Using data assimilation to improve climate models: The stratosphere

Manuel Pulido

Department of Physics, Universidad Nacional del Nordeste,

and IMIT, UMI-IFAECI/CNRS, CONICET, Argentina

Acknowledgements: S. Polavarapu, J. Ruiz, T. Shepherd, G. Scheffler, P. Tandeo, J. Thuburn.

Outline

1. **Motivation: Systematic Bias in the Stratosphere.**
2. **Estimation of the Missing Forcing. 4DVar.**
   - **Twin experiments**
   - **Realistic experiments**
3. **Offline Parameter Estimation.**
   - **Variational data assimilation**
   - **Genetic algorithm**
   - **EnKF with Maximum likelihood estimation**
4. **Conclusions.**

Source of errors in climate models: subgrid effects

Small-scale waves produce a systematic forcing to the general circulation. However, GCMs can not resolve all the spectrum of small-scale waves.

How can we infer this systematic momentum deficit (missing forcing) in a GCM?

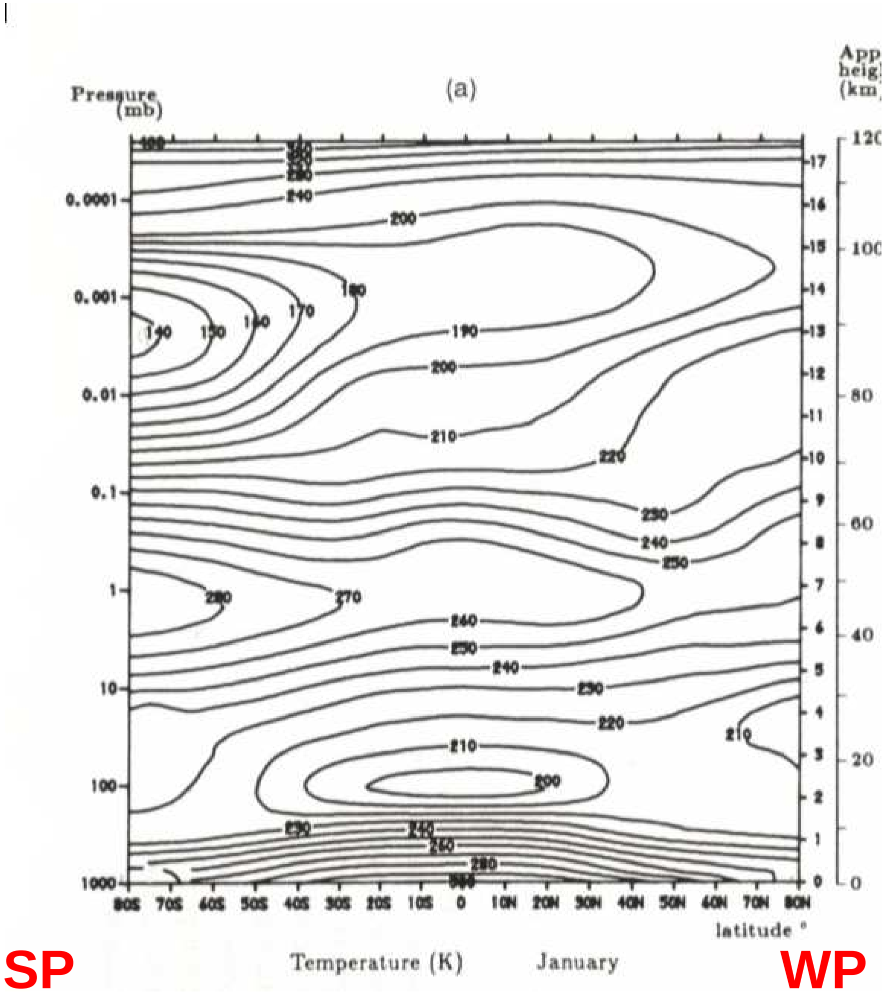
It is not easy. If one computes the difference between the true state (observation) and the GCM state, the result is a combination of different source of errors, recent and past, which once they are generated are advected and interact with other parts of the system.

1. Is there an objetive technique to determine the source of the momentum deficit, i.e., the exact time and position where the momentum error was produced?

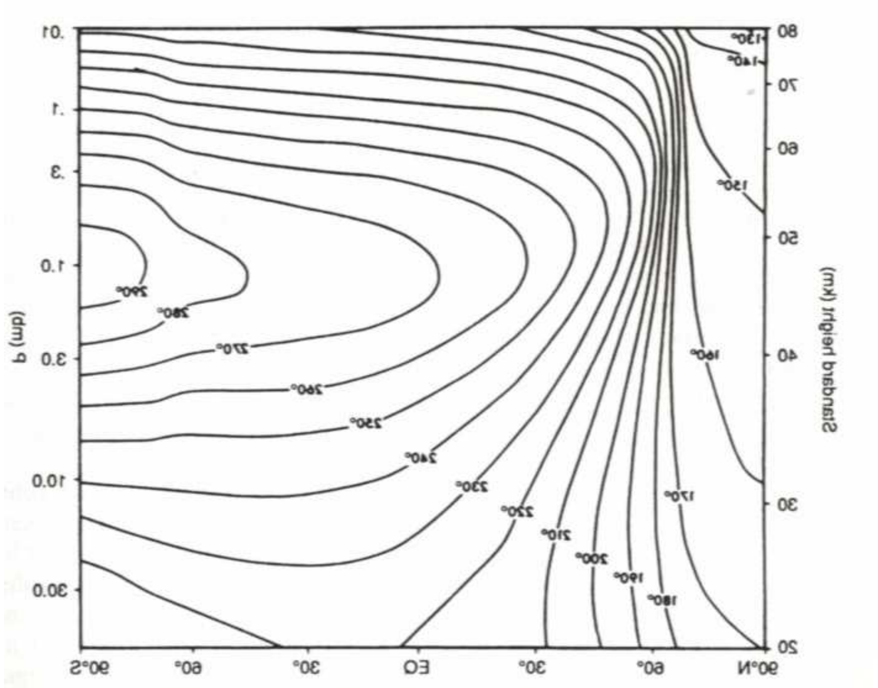
2. Can we use this information to improve parameterizations of small-scale effects?

Motivation

Observed temperature vs radiative temperature in the middle atmosphere.



SP WP



SP WP

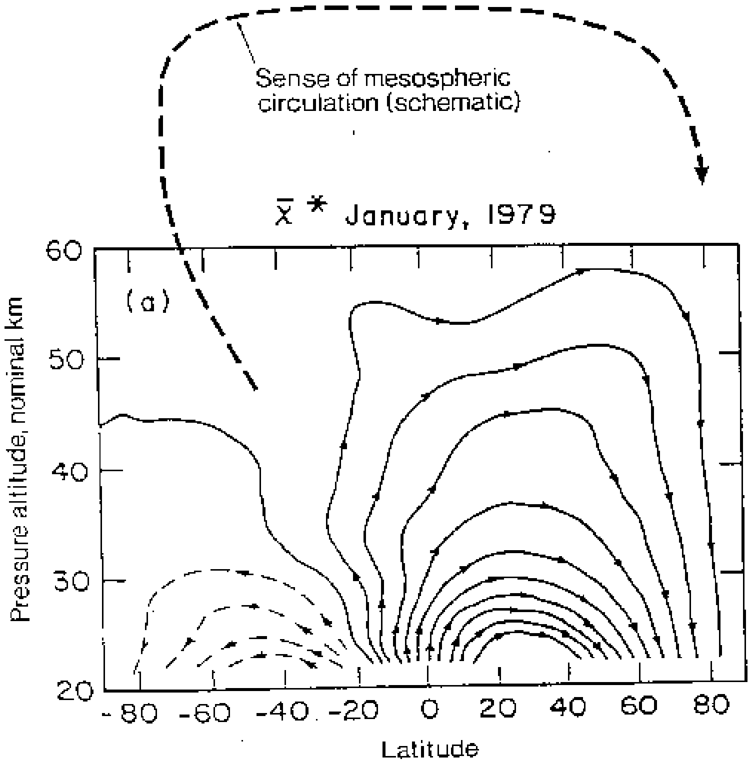
Observed temperature for January. Andrews, et al. 1987.

Radiative temperature. Fels et al. JGR85.

What effect produces the inversion in the horizontal temperature gradient?

What process cools the summer hemisphere and heats the winter hemisphere?

