by incorrect and inconsistent (in the sense that there are pole values for each longitude but their vector magnitude and directions vary) values archived in this dataset at the pole. This is discussed by Trenberth (1992) as "the pole problem" (see his section 3.2.1). There seems to be no way to avoid the high-latitude problem given the data available but it can be ameliorated if the data are smoothed, further degrading the resolution. In the T106 spectral archive of ECMWF data, such problems are not present. Consequently, because the high-latitude problems are unique to this dataset, we have not considered them as part of the more universal problems associated with attaining a global mass balance in three dimensions.

3. Conservation of mass

a. Continuous equations

We first consider principles of conservation of any quantity M integrated in the vertical over the mass of

the atmosphere from the bottom ($p = p_s$) to the top ($p = p_t = 0$). For practical reasons, it may often be necessary to recognize some other value than zero for p_t:

$$\tilde{M} = \frac{1}{g}\int_{p_t}^{p_s} M dp$$

results in the expression

$$\frac{\partial \tilde{M}}{\partial t} + \nabla \cdot \frac{1}{g}\int_{p_t}^{p_s} M\mathbf{v} dp = S, \tag{1}$$

where $\mathbf{v} = (u, v)$, ∇ is the two-dimensional horizontal operator and S is the source minus sink of M, vertically integrated in the column. This equation states that the flux of M out of the column is balanced either by a change in $\tilde{M}$ or by S.

The total mass of the atmosphere m in a column is

$$\tilde{m} = \tilde{m}_d + w,$$

where $\tilde{m}_d$ is the mass of dry air, the precipitable water $w = (1/g)\int_0^{p_s} q dp$, where q is the specific humidity, and

$$\tilde{m} = \frac{1}{g}\int_0^{p_s} dp = \frac{p_s}{g}.$$

Only the mass of dry air is conserved, so that $S = 0$ in that case.

Accordingly,

$$\tilde{m}_d = \frac{p_s}{g} - w, \tag{2}$$

and

$$\frac{\partial \tilde{m}_d}{\partial t} + \nabla \cdot \frac{1}{g}\int_0^{p_s} (1 - q)\mathbf{v} dp = 0. \tag{3}$$

For water vapor,

$$\frac{\partial w}{\partial t} + \nabla \cdot \frac{1}{g}\int_0^{p_s} q\mathbf{v} dp = \tilde{E} - \tilde{P}, \tag{4}$$

where E is the rate of evaporation and P is precipitation per unit mass, and we have ignored other forms of liquid and frozen water in the atmosphere. Combining these gives

$$\frac{\partial p_s}{\partial t} + \nabla \cdot \int_0^{p_s} \mathbf{v} dp = g(\tilde{E} - \tilde{P}). \tag{5}$$

This equation is more accurate than that commonly used in meteorology where the right-hand side is typically ignored. Trenberth (1991) discussed the magnitude of the errors in this approximation.

The commonly used equation of continuity is written

$$\nabla \cdot \mathbf{v} + \frac{\partial \omega}{\partial p} = 0. \tag{6}$$

Integrating over the entire column of the atmosphere gives

$$\omega_s + \int_0^{p_s} \nabla \cdot \mathbf{v} dp = 0, \tag{7}$$

and recognizing that

$$\omega_s = \frac{\partial p_s}{\partial t} + \mathbf{v}_s \cdot \nabla p_s \quad \text{at} \quad p = p_s, \tag{8a}$$

and

$$\nabla \cdot \int_0^{p_s} \mathbf{v} dp = \int_0^{p_s} \nabla \cdot \mathbf{v} dp + \mathbf{v}_s \cdot \nabla p_s \tag{8b}$$

gives Eq. (5) if the $E - P$ term is ignored. Because this assumption is widespread in atmospheric science, we will use Eq. (6) as the expression that determines whether or not mass is conserved by the analyses locally in three dimensions.

If all the terms in Eq. (3) or Eq. (6) are evaluated from real data or analyses, it is unlikely that the terms in the equations will balance exactly (Trenberth 1991). Even if they did hold for the NWP model, the interpolation of variables onto constant pressure surfaces and the need to approximate the integral with finite differences ensures that an imbalance will arise.

As will be seen later, Eq. (6) is not in a form that lends itself readily to correction because of the differing units and scalings of the vertical (p) versus horizontal coordinates. Instead a quasi-height coordinate is introduced in which we define $H = RT_{oo}/g$, where T_{oo} is a constant, and we change the vertical coordinate to

$$z = H \ln \frac{p_{oo}}{p}, \tag{9a}$$

where $p_{oo} = 1000$ mb, and thus define

$$w \equiv \frac{dz}{dt} = -\frac{H}{p}\omega \tag{9b}$$

as the vertical velocity in these $\ln p$ coordinates. It is also convenient to define a density factor in these coordinates:

$$\rho_o = \rho_{oo} e^{-(z/H)} = \rho_{oo}\frac{p}{p_{oo}} = \frac{p}{gH}, \tag{9c}$$

and ρ_{oo} is a constant. The thickness of a pressure layer can then be written $dp = pd\ln p = -\rho_o g dz$, provided $\rho_{oo} = p_{oo}/(RT_{oo})$.

Then Eq. (6) becomes

$$\nabla \cdot \rho_o \mathbf{v} + \frac{\partial(\rho_o w)}{\partial z} = 0. \tag{10}$$

In practice, evaluation of the terms in Eq. (10) results in a residual R as noted above.

b. Vertical levels

Following Trenberth (1991), the standard levels are denoted by $p_1, p_3, \ldots p_{2J-1}$, where $p_1 = 1000$ mb. For the seven-level ECMWF archive $J = 7$, then $p_{13} = 100$ mb. In addition, is it useful to define a level at $p = 0$ mb as the $2J + 1$ level. We define the intermediate half levels as

$$p_j = \frac{1}{2}(p_{j-1} + p_{j+1}), \quad j = 2, 4, 6, \cdots 2J.$$

Note that although the vertical coordinate is now z, the half level should be defined in terms of p to take account of mass weighting. Otherwise, for example, at the top of the model $z \to \infty$ and any half level also $\to \infty$. The upper boundary has to be treated so as to properly define the layer thickness. Thus, an upper boundary is imposed at $p = p_t = p_{2J}$, and, thus, $dz = -Hd \ln p$ must be written as $\Delta z_{2J} = H \ln 2$. Note that the thickness of the upper layer is the same regardless of the positioning of the pressure levels and is a manifestation of the infinite height of the atmosphere.

c. Vertical integrals

The mass-weighted vertical integral of any quantity M

$$I = \frac{1}{g}\int_{p_t}^{p_s} Mdp \tag{11a}$$

can be written as

$$I = \frac{1}{g}\int_{p_t}^{p_o} \beta Mdp, \tag{11b}$$

where p_o is now some fixed value of pressure greater than p_s everywhere, and we define $\beta = 0$, $p > p_s$; $\beta = 1, p \leq p_s$. This has the potential advantage of having fixed integration limits (e.g., Boer 1982).

Because data are available only at fixed pressure levels, Eq. (11) must be replaced with finite differences, and it is desirable to work in the ln p coordinates with $dp = pd \ln p = -\rho_o g dz$. In each case, the vertical integration will be carried out using a trapezoidal rule that is equivalent to

$$I = \sum_{j=1,2J-1,2} \beta_j \rho_{oj} M_j \Delta z_j, \tag{12a}$$

and we define

$$\beta_j = 1 \quad \text{if} \quad z_{j+1} > z_s$$

$$\beta_j = 0 \quad \text{if} \quad z_{j-1} < z_s$$

$$\beta_j = \frac{z_{j+1} - z_s}{z_{j+1} - z_{j-1}} \quad \text{if} \quad z_{j+1} > z_s > z_{j-1}, \tag{12b}$$

so that $0 \leq \beta \leq 1$.

