Estimation of optimal gravity wave parameters for climate models using data assimilation

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

Quarterly Journal of the Royal Meteorological Society

Q. J. R. Meteorol. Soc. 138: 298–309, January 2012 B

RMetS

Royal Meteorological Society
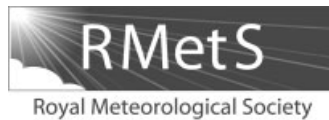
Estimation of optimal gravity wave parameters for climate models using data assimilation

M. Pulido,[a,b]* S. Polavarapu,[c†] T. G. Shepherd[a] and J. Thuburn[d]

[a]*Department of Physics, University of Toronto, Canada*
[b]*Department of Physics, FACENA, Universidad Nacional del Nordeste and CONICET, Corrientes, Argentina*
[c]*Environment Canada, Toronto, Ontario, Canada*
[d]*College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK*

*Correspondence to: M. Pulido, Department of Physics, FACENA, Universidad Nacional del Nordeste, Av. Libertad 5400, (3400) Corrientes, Argentina. E-mail: pulido@unne.edu.ar

[†]The contribution of this author to this article was prepared as part of her official duties as a Canadian government employee.

There is a current need to constrain the parameters of gravity wave drag (GWD) schemes in climate models using observational information instead of tuning them subjectively. In this work, an inverse technique is developed using data assimilation principles to estimate gravity wave parameters. Because most GWD schemes assume instantaneous vertical propagation of gravity waves within a column, observations in a single column can be used to formulate a one-dimensional assimilation problem to estimate the unknown parameters. We define a cost function that measures the differences between the unresolved drag inferred from observations (referred to here as the 'observed' GWD) and the GWD calculated with a parametrisation scheme. The geometry of the cost function presents some difficulties, including multiple minima and ill-conditioning because of the non-independence of the gravity wave parameters. To overcome these difficulties we propose a genetic algorithm to minimize the cost function, which provides a robust parameter estimation over a broad range of prescribed 'true' parameters. When real experiments using an independent estimate of the 'observed' GWD are performed, physically unrealistic values of the parameters can result due to the non-independence of the parameters. However, by constraining one of the parameters to lie within a physically realistic range, this degeneracy is broken and the other parameters are also found to lie within physically realistic ranges. This argues for the essential physical self-consistency of the gravity wave scheme. A much better fit to the observed GWD at high latitudes is obtained when the parameters are allowed to vary with latitude. However, a close fit can be obtained either in the upper or the lower part of the profiles, but not in both at the same time. This result is a consequence of assuming an isotropic launch spectrum. The changes of sign in the GWD found in the tropical lower stratosphere, which are associated with part of the quasi-biennial oscillation forcing, cannot be captured by the parametrisation with optimal parameters. Copyright © 2011 Royal Meteorological Society and Crown in the right of Canada.

1. Introduction

A major challenge in climate prediction is the proper representation of the impact of small-scale processes on larger scales (Hurrell *et al.*, 2009). In climate models, subgrid-scale processes are frequently parameterized in schemes that involve tunable parameters. Yet climate model simulations can be sensitive to the values specified for such parameters. Therefore, there is a need for an objective means of choosing optimal parameter values as well as obtaining a range of physically plausible values for ensemble simulations involving perturbed parameters (e.g. Stainforth *et al.*, 2007).

An example of a subgrid-scale process that requires parametrisation in weather and climate models that resolve the middle atmosphere is the forcing of the large-scale flow by unresolved gravity waves (referred to as 'gravity wave drag' or GWD). Models usually have at least two GWD parametrisations. One represents forcing due to gravity waves generated by orography (e.g. Palmer *et al.*, 1986; McFarlane, 1987), where the characteristics of the so-generated waves are partly known. A second parametrisation in the models (e.g. Warner and McIntyre, 1996; Hines, 1997) represents forcing from the remaining potential sources of gravity waves such as convection, shear instability and geostrophic adjustment, and here *a priori* knowledge is minimal.

Gravity waves help drive the Brewer–Dobson circulation (Randel *et al.*, 2008), and an important scientific question concerns the effect of climate change on this circulation. Orographic GWD has a strong and robust impact on changes in the Brewer–Dobson circulation due to climate change (Li *et al.*, 2008; McLandress and Shepherd, 2009), although there are large quantitative differences between individual models that at least partly arise from differences in implementation of orographic GWD (Butchart *et al.*, 2006, 2010). In addition, Sigmond and Scinocca (2010) found that the response of Arctic surface pressure to doubled CO_2 differs enormously between two versions of their model which differ only in the parameter setting in their orographic GWD scheme, because of the impact of orographic GWD on the climatological winds and thus on the response of planetary wave drag to doubled CO_2. Non-orographic GWD impacts the seasonal cycle of polar temperatures and ozone (Alexander *et al.*, 2010).

Though climate simulations may be sensitive to GWD, the tuning of GWD parameters can involve guesswork. Various parameter settings are selected and the resulting simulations are then evaluated in terms of their realism. However, values that lead to reduced biases in surface pressure may not give the best wind speeds in the lower stratosphere (Sigmond and Scinocca 2010). Thus a subjective choice is made. Here we consider an objective method of obtaining parameter values using data assimilation. The use of data assimilation to identify biases in climate model forecasts on short (weather forecasting) time-scales has been advocated by Phillips *et al.* (2004), Palmer *et al.* (2006), and Rodwell and Palmer (2007) as part of a 'seamless' approach to weather and climate prediction. Brunet *et al.* (2010) further acknowledge the value of tuning parameters in a climate model using data assimilation in order to reduce climate model errors. The use of data assimilation and forecasts on short time-scales to identify climate model biases allows different sources of model errors to be disentangled and allows model errors to be traced back to the specific time and location where they are produced. In this way, the estimation of optimal parameters using data assimilation may be useful not only for short-term forecasts but also for climate modelling.

Most GWD schemes are based on a representation of the gravity wave momentum flux. Currently, observations of gravity wave momentum flux are very limited as measurements only detect a part of the gravity wave spectrum (Preusse *et al.*, 2008), although considerable progress is expected with the launch of new satellites and the development of data processing techniques (Alexander *et al.*, 2010). While the gravity wave spectrum is not well observed, the large-scale flow which is derived in part from the forcing due to dissipating gravity waves is easier to observe. Thus data assimilation can be used to infer the GWD from observed large-scale fields, if the large-scale GWD response is predictable. Evidence of this is found in Karlsson *et al.* (2009) for a climate model and Ren *et al.* (2008) for a data assimilation system. Pulido and Thuburn (2005, 2006, 2008) proposed an inverse technique to estimate the missing GWD in climate models from the large-scale flow based on variational data assimilation principles. In their work, the GWD is estimated using the mismatch between the climate model and the observations. The main assumption is that the estimated forcing can be attributed mainly to under-resolved gravity waves rather than to initial-condition errors or other sources of model error. This is regarded as a reasonable assumption in the upper part of the middle atmosphere once the errors are averaged in space and time.

Pulido and Thuburn (2008) focussed on estimating the missing forcing due to GWD, which we refer to as the 'observed' GWD. In this work the aim is to go further and to determine an optimal set of parameters in a GWD scheme such that the drag given by the parametrisation can account for the observed GWD. A cost function which measures the departure of the observed GWD from the drag predicted with a given set of parameters is minimized to obtain an optimal set of parameter values. Because all current GWD schemes assume that gravity waves propagate straight up and instantaneously, it is entirely consistent to formulate the parameter estimation problem in one dimension, namely a single column. However, even the one-dimensional problem can be challenging. If the response to parameter changes involves switches (discontinuities) or is highly nonlinear, a meaningful solution is not guaranteed since common assimilation techniques (e.g. variational assimilation or the ensemble Kalman Filter) are designed for smooth and weakly nonlinear problems. While the results of parameter estimation will always be applicable only to the scheme considered, the process of applying data assimilation to a new problem (as is done here) may lead to insights into the formulation or assumptions of GWD parametrisation schemes. Although we consider only one GWD scheme here, McLandress and Scinocca (2005) have shown that different GWD schemes behave similarly when constrained by the same source spectrum. Thus, our results are expected to be relevant to other GWD schemes.

The plan of the paper is as follows. Technical aspects of the GWD scheme and the inverse technique are provided in section 2. In section 3, the optimization techniques are assessed using twin experiments in which the 'observed' GWD is produced using the parametrisation with a set of known parameters. This allows an assessment of the smoothness and nonlinearity of the problem. In section 4,

the GWD field obtained by Pulido and Thuburn (2008) using observations is used to estimate the optimal parameters. A summary of the results and a discussion of their generality is given in section 5.

