Q. J. R. Meteorol. Soc. (2006), **132**, pp. 1527–1543 doi: 10.1256/qj.05.43

Gravity-wave drag estimation from global analyses using variational data assimilation principles. II: Case study

By M. PULIDO* and J. THUBURN
Department of Meteorology, The University of Reading, UK

(Received 16 March 2005; revised 18 October 2005)

SUMMARY

An estimate of the three-dimensional gravity-wave drag field derived from middle-atmosphere analyses is presented. It is obtained using a novel technique based on variational data assimilation principles described in Part I. The assimilation is performed for a one-week period starting on 1 July 2002 using Met Office middle-atmosphere analyses as the observations as well as the initial condition. The rotational component of the gravity-wave drag is robustly estimated by the technique. The zonal-mean drag estimates reach maximum amplitudes of 50 m s^{-1}day^{-1} at 0.4 hPa near the winter-hemisphere polar jet. In the middle stratosphere, drag estimates are much weaker with zonal-mean peak values of 5 m s^{-1}day^{-1} and show transient small-scale features not obviously correlated with the local background flow. Vertical variations in the sign and pattern of the estimated drag suggest filtering of the gravity-wave spectrum by the background flow.

KEYWORDS: Adjoint model Middle atmosphere Parameter estimation Twin experiments

1. INTRODUCTION

In Part I (Pulido and Thuburn 2005) we described a technique based on variational data assimilation principles to estimate the irreversible forcing exerted by small-scale gravity waves on the larger-scale flow.

The estimated gravity-wave drag (GWD) field is the one that minimizes the difference between the observed state and the state of a middle-atmosphere model as measured by a suitably defined cost function. The minimization calculation is carried out with the aid of an adjoint model, which calculates the sensitivity of the flow to the GWD.

The twin experiments presented in Part I show that the technique can estimate a prescribed drag accurately, given a perfect model and perfect data. The technique is able to estimate both meridional and zonal components of the three-dimensional GWD field, and its day-to-day variations. This represents a significant improvement compared with previous techniques used to infer the drag, which in general are able to estimate only a zonally or temporally averaged GWD field, and in some cases only the zonal component (Hamilton 1983; Marks 1989; Klinker and Sardeshmukh 1991).

Here, in Part II, we apply the assimilation technique using Met Office middle-atmosphere analyses (Swinbank and O'Neill 1994) as observations in order to estimate the GWD in the real atmosphere. The GWD is estimated daily for a one-week period. The aims are to present the estimated GWD in a case study, and also to illustrate the performance and reliability of the technique when applied to real data.

The GWD field in the middle atmosphere depends on several factors. These include the distribution of gravity-wave sources, the propagation of gravity waves, including reflection, refraction and filtering of the wave spectrum by critical-level absorption, and wave breaking and dissipation. Knowledge of the evolving three-dimensional distribution of GWD can give important information about these processes.

In general, the response over time of the atmosphere to a localized GWD will be non-local and nonlinear. However, as shown in Part I, for drag of realistic amplitude and

* Corresponding author: Department of Physics, FACENA, Northeastern National University, Av. Libertad 5460, (3400) Corrientes, Argentina. e-mail: pulido@exa.unne.edu.ar

over a timescale of one day, linearity is an excellent approximation. In the linear case, non-locality of the response can be expressed formally in terms of Green's functions. In principle, if we know the response we should be able to trace the information backwards and deduce the drag that caused it. The use of the adjoint model in our technique is doing exactly this.

The atmospheric flow responds to a transient GWD in two different ways (Blumen 1972; Zhu and Holton 1987; Weglarz and Lin 1998). The rotational component of the GWD forces a geostrophic mode, which grows with time; it leads to an irreversible change in the potential vorticity of the larger-scale flow. Both the rotational and divergent components of GWD also trigger larger-scale inertia–gravity waves; observations consistent with such waves have been found in radiosonde analyses (Scaruzzo *et al.* 1998). Since only the rotational component of GWD may irreversibly change the local larger-scale flow, it is this component that is of most dynamical interest, and we put most emphasis on the estimation of this component. The use of a numerical model based on potential vorticity and divergence as state variables, and the curl and divergence of GWD as forcing variables, allows these two different physical aspects to be distinguished. Indeed the potential-vorticity field is a clean field in which the inertia–gravity waves generated by the transient forcing do not appear to first order (see, for example, the appendix of Part I).

Our drag estimation technique makes no explicit allowance for errors in the observations or the middle-atmosphere model used. For the twin experiments described in Part I the observations and model are indeed perfect. However, when the technique is applied to observations of the real atmosphere these conditions are no longer satisfied.

First, the model's dynamics and radiation scheme might predict tendencies different from those due to the same processes in the actual atmosphere. The estimation technique will try to compensate for these differences through the estimated drag field, thus introducing errors in the estimated drag. Experiments using the middle-atmosphere model at a different resolution suggest that the dynamical errors are small so that, at least in the mesosphere where the drag is large (e.g. Hamilton 1983; Shine 1989; Marks 1989), these dynamical errors should not be significant. Other experiments reported in Part I show that radiation errors are also small; their effects can accumulate over time, but even then they tend to manifest themselves in the divergent component of the estimated drag field and to have a recognizably unphysical pattern.

In addition, errors in the observations will also lead to errors in the estimated drag. In particular, middle-atmosphere analyses are based mainly on temperatures derived from satellite observations; the balanced rotational component of the wind can be analysed more reliably than the unbalanced divergent component. Twin experiments described in Part I show that the dynamically important rotational component of the drag can be accurately estimated using only rotational wind information or, in the extratropics, only temperature information. Some issues related to model and data errors are discussed further in section 4.

2. Technique details

The assimilation system we use is described in Part I. It comprises a middle-atmosphere dynamical model, its adjoint, and a minimization algorithm. The assimilation system for drag estimation is referred to as ASDE.

The dynamical model integrates the spherical hydrostatic primitive equations, predicting the potential vorticity $Q = \sigma^{-1}(f + \zeta)$, the pseudo-density $\sigma = \rho \partial z / \partial \theta$ and the divergence δ (where f is the Coriolis parameter, ζ is the relative vorticity, ρ is the

density and θ is the potential temperature), and using an isentropic vertical coordinate and a hexagonal–icosahedral horizontal grid (Gregory 1999). The radiative processes are taken into account through a realistic scheme (Shine 1987) which models diurnally varying solar heating and the long-wave effects of CO_2, O_3 and H_2O. The ozone distribution used for the radiation calculation is prescribed using monthly means from a zonally averaged climatology. A time-dependent Montgomery potential calculated from Met Office analyses (Swinbank and O'Neill 1994) at a potential temperature of 414 K ($\approx$100 hPa) is imposed as a bottom boundary condition.

As 'observations' we use Met Office middle-atmosphere analyses. The analysed variables are u, v and T on pressure levels and on a regular longitude–latitude grid. These variables are interpolated to the hexagonal–icosahedral grid and to the isentropic levels used in the model. To initialize the model we also need data up to the model's highest level, about 0.018 hPa. Above the maximum height of the Met Office analyses ($\approx$0.4 hPa) the data are merged smoothly with data from the CIRA* climatology (Fleming *et al.* 1986). On the model grid the model state variables Q and σ are computed from u, v and T.