Meridional circulation and the Coriolis torque



SP **WP**

Solomon,*et al.* 1986

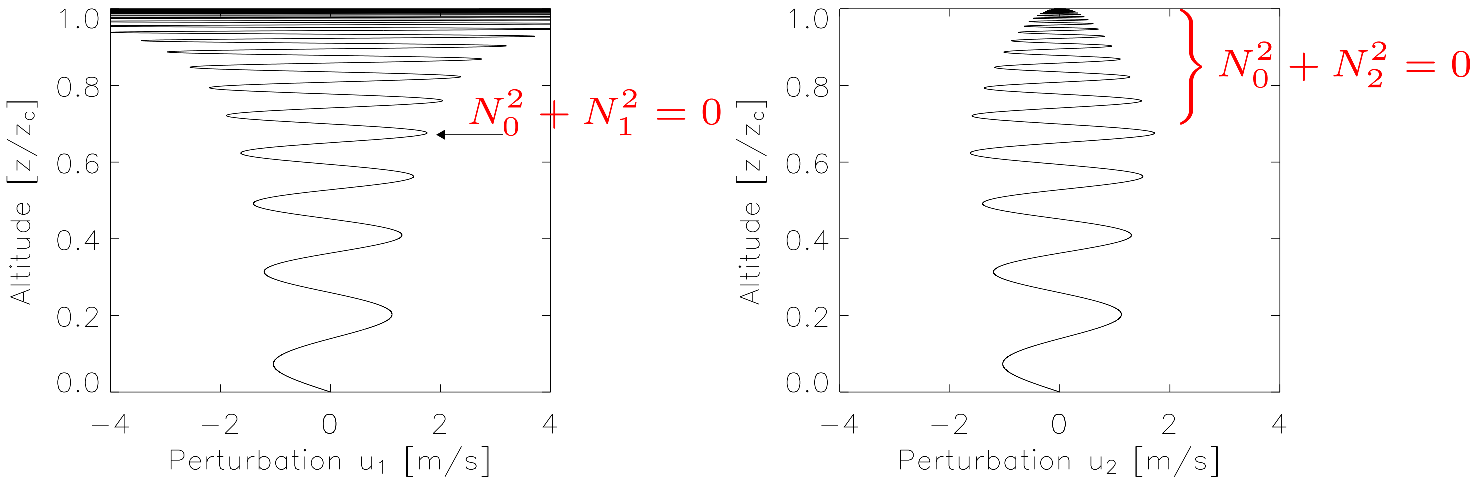
The meridional circulation needed to keep the observed temperature is $v \approx 9\mathrm{m\,s^{-1}}$.

This circulation produces a Coriolis torque of

$$-fv = 70\mathrm{m\,s^{-1}\,day^{-1}}.$$

What force balances the Coriolis torque in the mesosphere?

Small-scale gravity waves



Gravity wave amplitude increases with height due to density decreasing. When a wave reaches the instability threshold, starts to break generating turbulence.

Lindzen (1981) showed that if the wave amplitude is kept at the convective instability threshold: $N_0^2 + N_2^2 = 0$, the divergence of the momentum flux is

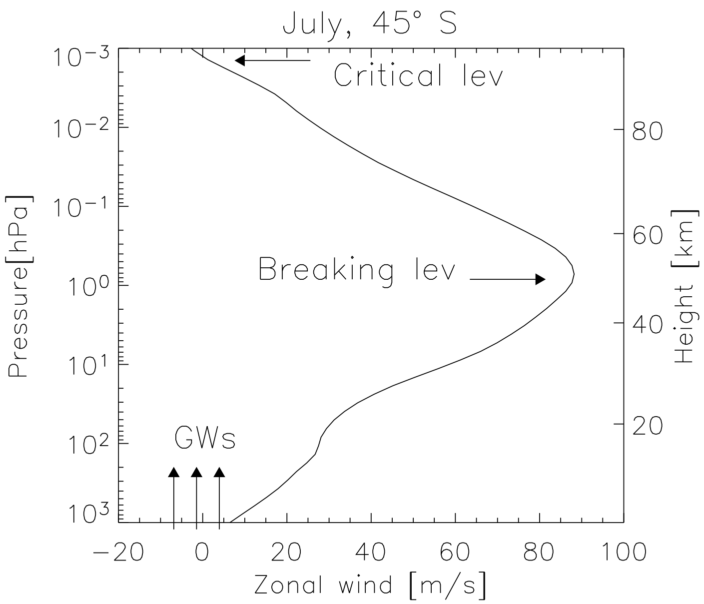
$$\rho_0^{-1} \frac{d}{dz} \overline{u_2 w_2} = -\frac{k}{2HN} (u_0 - c)^3$$

Small-scale gravity waves produce a nonreversible forcing to the mean flow.

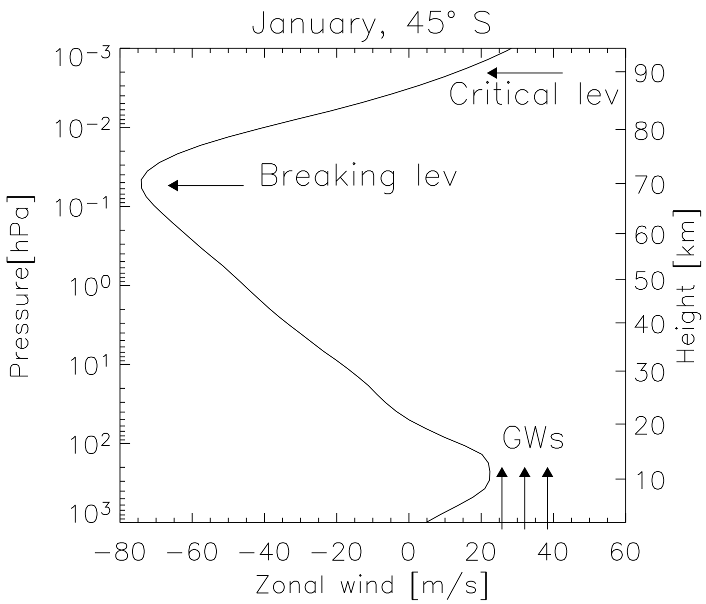
In particular: the breaking of waves produce momentum flux divergence that forces the mean flow.

Systematic gravity wave forcing

How can waves, that are statistically isotropic, produce a systematic forcing in one direction?



Mean zonal wind at 45^oS in winter.



Mean zonal wind at 45^oS in summer.

Ans. Part of the wave spectrum is filtered by the mean zonal wind (Lindzen, 1981).

$==>$ the breaking produces a positive zonal forcing (negative) in the summer hemisphere (winter) for an isotropic spectrum.

What is the actual GW forcing in the atmosphere?

High vertical and horizontal resolution measurements of u and w + They must be global. $\longrightarrow$ Unachievable.

Alternative: an inverse technique.

Through the effects that GW forcing produces on the large-scale flow (i.e. low-resolution large-scale observations), we want to estimate the source of this large-scale response.

Can we apply 4DVar concepts to estimate the missing force due to the unresolved waves in a model?

4DVar under perfect ignorance hypothesis

There is no background information (perfect ignorance), so the cost function is defined as

$$J = \frac{1}{2}\sum_{i=1}^{n}(H[\mathbf{y}_i] - \mathbf{x}_i)^T \mathbf{R}^{-1}(H[\mathbf{y}_i] - \mathbf{x}_i)$$

where $\mathbf{x}_i$ is the model state, $\mathbf{y}_i$ are the observations. The state is given by the model evolution

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} + M(\mathbf{x}, t) = \boldsymbol{X}$$

from t_0 to t_i, we have $\mathbf{x}_i = F(\mathbf{x}_0, \boldsymbol{X}, t_i)$

The model state is a function of the initial condition and also of the 'missing forcing'. Then $J = J(\mathbf{x}_0, \boldsymbol{X})$

Therefore, if we know $\mathbf{x}_0$ the control space of the cost function is only the field $\mathbf{X}$. The minimum of the cost function gives the 'missing forcing' (Pulido and Thuburn, QJ 2005).

Middle atmosphere dynamical model, $F(\mathbf{x}_0, \boldsymbol{X}, t_i)$

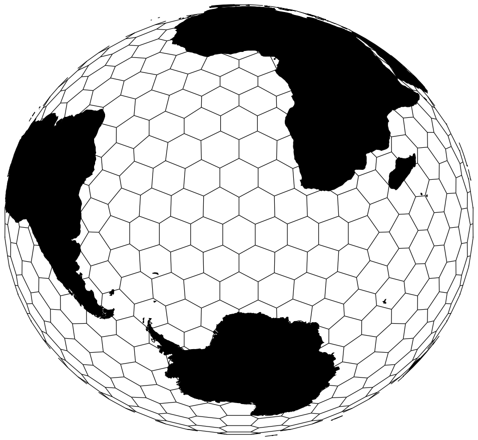
The dynamical model is based on the fully nonlinear, hydrostatic primitive equations, with an isentropic vertical coordinate and a hexagonal-icosahedral horizontal grid (Thuburn 1994).

$$\partial_t \sigma + \nabla \cdot (\sigma \mathbf{u}) + \partial_\theta (\sigma \dot{\theta}) = 0$$

$$\partial_t (\sigma Q) + \nabla \cdot (\sigma Q \mathbf{u} - \hat{\mathbf{k}} \times \dot{\theta} \partial_\theta \mathbf{u}) = X_\zeta$$

$$\partial_t \delta + \nabla \cdot [\sigma Q \hat{\mathbf{k}} \times \mathbf{u} + \nabla (\Psi + \frac{\mathbf{u}^2}{2}) + \dot{\theta} \partial_\theta \mathbf{u}] = X_\delta$$

$$\theta_t = H$$



Horizontal icosahedral grid.