This approach correctly weights the values at each level in the vertical with the appropriate mass of the layer and gives exact results in the case where ρM is a constant in z, for example.

d. Vertical finite differences

Equivalency between so-called flux formulations of the equations of motion and thermodynamics and the advective forms relies on the validity of the equation of continuity. Thus, in the thermodynamic equation, for example, the equality

$$\mathbf{v}\cdot\nabla T + \omega\frac{\partial T}{\partial p} = \nabla\cdot\mathbf{v}T + \frac{\partial \omega T}{\partial p} \tag{13}$$

relies on $\mathbf{v}$ and ω satisfying the equation of continuity Eq. (6). In addition, it relies on

$$\omega\frac{\partial T}{\partial p} = \frac{\partial \omega T}{\partial p} - T\frac{\partial \omega}{\partial p}, \tag{14}$$

then replacing $\partial\omega/\partial p$ with $-\nabla\cdot\mathbf{v}$ from the equation of continuity. Simple finite-difference schemes typically will not preserve this identity. One that does, so that Eq. (13) holds, is to formulate it as follows:

$$\frac{[\omega_{j+1} + \omega_{j-1}]}{2}\frac{[T_{j+1} - T_{j-1}]}{p_{j+1} - p_{j-1}} = \frac{[\omega_{j+1}T_{j+1} - \omega_{j-1}T_{j-1}]}{p_{j+1} - p_{j-1}} - \frac{[T_{j+1} + T_{j-1}]}{2}\frac{[\omega_{j+1} - \omega_{j-1}]}{p_{j+1} - p_{j-1}}, \tag{15}$$

in which case the thermodynamic equation should be applied not to levels but rather to layers, and the state variables for the layer are the averages of the adjacent level values, for example, $\mathbf{v}_j = 0.5(\mathbf{v}_{j+1} + \mathbf{v}_{j-1})$ for $j = 2, 4, 6, \cdots 2J - 2$, and vertical finite differences are applied as in Eq. (15). This has implications for how the finite differences should be set up in solving for this final correction, as follows in section 4d.

4. Correction methods and issues

Hantel (1986) addressed the problem of creating the three-dimensional velocity field such that the three-dimensional divergence is zero:

$$\nabla_3\cdot\mathbf{u} = 0, \tag{16a}$$

subject to

$$\int (\mathbf{u}^c)^2 dV = \text{minimum}, \tag{16b}$$

where $\mathbf{u}^c$ is the three-dimensional correction velocity and V is the total volume of the atmosphere. With this constraint, Hantel suggested that this could be achieved only if the modification is irrotational, and this was proved by Ehrendorfer et al. (1994). The corrections could be found by defining a potential field whose gradient gives the three velocity components and amounts to solving a Poisson equation for the correction velocity potential with the computed residuals on the right-hand

side. However, Hantel did not solve this problem for (u, v, ω), and Ehrendorfer et al. used a complicated local procedure involving mass fluxes through grid boxes. In fact, there is ambiguity involved in the latter paper over whether the correction is really to the velocity or the mass flow.[1] We will see the same problem emerging here. Below, we present an alternative means of doing the three-dimensional adjustment as well as addressing the other problems noted above.

There are several distinct steps to dealing with the problems discussed earlier. In practice it proves desirable to do the steps in the following order:

(a) Correct the global mean ω to be zero.

(b) Correct for the below-ground extrapolation locally in three-dimensional space (correction 1 to u, v).

(c) Set the lower-boundary condition to $\omega = 0$ at $p = 1000$ mb to remove the diurnal cycle effects (correction 2).

(d) Carry out a variational adjustment of the $\mathbf{v}$ and ω fields to minimally correct for three-dimensional mass imbalances (correction 3).

a. Global mean correction

The global mean ω in the analyses is not zero, as it should be if there is to be no net mass flow across pressure layers [again ignoring water vapor condensation etc., as in Eq. (6)]. Therefore, the first step is to correct the global mean value for ω. This is simple in spherical harmonic expansions as it means resetting the first ($m = 0$, $n = 0$) term to zero [cf. section 4d, Eq. (23)].

b. Below-ground corrections

For the 1000–850-mb layer for May 1988, Fig. 1 shows the monthly mean mass horizontal and vertical divergence terms and their sum, which is the residual. The contour interval is large (2 or 4 $\times 10^{-6}$ s^{-1}), yet values for the divergence and the residual exceed 20 $\times 10^{-6}$ s^{-1}, especially in the vicinity of mountain areas. The $\partial(\rho_o w)/\partial z$ term looks somewhat more reasonable.

The problems seen here originate directly from the algorithms used to extrapolate the various fields below ground. It is not clear what algorithms are used in practice. The ECMWF documentation claims that u, v are held constant below ground, while checks show that this is not quite the case. (The small discrepancy may arise from postprocessing to the 2.5° grid.) However, any procedure related to this will create major horizontal gradients between those regions where no such extrapolation occurs and where values from levels substantially above 1000 mb are carried down. Hence, the extremely large and unreasonable horizontal divergence field (Fig. 1). However, the vertical divergence term is well behaved under a constant ω below-ground scenario since the vertical gradients would be zero. In fact, ω is split into two parts, and only the part associated with the vertical column mass divergence is retained below ground, with the orographic lifting contribution excepted.

This finding provides a basis for making a reasonable correction in the regions affected by the below-ground extrapolation. We assume the vertical divergence term is correct and reduce the residual to exactly zero in selected areas by recomputing locally what the horizontal divergence should be. The key then is to choose the "selected areas." A priori, it seems logical to choose areas that are influenced by the high topography. However, another method is simply to select the areas where the residual is larger than a prespecified amount (e.g., 1×10^{-6} s^{-1}). Both methods have been used. Ultimately, there is a considerable advantage in designing a single fixed mask to indicate the regions to be subjected to this correction. The mask is given in Fig. 2, and its derivation is described below. The goal is to make adjustments so that the influence of the artificial below-ground atmosphere is benign, and its effects are not propagated throughout the rest of the atmosphere, as they will be if such local balances are not achieved. Note that the velocity field is corrected to fit the ω field, whereas it has been widespread in atmospheric science to do the opposite following the work of O'Brien (1970).

The procedure is to assume that the 500-mb level is uncorrupted. The mass residuals are computed in the next layer down (700–500 mb for our case), and where the mask for this layer applies (Fig. 2), the horizontal divergence is reset to make the residual zero. Because the layer divergence involves the velocities at both 500 and 700 mb and we have assumed the 500-mb level is not corrupted, the entire adjustment occurs in the 700-mb velocity. Note ω is not changed at either level. Then the next layer down (850–700 mb) is similarly adjusted, with all the adjustment taking place in the 850-mb wind, and so on.

The mask in Fig. 2 was designed reasonably objectively but with some empirical adjustments based on the observed residuals. When the magnitude of the residual alone is used, regions are identified over the tropical oceans that are associated with problems in the divergence field there [see Trenberth 1992, his Figs. 30–32], but these should more properly be addressed in following stages of the procedure. Therefore, the mask was based on the distribution of surface pressure and topography, plus the nearby regions influenced by the strong gradients in regions of steep topography. For the 1000–850 mb layer, there are extensive regions over the southern oceans in all months of the year where the mean surface pressure is less than 1000 mb, and thus the mass balance is corrupted in these regions. This region is even more extensive if we search for the lowest surface pressure during the month in question.

[1] Ehrendorfer et al. (1994) describe a procedure that corrects velocity but, in fact, is based upon a correction to the mass flow.