The period studied here covers 1–8 July 2002. These dates were chosen because we expect the strongest GWD near the solstice (e.g. Marks 1989). Furthermore, at that time there is a strong polar vortex in the southern hemisphere and high planetary-wave activity, which presents ideal conditions to examine the role of GWD in the middle atmosphere.

The cost function definition used in the minimization is

$$J = \frac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{L} \overline{\sigma}_l^2 (Q_{kl} - Q_{kl}^*)^2 + (\tau \overline{\sigma}_l)^{-2} (\sigma_{kl} - \sigma_{kl}^*)^2, \tag{1}$$

where k is an index to the horizontal position on the model grid and l the vertical level, the asterisk indicates an observed value, and the overbar indicates a horizontal mean. The choice of this particular form of the cost function is such that the minimization problem becomes well-conditioned for small perturbations to a resting background flow, as shown in Part I. The characteristic timescale is taken to be $\tau = 4 \times 10^4$ s, as in Part I.

As discussed above and in Part I, the divergent component of the GWD is dynamically less important than the rotational component because it does not produce a permanent change in the larger-scale flow. Also, the estimated divergent drag is more sensitive than the rotational component to model and data errors (see Part I and section 3(b) below). For these reasons, our main results are computed using only the rotational drag as a control variable, with the divergent drag fixed at zero. For comparison, we also present an estimation using both rotational and divergent drag as control variables (section 3(b)).

The top of the Met Office analyses is at 0.4 hPa; therefore we include contributions only from pressures greater than this in (1). However, we allow the control variables to be non-zero anywhere up to the model top at about 0.018 hPa, since effects at lower altitudes may come from GWD at higher altitudes (Haynes *et al.* 1991). The results show a very small GWD estimated above the top of observations; this is discussed further in section 3.

The assimilation window length used in these experiments was 24 hours, so that seven assimilation windows cover the one-week period studied. For the first assimilation window, the initial conditions are taken from the Met Office analyses. Since divergence from analyses is not expected to be a reliable field we take $\delta = 0$ as initial condition.

* Cooperative Institute for Research in the Atmosphere.

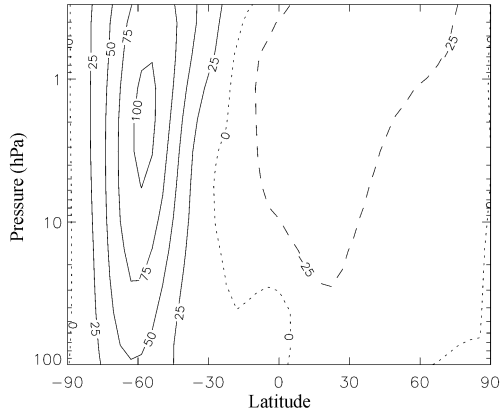

Figure 1. Weekly averaged zonal-mean zonal wind (m s^{-1}) from Met Office analyses.

For subsequent windows the initial conditions (including divergence) are taken as the final model state from the previous assimilation window when the model is evolved with the best estimate of the GWD (see Part I). In this way, the model evolves continuously during the one week period. We find that, when the model uses the best estimate of GWD in each 24 hour window, it is able to track the Met Office analyses closely for the whole week. (Other experiments confirm that this remains true for at least one month, which is the longest assimilation period tested.)

We find a similar good convergence of the assimilation scheme using Met Office analyses to that found in the twin experiments described in Part I. The results presented here use 25 iterations of the minimization scheme in each assimilation window.

3. Results

(*a*) *Estimates of GWD*

Figure 1 shows the weekly and zonally averaged zonal wind from Met Office analyses from 1–8 July 2002. There is a strong zonal jet that exceeds 100 m s^{-1}. In the year 2002, the polar vortex was particularly located at high latitudes. Preliminary tests suggest that strong GWD is found for years with a high-latitude polar vortex at the winter stratopause. The temperature field has, among other features, a maximum at the winter stratopause (Fig. 2), which is believed to be produced by forcing both by planetary and small-scale waves (e.g. Andrews *et al.* 1987).

First, the dynamical model was run for seven days starting from 1 July 2002 with no GWD parametrization except for a linear damping of the winds in the two highest layers (coefficients were 2.2×10^{-6} s^{-1} at $\theta = 4500$ K and 5.3×10^{-6} s^{-1} at $\theta = 6400$ K), which acts as a 'sponge layer' avoiding spurious reflections at the top of the model. As in Part I, we call the evolution of the model without GWD the 'control' case. Figure 3 shows the differences between the Met Office analysis and the control case. As is usually found in models with no representation of GWD, the model jet is too strong and it is located at too high a latitude. Maximum differences reach -40 m s^{-1} at high altitudes around 75°S. Note also that the strength of the summer hemisphere jet is overestimated at low latitudes above 1 hPa.

Figure 4 shows the temperature differences between the Met Office analysis and the control case. The model without drag cannot reproduce the warm temperatures

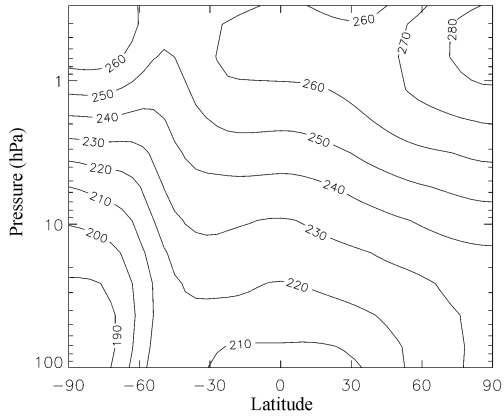

Figure 2. Weekly averaged zonal-mean temperature (K) from Met Office analyses.

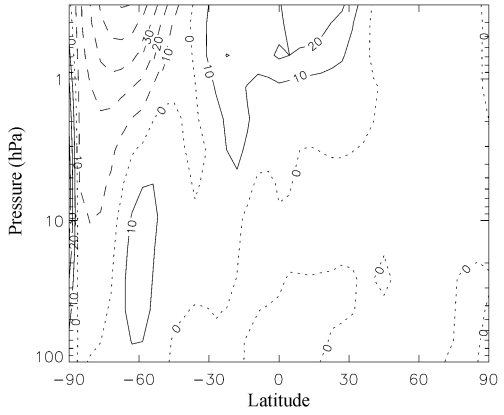

Figure 3. Zonal wind difference (m s^{-1}) between Met Office analysis and the control run on 8 July 2002.