The bottom boundary condition is set at $p \approx 100mb$, where a time dependent observational Montgomery potential is imposed.

A realistic parametrisation of radiative transfer is used (Shine 1987; Shine and Rickaby 1989). GW scheme is switched off.

4DVar implementation

Details of the implementation:

- The gradient of the cost function is calculated with the adjoint model. The full adjoint model of the middle atmosphere dynamical model was developed.
- Augmented State. Initial condition + Forcing term.
- The initial condition is assumed to be known exactly.
- The missing forcing is assumed constant within an assimilation window.
- A conjugate gradient algorithm coupled with the secant method is used to perform the minimization of the cost function.

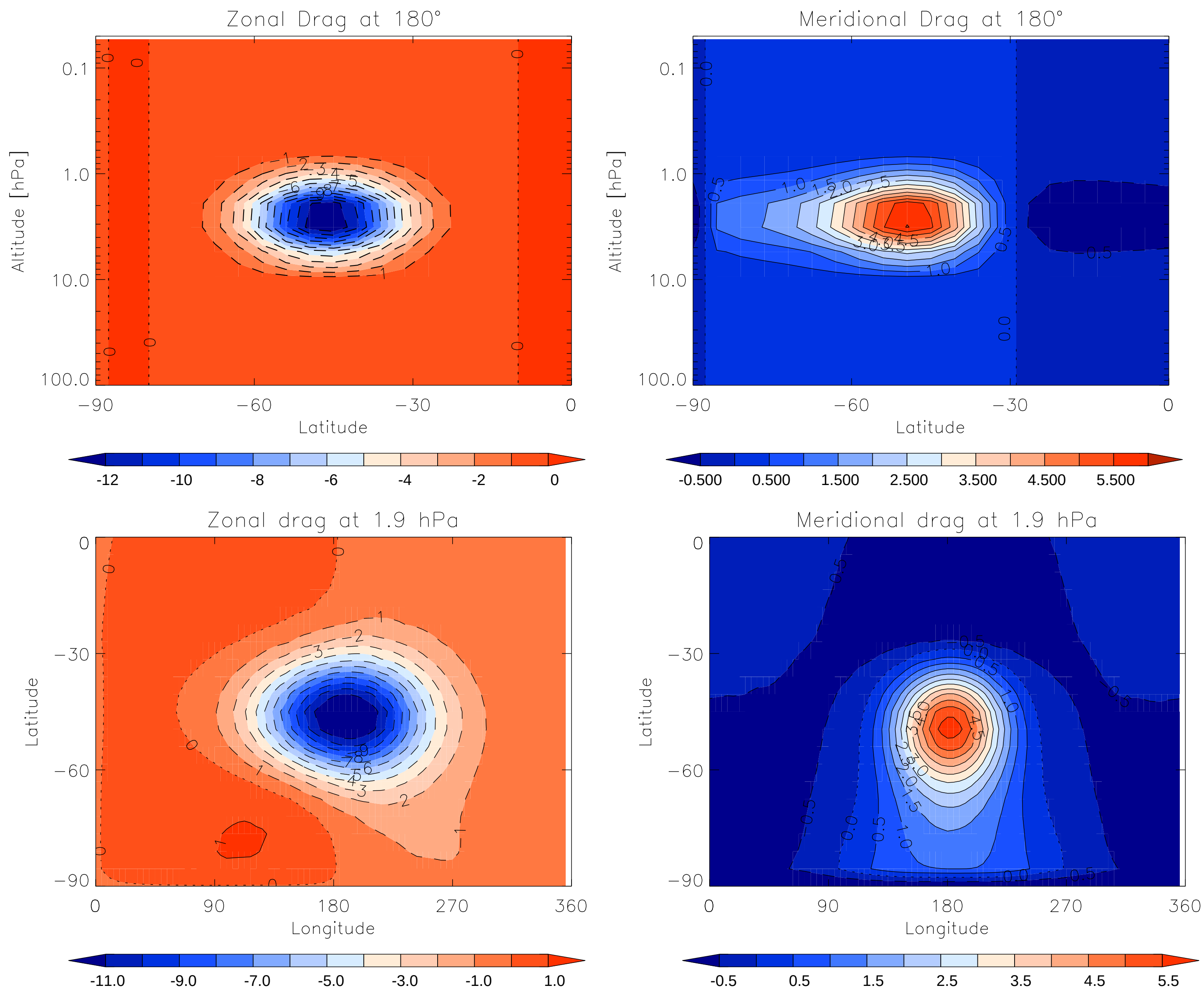
Twin experiments

Experiment:

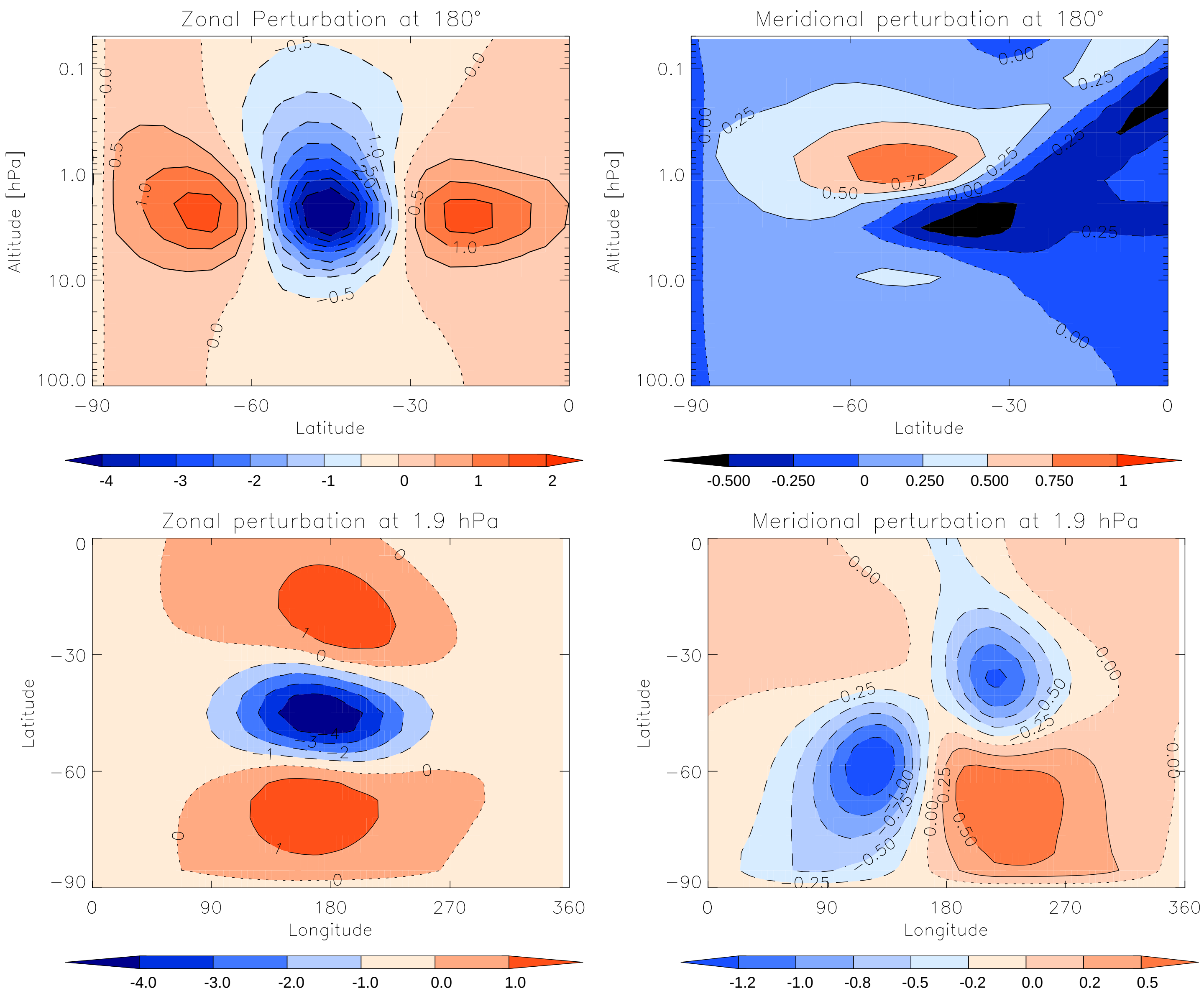
- A Gaussian forcing is used as the prescribed forcing to generate "observation".