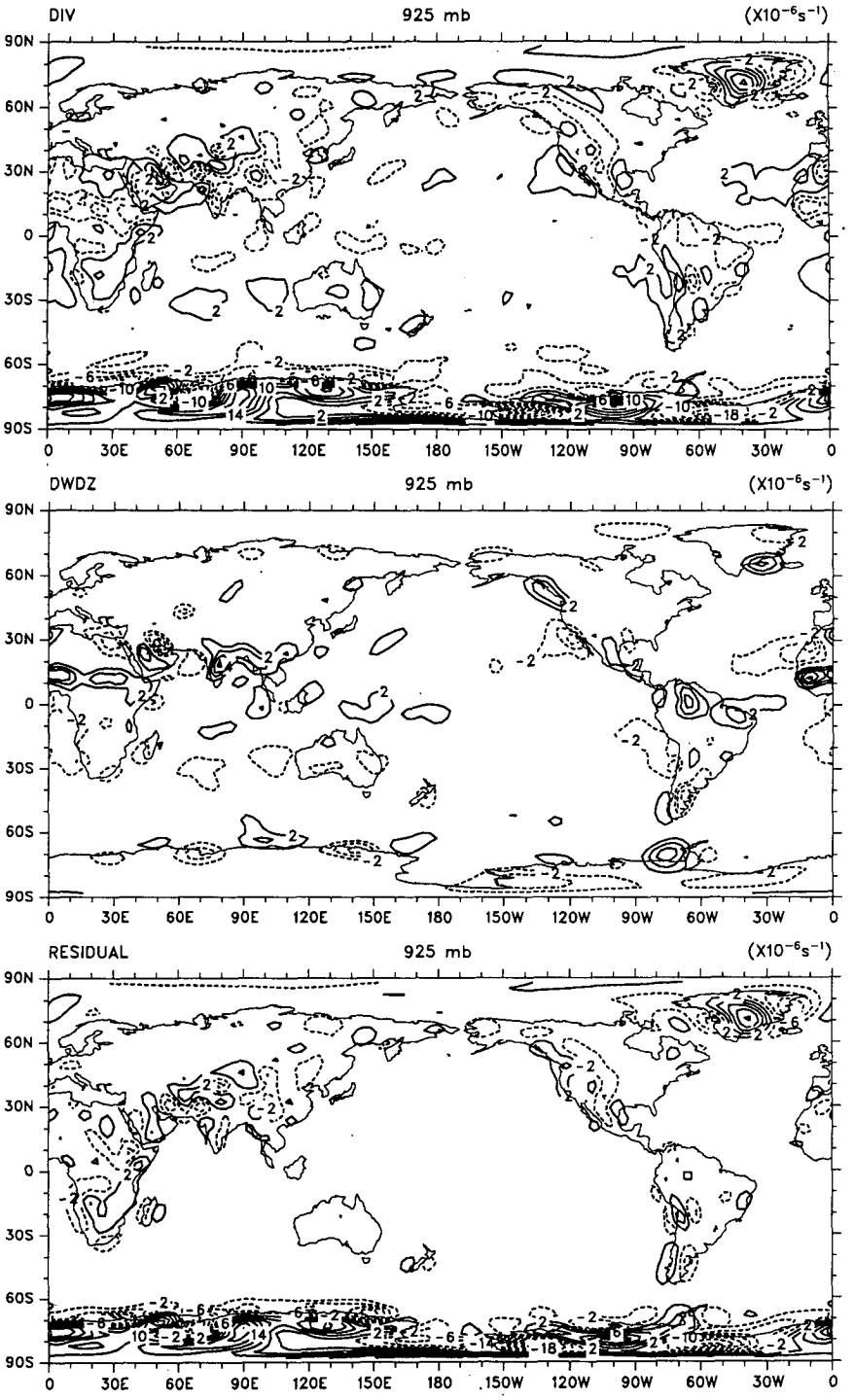

FIG. 1. For May 1988, the 1000–850-mb layer mean horizontal divergence $\nabla \cdot \rho_o \mathbf{v}$ (top), vertical divergence $\partial(\rho_o w)/\partial z$ (middle), and their sum, the residual R (bottom). The contour intervals are 4, 2, and 4 ($\times 10^{-6}$ s^{-1}), respectively. The contours start at ± 2 in the top and bottom panels, and the zero contour is omitted from the middle panel. Negative values are dashed.

However, we assume that small extrapolations for short periods are acceptable so that the monthly mean p_s is used. In the Northern Hemisphere, the only time of year when a similar problem occurs is in the winter when surface pressures are commonly below 1000 mb in the North Pacific and Atlantic. Accordingly, to ob-

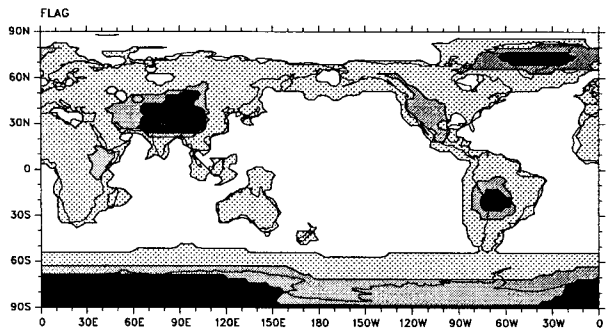

FIG. 2. Mask used for the below-ground regions for the first correction step for each layer. Successively dense stippling indicates the masks for the layers 1000–850, 850–700, and 700–500 mb.

tain a single mask, we use the mean surface pressure field for December–January–February. [Results show that it is desirable to be conservative and extend the mask farther afield, rather than underestimate the domain for this correction.] In addition, to objectively assess the influence of the topography near steep mountains on the divergence field, we make use of the magnitude of the Laplacian of the surface pressure field. At 1000 mb, for example, the mask is extended to regions where $|\nabla^2 p_s| > 5 \times 10^{-11}$ mb m^{-1}. This includes regions off the coast of South America where the Andes have a corrupting influence and also near Greenland, Antarctica, and the Himalayan plateau. For other layers, the same procedure was employed, except the Laplacian was applied to the minimum of the surface pressure and the value of the bottom of the layer, and the critical magnitude of the Laplacian was reduced somewhat so that the influence zone appropriately shrinks with height.

While this procedure sets the local below-ground residual to zero, because of the constraint that the global mean divergence must be zero and following spectral truncation, there is a global influence that

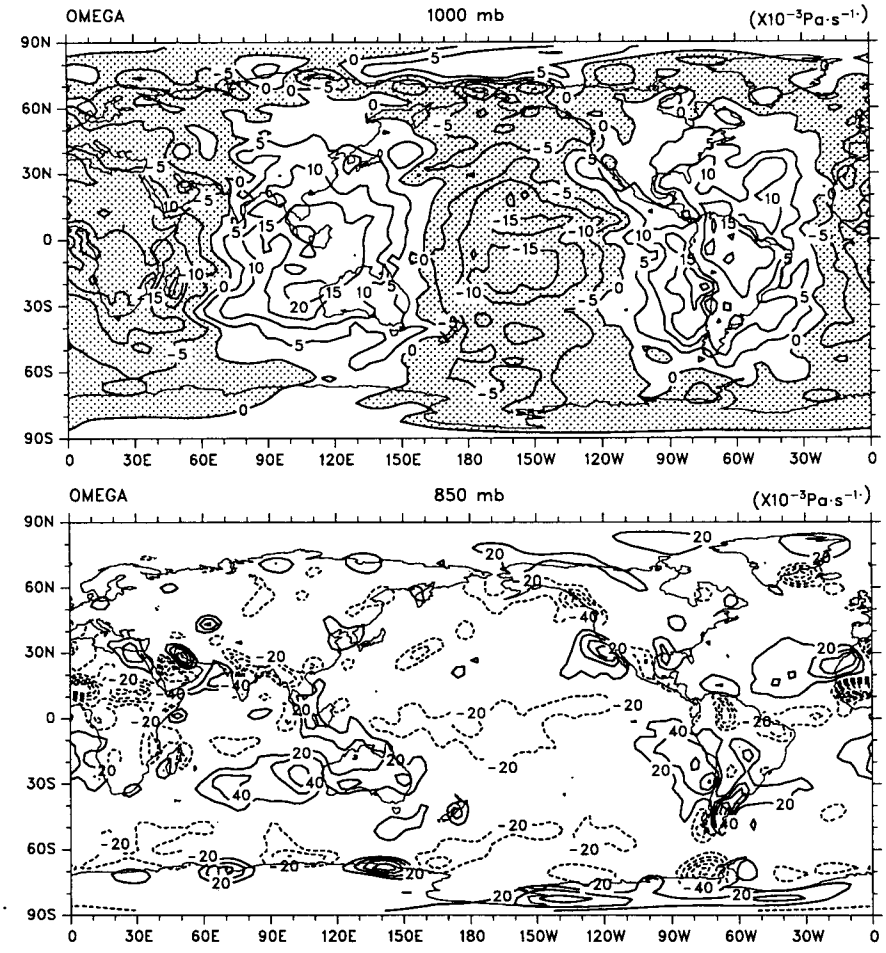

FIG. 3. Mean ω for May 1988 at 1000 and 850 mb. The contour intervals are 5 and 20 ($\times 10^{-3}$ Pa s^{-1}). Negative values are stippled at 1000 mb and dashed with the zero contour omitted from the 850-mb field.