2. Technical details

2.1. Gravity wave drag scheme

In this study we use the Scinocca (2003) GWD scheme. This is the operational scheme in the Canadian Middle Atmosphere Model (Scinocca *et al.*, 2008) and is also operational in the ECMWF model since September 2009 (Orr *et al.*, 2010). Here we summarize the relevant characteristics of the scheme, emphasising the free parameters and their role in the scheme. In order to reduce the number of free parameters and to concentrate on the most physically relevant ones, we have considered the simplest version of the gravity wave scheme which represents the main physical processes of momentum deposition. In this simple version the reflection of waves is assumed to be negligible (Scinocca, 2002).

Scinocca (2003) suggests that an efficient non-orographic gravity wave scheme is obtained if the spectrum of the Eliassen–Palm (EP) flux is expressed as a function of the horizontal phase speed c in each azimuthal direction, ϕ. This produces an important conceptual and numerical simplification since the phase speed spectrum is not altered by horizontal background wind refraction under conservative propagation. In contrast, the vertical wavenumber spectrum is affected by changes in horizontal background wind. The spectrum is launched at some height z_l, which is usually assumed to be between the surface and the tropopause, i.e. within the region where the actual sources are expected. The launch EP flux spectral density of the Desaubies form in terms of the phase speed is

$$E_l(\widehat{c}, \phi) = E(\widehat{c}, \phi, z_l) = \frac{E_*}{\widetilde{E}} \frac{\widehat{c}}{1 + \left(\frac{\widehat{c}}{c_*}\right)^4}, \quad (1)$$

where $\widehat{c}$ is the horizontal phase speed with respect to the horizontal wind at the launch height ($\widehat{c} = c - u(z_l, \phi)$; this being the intrinsic phase speed at launch height, $\widehat{c}$ is constant with height), $c_* \equiv N_l/m_*$ is the characteristic phase speed (N_l is the Brunt–Väisälä frequency at the launch height and m_* is the characteristic vertical wavenumber), and $\widetilde{E}$ is given by

$$\widetilde{E} = \int_0^\infty \frac{\widehat{c}}{1 + \left(\frac{\widehat{c}}{c_*}\right)^4} \, \mathrm{d}\widehat{c} = \frac{\pi}{4} c_*^2. \quad (2)$$

Note from (1) and (2) that E_* is then the total momentum flux, i.e.

$$E_* \equiv \int E(\widehat{c}, \phi, z_l) \, \mathrm{d}\widehat{c}$$

(for any azimuthal direction). The momentum flux is assumed to be isotropic in intrinsic phase speed at the launch height.

For the waves that propagate conservatively, the EP flux spectral density is constant, i.e.

$$E(\widehat{c}, \phi, z) = E(\widehat{c}, \phi, z_l). \quad (3)$$

Dissipation of the waves is activated when a component of the spectrum exceeds a saturation threshold given by

$$E_s(\widehat{c}, \phi, z) = S_* \frac{E_*}{\widetilde{E}} \frac{\rho(z)}{\rho(z_l)} \frac{N(z_l)}{N(z)} \frac{\{\widehat{c} - \widehat{u}(z, \phi)\}^{3/2}}{\widehat{c}^{1/2}}, \quad (4)$$

where ρ is density, S_* is the saturation amplitude, and $\widehat{u}(z, \phi)$ is the horizontal wind in the azimuthal direction ϕ relative to the horizontal wind at the launch height, viz. $\widehat{u}(z, \phi) = u(z, \phi) - u(z_l, \phi)$. Note that $E_s(\widehat{c}, \phi, z)$ diminishes with height because of the decreasing density. On the other hand, the undissipated EP flux is constant with height, so waves that are propagating upwards will eventually become saturated at some height. If the saturation amplitude is $S_* = 1$, then the amplitude of the launch spectrum for $\widehat{c} = c_*$, $E_l(c_*, \phi)$, is at the threshold value $E_s(c_*, \phi, z_l)$. For $S_* > 1$, the launch spectrum amplitude for $\widehat{c} = c_*$ is smaller than the saturation threshold. On the other hand, $S_* < 1$ would indicate an oversaturated spectrum; in this case the parametrisation would produce a physically unrealistic drag since the EP flux that exceeds the threshold is deposited in the first layer. Note that the parameter S_* was not present (i.e. a fixed $S_* = 1$ saturation amplitude was taken) in the original Scinocca (2003) parametrisation; it was introduced in McLandress and Scinocca (2005).

The other mechanism that affects the propagation of the waves is critical-level filtering by the background flow; the components of the spectrum that are eliminated by filtering are the ones with slow intrinsic phase speed.

The zonal component of the total momentum flux is given by

$$E_x(z) = \int_0^{2\pi} \int_0^\infty E(\widehat{c}, \phi, z) \, \cos\phi \, \mathrm{d}\widehat{c} \, \mathrm{d}\phi, \quad (5)$$

and the meridional component of the total momentum flux by

$$E_y(z) = \int_0^{2\pi} \int_0^\infty E(\widehat{c}, \phi, z) \, \sin\phi \, \mathrm{d}\widehat{c} \, \mathrm{d}\phi. \quad (6)$$

The GWD, i.e. the EP flux divergence divided by density, is

$$(X, Y) = (\rho^{-1}\partial_z E_x, \, \rho^{-1}\partial_z E_y). \quad (7)$$

To understand the results of the 1D assimilation problem, it is necessary to consider how the drag is affected by changes in the tunable parameters of the scheme. For this scheme, the tunable parameters are: the launch total momentum flux E_*, the characteristic phase speed c_*, saturation amplitude S_*, and launch height z_l. Although Manzini and McFarlane (1998) find important sensitivity to the launch height, in this work we consider a fixed launch height. That is, z_l is not considered in the control space only because it is a non-smooth parameter and it is not trivial to vary this parameter in the Scinocca (2003) scheme. Although the computational code could also work with negative parameters, the physical interpretation of the four free parameters requires that they be positive.

Since the tunable parameters are independent of height, they cannot directly affect the shape of the drag profile. They only do so indirectly. The EP flux is a linear function of E_* according to (1) so that the drag is also a linear function of

E_* according to (7). The vertical gradient of EP flux depends on S_* only when saturation occurs. In this case, the amount deposited will depend linearly on S_* (see (4)) for a given wave. Since all three tunable parameters can control the vertical gradient of EP flux, changes in one parameter can be offset by changes in one or more of the other parameters. Thus, there is no unique set of parameters that will provide a given drag. This will be manifested in practice as multiple minima and as long valleys in the cost function of the 1D assimilation problem.

2.2. Inverse technique

GWD resulting from the parametrisation is used as a forcing term in the momentum equations of the climate model. A general inverse technique in which a cost function measuring the differences between the model state and observations could be used. Assuming the initial conditions are known, the cost function can be considered a function only of the gravity wave parameters. Therefore the minimum of the cost function determines the optimal parameters –namely those values that produce the closest match of the model to the observations.

Because the gravity wave sources in the GWD scheme are considered constant and do not depend on the background flow, the estimation problem simplifies greatly. The problem can then be split into two stages. First, the GWD forcing term in the momentum equations that drives the model towards the observations is estimated. Second, the gravity wave parameters that produce the closest match to the observed GWD found in the first stage are identified. Because the resolved wave forcing is already represented by the model, we can attribute the missing drag to unresolved waves. This largely removes the feedback between the resolved wave forcing and the subgrid-scale wave forcing that plagues other methods of GWD parameter estimation.

In the first stage of the inverse technique, the GWD estimation is conducted using ASDE (Assimilation System for Drag Estimation) as in Pulido and Thuburn (2008). The technique uses 4D variational assimilation principles. Analyses from the Met Office are taken as the observations. The height range of the GWD, wind and temperature fields is from a potential temperature of 414 K (100 hPa) up to 2400 K (0.3 hPa). Note that ASDE does not employ any GWD parametrisation; it only estimates the missing forcing, i.e. GWD.

The GWD parametrisation assumes that the wave propagation is in a vertical column. This implies that a 1D inverse problem can be formulated for the second stage of the inverse technique. The cost function of the second stage is defined as

$$J = \frac{1}{2} \sum_n (\mathbf{y}^o - H[\mathbf{x}])^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y}^o - H[\mathbf{x}]), \qquad (8)$$

where $\mathbf{x}$ is a vector with the set of free parameters, H is the GWD scheme and $\mathbf{y}^o$ is a vertical profile of the estimated GWD given by ASDE. This vertical profile is taken from the drag field which was determined in the first stage of the inverse technique. In what follows, in order to use the standard terminology in data assimilation we refer to this estimated GWD as the 'observed' GWD. The matrix $\mathbf{R}$ should be positive definite, but otherwise may be chosen by the user to give different weights to different observations

or to improve the convergence of the minimization. Here $\mathbf{R}$ is taken to be the identity matrix. We return to this point in a subsequent section.