- The adiabatic evolution is started from resting condition with an isothermal atmosphere.

- The model evolution with the prescribed forcing is taken as the observation.



Flow response. 'The observations'



Flow response to the applied forcing at t $=$1 day.

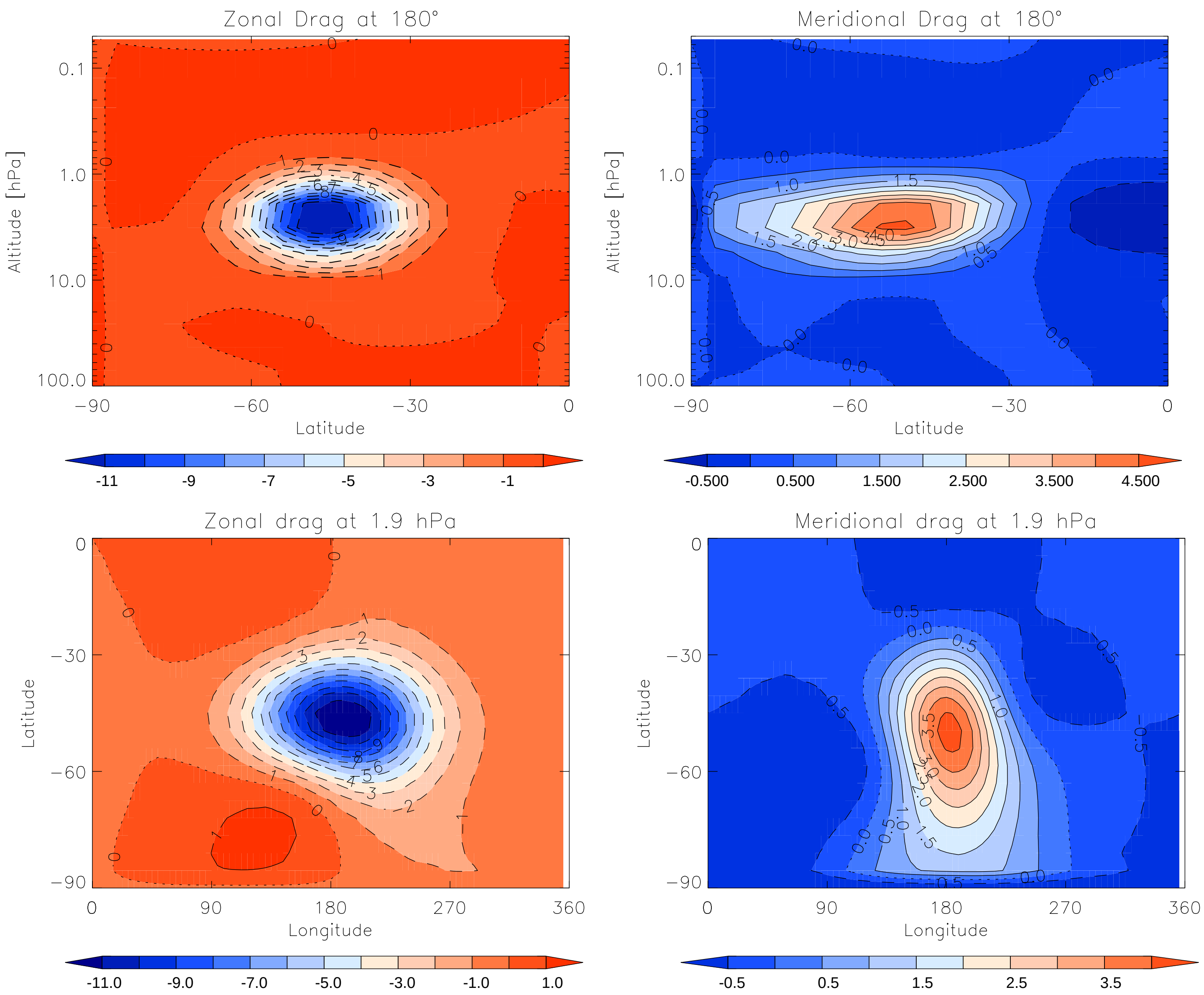
Geostrophic adjustment.

This could be interpreted as a crude budget calculation:

$$\mathbf{X} = [\mathbf{u}_F(1\mathrm{d}) - \mathbf{u}_H(1\mathrm{d})]/1\mathrm{d}$$

$\mathbf{u}_F(1\mathrm{d})$ is the evolution of the model with the forcing term. $\mathbf{u}_H(1\mathrm{d})$ is the evolution of the model without the forcing term.

Estimated 'missing' forcing with 4DVar



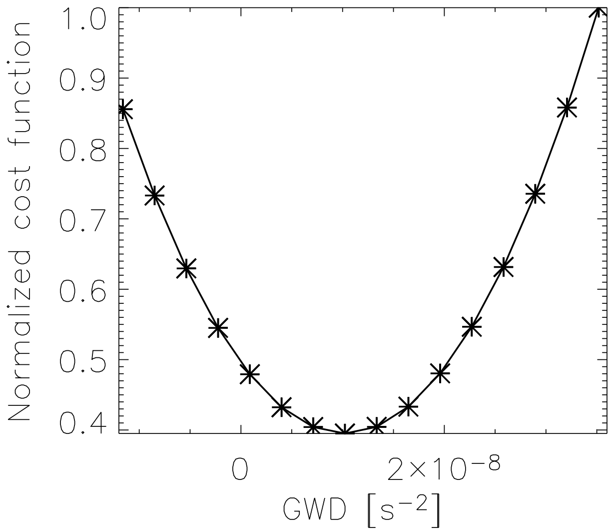
Estimated forcing after 25 minimisation iterations.

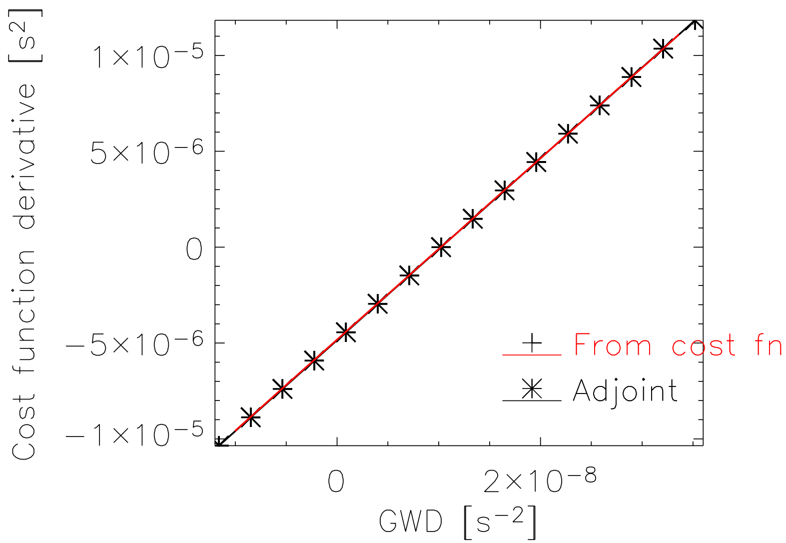
Observations are: $\sigma^*(1\mathrm{d})$, $Q^*(1\mathrm{d})$ and $\delta^*(1\mathrm{d})$. So that

$$J = \sum(\delta - \delta^*)^2 + \overline{\sigma}^2(Q - Q^*)^2 + (\tau\overline{\sigma})^{-2}(\sigma - \sigma^*)^2$$

The error in the forcing estimation is smaller than 1 m/s/day (Pulido and Thuburn QJ 2005).

Is the response linear?