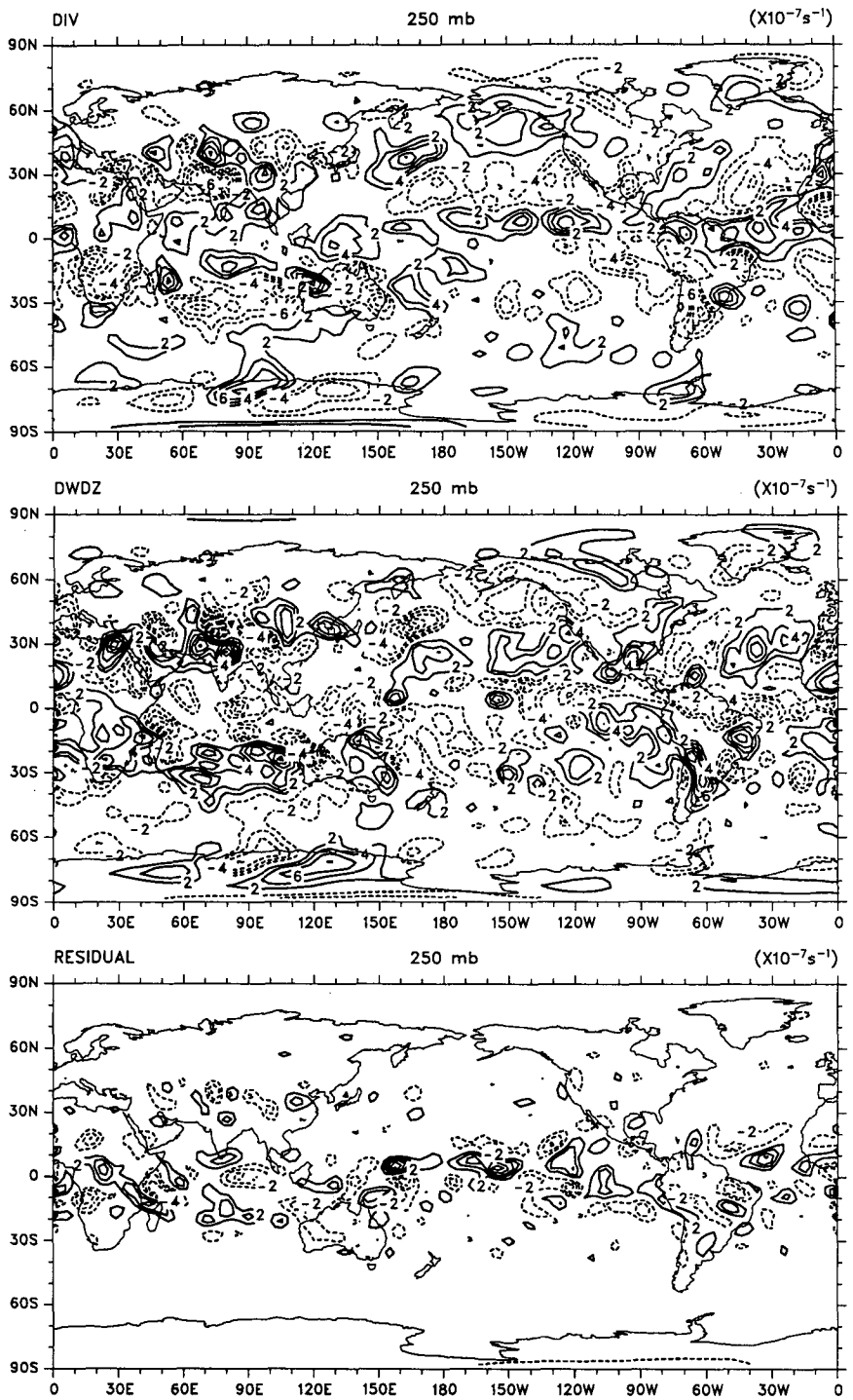

FIG. 4. For May 1988, the 300–200-mb layer mean horizontal divergence $\nabla \cdot \rho_o \mathbf{v}$ (top), vertical divergence $\partial(\rho_o w)/\partial z$ (middle), and their sum, the residual R (bottom), contour interval 2×10^{-7} s^{-1}. The zero contour is omitted, and negative values are dashed.

feeds back and creates a continuing small imbalance below ground. This is appropriately taken care of in step 4d.

c. The lower boundary and semidiurnal tide

The correct lower boundary condition for the atmosphere is Eq. (8a). This is built into the omega fields

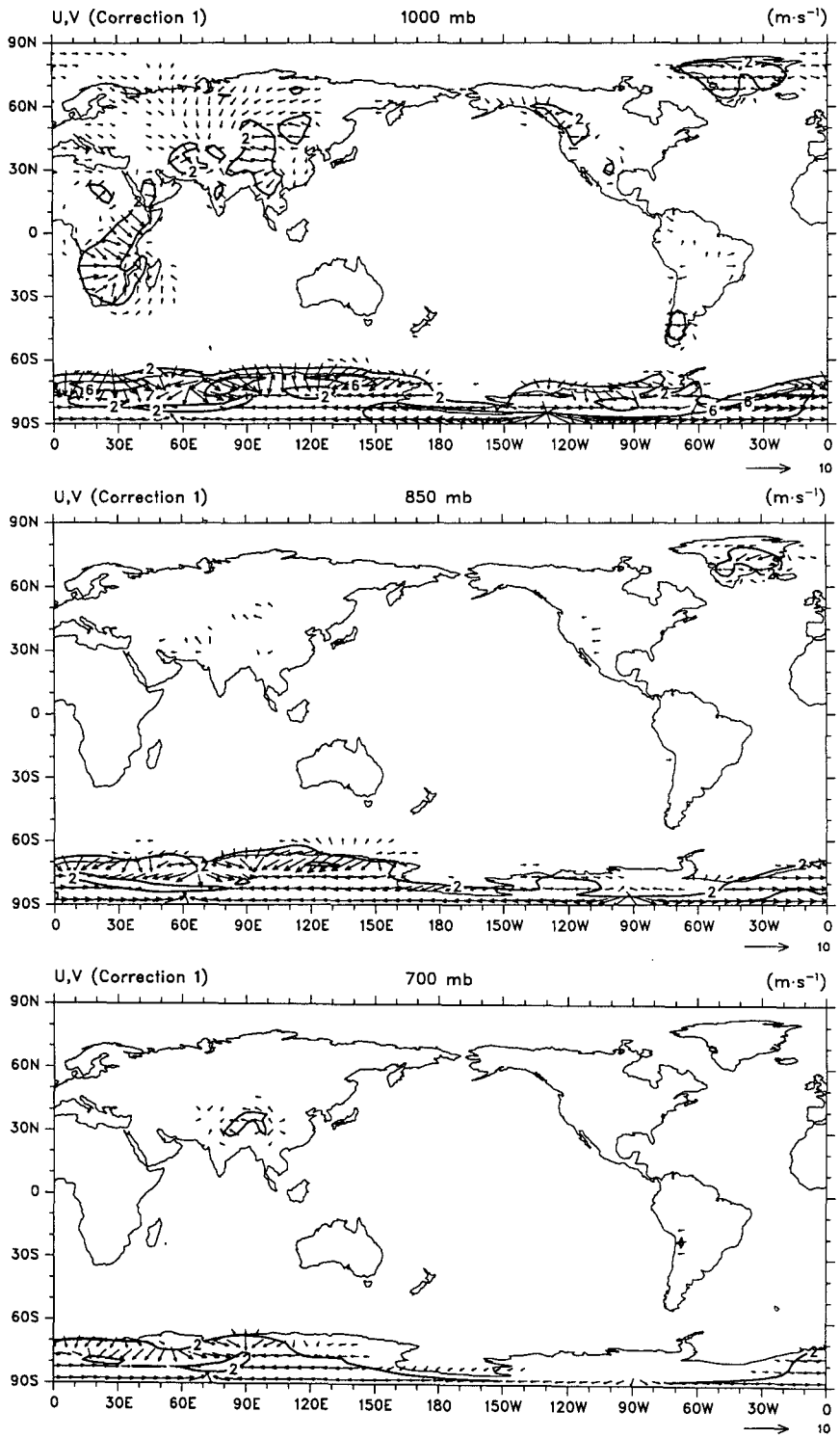

FIG. 5. For May 1988, the corrections to u and v for correction 1 in regions of high topography at 1000, 850, and 700 mb. Shown are vector plots, with magnitudes less than 1 m s^{-1} omitted and with the magnitude of the vector given as contours for 2 and 6 m s^{-1}. The scale vector is at lower right.