The control space is formed by $\mathbf{x} = (E_*, \lambda_*, S_*)$, where $\lambda_* = 2\pi c_*/N_l$ is the characteristic vertical wavelength. We decided to use λ_* instead of c_* because λ_* is the parameter usually determined in gravity wave observations (e.g. Allen and Vincent, 1995).*

(i) Variational data assimilation. One of the techniques that we use to estimate the optimal parameters is based on variational data assimilation principles. The idea is to use the same framework as the one used for the GWD estimation (Pulido and Thuburn, 2005) in which the optimization module uses conjugate gradients with a secant method to find the root of the cost function gradient in each search direction. This optimization requires the gradient of the cost function. The adjoint of the GWD scheme was built with an automatic adjoint generator: the Tangent and Adjoint Model Compiler (Giering and Kaminski, 1997).

(ii) Genetic algorithm. Given the problems found with the variational assimilation technique, which are described in section 3, a genetic algorithm was also implemented (Golberg, 1989). This technique is particularly suitable for constrained minimization problems in which the model behaves nonlinearly with respect to the parameters. Convergence towards the global minimum is possible even if the cost function contains multiple local minima. The genetic algorithm does not require the evaluation of the derivative of the cost function with respect to the model parameters. A drawback of the technique is that it needs a large number of model evaluations; this is particularly so if the structure of the cost function is highly complex or for a large-dimension control space.

The genetic algorithm implemented in this study (Pikaia) was developed by Charbonneau (2002). This is a basic algorithm which applies the essential ideas of natural selection. The algorithm behaves robustly in our application.

3. Twin experiments

There are currently no global estimates of gravity wave parameters directly from observations (Alexander *et al.*, 2010). The inverse technique is therefore evaluated by means of twin experiments in which the observed GWD is calculated with the Scinocca scheme with a prescribed set of true parameters using standard winter and summer midlatitude wind and temperature profiles. Then the optimization technique, e.g. the variational technique, is used to estimate the best parameters. These parameters give the GWD that optimally fits the synthetic observed GWD. The convergence of the technique towards the known true parameters can be readily evaluated in these experiments.

The parameters are normalized to the standard values used in the Scinocca scheme: $\widehat{E}_* = E_*/E_R$, $\widehat{\lambda}_* = \lambda_*/\lambda_R$ and $\widehat{S}_* = S_*/S_R$ where $E_R = 2.5\sqrt{2} \times 10^{-4}$ Pa, $\lambda_R = 2$ km, and $S_R = 1$. The launch height z_l has been fixed to a potential temperature of 440 K (approximately 100 hPa).

In a first set of experiments, we take standard summer midlatitude wind and temperature profiles and prescribe as true parameters the standard values, i.e. $\widehat{E}_*^{\mathrm{T}} = 1$, $\widehat{\lambda}_*^{\mathrm{T}} = 1$ and

---

*The use of $m_* = 2\pi\lambda_*^{-1}$ instead of λ_* as free parameter was also evaluated, but resulted in a slower convergence rate.
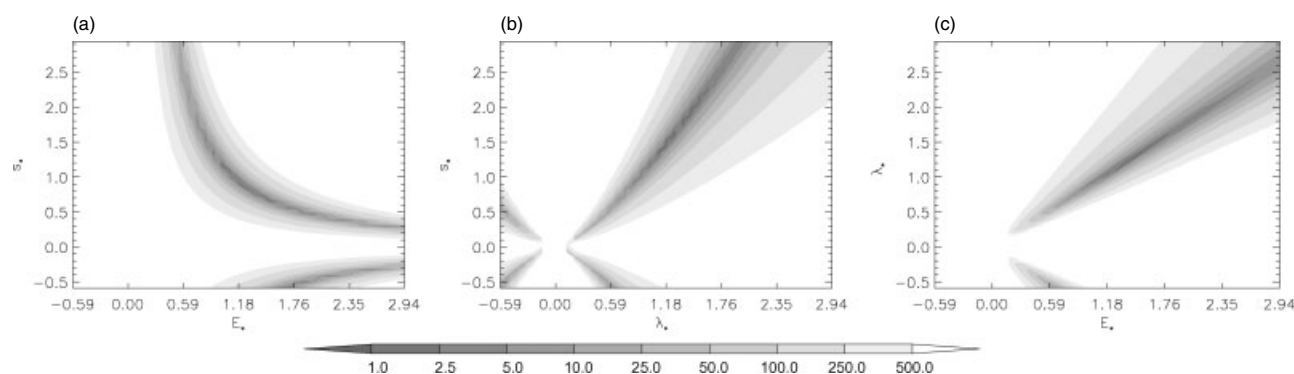
Figure 1. Cross-sections of the cost function without scaling at (a) $\widehat{\lambda}_* = 1$, (b) $\widehat{E}_* = 1$, and (c) $\widehat{S}_* = 1$.

$\widehat{S}_*^T = 1$. Figure 1 shows the geometry of the cost function in three cross-sections: $\lambda_* = \lambda_*^T$, $E_* = E_*^T$ and $S_* = S_*^T$. The cost function is rather convex near the true parameters. However, there are multiple minima related to unphysical (negative) parameters. These minima may cause problems in the convergence with variational optimization schemes. Note also that elongated valleys are present in the cost function which make the parameter estimation difficult in some search directions.

Four optimization techniques were implemented and evaluated: conjugate gradients (CGs), quasi-Newton with bounded domain (BQN), the genetic algorithm (GA), and a hybrid scheme using the genetic algorithm and conjugate gradients (Mixed). The convergence of the conjugate gradients method depends on the initial guess parameters. As shown in Table I, using an initial guess of $\widehat{E}_* = 2.08$, $\widehat{\lambda}_* = 2.08$, $\widehat{S}_* = 0.11$, the variational technique converges towards a negative S_* value. For the BQN technique we chose the parameter range $[0.1, 2.1]$. This optimization algorithm also fails to converge towards the true parameters using the same initial guess (Table I). The cost function contains local minima at the low S_* boundary, i.e. $\widehat{S}_* = 0.1$, and the chosen initial guess parameters, $\widehat{E}_* = 2.08$, $\widehat{\lambda}_* = 2.08$, $\widehat{S}_* = 0.11$, are close to these local minima. Note that $\widehat{S}_* < 1.0$ values may be unphysical, since they indicate oversaturation. The algorithm also does not converge when the lower limit is changed to $S_* = 0.9$.

The GA is particularly suited for minimization in a constrained domain. As before, we choose the parameter range $[0.1, 2.1]$. Table I shows that the GA gives a good coarse estimation of the parameters. Moreover, this estimation is independent of the initial guess parameters. The GA was also evaluated changing the prescribed true parameters within the parameter range $[0.1, 2.1]$. In all cases the algorithm converged toward the true parameters.

Table I also shows the result of a newly proposed technique that first uses the genetic algorithm and then, once the estimation is close to the global minimum, the CG method is used. This hybrid technique presents very good convergence taking advantage of both techniques. The GA is able to select the deepest minimum of the cost function despite its complex structure, then the CG algorithm is able to exploit the near-quadratic structure of the cost function near the minimum to achieve quick and accurate convergence. If performance is an issue in future applications, then a small number of generations, say 50, could be used and then the CGs could be applied to improve the 'coarse' estimation of the GA. For the case-studies shown, we use 200 generations and 100 individuals as the default configuration. In general,

Table I. Estimation of the parameters for different optimization techniques.

Technique	Parameter		
	$\widehat{E}_*$	$\widehat{\lambda}_*$	$\widehat{S}_*$
True	1	1	1
Initial	2.08	2.08	0.11
Est CG	0.70	46.16	−8.93
Est BQN	1.44	2.1	0.1
Est GA	1.018	1.001	0.975
Mixed	1.00001	0.99999	0.99999

the GA was stopped because the generation number reached 200 instead of reaching the precision criterion (in particular for the results in Table I).

In summary, the 1D assimilation problem for estimating GWD parameters is highly nonlinear, resulting in a poorly conditioned problem, viz. long elongated valleys are found in the cost function making the determination of optimal parameter values difficult and the associated errors of the estimation large. The problem also contains multiple minima. Nevertheless, a solution can be obtained with a new technique which combines variational techniques with a GA. With these twin experiments, we were able to assess the potential difficulty of the mathematical problem by assuming the parametrisation scheme is perfect. In reality, the assumptions made by non-orographic GWD schemes are not entirely justifiable so in the next section we relax this assumption. The parameter estimation problem in this more realistic context may then lead to insights into the deficiencies and assumptions of the GWD scheme.