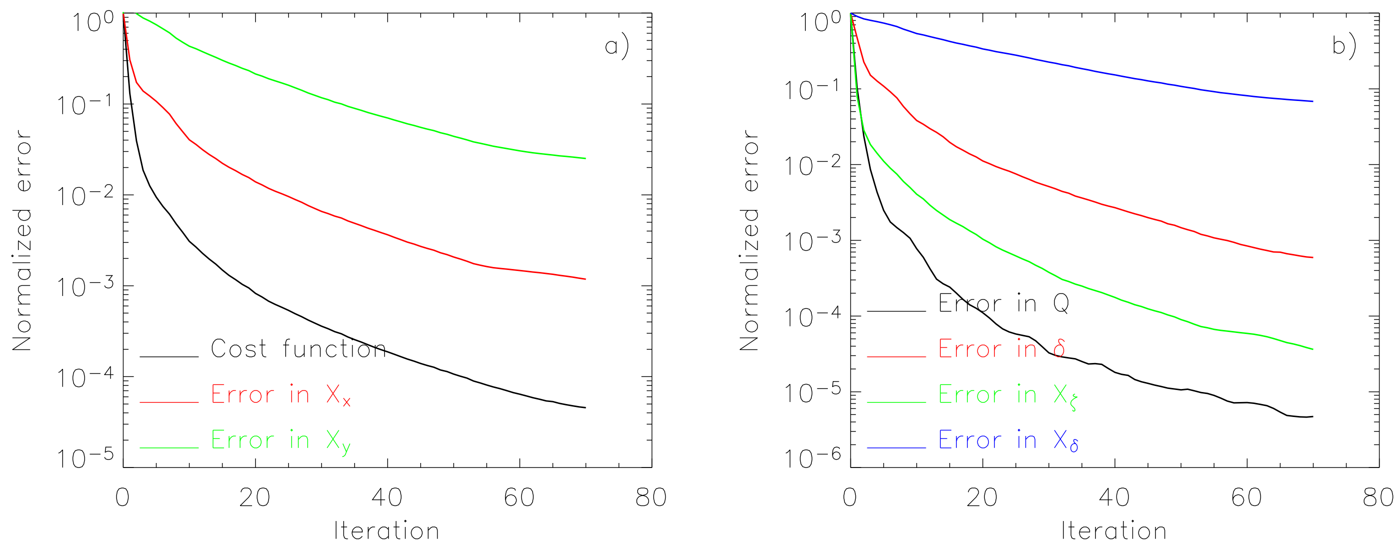


Cost function shape at the 10th minimisation direction.

Derivative of the cost function calculated with the adjoint model and directly from the cost function.

The gradient of the cost function is approximatly linear in all the minimisation directions.

Convergence



Error as a function of minimisation iteration.

25 minimisation iterations are enough to find a good forcing estimate.

The rotational component of forcing is better estimated than divergence component.

Adjustment process

We assume an isothermal background state on an f-plane. Linearizing about a state of rest gives,

$$[(\partial_{tt}^2 + f^2)\mathcal{H} + \nabla^2]\partial_t(\overline{\sigma}^{-1}\sigma') = -\mathcal{H}[\partial_t X_\delta + f X_\zeta]$$

$$[(\partial_{tt}^2 + f^2)\mathcal{H} + \nabla^2]\partial_t\zeta' = \nabla^2 X_\zeta + \mathcal{H}[\partial_{tt}^2 X_\zeta - f\partial_t X_\delta]$$

$$[(\partial_{tt}^2 + f^2)\mathcal{H} + \nabla^2]\delta' = \mathcal{H}[\partial_t X_\delta + f X_\zeta]$$

where $\mathcal{H} = (g\overline{\sigma})^{-1}\partial_\theta(\overline{\rho}\theta\partial_\theta)$.

The solution can be expressed in Fourier components with the fields given by

$$(\sigma'/\overline{\sigma}, \zeta', \delta') = (\hat{\sigma}(t)/\overline{\sigma}, \hat{\zeta}(t), \hat{\delta}(t)) \exp[i(kx + ly) + (1/2H + im)z]$$

The solutions of the homogeneous equations are free inertia-gravity waves which satisfy the dispersion relationship

$$\omega^2 = f^2 + \frac{(k^2 + l^2)N^2}{1/4H^2 + m^2}$$

and a geostrophic mode of frequency, $\omega = 0$.

Solution of the adjustment process

The forced solution is

$$\frac{\hat{\sigma}}{\overline{\sigma}} = -\frac{f}{\omega^2}\hat{X}_\zeta t - \frac{\hat{X}_\delta}{\omega^2}[1-\cos(\omega t)] + \frac{f}{\omega^3}\hat{X}_\zeta \sin(\omega t)$$

$$\hat{\delta} = \frac{f\hat{X}_\zeta}{\omega^2}[1-\cos(\omega t)] + \frac{\hat{X}_\delta}{\omega}\sin(\omega t)$$

$$\hat{\zeta} = \left(1-\frac{f^2}{\omega^2}\right)\hat{X}_\zeta t - \frac{f\hat{X}_\delta}{\omega^2}[1-\cos(\omega t)] + \frac{f^2}{\omega^3}X_\zeta \sin(\omega t)$$

$\hat{X}_\zeta$ produces a geostrophically balanced growing anomaly in Q, ζ and σ, along with some inertia-gravity waves in ζ and σ.

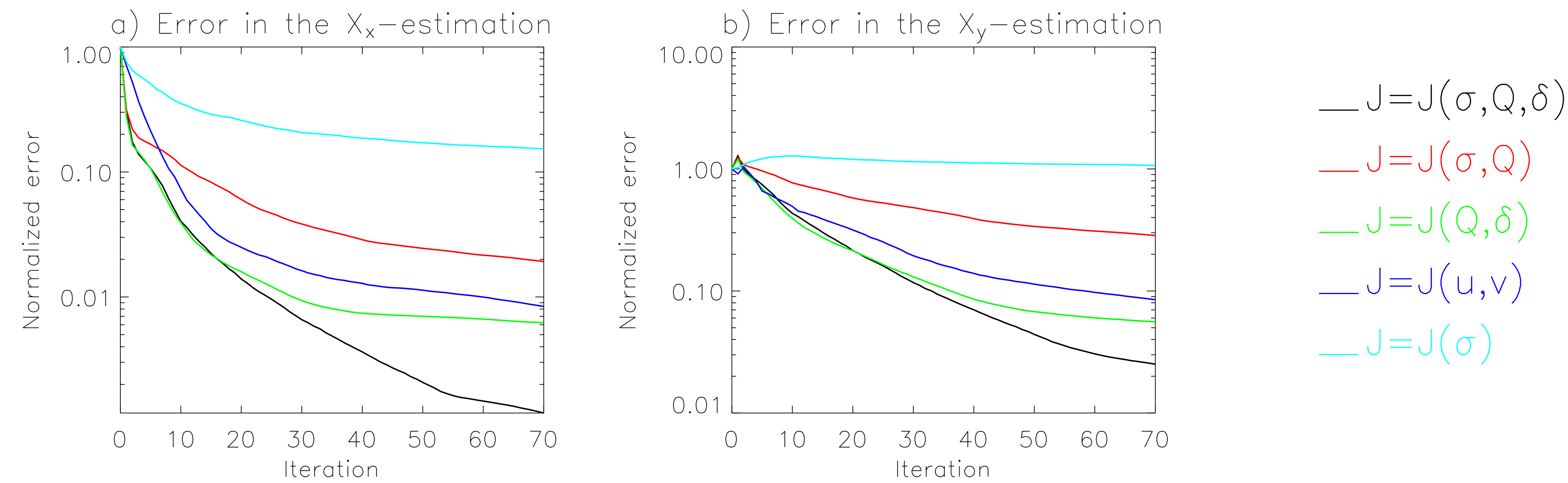
$\hat{X}_\delta$ produces steady ζ and σ anomalies along with some inertia-graity waves.

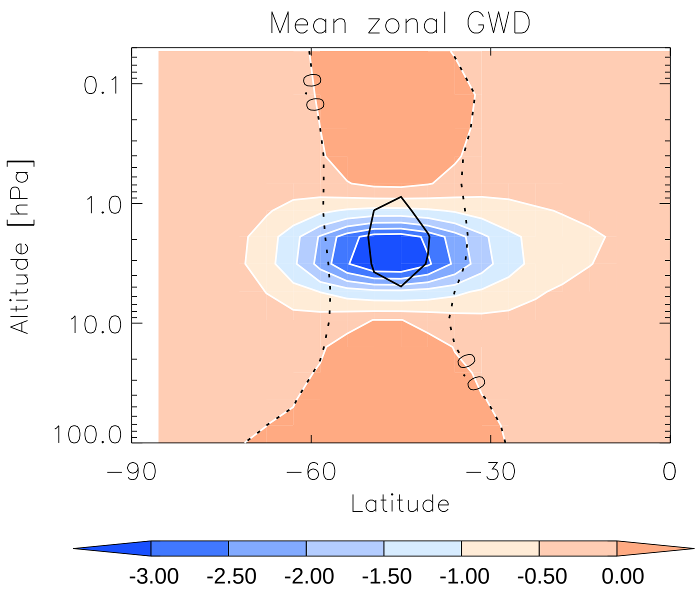
PV is given by:

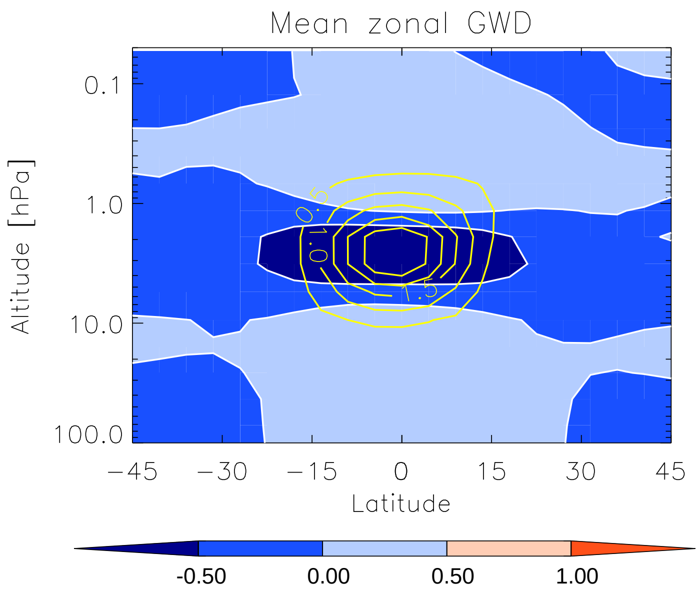
$$Q = \overline{\sigma}^{-1}X_\zeta t$$

Q is affected only by X_ζ, not X_δ. Q does not have an IGW component.