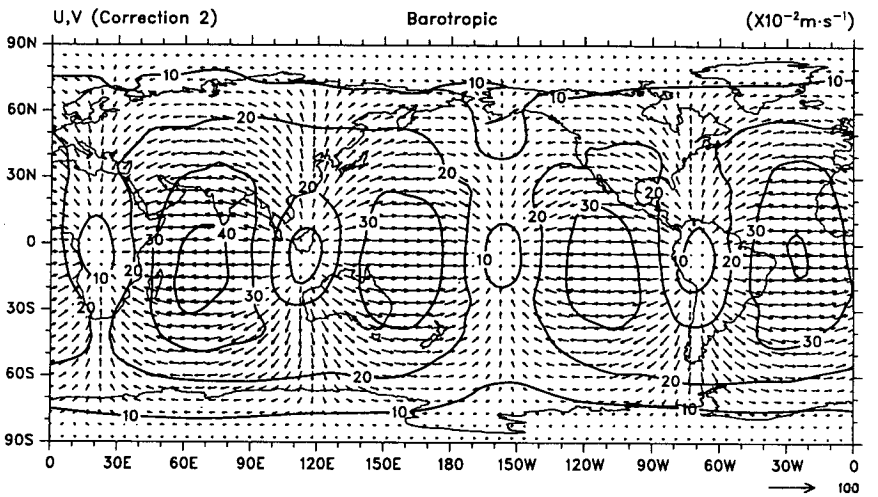

FIG. 6. For May 1988, vector corrections to u and v for correction 2, which is barotropic in height. Contours are magnitudes in cm s^{-1}.

from the analyses, but it is not practical for application in p coordinates. Trenberth (1991) shows that the semidiurnal tide is a prominent part of any apparent mass residual because of inadequate sampling throughout the day, so that the diurnal surface pressure tendency is aliased as a large-scale mass imbalance. Figure 3 shows the monthly mean May 1988 ω at 1000 and 850 mb. The 1000 mb ω is dominated by the semidiurnal tide pattern with large-scale wave 2 centered in the Tropics. Much of this is associated with the average diurnal surface pressure tendency at 0000 UTC plus 1200 UTC (see also Hsu and Hoskins 1989). The same pattern is also present, however, at other levels, although it is camouflaged by the larger and more regional vertical motions associated with convergence zones and orographic effects (e.g., at 850 mb). Also, because of the way ω is extrapolated below ground, the orographic component is missing at 1000 mb but is mostly present at 850 mb.

It seems the only way to remove the undesirable influence from the analyses is to require no mass flow through the lower boundary, and in p coordinates, the lower boundary applies at 1000 mb. Accordingly, the appropriate lower boundary condition is $\omega = 0$ at $p = 1000$ mb. If properly implemented, this will remove the semidiurnal tide and other diurnal effects.

Because we are considering only the adjustment for a change in the lower boundary condition, the problem is reduced to a two-dimensional one such that

$$\int (\mathbf{v}^c)^2 dV = \text{minimum},$$

where $\mathbf{v}^c$ is the horizontal correction velocity. Clearly, this is a minimum if it is identical at all levels and thus is barotropic. This step therefore requires an adjustment to the horizontal divergence of

$$\delta^c = \frac{\omega_1}{p_{oo} - p_t}, \tag{17}$$

and an adjustment to the vertical motions of

$$\omega^c = -\omega_1 \frac{p - p_t}{p_{oo} - p_t}, \tag{18}$$

where p_t is the top of the atmosphere and where the correction term, given by the superscript c, is added to the original value. In this way, $\omega_1 = 0$ after correction.

We tried to implement the changed lower boundary condition as part of the three-dimensional correction described in section 4d. There is no problem in doing so, but it turns out that it is equivalent to making a barotropic correction to the mass flow $\rho_o \mathbf{v}$ rather than to $\mathbf{v}$. The problem with the former is that as $p \to 0$, then $\rho_o \to 0$, and the correction to $\mathbf{v}$ becomes infinite. This is clearly unacceptable in addition to the fact that it violates the idea of making a minimal correction to $\mathbf{v}$ as expressed above.

d. Correcting the three-dimensional velocity

The constraint Eq. (16b) requires that the corrections should be irrotational. Because Eq. (6) has the vertical coordinate of pressure, it does not immediately lend itself to the form suitable for defining a potential function. Instead, it is necessary to adopt the $\ln p$ coordinates and the equation of continuity as in Eq. (10).

We define a potential function χ such that

$$\nabla_3 \chi = \rho_o(u, v, w), \tag{19}$$

so that from Eq. (10)

$$\left(\nabla^2 + \frac{\partial^2}{\partial z^2}\right)\chi = R, \tag{20}$$

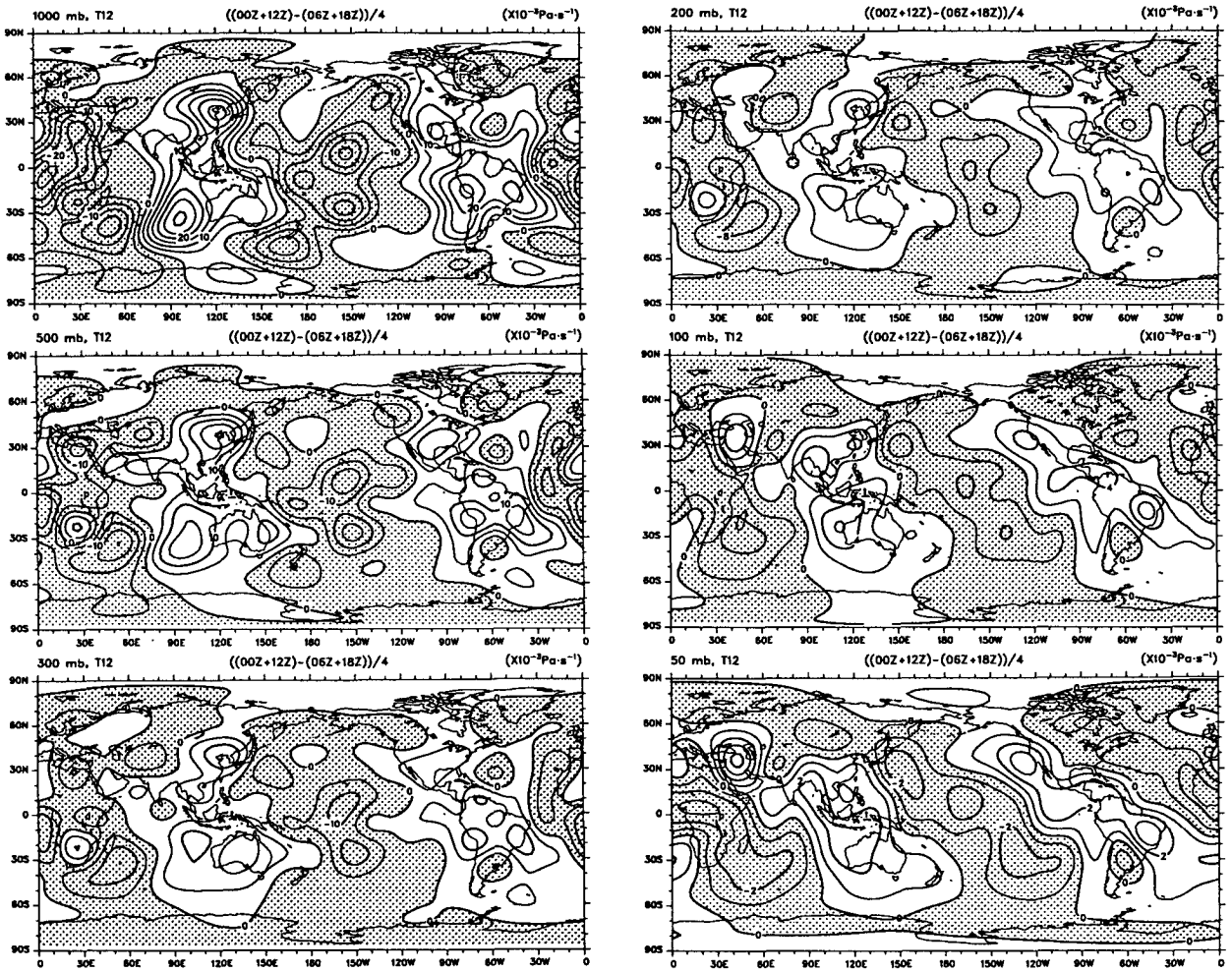

FIG. 7. For May 1988, the semidiurnal component of ω highlighted by taking the mean of the (0000 + 1200) − (0600 + 1800) UTC analyses at 1000-, 500-, 300-, 200-, 100-, and 50-mb levels. Negative values are shaded, and the fields have been truncated at T12. The contour intervals are 5, 5, 5, 4, 2, and 1 × 10^{-3} Pa s^{-1}, respectively.

where R is the residual, which should be zero, and the operator on the left-hand side has the appearance of a three-dimensional Laplacian. This is the advantage of the $\ln p$ coordinate system.