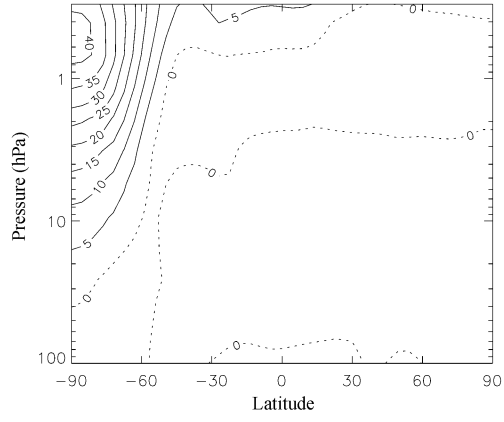

Figure 4. Temperature difference (K) between Met Office analysis and the control run on 8 July 2002.

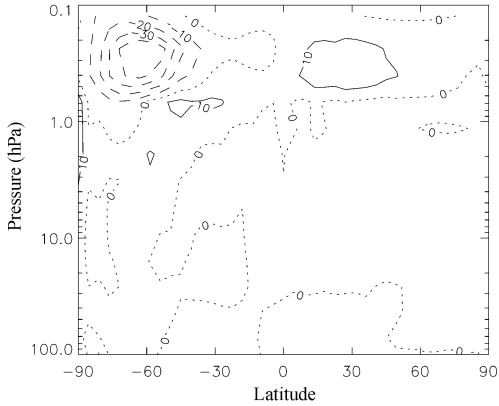

Figure 5. Weekly averaged zonal-mean estimated zonal gravity-wave drag (m $s^{-1}$$day^{-1}$).

at the winter pole where temperatures are 40 K colder than analyses. Again, this problem is often found in middle-atmosphere models and is known as the cold pole bias (e.g. Andrews *et al.* 1987).

The zonal-mean zonal GWD estimated using the ASDE is shown in Fig. 5; it shows a strong drag centred at 0.4 hPa and 65°S, which is decelerating the zonal jet, but peaking above and poleward of the jet core. There is also a weaker decelerating centre at low latitudes in the summer hemisphere that is reducing the strength of the summer jet. Interestingly, it is also found above the jet core. This decelerating centre is related with the positive bias observed in Fig. 3.

Previous estimates of the GWD using mean equations are consistent with the results presented here. Marks (1989) showed estimates that reach peak values of 55 m $s^{-1}$$day^{-1}$ found at 0.1 hPa, while Fig. 5 shows peak values of 50 m $s^{-1}$$day^{-1}$ at 0.4 hPa. In the summer hemisphere similar estimates are also found; Marks's estimates show a maximum of 15 m $s^{-1}$$day^{-1}$ centred at 30°N while ASDE also estimates a centre at 30°N of 10 m $s^{-1}$$day^{-1}$.

At lower altitudes the GWD weakly accelerates the jets in both hemispheres. This result coincides with a similar feature found in a zonal-mean budget study by Alexander and Rosenlof (1996). Such a pattern of drag, with weak acceleration below the jet core and strong deceleration above, is predicted on the basis of filtering of gravity waves with a broad phase-velocity spectrum by their respective critical layers (e.g. Warner and McIntyre 1996; Hines 1997).

The estimated drag notably reduces the wind and temperature errors seen in the control experiment. Figures 6 and 7 show the zonal-wind and temperature differences between Met Office analyses and the model evolution with the estimated drag. The strong polar jet and the cold pole bias problems have been completely solved. In particular, the jet presents its maximum strength at the same height as the Met Office analyses. The biggest remaining differences are now found in the tropics, where the variational data assimilation technique does not perform so well (see Part I).

The pattern of the estimated GWD is not directly proportional to the differences between the control case and the analyses (compare Fig. 5 with 3); it is much more concentrated in height. However, the effects of the drag are spread in height, so that the estimated zonal wind and temperature are very similar to the analyses (Figs. 6 and 7). Thus, there is a significant non-local response to the estimated drag, even for these short

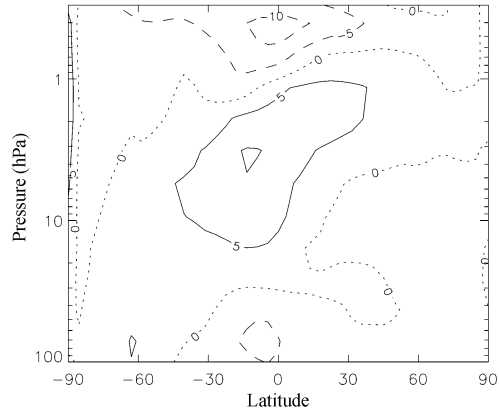

Figure 6. Zonal wind difference ($m\,s^{-1}$) between Met Office analysis and the assimilation with ASDE on 7 July 2002.

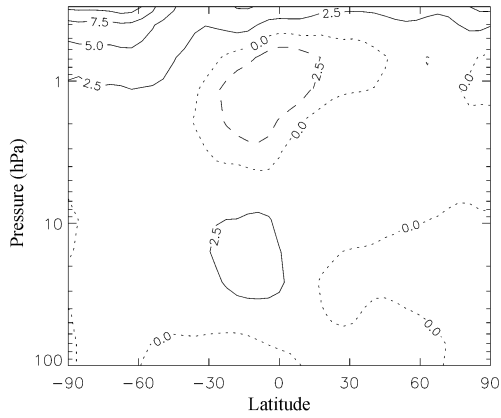

Figure 7. Temperature difference (K) between Met Office analysis and the assimilation with ASDE on 7 July 2002.

timescales (Haynes *et al.* 1991), and the assimilation system is able to capture this non-locality.

The maximum estimated drag is found at the top of the region of observations. It is not clear whether even larger drag values would be found at higher altitudes if higher-altitude observations could be included in the cost function (1). In principle, the technique is able to estimate the GWD at altitudes above the observations since the effects of that GWD will be felt in the region of observations through the downward control mechanism (Haynes *et al.* 1991). However, in idealized twin experiments to test this, the technique was found to greatly underestimate drag values above the region of observations. The problem seems to be that downward control is a non-local effect operating through the divergent part of the circulation; therefore, estimating drag from only its downward control effect involves a poorly conditioned minimization problem (see Part I), which would require very many iterations for an accurate solution. Note, however, that the ASDE system does capture the downward control effect of GWD, as noted in the preceding paragraph, provided that the drag is located within the observation

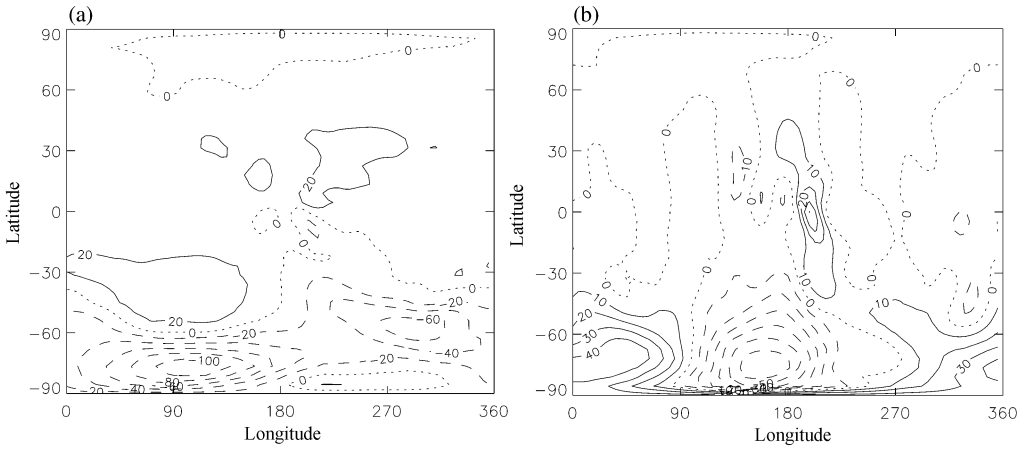

Figure 8. Estimated (a) zonal and (b) meridional gravity-wave drag (m $s^{-1}$$day^{-1}$) at 0.4 hPa on 1–2 July 2002.