Limited observational information







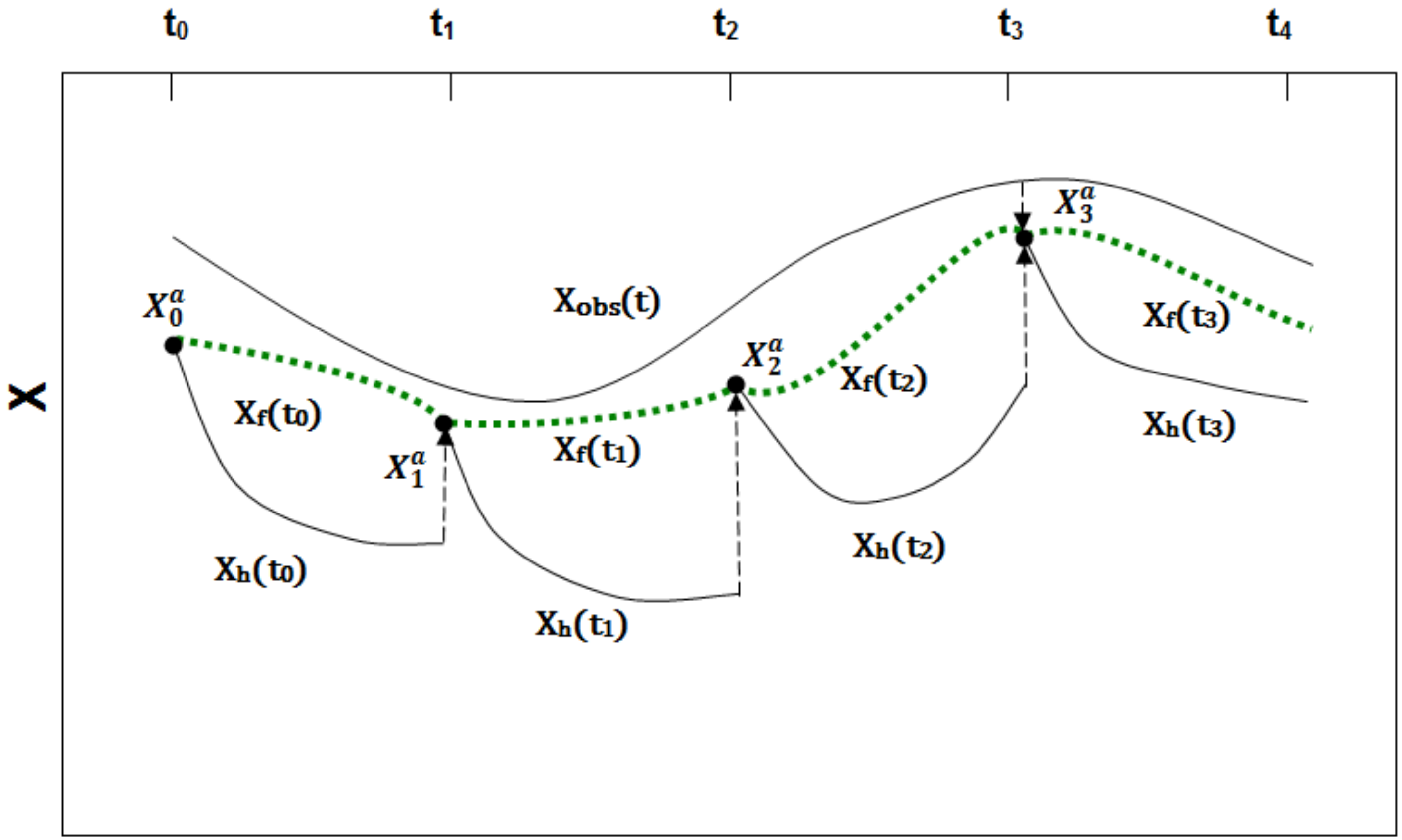
Experiment $J(\sigma)$ with prescribed forcing at midlatitudes (left panel) and at the equator (right panel).

Shading contours are the estimations and contours are the errors.

Actual estimations: Met Office analysis

Observations: Met Office middle atmosphere analyses.

Initial condition: for the first assimilation window of each month is taken from MO analyses, for subsequent windows we use our analyses.

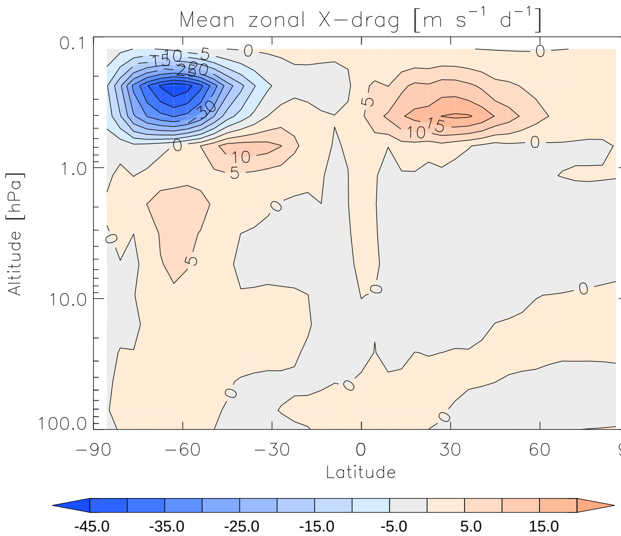


Data assimilation cycle with an optimal forcing.

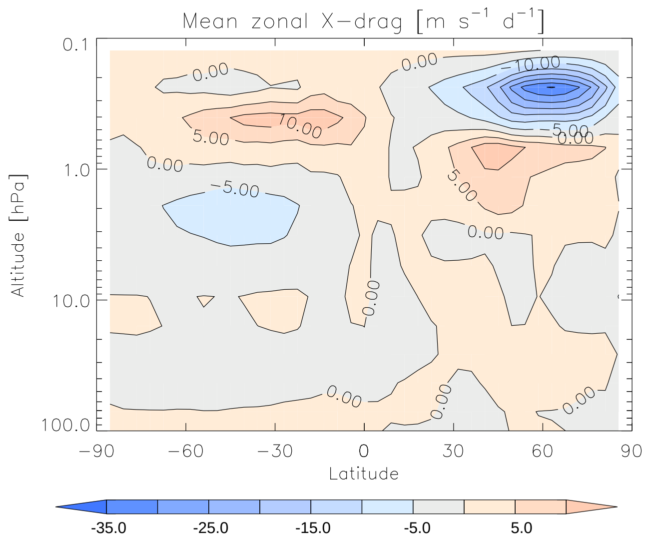
Actual estimations: Met Office analysis

Cost function: potential vorticity and pseudo-density (function of temperature only) are used as observed variables which are taken from MO analyses.

Control space: Curl of forcing only.