In practice, these quantities must be dealt with in terms of finite differences. To evaluate the equation of continuity Eq. (10) as exactly as possible, we have computed all terms at half levels, $j = 2, 4, 6, \cdots 2J - 2$, with

$$\mathbf{v}_j = \frac{1}{2}(\mathbf{v}_{j+1} + \mathbf{v}_{j-1}) \tag{21}$$

and $\mathbf{v}_{2J} = \mathbf{v}_{2J-1}$, so that Eq. (10) becomes

$$R_j = \frac{1}{2}(D_{j-1} + D_{j+1}) + \frac{W_{j+1} - W_{j-1}}{z_{j+1} - z_{j-1}}, \tag{22}$$

where $D_j = \nabla \cdot \rho_{oj}\mathbf{v}_j$, $W_j = \rho_{oj}w_j$, and R_j is the residual. The residuals of the similar equation in p coordinates evaluated locally at all levels were presented in Trenberth (1991), and as discussed earlier, the residuals are a substantial fraction of either term.

We solve Eq. (20) by using spherical harmonics and putting

$$\chi(\lambda, \phi, z, t) = \sum_{m=-N}^{N} \sum_{n=|m|}^{N} \chi_n^m(z, t) Y_n^m(\lambda, \phi), \tag{23a}$$

where λ is longitude, ϕ is latitude, t is time, a is the radius of the earth, and

$$Y_n^m(\lambda, \phi) = P_n^m(\phi) e^{im\lambda} \tag{23b}$$

are spherical harmonics of order m and degree n, and N is the truncation wavenumber. Similar expansions are used for R. Then Eq. (20) becomes

$$\left[\frac{-n(n+1)}{a^2} + \frac{\partial^2}{\partial z^2}\right]\chi_n^m(z) = R_n^m(z).$$

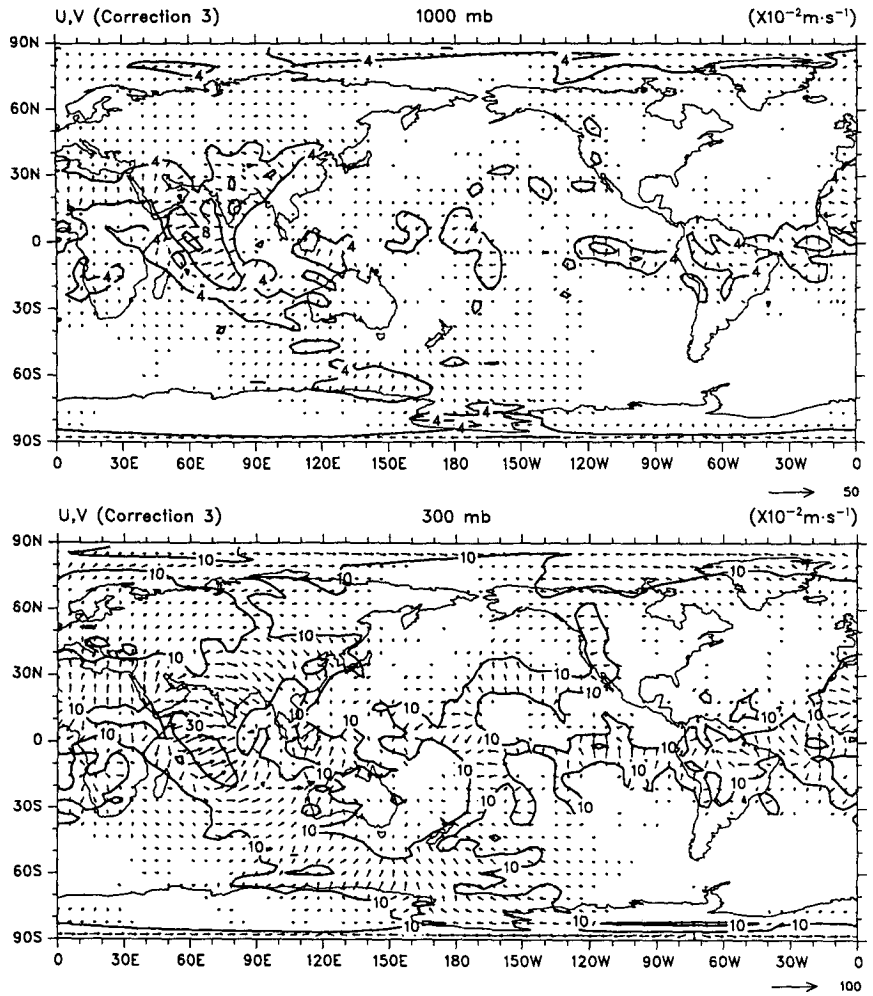

FIG. 8. For May 1988, vector plot correction 3 for u, v at 1000 and 300 mb. Contours are magnitudes in cm s^{-1}. Vectors less than 2 cm s^{-1} at 1000 and 5 cm s^{-1} at 300 mb are not plotted.

In this way we have reduced the dimensionality of the problem to 1, and this expression is readily solved in finite differences with appropriate boundary conditions.

Using centered finite differences, dropping the (m, n) superscripts and subscripts for the moment, and defining $\Delta z_j = z_{j+1} - z_{j-1}$, then

$$\frac{\chi_{j+2}}{\Delta z_{j+1}} + \frac{\chi_{j-2}}{\Delta z_{j-1}} - \left[\frac{1}{\Delta z_{j+1}} + \frac{1}{\Delta z_{j-1}} + \frac{n(n+1)}{a^2}\Delta z_j\right]\chi_j = R_j \Delta z_j \quad (24a)$$

for $j = 4, 6, \cdots 2J-2$. The upper boundary condition is $\omega = 0$ at $p = p_t$, which becomes

$$\frac{\chi_{j-2}}{\Delta z_{j-1}} - \left[\frac{1}{\Delta z_{j-1}} + \frac{n(n+1)}{a^2}\Delta z_j\right]\chi_j = R_j \Delta z_j \quad (24b)$$

at $j = 2J$. For the lower boundary, ω_1 is fixed and equal to zero. In the absence of the third step, however, the diurnal effects could be removed by setting $\omega_1 = 0$ and modifying the residual in Eq. (22) for the lowest layer, and this produces the case where the third adjustment is constant in height in the mass flow, rather than the velocity. For $j = 2$, the proper lower boundary condition becomes

$$\frac{\chi_4}{\Delta z_3} - \left[\frac{1}{\Delta z_3} + \frac{n(n+1)}{a^2}\Delta z_2\right]\chi_2 = R_2 \Delta z_2. \quad (24c)$$

Note that because we are solving a three-dimensional Poisson equation in Eq. (20), there are constants of integration that translate into lack of information on the global mean ω at each level. As noted earlier, we must specifically reset the global mean $m = 0$, $n = 0$ component for ω to zero to conserve mass.