4. Results in real cases

In this section the observed GWD is taken from GWD estimations constrained by observations using ASDE (Pulido and Thuburn, 2005). This field represents the drag that the GWD scheme should approximate. The 3D GWD fields are averaged zonally, and the resulting GWD fields are also monthly averages for July 2002. Zonal mean zonal wind and temperature fields are used as inputs to the GWD scheme to determine the characteristics of wave propagation. The GWD estimations used in this work are similar to the ones shown in Pulido and Thuburn (2008). However, there were slight changes in ASDE which did not affect appreciably the
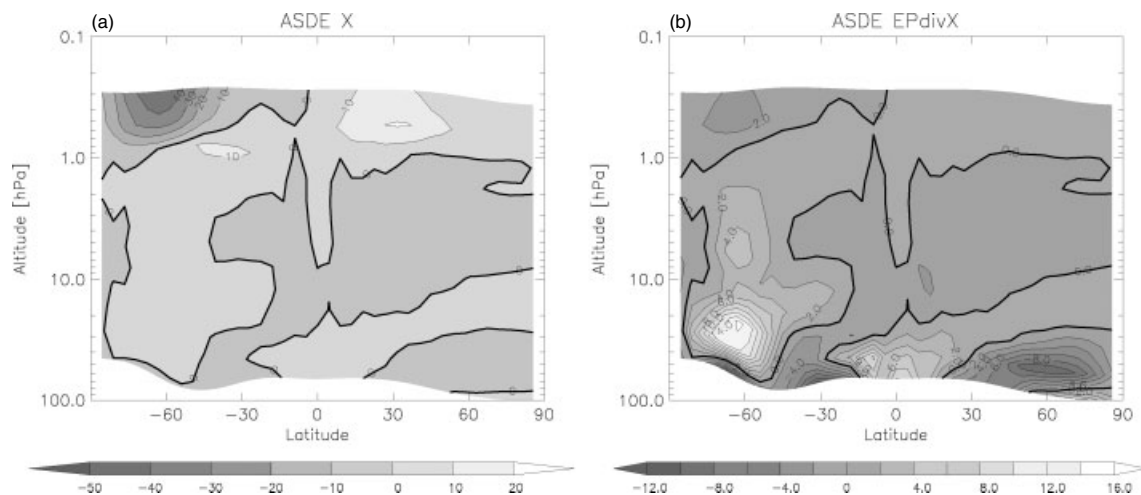
Figure 2. (a) Observed zonal mean zonal gravity wave drag (m s^{-1}day^{-1}) estimated by ASDE for July 2002, and (b) the corresponding zonal EP flux divergence (10^7 Pa m^{-1}). The zero drag contours are shown in bold.

estimated GWD field for the configuration used. Figure 2 shows the zonal GWD used in this study as the 'observation' and the corresponding zonal EP flux.

In this realistic experiment, there is no guarantee that a set of parameters achieves a zero cost function. Parametrisations are a simplification of the complex nonlinear wave–mean flow interactions, so that they might not represent exactly the drag field. Furthermore, the observed drag may be contaminated with model errors at some heights which the data assimilation technique will identify as missing forcing and therefore GWD. Experiments to evaluate the shape of the cost function in realistic cases were conducted. In all the cases, the cost function was smooth with a global minimum in the physical parameter range (not shown).

For the twin experiments, changes in the $\mathbf{R}^{-1}$ matrix can only improve/degrade the convergence rate since the observation is exactly reproduced by the model (the parametrisation). On the other hand, in this realistic experiment changes in $\mathbf{R}^{-1}$ may give a different set of optimal parameters since the parametrisation cannot match exactly the whole observed profile. There are two natural options for the observed variable $\mathbf{y}^o$: GWD, i.e. $\mathbf{X}$ from (7), or EP flux divergence, i.e. $\partial_z E_x = \rho X$. The latter could also be thought of as an observed variable of $\mathbf{y}^o = \mathbf{X}$ and a $\mathbf{R}^{-1}$ matrix whose diagonal elements are $\rho(z_i)^2$ (see (8)).

Unconstrained positive parameter estimations give extremely large parameter values which produce a good fit in the upper part of the drag profile, however breaking and saturation are completely absent at low levels because of the combination of large $\widehat{S}_*$ and $\widehat{\lambda}_*$ values. To avoid this artefact, we conducted constrained optimization experiments, in which $\widehat{S}_*$ was constrained to the range 0.9 to 10, and $\widehat{\lambda}_*$ to the range 0.3 to 10.

Optimal parameters are estimated independently for each latitude, using the zonal mean profiles of GWD, zonal wind and temperature. Figure 3 shows the estimated parameters normalized with the standard values, E_*/E_R, λ_*/λ_R and S_*/S_R, as a function of latitude. The estimated parameters depend on the choice of observed variable. The reason for this will be explained in detail shortly. A brief explanation is that neither choice can yield a good fit to both the lower and upper part of the observed drag profile. Note that both choices yield parameters that vary considerably with latitude. We shall show subsequently that allowing the

parameters to vary with latitude is important for obtaining a good estimation of the observed drag.

For the $\mathbf{y}^o = \mathbf{X}$ case, the total EP flux $\widehat{E}_*$ shows large values at high latitudes and small values in the Tropics (Figure 3). The maximum total EP flux in the winter hemisphere at 100 hPa is 4.6×10^{-3} Pa at 60°S, while it is 2.5×10^{-3} Pa in the summer hemisphere at 30°N. The characteristic vertical wavelength, λ_* is about 2 km (4 km) in mid (high) latitudes, while a longer characteristic wavelength is found in the Tropics. The optimal parameters shown in Figure 3 are estimated for monthly mean zonally averaged profiles in July 2002. However the parameters may depend on the seasonal cycle and may vary interanually, so a multiyear analysis should be performed as future work for a comprehensive study of parameter statistics.

Figure 4(a) shows the GWD that results from the optimization using the hybrid algorithm, i.e. the GWD from the parametrisation with the optimal parameters. This case uses GWD directly as observed variable $\mathbf{y}^o = \mathbf{X}$. The parametrisation does a remarkably good job in reproducing the zonal mean drag. The winter and summer deceleration centres are well located and their amplitudes are rather close to the observed ones (Figure 2(a)). At lower heights, the change of sign in the drag is also well reproduced. At low latitudes, the positive zonal forcing that is observed above 10 hPa is also captured by the parametrisation. On the other hand, the positive zonal forcing at low heights in the Tropics and the summer hemisphere is not reproduced by the parametrisation with the optimal parameters.