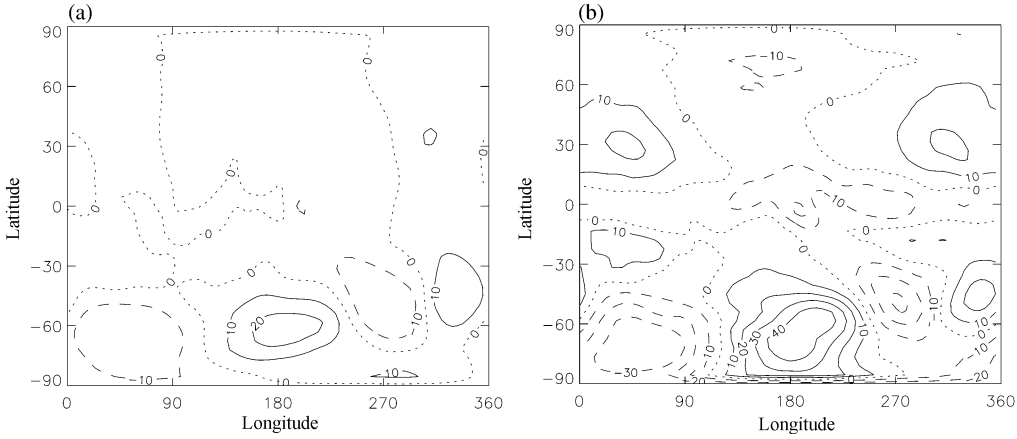

Figure 9. (a) Observed meridional wind (m s^{-1}) and (b) control meridional wind at 0.4 hPa on 2 July 2002.

region so that the local effect of the drag on the potential vorticity is observed, the minimization problem is well-conditioned (Part I), and the drag itself is well estimated.

Figure 8 shows the weekly averaged zonal and meridional components of the estimated GWD in a horizontal section at 0.4 hPa. Both components of the drag are strongly, though not perfectly, anticorrelated with the local background flow. In particular, not only does the drag decelerate the jet, but it also tends to damp the planetary wave activity (see Fig. 9(a)), which is roughly twice as strong in the control integration as in the analyses (compare with Fig. 9(b)).

At lower altitudes the zonal and meridional components of estimated GWD (Fig. 10) are not correlated with the local flow, and the scale of features in the estimated drag field is much smaller. However, these small-scale GWD features are found at high latitudes where the polar vortex is located.

The assimilation system assumes that GWD is constant with time in each one-day assimilation window. Apart from this, it imposes no further constraints on the time variation, since the drag values in successive assimilation windows are completely

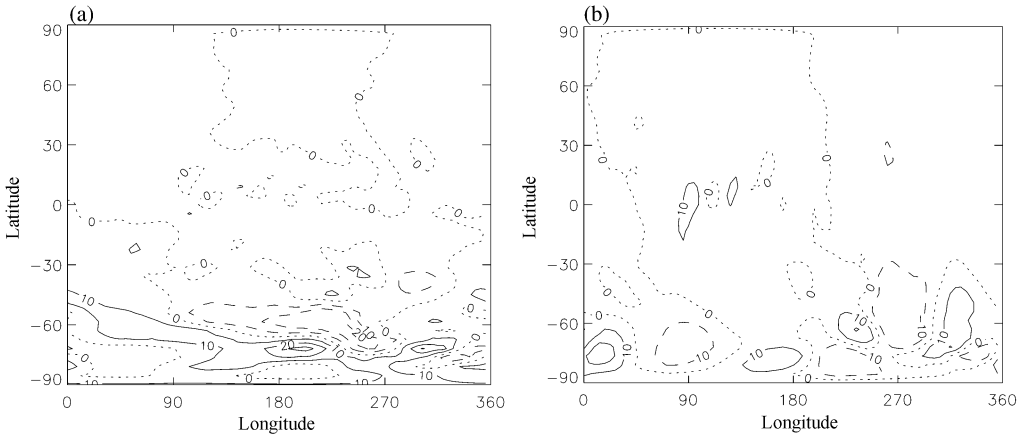

Figure 10. Estimated (a) zonal and (b) meridional gravity-wave drag (m $s^{-1}$$day^{-1}$) at 9 hPa on 1–2 July 2002.

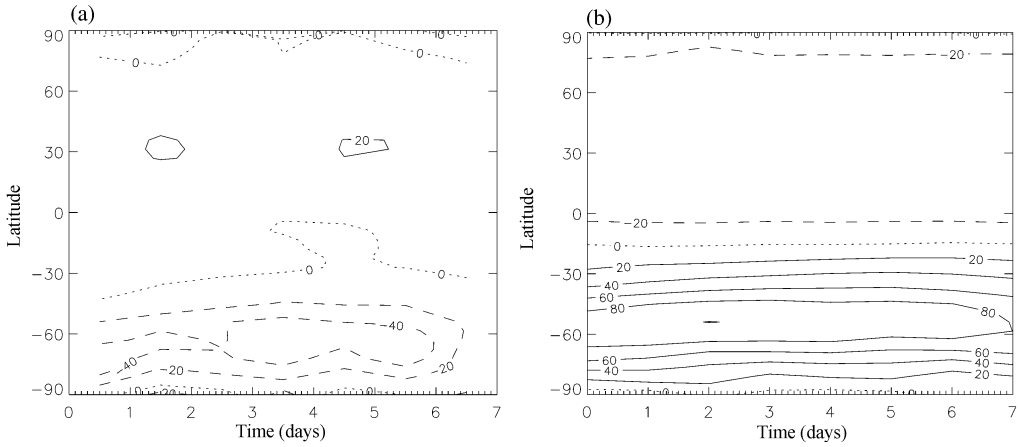

Figure 11. Time dependence of (a) the estimated zonal-mean zonal gravity-wave drag (m $s^{-1}$$day^{-1}$) and (b) the zonal-mean zonal wind (m s^{-1}) at 0.4 hPa.

independent parameters. Nevertheless, at 0.4 hPa the assimilation system estimates a rather smooth time evolution of the drag. The zonal component is shown in Fig. 11; the meridional component also evolves smoothly (not shown). The GWD pattern in Fig. 11 seems to be strongly linked to the background flow patterns which are also evolving smoothly (Fig. 11(b)). In particular, in the two last days of the assimilation (6–7 July) there is a decrease in the zonal wind, which coincides with a reduction of the zonal drag.