The equations (24) can be solved for χ using a matrix inversion for each m, n and thus the total χ field may be constructed using Eq. (23). The correction velocity components for this step can then be recovered from

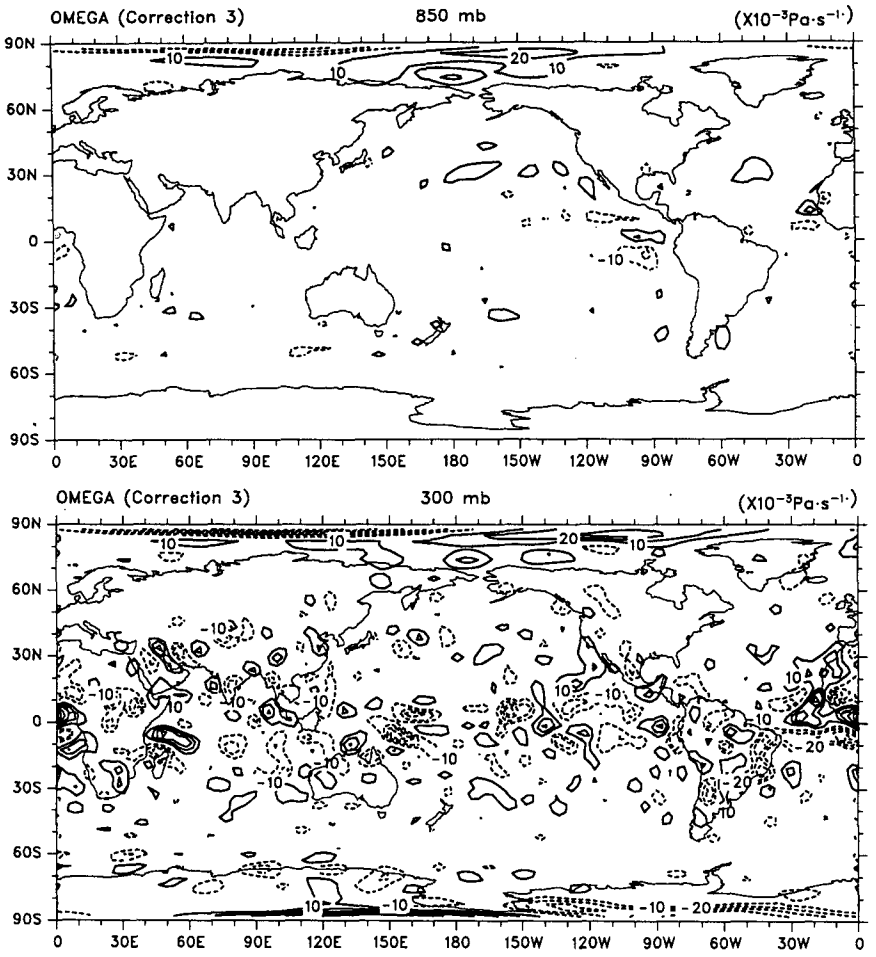

FIG. 9. For May 1988, correction 3 for ω at 850 and 300 mb. The zero contour is omitted, and negative values are dashed. Contour interval is 10×10^{-3} Pa s^{-1}.

$$\rho_o \mathbf{v}_j^c = \nabla \chi; \quad j = 2, 4, \cdots, 2J. \tag{25a}$$

$$\rho_o w_j^c = \frac{\chi_{j+1} - \chi_{j-1}}{\Delta z_j}; \quad j = 1, 3, \cdots, 2J - 1. \tag{25b}$$

This provides the desired correction for w, and thus ω, but it provides correction velocities only at the even levels, whereas the original fields are on odd levels. The series of expressions (21) provide J equations for J unknowns $\mathbf{v}_j$, $j = 1, 3, \cdots 2J - 1$ and can be easily solved starting at the top where $\mathbf{v}_{2J} = \mathbf{v}_{2J-1}$.

5. Application to May 1988

The first correction for the global mean ω is very small but important as it has such large scale. The other three corrections are larger and change the u and v fields and are referred to as corrections 1, 2, and 3.

The degree of mass imbalance in the lowest layer has already been shown in Fig. 1. Figure 4 presents the corresponding terms for the 300–200-mb layer, which is above the influence of the below-ground extrapolation but lies in the upper troposphere in the Tropics and lower stratosphere at higher latitudes near where the divergence reaches a maximum in the atmosphere. The magnitudes of the individual terms are large and reveal realistic features, such as the tropical convergence zones as upper-level divergence and subtropical convergence into the subtropical high pressure systems. The third panel shows that considerable but far from complete compensation occurs between the two terms. The residuals of consequence are found in the tropical regions and it is this imbalance that needs to be taken care of by the final correction procedure.

Correction 1 for the below-ground regions shown by the mask in Fig. 2 is given for May 1988 in Fig. 5. The corrections are largest at 1000 mb and are typically several m s^{-1} and, thus, are almost the same size as the fields themselves in these regions. Here we have chosen to provide vector plots with only vectors greater than 1 m s^{-1} in magnitude shown and with contours showing where the magnitude exceeds 2 and 6 m s^{-1}.

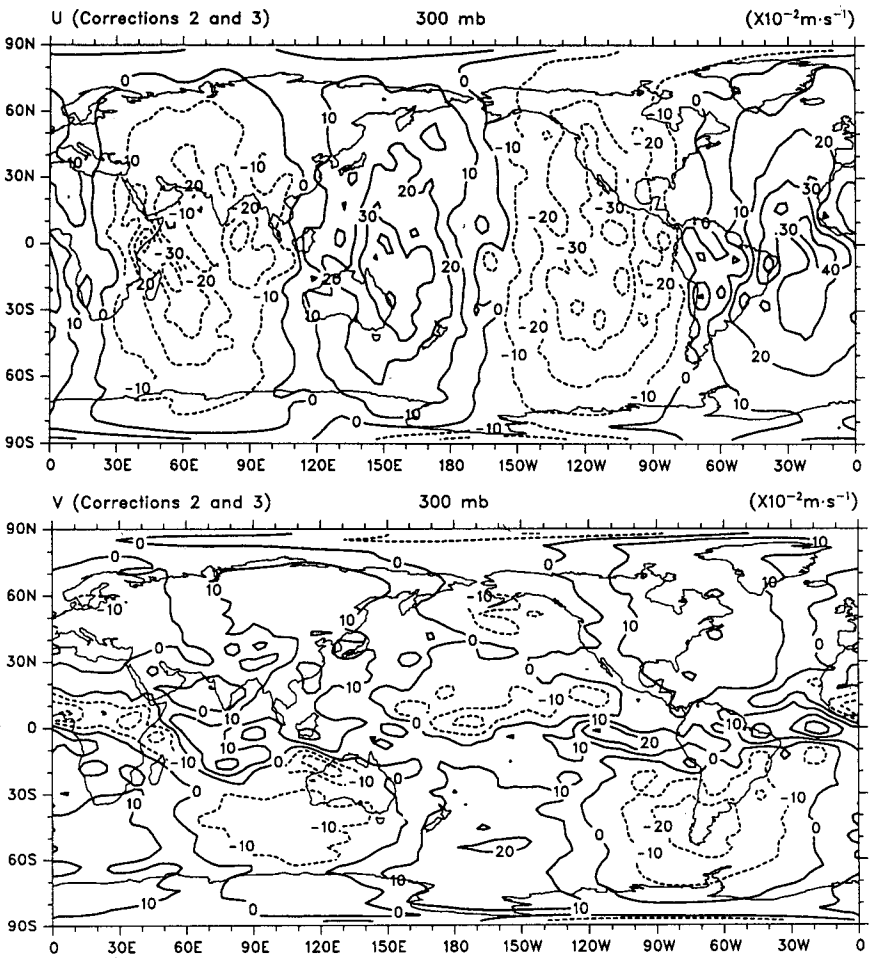

FIG. 10. For May 1988, total corrections for u, v at 300 mb. Negative values are dashed, and the contour interval is 10 cm s^{-1}.

Correction 2, which is barotropic in u, v, is shown in Fig. 6. It features a large-scale wave 2 pattern that is remarkably free from small-scale structure and typical magnitudes are 10–30 cm s^{-1}, at least an order of magnitude less than correction 1. The correction to ω is the reverse of Fig. 3 at 1000 mb and has exactly the same pattern with height but falls off with height according to Eq. (15).

Because the wave 2 structure is so obviously associated with the semidiurnal tide, we have used the four-times daily ECMWF analyses to examine the vertical structure of ω associated with the semidiurnal tide. These analyses differ from the ones used in most of this paper by being uninitialized and therefore considerably noisier. To highlight the semidiurnal component, we have taken the mean of the (0000 + 1200) − (0600 + 1800) UTC ω analyses (Fig. 7). Therefore, the results reveal the phasing corresponding to 0000 or 1200 UTC. Note that unlike the initialized analyses, it has been necessary to truncate at T12 in Fig. 7 to eliminate the otherwise considerable noise.