The EP flux divergence resulting from the parametrisation with optimal parameters is shown in Figure 4(b). This is again for the case with GWD as observed variable $\mathbf{y}^o = \mathbf{X}$. The observed EP flux divergence presents strong positive forcing in the winter lower stratosphere and negative forcing in the summer hemisphere (Figure 2(b)). The positive forcing in the winter lower stratosphere is underestimated by a factor of two, while the centres of enhanced forcing in the Tropics and the summer lower stratosphere are not reproduced by the GWD parametrisation. (Note that the shading scales in Figures 4(b) and 2(b) are different.) The parametrisation does a good job fitting the observed GWD and EP flux divergence in the upper levels but differs in the lower levels where enhanced centres of forcing are found in the observed EP flux divergence.
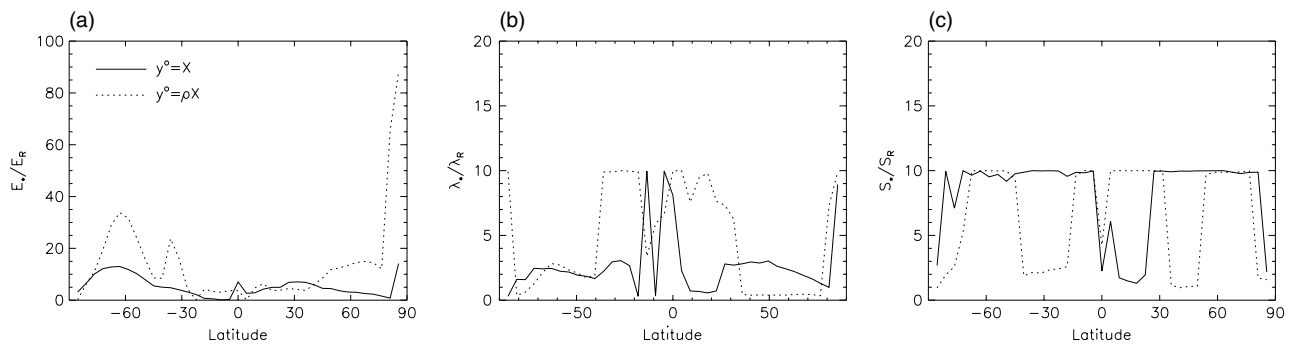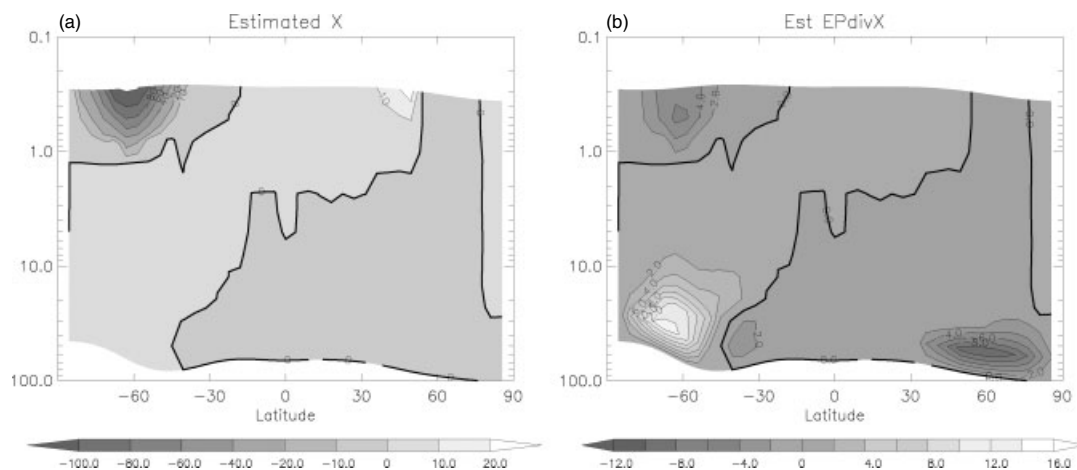
Figure 3. Estimated parameters normalized with the reference values: (a) E_*/E_R, (b) λ_*/λ_R, and (c) S_*/S_R, using gravity wave drag ($\mathbf{y}^o = \mathbf{X}$, solid line) and EP flux divergence ($\mathbf{y}^o = \partial_z E_x$, dotted line) as the observed variable.
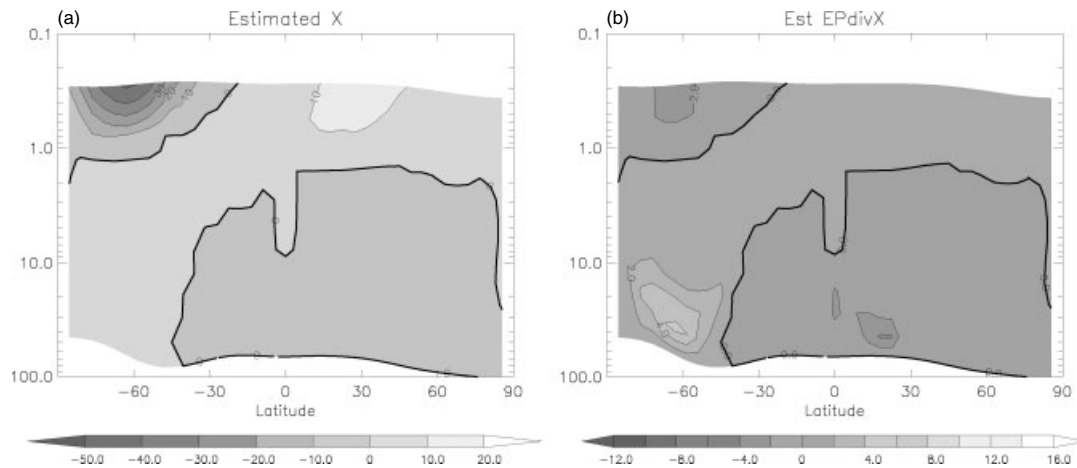
Figure 4. (a) Zonal mean zonal drag given by the GWD scheme with the optimal parameters using drag as the observed variable and (b) the corresponding zonal EP flux divergence.


Figure 5. (a) Zonal mean zonal drag given by the GWD scheme with the optimal parameters using EP flux divergence as the observed variable, and (b) the corresponding zonal EP flux divergence.

These results suggest that a larger weight at the lower levels of the observed variable in the cost function could resolve the differences found between the observed and estimated EP flux divergence (cf. Figures 2(b) and 4(b)). Thus we repeated the experiment with $\mathbf{y}^o = \rho\mathbf{X}$ (EP flux divergence) as the observed variable. The EP flux divergence that results from this experiment is shown in Figure 5(b). The positive forcing centre at the lower stratosphere of the winter hemisphere is well reproduced in this experiment. Furthermore, the negative forcing centre in the summer hemisphere is also nicely reproduced. On the other hand, in

the Tropics the parametrisation cannot capture the positive centre at low altitudes even in this experiment where the lower stratosphere is emphasised in the cost function with the ρ weight. To examine the estimation at higher altitudes, GWD is shown in Figure 5(a). The high deceleration centre in the winter hemisphere and the acceleration centre in the summer hemisphere are both captured, although the deceleration centre is stronger for the $\mathbf{y}^o = \rho\mathbf{X}$ experiment reaching over 100 m s^{-1}day^{-1}.

The GWD calculated with the parametrisation using the standard set of parameters for wind and temperature profiles
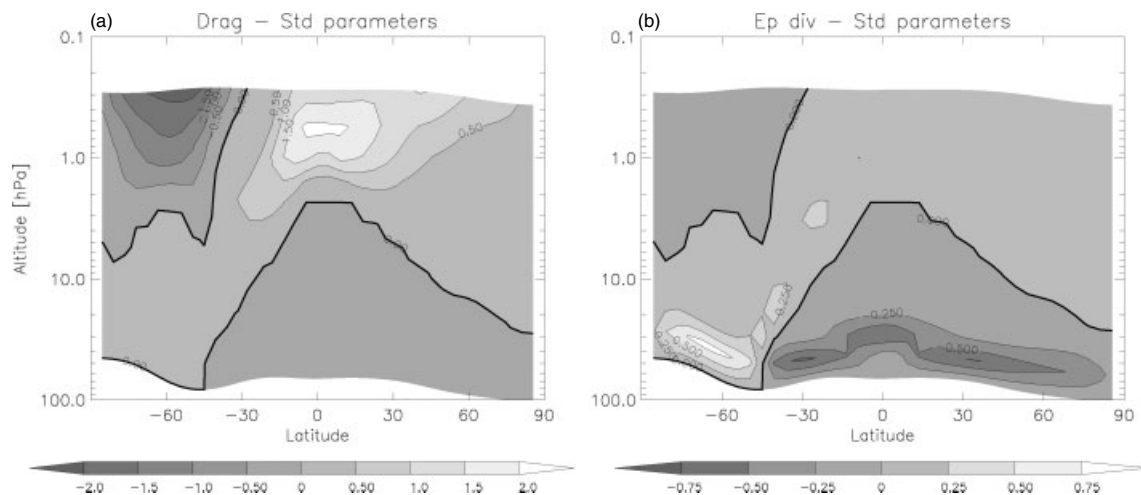
Figure 6. (a) Zonal gravity wave drag given by the parametrisation with the standard set of parameters using wind and temperature for July 2002 and (b) the corresponding zonal EP flux divergence. Note that the contours in both panels are different from those used in Figure 2.

in July 2002 is shown in Figure 6. We should not expect that constant parameters in a gravity wave parametrisation can exactly reproduce the spatial distribution of gravity wave sources and therefore of the actual GWD. However, we do expect that a fairly realistic latitudinal distribution of the drag given by the parametrisation be found if the filtering mechanism is the dominant process. The parameter estimation can give us an objective measure of the impact of latitudinal variation of the sources compared to the dependencies produced by the filtering mechanism. Quantitatively, the magnitude of the forcing found with the standard set of parameters (Figure 6) is over an order of magnitude weaker than the observed GWD. Comparing qualitative features, two drawbacks of the standard set are evident when this GWD (Figure 6) is compared to the observed one (Figure 2): the summer deceleration centre is not well represented, and a new forcing centre appears in the Tropics which is not present in the observed drag. These deficiencies show the benefits of objectively estimating parameter values. Also, the deceleration centres in winter and summer extend to very low heights. In contrast, the use of optimal λ_* and S_* parameters gives a realistic drag distribution with a change of sign at 1 hPa (Figure 4) as found in the observed GWD (Figure 2).

A question arises as to why the inverse technique can capture the higher part of the drag profile $\mathbf{y}^o = \mathbf{X}$ or the lower part of the drag profile $\mathbf{y}^o = \rho\mathbf{X}$, but cannot capture the whole drag profile in a single experiment. This is related to the physical mechanisms that produce the momentum deposition. For a standard winter jet profile, the filtering mechanism is the dominant mechanism between the launch level and the level of maximum eastward zonal wind. In this region, eastward waves find their critical levels, producing an eastward forcing. At higher altitudes the most important mechanism is wave saturation of the remaining westward waves which produces a westward forcing. With the dominant weight in the inverse technique at high altitudes, $\mathbf{y}^o = \mathbf{X}$, the estimation is dominated by the fit to the westward drag, by setting the optimal characteristics of the westward waves at launch height. Then, since the spectrum is assumed to be isotropic, the eastward waves will have the same characteristics (parameters) and therefore a wrong lower drag profile will likely result. On the other hand, if the dominant weight is in the lower part of the

drag profile, $\mathbf{y}^o = \rho\mathbf{X}$, the estimation will give optimal parameters appropriate for the eastward waves (eastward drag) and because of isotropy, the westward drag will be the result of these 'eastward optimal' parameters.