The time dependence of the zonal GWD at 9 hPa is shown in Fig. 12. At this altitude the drag varies on a shorter timescale than it does at 0.4 hPa. Some structures last at most two or three days and then disappear.

There are striking differences between the estimated drag in the middle stratosphere (9 hPa), which is noisy and transient, and in the lower mesosphere (0.4 hPa), which is large scale and smoothly evolving. If we can assume that the differences between the estimated drag in the middle stratosphere and lower mesosphere are indicative of differences in the real atmosphere (see discussion in section 4) then this suggests

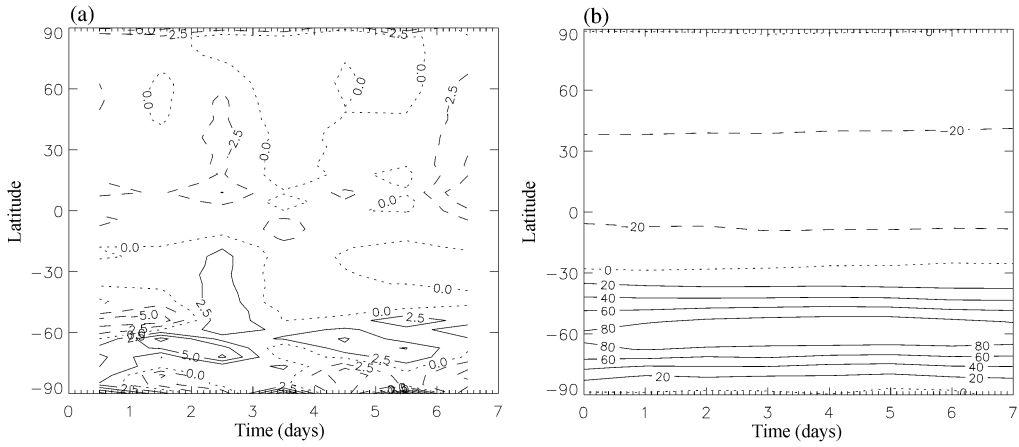

Figure 12. Time dependence of (a) the estimated zonal-mean zonal gravity-wave drag (m $s^{-1}$$day^{-1}$) and (b) the zonal-mean zonal wind (m s^{-1}) at 9 hPa.

an interesting physical interpretation. As discussed in section 1, the pattern of drag will reflect both the gravity-wave sources and filtering by the flow through which the gravity waves have propagated. In the middle stratosphere, relatively close to the tropospheric gravity-wave sources, the drag pattern might reflect the expected small scales and transience of the gravity-wave sources. In the lower mesosphere, much further from the gravity-wave sources, part of the broad gravity-wave spectrum should have been filtered by the slowly evolving background flow through which the gravity waves propagate. Therefore, the wave-drag pattern should reflect the persistent forcing by the small portion of the spectrum with one-signed zonal phase speed that can reach the mesosphere.

If we can neglect the horizontal divergence of the pseudo-momentum flux (as all GWD parametrizations do) then the zonal component of gravity-wave drag X_x is given by

$$X_x = -\frac{1}{\rho}\frac{\partial F}{\partial z}, \tag{2}$$

where F is the upward flux of zonal pseudo-momentum. A similar expression holds for the meridional component. We can vertically integrate

$$F_\mathrm{b} - F_\mathrm{t} = \int_{z_\mathrm{b}}^{z_\mathrm{t}} \rho X_x \,\mathrm{d}z = \int_{\theta_\mathrm{b}}^{\theta_\mathrm{t}} \sigma X_x \,\mathrm{d}\theta \tag{3}$$

to obtain an expression for the gravity-wave source F_b, assuming that the pseudo-momentum flux out of the top of the domain F_t is negligible.

Figure 13 shows the weekly averaged zonal and meridional components of the mass-weighted vertical integral of the estimated drag. It does not appear to be related in any obvious way to the expected pattern of gravity-wave sources such as orography, convection, and weather systems. In particular, the global wave-number 1 pattern in the meridional component is unlikely to reflect real gravity-wave sources. One possibility is that it might arise because of differences in the tidal signal between the dynamical model and the Met Office analyses, which the estimated drag tries to compensate for. However, repeating the experiment with the diurnal cycle of radiation switched off in

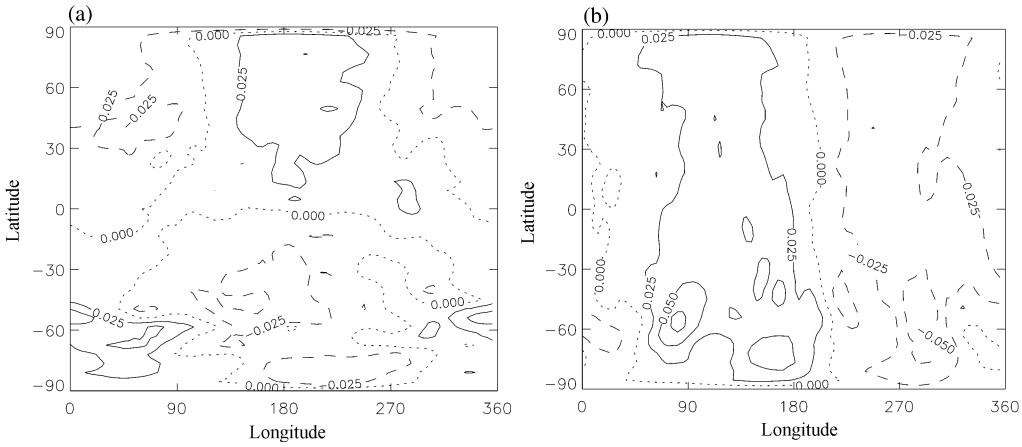

Figure 13. Mass-weighted height integral of (a) the zonal and (b) the meridional gravity-wave drag (N m^{-2}).

the model shows very similar results to Fig. 13, implying that tidal differences are not the explanation. Because of the mass weighting, the vertical integral is dominated by contributions from the lowest altitudes, which, as shown in Fig. 10, are less coherent in space and time. In summary, it appears that the technique is not currently able to give a useful estimate of the gravity-wave source distribution. It is not clear whether the inaccuracies are dominated by data errors, model errors, limitations of the assimilation system itself, or simply the assumption that horizontal propagation is negligible.

(*b*) *Sensitivity of the results to the technique*

For the results presented so far, only the rotational component of the drag was estimated. In principle, as we showed in Part I, the technique is also able to obtain information about the divergent component, even when the horizontal wind divergence is not available from observations. Now we examine the results of experiments in which both the curl and divergence of the GWD are used as control variables to see if the technique is able to obtain information about the divergent component of drag using middle-atmosphere analyses as input data.