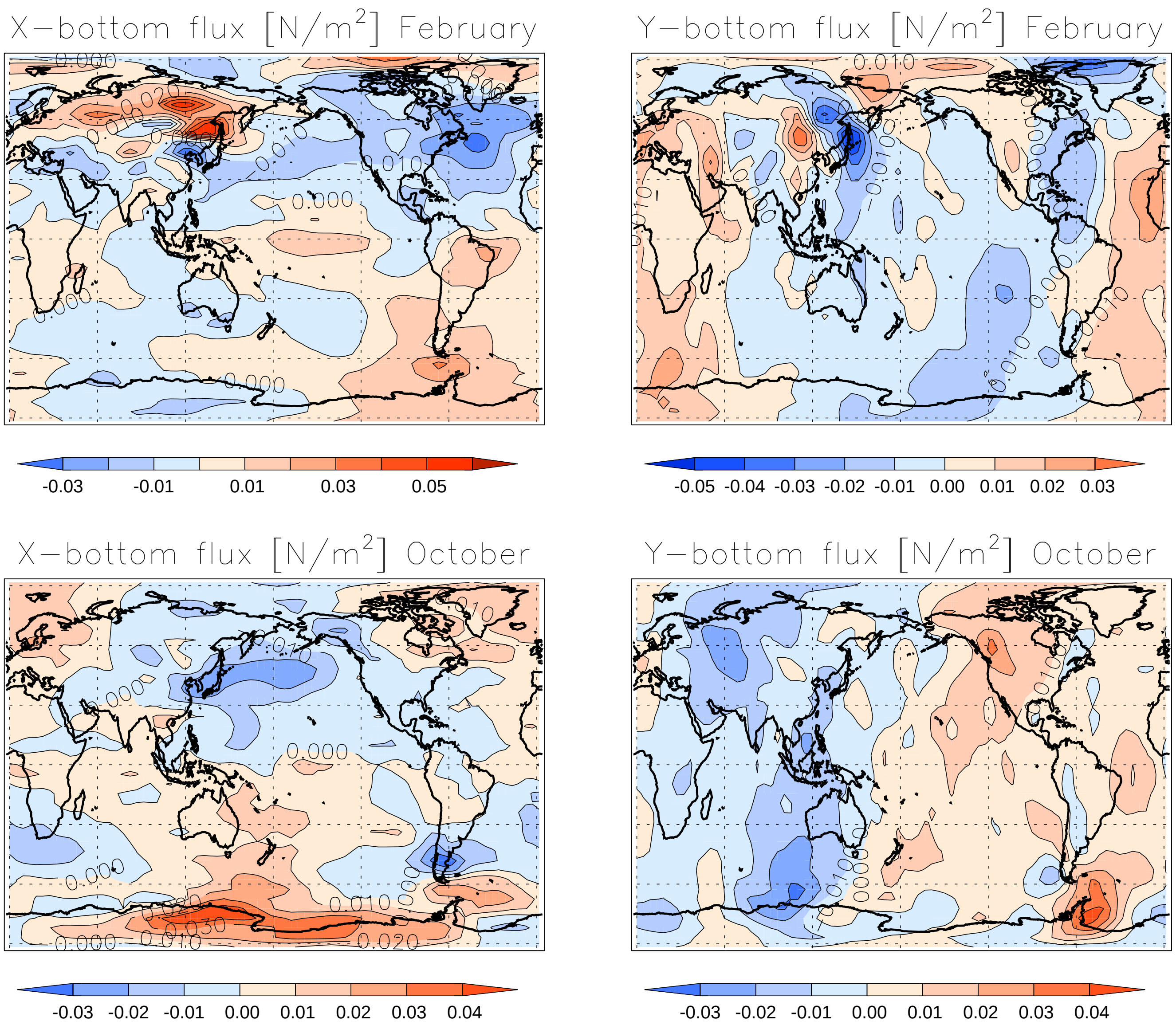
July



December

Estimated zonal mean monthly averaged zonal forcing.

Bottom momentum flux: Sources?



Integrating forcing and neglecting the top momentum flux:

$$F_b = \int_{\theta_b}^{\theta_t} \sigma X_x \mathrm{d}\theta.$$

For February (upper panels) and October (Lower panels)

Pulido and Thuburn (2008).

A further step: Parameter estimation

The estimated missing forcing should be reproduced by GW schemes.

Can GW schemes with optimum parameters reproduce the estimated missing forcing?

$==>$ **Offline estimation: We want to match the “observed” missing field by a GW scheme with some unknonwn parameters.**

Scinocca GW scheme

Scinocca (2002) scheme assumes the launch EP momentum flux spectrum is given by

$$E(c, z_l) = \frac{4E*}{\pi c_*^2} c \left[1 + \left(\frac{c}{c_*} \right)^4 \right]^{-1}$$

$c_* \equiv \frac{N_l \lambda_*}{2\pi}$ is the characteristic phase speed and E_* the total momentum flux.

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

$$E_s(c, z) = \frac{S_* E_*}{c_*^2} \frac{\rho(z) N_l}{\rho_l N(z)} \frac{[c - u(z)]^2}{c}$$

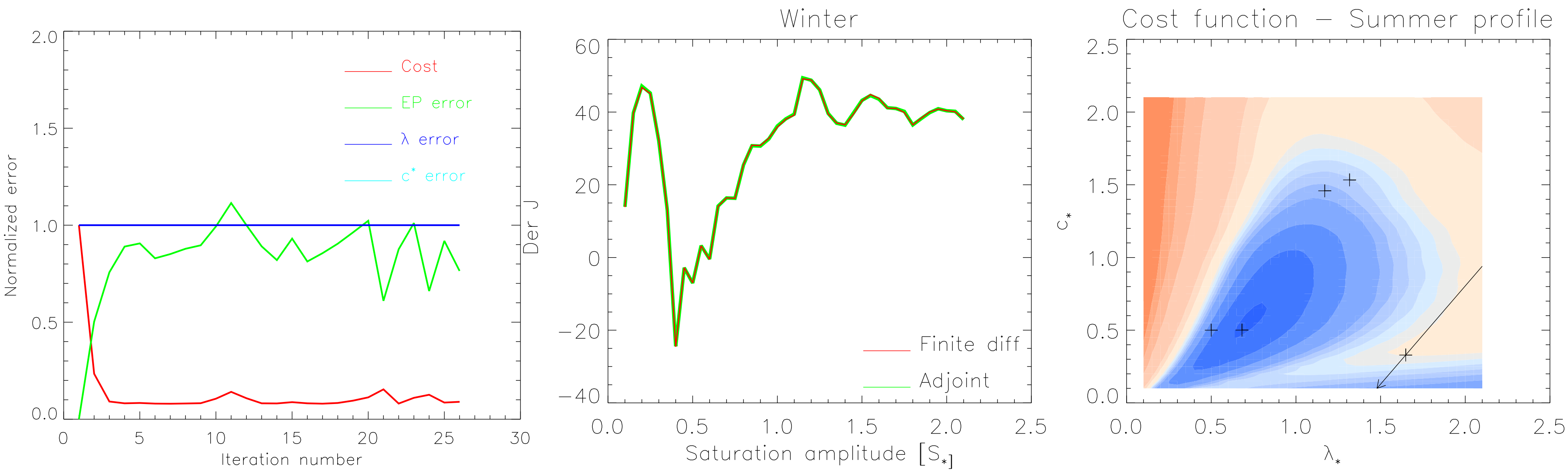
The momentum flux that is eliminated and the forcing are given by

$$E_T(z) = E_* - \int_0^{c_c} [E(c, z_l) - E_s(c, z)] \,\mathrm{d}c \qquad X = \rho^{-1} \partial_z E_T.$$

We consider as free parameters: (E_*, λ_*, S_*). The launch height z_l is considered fixed.

Optimum parameters: Variational data assimilation

The cost function is defined as: $J = (\mathbf{x} - \mathbf{y})^T \mathrm{R}^{-1} (\mathbf{x} - \mathbf{y})$ where $\mathbf{y}$ is the observed GWD profile and $\mathbf{x} = X(E_*, \lambda_*, S_*)$ is the forcing resulting from the GW scheme.

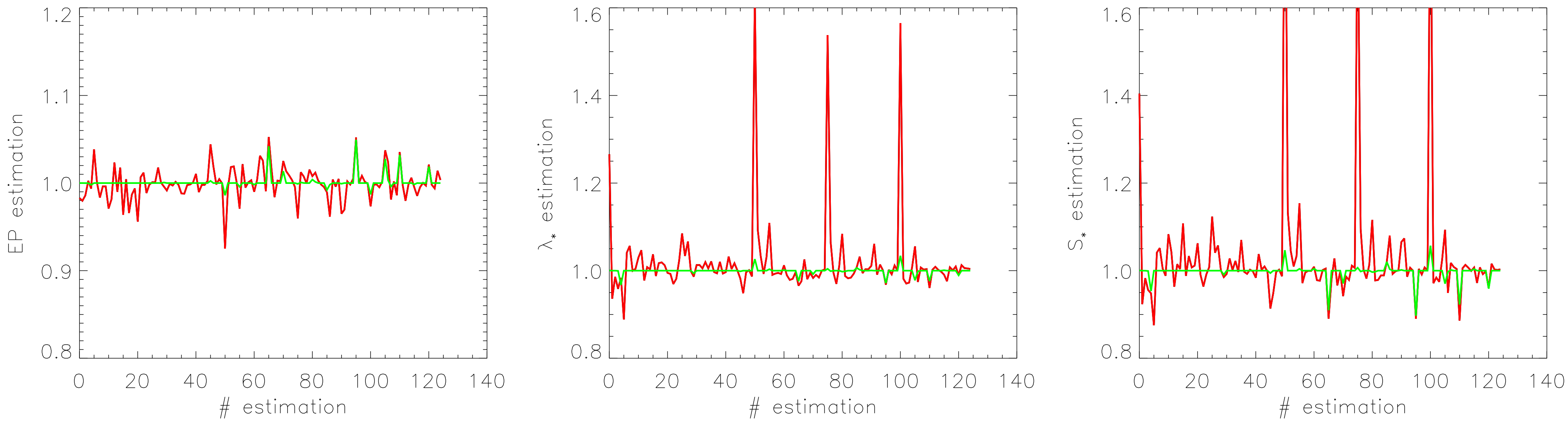


We applied the variational data assimilation technique similar to the forcing estimation technique. The minimization is also performed by a conjugate gradient method. The adjoint model of the scheme was developed.