To illustrate this point, Figure 7 shows zonal wind and drag profiles at 63°S. The $\mathbf{y}^o = \mathbf{X}$ case presents a close fit at high altitudes to the observed drag profile. The saturation mechanism of westward propagating waves is fitted so that it closely matches the deceleration forcing found above the jet. The eastward wave spectrum, as a result of isotropy, gives the characteristic eastward forcing at and below the jet maximum. However it does not capture the lower forcing centre (Figure 7(d)) which is also clearly visible in Figure 2(b). On the other hand, the $\mathbf{y}^o = \partial_z E_x$ case closely fits this lower part of the drag profile (up to 10 hPa; Figure 7(d)) where the eastward wave spectrum is strongly filtered due to the strong eastward wind shear. In this case the estimation of parameters is dominated by these lower-level features and therefore the westward forcing produced at upper levels due to the saturation mechanism is the result of the remaining waves and is not accurate (Figure 7(b)).

In the summer jet, one would in principle expect a similar picture to the winter jet: namely a dominant effect of the saturation mechanism for the $\mathbf{y}^o = \mathbf{X}$ case, and of the filtering mechanism for the $\mathbf{y}^o = \partial_z E_x$ case. The optimal parameters for the $\mathbf{y}^o = \mathbf{X}$ case are similar to the winter jet, $\widehat{\lambda_*} \approx 2$ and $\widehat{S_*} \approx 10$ (Figure 3), and these optimal parameters give a reasonable fit to the upper part of the drag distribution (Figure 8(b)). For the $\mathbf{y}^o = \partial_z E_x$ case, there is a good agreement in the lower part of the EP flux divergence (Figure 8(d)). The optimal parameters, $\widehat{\lambda_*} \approx 0.5$ and $\widehat{S_*} \approx 10$, produce the maximum EP flux divergence at low heights. However there is no sign of the summer deceleration centre at the upper part of the profile. Apparently the eastward waves do not saturate within the estimation height range (Figure 8(c)) and therefore no positive forcing (deceleration centre) is found in this case up to the 0.3 hPa top. This behaviour is in fact expected on physical grounds, since the summer jet core is located at higher altitudes, thus the waves are expected to saturate at higher altitudes (e.g. Lindzen, 1981).

Keeping the assumption of an isotropic spectrum raises the question of whether $\mathbf{y}^o = \mathbf{X}$ or $\mathbf{y}^o = \partial_z E_x$ gives the most appropriate set of parameters. There are two major
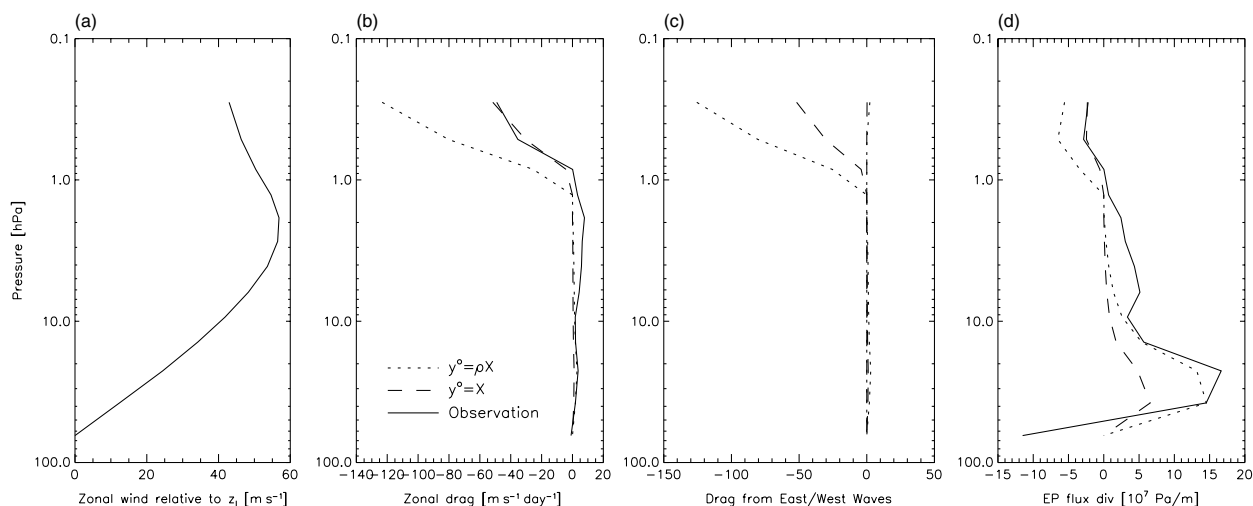
Figure 7. (a) Monthly averaged zonal wind profile $\widehat{u}$ at 63°S for July 2002. (b) Observed gravity wave drag profile at 63°S (solid line), estimated drag using $\mathbf{y}^o = \mathbf{X}$ (dashed line) and $\mathbf{y}^o = \partial_z E_x$ (dotted line). (c) Contribution to the GWD profile from westward and eastward waves (corresponding respectively to negative and positive drag) for $\mathbf{y}^o = \mathbf{X}$ (dashed lines) and $\mathbf{y}^o = \partial_z E_x$ (dotted lines). (d) Observed EP flux divergence (solid line), estimated EP flux divergence using $\mathbf{y}^o = \mathbf{X}$ (dashed line) and $\mathbf{y}^o = \partial_z E_x$ (dotted line).
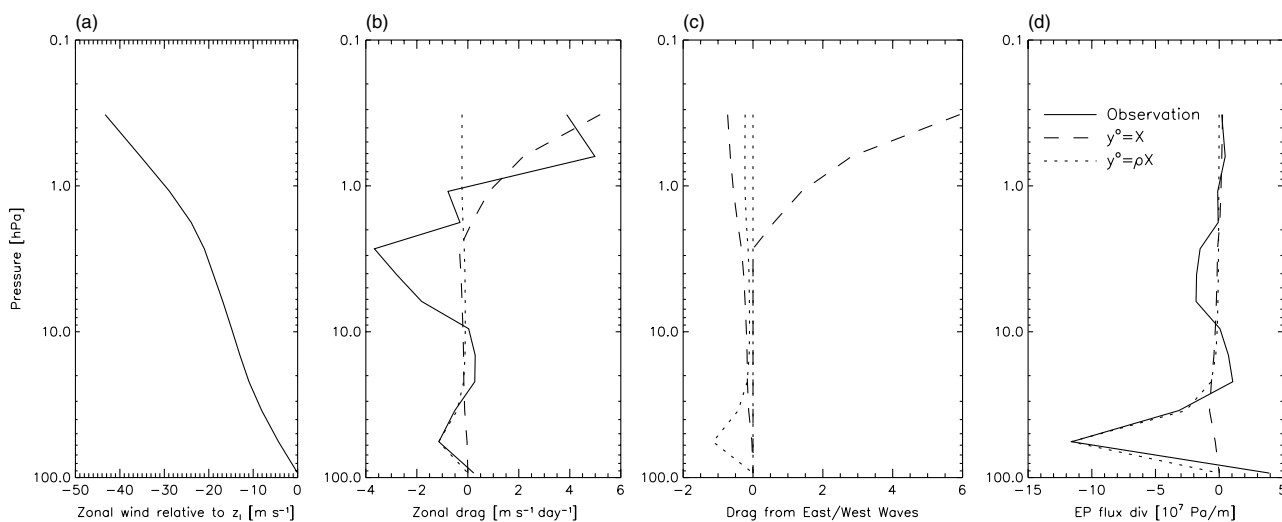

Figure 8. As Figure 7, but for 63°N.

points that lead to the conclusion that $\mathbf{y}^o = \mathbf{X}$ is likely to be the best choice. Firstly, the most important effect of non-orographic GWD parametrisations is that produced around and above the stratospheric jet maximum, while at lower heights the effect of the forcing by non-orographic GWD parametrisations generally plays a secondary role. Secondly, the filtering of the spectrum at low heights is most likely to be the mechanism responsible for the net (non-isotropic) momentum flux found in observations (e.g. Hertzog *et al.*, 2008) and in the data assimilation technique (Pulido and Thuburn, 2008); also Alexander *et al.* (2010). The part of the spectrum corresponding to waves with slow intrinsic phase speeds must be correlated with actual sources (these waves have vertical wavenumber higher than $m_* = 2\pi/\lambda_*$), while fast waves are expected to better satisfy the isotropic assumption. Thus, constraining the drag at high altitudes is more consistent with an isotropic launch spectrum. The choice of GWD or EP flux divergence as control variable could also be evaluated in future work examining the response to the different sets of optimal parameters in GCM simulations.