Because of the near linearity of the drag estimation problem, mentioned in section 1 and discussed in Part I, the rotational component of the drag is almost identical to that in the previous experiment. Since the zonal-mean zonal drag is non-divergent and is therefore defined completely by the rotational drag, this field too (Fig. 14(a)) is almost identical to that in the previous experiment (Fig. 5). The zonal-mean meridional drag, on the other hand, is irrotational and is defined completely by the divergent component of the drag. This field was, therefore, identically zero in the previous experiment. In the new experiment it has a global-scale pattern of very large values (Fig. 14(b)). Such a pattern of meridional drag is not expected on any physical grounds. Clearly the technique is not estimating this component well.

One candidate explanation for the unphysical estimates of mean meridional drag might be initialization of the model, which might generate zonally symmetric gravity modes, which the meridional drag then tries to correct. However, experiments with different ways of initializing the model show little sensitivity to initializing the divergence field from the Met Office analyses rather than setting it to zero. Moreover, in longer

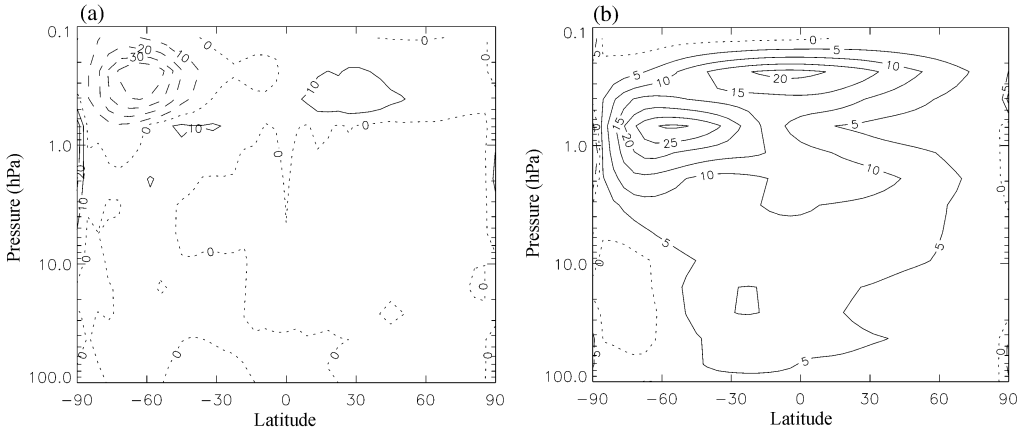

Figure 14. Weekly averaged (a) zonal-mean estimated zonal gravity-wave drag (GWD) ($\mathrm{m\,s^{-1}day^{-1}}$) and (b) meridional GWD from an estimation with the GWD curl and divergence as control variables.

assimilation experiments (up to one month) we would expect the effect of the initial conditions to diminish, but we find that the unphysical estimates of mean meridional drag actually grow with time. A second possible explanation is that the mean meridional drag might be associated in some way with a poor representation of the tides, either in the model or in the Met Office analyses. However, the estimated drag is found to show little sensitivity to switching off the diurnal cycle in the radiation scheme. Thus, neither the initialization nor the representation of the tides appear to be able to explain the unphysical estimates of mean meridional drag.

The poor estimation of the divergent drag appears to occur for the following reasons. First, the divergent wind does not appear in the cost function (1); the divergent wind, if it were reliable, would give the most direct information on the divergent drag. In principle, the σ field (and, at higher order, the Q field) do contain information about the divergent drag. However, the dynamical response to a divergent drag is primarily in the form of inertia–gravity waves; these propagate away from the region of the drag and are partly dissipated. Thus, some of the information is lost through dissipation, and the rest is in a form that can not be recovered efficiently by the iterative minimization scheme (i.e. it is less well-conditioned—see Part I). Experiments with different forms of the cost function show that the large-scale meridional drag pattern in Fig. 14 is estimated on the basis of information in the potential-vorticity data. If the control run has a certain potential-vorticity error, the assimilation scheme can try to correct this error through a combination of local rotational drag (directly modifying the potential vorticity) and a large-scale meridional circulation (driven by a large-scale meridional drag) that advects the air towards a latitude where its potential-vorticity value is correct. The lack of divergent-wind information in the cost function means that this unrealistic large-scale meridional circulation is not penalized.

The estimated rotational drag, on the other hand, appears to be very robust to changes in details of the technique. Figure 15 shows results of two extreme experiments; in panel (a) $\tau = 0$ so that all the weight is in the σ term of (1); in panel (b) $\tau = \infty$ so that the cost function is only defined by the Q term. The estimated zonal drag fields are remarkably similar except in some details; there is a weaker summer hemisphere centre in the $\tau = 0$ case and a stronger acceleration centre at the height of the jet core.

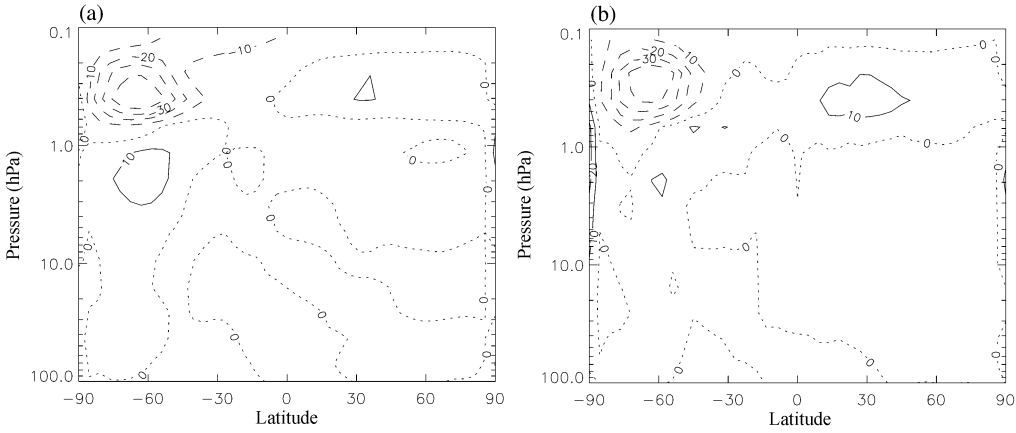

Figure 15. Weekly averaged zonal-mean estimated zonal gravity-wave drag m s^{-1}day^{-1} (a) with $J(\sigma)$ and (b) with $J(Q)$ (see text).

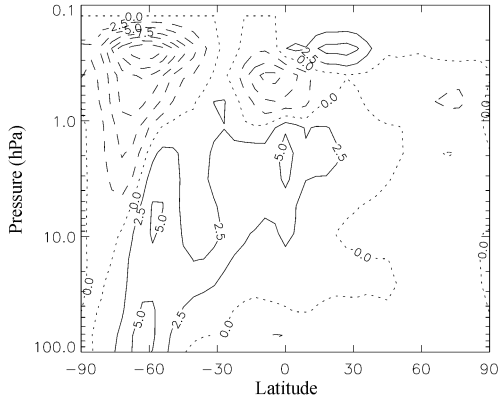

Figure 16. Weekly averaged zonal-mean zonal gravity-wave drag calculation with time-mean equations (contour intervals 2.5 m s^{-1}day^{-1}).