At 1000 mb, Fig. 7 shows that there is somewhat more structure and slightly stronger magnitudes to ω than for the initialized data in Fig. 3 (recall values are not truncated). The pattern is remarkably similar with height throughout the troposphere and falls off steadily in amplitude. We have computed the spatial standard deviation of the T12 ω field in Fig. 7 at each level to give an estimate of the amplitude with height. The ratio of this amplitude to the value at 1000 mb at 850, 700, 500, 300, 200, 100, and 50 mb is 0.93, 0.83, 0.68, 0.48, 0.35, 0.22, and 0.14, respectively. Thus, the falloff in amplitude with height is somewhat slower than the barotropic model would suggest. In the lower stratosphere, the spatial pattern changes. The wave 2 structure is still present, but it slopes westward with height in the Northern Hemisphere between about 200 and 50 mb and eastward with height in the Southern Hemisphere. The result is that whereas the phase of the wave 2 is roughly the same with latitude in the troposphere, there is a strong eastward slope from north to south in the stratosphere. At this time of year there

is a strong polar night westerly jet exceeding 50 m s^{-1} at 10 mb in the Southern Hemisphere, while weak easterlies are present in the Northern Hemisphere stratosphere. Thus, the slope with latitude in the semidiurnal tide follows the shear imposed by the prevailing winds.

Because we have specified this correction to be barotropic, whereas physically it contains more structure in the real atmosphere, some residual of the diurnal component will remain to be taken care of by the third correction. Nevertheless, it is clear that the simple barotropic correction does take care of most of it.

Results from correction 3 are shown in Fig. 8 for the horizontal wind components. This correction includes a lot of small-scale three-dimensional structure that has apparently arisen from the postprocessing. However, there are also some large-scale structures that emerge, notably wave 1 and wave 2 components that are most evident at 1000 mb and that are apparently also associated with diurnal variations. Whereas the second correction was based entirely on the 1000-mb ω, the pattern in Fig. 8 arises from the mutual adjustment of ω and u, v, and thus, it implies that the original 1000-mb ω is too simple in that it omits all the orographic influences. The magnitudes of correction 3 are of the order of several cm s^{-1} at low levels and increase somewhat to one- or two-tenths m s^{-1} in the lower stratosphere. The correction is comparable in magnitude to correction 2 but much smaller than correction 1.

The corresponding ω correction 3 is given in Fig. 9 for 850 and 300 mb. In the Tropics, this has a relationship at 300 mb to the residual in Fig. 4 for the 300–200-mb layer. In addition, at very high latitudes, an additional correction has emerged that propagates up from low levels. It arises from imbalances that exist below about 700 mb related to problems in the divergence field rather than the ω field, but these problems do not arise from below-ground extrapolations. Instead they arise from the way in which the ECMWF data were archived, as discussed in section 2.

The total correction is dominated by correction 1 at low levels. At 300 mb it consists of the sum of corrections 2 and 3 (Fig. 10). Although the correction varies somewhat with height in the free atmosphere, it now bears a striking resemblance to the two-dimensional correction, introduced by Trenberth (1991), based solely on the vertically integrated mass budget (except the correction, by definition, now has opposite sign). The latter dealt only with the layer from p_s to the top of the atmosphere. This agreement is very encouraging and indicates that we may have been successful in carrying out a three-dimensional implementation of mass correction that is reasonably compatible with the two-dimensional constraint.

6. Conclusions

A four-step procedure, as outlined in section 4, has been implemented to substantially improve the characteristics of the archived global analyses in p coordinates by ensuring that a mass balance occurs locally in three dimensions and for vertical columns extending both from the surface to the top of the atmosphere and from 1000 mb to the top of the atmosphere. This makes the analyses useful for other studies of heat, momentum, and moisture budgets and for advecting and determining the sources and sinks of trace gases.

Adjustments to the analyses using a single procedure will produce undesirable results by not properly accounting for the specific origins of the differing problems. It is important to first correct for the below-ground regions and make sure an exact local mass balance is achieved in those areas. Otherwise, effects from these regions can propagate throughout the atmosphere. Procedures for isolating the below-ground regions by applying a weighting of zero using Boer's (1982) β function as given in section 3c [see Eqs. (11) and (12)] can help but suffer from poor representation of a Heaviside function in the spectral domain. The second correction is designed to deal with the diurnal cycle, which is badly aliased using twice-daily data. Four-times daily data are highly desirable from this standpoint, as they can resolve the semidiurnal tide, but residual diurnal analysis components will still be present because of the diurnal distribution of data and other harmonics of the tidal components. The third correction is a universal one, designed to deal with all remaining problems by minimizing the three-dimensional flow adjustments and mutually adjusting the horizontal divergence and vertical motion fields.

The corrections are not negligible but they are fairly small. For u and v they amount to a few cm s^{-1} in the free atmosphere, well within the observational error. There is a strong barotropic component that mostly arises from the semidiurnal tide. In addition, the corrections below ground are substantial—several m s^{-1}—and imply that improvements in postprocessing to pressure coordinates are desirable.

The corrections have been such that $\nabla \cdot \int_{p_t}^{1000} \mathbf{v}\, dp$ is now identically zero, and this is desirable for many purposes. However, because it includes the below-ground artificial atmosphere, it is not the desirable physical quantity $\nabla \cdot \int_{p_t}^{p_s} \mathbf{v}\, dp$ that should go to zero or at least be about an order of magnitude less than it is. It will not be exactly zero, because we have ignored the monthly mean changes in pressure—the tendency term—which is small and given by Trenberth (1991), and the moisture contribution [cf. Eq. (5)]. However, correction 1 deals explicitly with the layer 1000 to p_s and sets that residual to zero. Thus, the remaining corrections deal with the layer from p_s to p_t.

These results have implications for the centers that produce the global analyses. A particular recommendation is that the divergent component of the horizontal velocity below ground on constant pressure surfaces should not be computed by extrapolation but should

make use of whatever procedures are in place for extrapolating ω to 1000 mb and computing the result from the equation of continuity.

Acknowledgments. This research was partially supported by the Tropical Oceans Global Atmosphere (TOGA) Project Office under Grant NA87AANRG-0208 and by NASA under NASA Order W-18 077. The data were supplied by ECMWF.

REFERENCES

Alexander, M. A., and S. D. Schubert, 1990: Regional earth–atmosphere energy balance estimates based on assimilations with a GCM. *J. Climate,* **3,** 15–31.

Boer, G. J., 1982: Diagnostic equations in isobaric coordinates. *Mon. Wea. Rev.,* **110,** 1801–1820.

——, and N. E. Sargent, 1985: Vertically integrated budgets of mass and energy for the globe. *J. Atmos. Sci.,* **15,** 1592–1613.

Ehrendorfer, M., M. Hantel, and Y. Wang, 1994: A variational modification algorithm for three-dimensional mass flux non-divergence. *Quart. J. Roy. Meteor. Soc.,* **120,** 655–698.

Hantel, M., 1986: Variational modification of the 3D-wind field. *Variational Methods in the Geosciences.* Y. Sasaki, Ed., Elsevier, 107–112.

Hsu, H-H., and B. J. Hoskins, 1989: Tidal fluctuations as seen in ECMWF data. *Quart. J. Roy. Meteor. Soc.,* **115,** 247–264.

O'Brien, J. J., 1970: Alternative solutions to the classical vertical velocity problem. *J. Appl. Meteor.,* **9,** 197–203.

Trenberth, K. E., 1991: Climate diagnostics from global analyses: Conservation of mass in ECMWF analyses. *J. Climate,* **4,** 707–722.

——, 1992: Global analyses from ECMWF and atlas of 1000 to 10 mb circulation statistics. NCAR Tech. Note. NCAR/TN-373+STR, 191 pp., plus 24 fiche.

——, 1995: Truncation and use of model-coordinate data. *Tellus,* **47A,** in press.

——, and J. G. Olson, 1988: An evaluation and intercomparison of global analyses from NMC and ECMWF. *Bull. Amer. Meteor. Soc.,* **69,** 1047–1057.

——, and A. Solomon, 1993: Implications of global atmospheric spatial spectra for processing and displaying data. *J. Climate,* **6,** 531–545.

——, J. C. Berry, and L. E. Buja, 1993: Vertical interpolation and truncation of model-coordinate data. NCAR Tech. Note NCAR/TN-396+STR. 54 pp.