4.1. Global estimation

An estimation of the optimal global parameters was performed. In this case the observation $\mathbf{y}^o$ is not a vertical column of GWD at a single latitude but all the columns. The optimal parameters are the ones that give the best fit to the observed GWD for all latitudes at the same time. Profiles at all latitudes have the same weight in the cost function. The optimal parameters are $E_*/E_R = 5.9$, $\lambda_*/\lambda_R = 2.2$ and $S_*/S_R = 9.5$. Figure 9 shows the GWD and the zonal EP flux given by the parametrisation with the optimal set of global parameters. Apparently the parametrisation cannot reproduce the observed GWD with a single set of parameters applied globally, whereas a latitude-by-latitude fit can reproduce the observed drag quite well (Figure 4). The optimal global parameter set does not give a good representation of the GWD in the tropical region, though it does improve the GWD representation compared with the standard set of parameters (Figure 6). In particular, the vertical distribution of the deceleration centre above and the acceleration centre below are better represented.
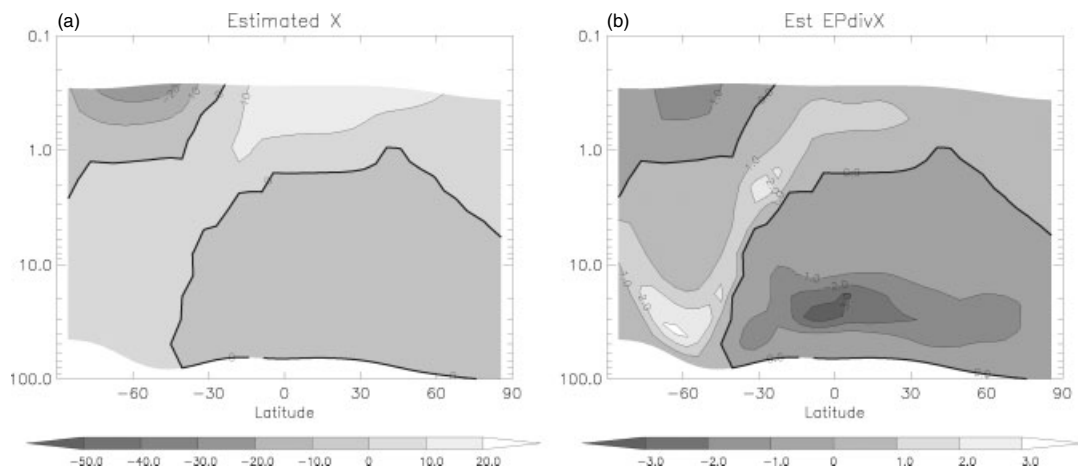
Figure 9. (a) Zonal mean zonal drag given by the GWD scheme with the optimum global parameters using drag in the observation space, and (b) the corresponding zonal EP flux divergence. Note that the contours in (b) are different from those in Figure 2(b).

Because even the latitude-by-latitude fit cannot reproduce the observed drag in the Tropics, it might seem inappropriate to include the Tropics within the global fit. Therefore, another global optimization was performed, where the cost function only involved profiles located at latitudes poleward of $30°$. In this case the optimal parameters are $E_*/E_R = 8.7$, $\lambda_*/\lambda_R = 2.2$, and $S_*/S_R = 9.8$. The latter two parameters are similar to the values in the full global optimization so the GWD patterns are the same, but the total momentum flux E_* is stronger. Therefore the resulting deceleration centre in the winter lower mesosphere is stronger, reaching a peak deceleration of 30 m s^{-1}day^{-1}, which is close to the observed GWD. However an important overestimation of the drag in the Tropics is found with values that exceed $20 \, \text{m s}^{-1}\text{day}^{-1}$. This suggests that tropical parameters should be set differently from extratropical parameters.

5. Conclusions

Physical parametrisations model the nonlinear small-scale processes in climate models, and they often contain switches given by threshold values of the state variables. In particular GWD parametrisations represent saturation and filtering of the gravity wave spectrum. The sensitivity of the GWD to changes in the parameters that characterize the launch spectrum behaves nonlinearly. Furthermore, certain directions in the parameter space have weak sensitivity, i.e. large changes in the parameter produce small changes in the GWD. The estimation problem under these conditions is challenging. The evaluation of a variational data assimilation technique shows that it does not converge towards the global minimum for some cases. The genetic algorithm does appear to have robust estimations. The use of a hybrid technique gives the best convergence results. It uses the genetic algorithm to localize the global minimum and then, given a good first guess, the variational technique does converge towards the minimum and it is quite efficient, the convergence rate being quadratic.

The Scinocca GWD parametrisation with optimal parameters does a remarkably good job in fitting the 'observed' GWD at high latitudes, where the 'observed' GWD is determined by an independent data assimilation technique (Pulido and Thuburn, 2008). At low latitudes, the parametrisation cannot capture the low positive GWD which is essential for the positive phase of the quasi-biennial

oscillation (QBO). The parametrisation cannot reproduce this feature even for a cost function that puts particular emphasis on the fit to the lower stratosphere. This could be partially because the fixed launch level of the waves, 100 hPa, in the parameter estimation may be too high in the Tropics. Also it could indicate a deficiency with the Scinocca scheme (for the given launch height), or it could reflect an incorrect representation (in the ASDE model) of the resolved equatorial planetary waves, especially Kelvin waves which are known to be crucial for driving the positive phase of the QBO.

In a sheared mean flow, an inverse technique can capture the characteristics of the launch spectrum from the filtering process. A particular drag at a particular height range due to the filtering process can only be produced by the waves which have their phase speed equal to the relevant component of the mean flow. The amplitude of the waves as a function of phase speed can then be recovered from an inverse technique. Once the characteristics of the launch spectrum are estimated by the filtering mechanism (say the eastward waves in a winter jet), the non-filtered part of the spectrum (westward waves) is then determined by the isotropy assumption, and therefore the momentum deposition aloft is completely determined and it may not fit the observed drag profile. This is the case when the EP flux divergence is used as observed variable. On the other hand, when drag is used as observed variable, the characteristics of the launch spectrum are determined from the saturation process that occurs at high altitudes (say for westward waves in a winter jet) and therefore the lower part of the drag profile cannot fit the observed drag since the eastward waves are already set by the isotropy assumption. Future work could relax the isotropy assumption in the parameter estimation, independently estimating momentum fluxes, say E_* and λ_*, for each direction. The impact of changing the launch height in an inverse problem context also needs to be evaluated. By changing the launch height, the anisotropy of the spectrum at a given level above will also change, so these two sensitivities are related.

The most reliable parameter in the estimation is E_*, the drag profile being directly proportional to this parameter. The global estimation gives a value of $\widehat{E}_*$ of 5.9 ($E_* = 2 \times 10^{-3}$Pa) which results in a GWD with a deceleration peak of 25 m s^{-1}day^{-1} at 0.3 hPa in the winter Southern Hemisphere. The estimated values of $\widehat{E}_*$

are in a range of 2 to 10 in midlatitudes when the parameters are assumed to depend on latitude for the $\mathbf{y}^{\mathrm{o}} = \mathbf{X}$ case. Although a completely unconstrained optimization may yield unrealistically large parameter values because of the low sensitivity of the drag to a combination of λ_* and S_* changes, by constraining one of the parameters within a physically reasonable range, i.e. $S_* < 10$, the estimation of the other parameters is then also within a physically reasonable range. (Although $\widehat{\lambda}_*$ was also constrained to values smaller than 10, this constraint is not needed for the $\mathbf{y}^{\mathrm{o}} = \mathbf{X}$ case except for two profiles at low latitudes.) This ability to approximate the observed drag profiles with physically reasonable parameter values demonstrates the basic physical self-consistency of Scinocca's scheme for the extratropics.

Scinocca's parametrisation, as with all non-orographic GWD parametrisations, has been designed with only a few free parameters since the launch momentum flux and the breaking mechanisms were almost completely uncertain in the past. A consequence of this strategy identified here is the weak sensitivity to certain combinations of the parameters and the resulting difficulty in fitting standard high-latitude drag profiles. This impacts not only the use of inverse techniques but also observational constraints on the launch momentum flux. Currently the available observational information on the GWD and also on the launch momentum flux, from indirect techniques or from direct observations (Alexander *et al.*, 2010), is rapidly increasing. Therefore, a redesign and some relaxation of assumptions in the parametrisations may be required in order to make optimal use of the observational information. The main results of this work in this sense have also been found in preliminary experiments (not shown) for the Warner and McIntyre (2001) and Hines (1997) parametrisations, which share the same philosophy on the launch momentum flux. As has been shown by McLandress and Scinocca (2005), the differences of the breaking mechanisms in these parametrisations do not produce important differences in the resulting GWD.