(*c*) *Comparison with a budget study*

In order to demonstrate the value of our variational technique for GWD estimation we compare it here with a much simpler and cheaper budget-based technique. The budget-based technique uses the time average over an assimilation window from t_0 to t_1 of length τ_w of the vorticity equation, on isentropic surfaces, for the true evolution (subscript o) and for a zero-drag control evolution (subscript c). Neglecting nonlinear terms and using $\sigma_o(t_0) = \sigma_c(t_0)$, $Q_o(t_0) = Q_c(t_0)$, we obtain

$$\mathbf{k} \cdot (\nabla \times \mathbf{X}) = \tau_w^{-1}\{\sigma_o(t_1)Q_o(t_1) - \sigma_c(t_1)Q_c(t_1)\}, \tag{4}$$

where $\mathbf{k}$ is the unit vertical vector and $\mathbf{X}$ is the GWD vector. Hence the rotational drag may be estimated by using Met Office analyses for $\sigma_o(t_0) = \sigma_c(t_0)$, $Q_o(t_0) = Q_c(t_0)$ and $\sigma_o(t_1)$, $Q_o(t_1)$ and a model integration for $\sigma_c(t_1)$, $Q_c(t_1)$. The results of this calculation are shown in Fig. 16.

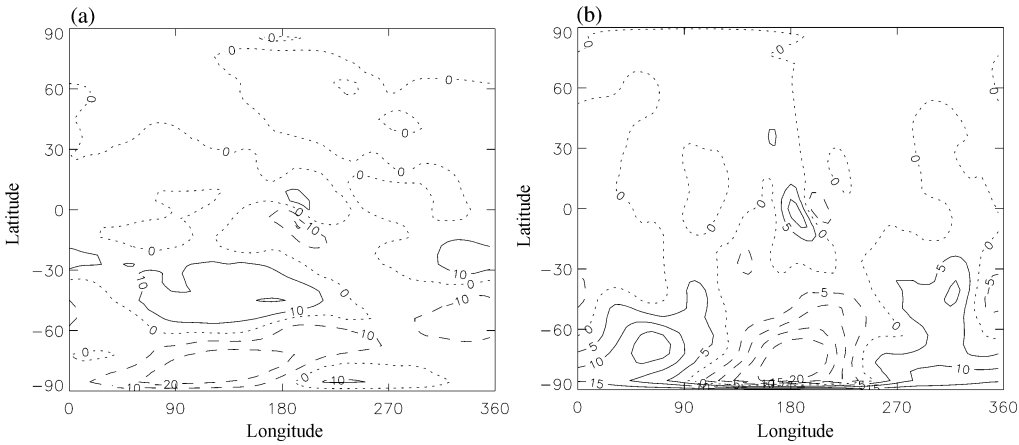

Figure 17. (a) Zonal and (b) meridional gravity-wave drag ($\mathrm{m\,s^{-1}day^{-1}}$) calculated using the budget equation (4) at 0.4 hPa on 1–2 July 2002.

Although there are some similarities between the results of the budget calculation and the ASDE estimate, there are also some notable differences. The maximum zonal-mean zonal GWD calculated with ASDE is $-50\ \mathrm{m\,s^{-1}}$ while it is $-15\ \mathrm{m\,s^{-1}}$ for the budget calculation. The accelerating region at 10 hPa is much stronger in the budget calculation. The decelerating region above it is quite concentrated in the ASDE estimate, while in the budget calculation the decelerating region extends down to $\approx$8 hPa. This emphasizes the advantage of the ASDE technique over the budget technique. The budget-technique estimate is simply proportional to the difference between the observed and control fields, while the ASDE technique, on the other hand, captures the non-locality of GWD effects. Note also, the budget calculation gives another region of negative zonal drag in the tropics, which is not present in the ASDE estimate, and in the summer hemisphere there is a very small acceleration centre for the budget calculation.

The longitude–latitude distribution of estimated drag (Fig. 17) shows that the estimate made by the budget technique has been significantly advected downstream compared with the estimate made by the ASDE technique (compare with Fig. 8). Again, this is because the budget estimate is proportional to the local wind differences between observations and control, whereas the ASDE technique is able to trace effects back to their causes.

4. Discussion

Our variational technique yields very plausible patterns and amplitudes for the estimated gravity-wave drag. Nevertheless, we must consider the possibility that the results may be contaminated by errors in the Met Office analyses or in the dynamical model used. For example, at this stage we cannot rule out the possibility that the small-scale transient drag features estimated for the middle stratosphere are in fact associated with model or data errors, though they also appear plausible on physical grounds. Similar small-scale transient features may be present at higher altitudes too, but there they would be swamped by a much larger amplitude large-scale smoothly evolving signal.

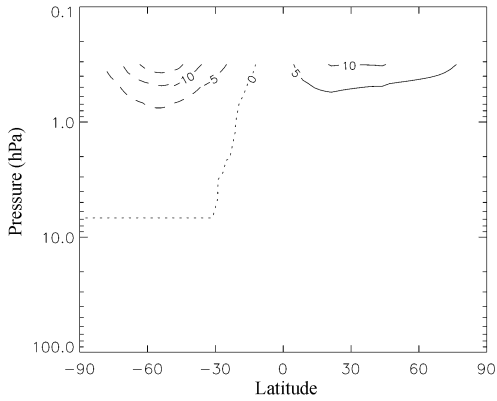

Figure 18. Weekly averaged zonal-mean linear drag used in the Met Office numerical model from 1–8 July 2002 (contour intervals 5 m $s^{-1}day^{-1}$).

One particular issue is that Met Office analyses are derived using a numerical model that includes a linear relaxation of the wind, or 'Rayleigh friction', with a height-dependent coefficient, in the upper stratosphere and mesosphere (Swinbank *et al.* 1998). If the Met Office analyses are dominated by their numerical model rather than observed data at those altitudes, then our estimated drag will be dominated by the Met Office model's linear drag, rather than 'real-world' gravity-wave drag.

Figure 18 shows the weekly averaged zonal-mean Rayleigh friction used in Met Office numerical model to perform the analyses, calculated using the drag coefficient given by Swinbank *et al.* (1998). There are qualitatively important differences between the linear drag (Fig. 18) and the GWD estimated with ASDE (Fig. 5), showing that the ASDE is not merely recovering the linear drag used in the Met Office model. First, the estimated drag peak at 0.4 hPa occurs poleward of the jet peak, not at the jet peak as linear drag does. This difference is related with characteristic wind errors in numerical models using linear drag in which the winter jet maximum is vertically aligned rather than sloping upwards and equatorwards; the fact that the ASDE system can track and maintain the observed jet structure implies that its estimated drag is doing something more realistic than mere linear relaxation. A second difference is that the estimated drag tends to accelerate the jets below the jet peaks, while linear drag does not do that. Finally, a third difference is the strength of the winter deceleration centre; the linear-drag peak is weaker, at only 15 m $s^{-1}day^{-1}$.

A related possibility, which we cannot rule out at this stage, could be that the large-scale smoothly evolving drag estimated at 0.4 hPa is excessively large-scale and smooth, and excessively well-correlated with the background wind, because of the influence of the linear drag in the Met Office analyses.

5. Conclusions

The experiments show that the assimilation technique gives a robust estimate of the rotational GWD. Using different observational data via the cost function leads to only small variations in the estimated rotational GWD. Also, increasing the number of iterations used in the minimization algorithm does not change significantly the estimated rotational GWD (not shown). The zonally averaged zonal drag has peak values of

50 m s^{-1}day^{-1} at 60°S and 0.4 hPa. In the summer hemisphere peak mean zonal values are 10 m s^{-1}day^{-1}. In both hemispheres the peak drag decelerates the zonal-mean jet. At lower heights, below about 1 hPa, the mean zonal GWD is in the same sense as the zonal-mean flow, but much weaker—of the order of 5 m s^{-1}day^{-1}. Such a pattern of weak acceleration below the jet core and strong deceleration above is predicted on the basis of filtering of gravity waves with a broad phase-velocity spectrum by their respective critical layers.

On the other hand, the divergent GWD does not appear to be properly estimated with our current set-up. This is partly because we do not use wind divergence information in our cost function since the analysed divergent wind is expected to be unreliable, but also partly because estimation of the divergent component is an inherently more difficult (less well-conditioned) problem than estimating the rotational component. Future work could concentrate on this limitation of the technique. One possible approach could be to include some constraints on the wind divergence to avoid this kind of problem, for example penalizing imbalance (while being careful not to penalize the true mean meridional circulation). Another might be to include observations at different times within a longer assimilation window instead of only at the final time. Fortunately, from the climatological point of view the rotational GWD is far more important than the divergent GWD, since permanent changes to the larger-scale flow are dominated by the rotational drag (e.g. Part I; Zhu and Holton 1987).

Our technique is able to estimate the longitudinal dependence of the GWD, and both the zonal and meridional components (of the rotational drag). In the lower mesosphere we found a large-scale smoothly evolving drag estimate with a clear planetary-wave pattern. This drag tends to damp the planetary-wave activity near the stratopause. This drag pattern may reflect the pattern of GWD in the real atmosphere, perhaps indicating selective filtering as suggested by Holton (1984). Alternatively, it may reflect model errors if planetary waves are too strong in the dynamical model, or data errors if planetary waves are too weak in the Met Office analyses. This topic merits further investigation, particularly since filtering of the gravity-wave spectrum at some altitude by a planetary-wave pattern in the background flow could lead to forcing of planetary waves by GWD at higher altitudes, as suggested by Holton (1984) and seen in some recent observational studies (Osprey and Lawrence 2001; Smith 2003).

At lower altitudes, the estimated drag field has smaller space and timescales, and is not obviously correlated with the local background flow. We found peak drag values of around 10 m s^{-1}day^{-1} at 9 hPa at high latitudes in the winter hemisphere. If these drag estimates in the middle stratosphere are believable, they suggest that the GWD pattern in the middle stratosphere reflects the intermittency and small spatial scales of tropospheric gravity-wave sources, while the GWD pattern in the mesosphere reflects the large spatial scales and smooth evolution of the background flow by which the wave spectrum has been filtered.

References

Alexander, M. J. and Rosenlof, K. H. 1996 Nonstationary gravity wave forcing of the stratospheric wind. *J. Geophys. Res.*, **101,** 23465–23474

Andrews, D. G., Holton, J. R. and Leovy, C. B. 1987 *Middle-atmosphere dynamics*. Academic Press

Blumen, W. 1972 Geostrophic adjustment. *Rev. Geophys.*, **10,** 485–528

Fleming, E. L., Chandra, S., Barnett, J. J. and Corney, M. 1986 Zonal mean temperature, pressure, zonal wind and geopotential height as functions of latitude. *Adv. Space Res.*, **10,** (12)11–(12)59

Gregory, A. R. 1999 'Numerical simulations of winter stratospheric dynamics'. PhD thesis, The University of Reading, UK

Hamilton, K. 1983 Diagnostic study of the momentum balance in the northern hemisphere winter stratosphere. *Mon. Weather Rev.*, **111,** 1434–1441

Haynes, P. H., Marks, C. J., McIntyre, M. E., Sheperd, T. G. and Shine, K. P. 1991 On the downward control of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.*, **48,** 651–678

Hines, C. O. 1997 Doppler spread parametrization of gravity-wave momentum deposition in the middle atmosphere. Part 1: Basic formulation. *J. Atmos. Solar Terrestr. Phys.*, **59,** 371–386

Holton, J. R. 1984 The generation of mesospheric planetary waves by zonally asymmetric gravity wave breaking. *J. Atmos. Sci.*, **41,** 3427–3430

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

Marks, C. J. 1989 Some features of the climatology of the middle atmosphere revealed by Nimbus 5 and 6. *J. Atmos. Sci.*, **46,** 2485–2508

Osprey, S. M. and Lawrence, B. N. 2001 A possible mechanism for *in situ* forcing of planetary waves in the summer extratropical mesosphere. *Geophys. Res. Let.*, **28,** 1183–1186

Pulido, M. and Thuburn, J. 2005 Gravity-wave drag estimation from global analyses using variational data assimilation principles. I: Theory and implementation. *Q. J. R. Meteorol. Soc.*, **131,** 1821–1840

Scaruzzo, C., Lamfri, M., Teitelbaum, H. and Lott, F. 1998 A study of the low-frequency inertio-gravity waves observed during Pyrex. *J. Geophys. Res.*, **103,** 1747–1758

Shine, K. 1987 The middle atmosphere in the absence of dynamical heat fluxes. *Q. J. R. Meteorol. Soc.*, **113,** 603–633

1989 Sources and sinks of zonal momentum in the middle atmosphere diagnosed using the diabatic circulation. *Q. J. R. Meteorol. Soc.*, **115,** 265–292

Smith, A. K. 2003 The origin of stationary planetary waves in the upper mesosphere. *J. Atmos. Sci.*, **60,** 3033–3041

Swinbank, R. and O'Neill, A. 1994 A stratosphere–troposphere data assimilation system. *Mon. Weather Rev.*, **122,** 686–702

Swinbank, R., Lahoz, W. A., O'Neill, A., Douglas, C. S., Heaps, A. and Podd, D. 1998 Middle atmosphere variability in the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.*, **124,** 1485–1525

Warner, C. D. and McIntyre, M. E. 1996 On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere, *J. Atmos. Sci.*, **53,** 3213–3235

Weglarz R. P. and Lin, Y. L. 1998 Nonlinear adjustment of a rotating homogeneous atmosphere to zonal momentum forcing. *Tellus*, **50,** 616–636

Zhu, X. and Holton, J. R. 1987 Mean fields induced by local gravity-wave forcing in the middle atmosphere. *J. Atmos. Sci.*, **44,** 620–630