The estimation of the gravity wave parameters has been performed with an offline optimization algorithm so that it did not require the use of a general circulation model. The impact of the different sets of optimal parameters in simulations using a general circulation model needs to be evaluated in future work. Other optimization techniques, apart from the genetic algorithm used here, that are suitable for highly nonlinear problems are multiple very fast simulated annealing or Markov chain Monte Carlo. Future work could also involve the estimation of the parameters directly in the general circulation model instead of using a two-stage estimation. However neither the genetic algorithm nor Markov chain Monte Carlo appear to be appropriate for such experiments. For the estimation shown in Figure 3 about 5 000–20 000 evaluations of the 'model' were required. Alternative techniques that are particularly suitable for large-dimension systems and that converge faster are therefore needed. Techniques based on the ensemble Kalman filter (e.g. Evensen, 2003) could be appropriate, however evaluation of the impact of nonlinearities and ill-conditioning in these techniques for the gravity wave problem is required first.

Acknowledgements

References

Alexander MJ, Geller M, McLandress C, Polavarapu S, Preusse P, Sassi F, Sato K, Eckermann S, Ern M, Hertzog A, Kawatani Y, Pulido M, Shaw TA, Sigmond M, Vincent RA, Watanabe S. 2010. Recent developments in gravity wave effects in climate models, and the global distribution of gravity wave momentum flux from observations and models. *Q. J. R. Meteorol. Soc.* **136**: 1103–1124.

Allen SJ, Vincent RA. 1995. Gravity wave activity in the lower atmosphere: Seasonal and latitudinal variations. *J. Geophys. Res.* **100**: 1327–1350.

Brunet G, Shapiro M, Hoskins BJ, Moncrieff M, Dole R, Kiladis GN, Kirtman B, Lorenc AC, Mills B, Morss R, Polavarapu S, Rogers D, Schaake J, Shukla J. 2010. Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bull. Amer. Meteorol. Soc.* **91**: 1397–1406.

Butchart N, Scaife AA, Bourqui M, de Grandpré J, Hare SHE, Kettleborough J, Langematz U, Manzini E, Sassi F, Shibata K, Shindell D, Sigmond M. 2006. Simulations of anthropogenic change in the strength of the Brewer–Dobson circulation. *Climate Dyn.* **27**: 727–741, DOI: 10.1007/s00382-006-0162-4

Butchart N, Cionni I, Eyring V, Shepherd TG, Waugh DW, Akiyoshi H, Austin J, Brühl C, Chipperfield MP, Cordero E, Dameris M, Deckert R, Frith SM, Garcia RR, Gettelman A, Giorgetta MA, Kinnison DE, Li F, Mancini E, McLandress C, Pawson S, Pitari G, Plummer DA, Rozanov E, Sassi F, Scinocca JF, Shibata K, Tian W. 2010. Chemistry-climate model simulations of 21st century stratospheric climate and circulation changes. *J. Climate* **23**: 5349–5374.

Charbonneau P. 2002. 'An introduction to genetic algorithms for numerical optimization'. Technical Note TN-450+IA. NCAR: Boulder, MA.

Charron M, Manzini E. 2002. Gravity waves from fronts: Parameterization and middle atmosphere response in a general circulation model. *J. Atmos. Sci.* **59**: 923–941.

Evensen G. 2003. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **53**: 343–367. DOI: 10.1007/s10236-003-0036-9

Giering R, Kaminski T. 1997. Recipes for adjoint code construction. *ACM Trans. Math. Software* **24**: 437–474.

Golberg DE. 1989. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley: Boston, USA.

Hertzog A, Boccara G, Vincent RA, Vial F, Cocquerez P. 2008. Estimation of gravity wave momentum flux and phase speeds from quasi-Lagrangian stratospheric balloon flights. Part II: Results from the Vorcore campaign in Antarctica. *J. Atmos. Sci.* **65**: 3056–3070.

Hines CO. 1997. Doppler spread parametrization of gravity-wave momentum deposition in the middle atmosphere. Part 1: Basic formulation. *J. Atmos. Sol. Terr. Phys.* **59**: 371–386.

Hurrell J, Meehl GA, Bader D, Delworth TL, Kirtman B, Wielicki B. 2009. A unified modeling approach to climate system prediction. *Bull. Amer. Meteorol. Soc.* **90**: 1819–1832.

Kalnay E. 2002. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press: Cambridge, UK.

Karlsson B, McLandress C, Shepherd TG. 2009. Inter-hemispheric mesospheric coupling in a comprehensive middle atmosphere model. *J. Atmos. Solar-Terr. Phys.* **71**: 518–530. DOI: 10.1016/j.jastp.2008.08.006

Li F, Austin J, Wilson J. 2008. The strength of the Brewer–Dobson circulation in a changing climate: Coupled chemistry-climate model simulations. *J. Climate* **21**: 40–57.

Lindzen RS. 1981. Turbulence and stress owing to gravity wave and tidal breakdown. *J. Gheophys. Res.* **86**: 9707–9714.

Manzini E, McFarlane NA. 1998. The effect of varying the source spectrum of a gravity wave parameterization in a middle atmosphere general circulation model. *J. Geophys. Res.* **103**: 31523–31539.

McFarlane NA. 1987. The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.* **44**: 1775–1800.

McLandress C, Scinocca JF. 2005. The GCM response to current parameterizations of non-orographic gravity wave drag. *J. Atmos. Sci.* **62**: 2394–2413.

McLandress C, Shepherd TG. 2009. Simulated anthropogenic changes in the Brewer–Dobson circulation, including its extension to high latitudes. *J. Climate* **22**: 1516–1540.

Orr A, Bechtold P, Scinocca JF, Ern M, Janisková M. 2010. Improved middle atmosphere climate and forecasts in the ECMWF model through a non-orographic gravity wave drag parametrization. *J. Climate* **23**: 5905–5926.

Palmer TN, Shutts GJ, Swinbank R. 1986. Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through on orographic gravity wave drag parametrization. *Q. J. R. Meteorol. Soc.* **112**: 1001–1039.

Palmer TN, Doblas-Reyes FJ, Weisheimer A, Rodwell MJ. 2008. Toward seamless prediction: Calibration of climate change projections using seasonal forecasts. *Bull. Amer. Meteorol. Soc.* **89**: 459–470.

Phillips TJ, Potter GL, Williamson DL, Cederwall RT, Boyle JS, Fiorino M, Hnilo JJ, Olson JG, Xie S, Yio JJ. 2004. Evaluating parameterizations in general circulation models. Climate simulation meets weather prediction. *Bull. Amer. Meteorol. Soc.* **85**: 1903–1915.

Preusse P, Eckermann SD, Ern M. 2008. Transparency of the atmosphere to short horizontal wavelength gravity waves. *J. Geophys. Res.* **113**: D24104. DOI: 10.1029/2007JD009682

Pulido M, 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.

Pulido M, Thuburn J. 2006. Gravity wave drag estimation from global analyses using variational data assimilation principles. II: A case-study. *Q. J. R. Meteorol. Soc.* **132**: 1527–1543.

Pulido M, Thuburn J. 2008. The seasonal cycle of gravity wave drag in the middle atmosphere. *J. Climate* **21**: 4664–4679.

Randel WJ, Garcia R, Wu F. 2008. Dynamical balances and tropical stratospheric upwelling. *J. Atmos. Sci.* **65**: 3584–3595.

Ren S, Polavarapu S, Shepherd TG. 2008. Vertical propagation of information in a middle atmosphere data assimilation system. *Geophys. Res. Lett.* **35**: L06804, DOI: 10.1029/2007GL032699

Rodwell MJ, Palmer TN. 2007. Using numerical weather prediction to assess climate models. *Q. J. R. Meteorol. Soc.* **33**: 129–146.

Scinocca JF. 2002. The effect of back-reflection in the parametrization of non-orographic gravity-wave drag. *J. Meteorol. Soc. Japan* **80**: 939–962.

Scinocca JF. 2003. An accurate spectral non-orographic gravity wave drag parameterization for general circulation models. *J. Atmos. Sci.* **60**: 667–682.

Sigmond M, Scinocca JF. 2010. The influence of the basic state on the Northern Hemisphere circulation response to climate change. *J. Climate* **23**: 1434–1446.

Song I-S, Chun H-Y. 2008. A Lagrangian spectral parameterization of gravity wave drag induced by cumulus convection. *J. Atmos. Sci.* **65**: 1204–1224.

Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, Piani C, Sexton D, Smith LA, Spicer RA, Thorpe AJ, Allen MR. 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**: 403–406.

Warner CD, McIntyre ME. 1996. On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere. *J. Atmos. Sci.* **53**: 3213–3235.

Xu Q. 1996. Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples. *J. Atmos. Sci.* **53**: 1123–1142.