Optimum parameters: Genetic algorithm

A genetic algorithm developed in NCAR by Charbonneau and Knapp (1995) is used to minimize the cost function.

- The minimization is perfomed in a constrained domain.
- We set the number of individuals in a population to 100 and the number of generations to 200.

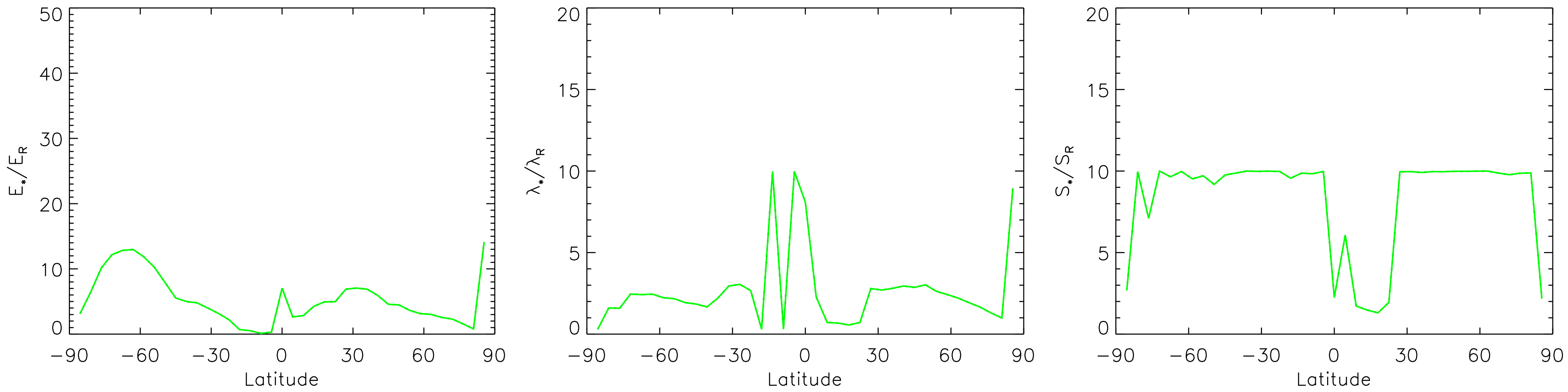


Experiments with different true parameters. Genetic algorithm (red) and hybrid genetic-variational algorithm (green).

Estimated parameters

Zonal wind and temperature is taken from Met Office analysis.

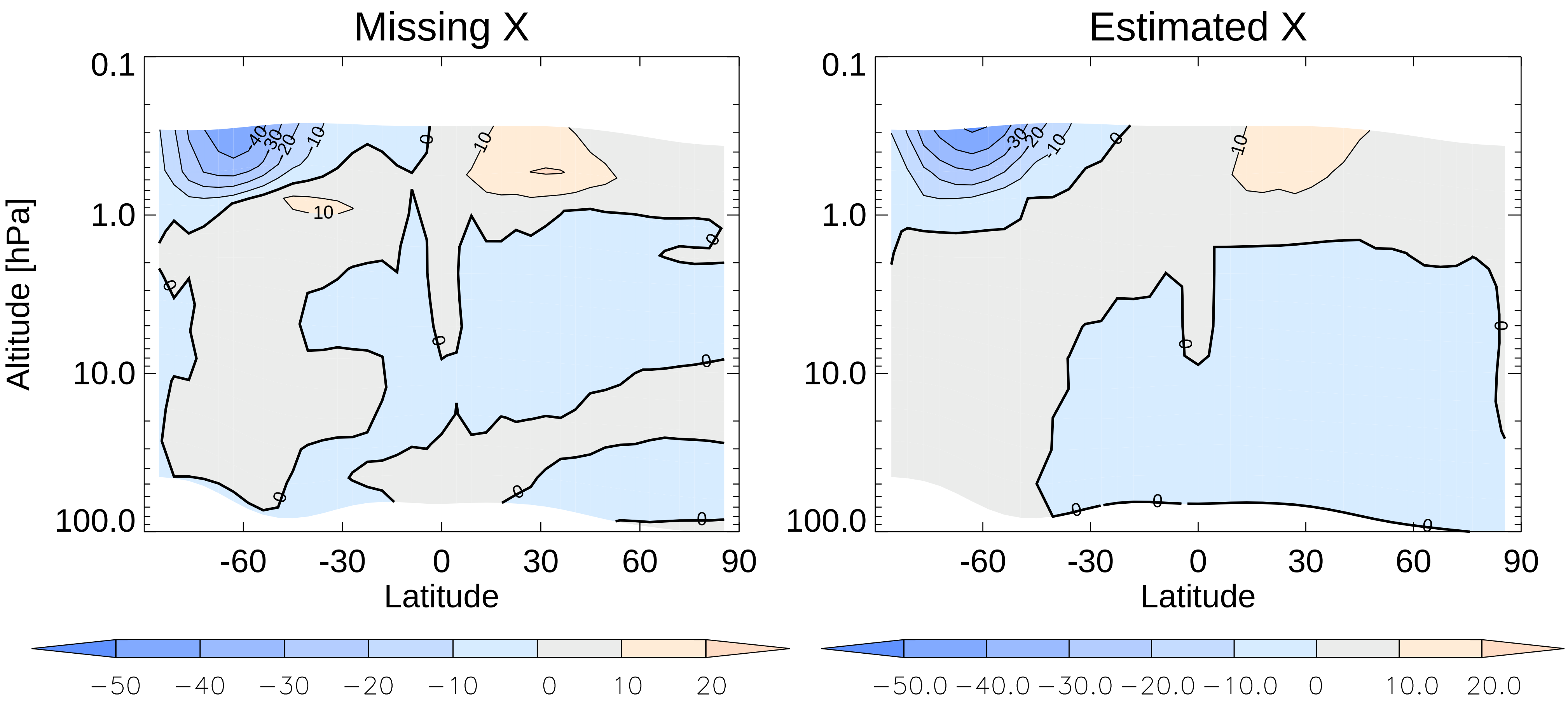
The GWD field estimated with the ASDE-4DVar technique (Pulido and Thuburn, JC 2008) for July 2002 is used as observational forcing profile $\mathbf{y}$.



Parameters E_* (left) λ_* (middle) and S_* (right) estimated for Met Office analysis in July 2002. Pulido et al. QJ 2012.

Parameter λ_* appears to agree in midlatitudes with measurements.

Estimated and optimal forcing



Missing forcing (momentum flux divergence) from observations and the estimated forcing using GW Scinocca scheme with optimum parameters (right panel).

4DVar works really well to estimate GW forcing, however it is model dependent. If we want to estimate the missing forcing in other models, full adjoint models of each model has to be developed.

Ensemble-based data assimilation: Kalman filtering

This is a model independent technique so that it could be very useful for an intercomparison project of the "missing forcing" in different GCMs.

Estimation of parameters in a subgrid orographic scheme

Offline optimization of the subgrid orographic scheme (Lott 1998, operational in ECMWF, LMD-Z).

In the twin experiments, ensemble Kalman filter fails to converge towards the known true parameters. The parameter error covariance is unknown.

New proposed technique: EnKF + Maximum likelihood error covariance estimation. Tandeo, Pulido and Lott (2012) in preparation.

The state, model parameters, is governed by a Gaussian random walk which is given by

$$\mathbf{x}^t(t_k) = \mathbf{x}^t(t_{k-1}) + \boldsymbol{\eta}_k,$$

Estimation of parameters in a subgrid orographic scheme

The observation eq is

$$\mathbf{y}_k^o = \mathcal{H}_k\left(\mathbf{x}^t(t_k)\right) + \boldsymbol{\epsilon}_k.$$

where the observation operator $\mathcal{H}_k$ is the nonlinear function defined by

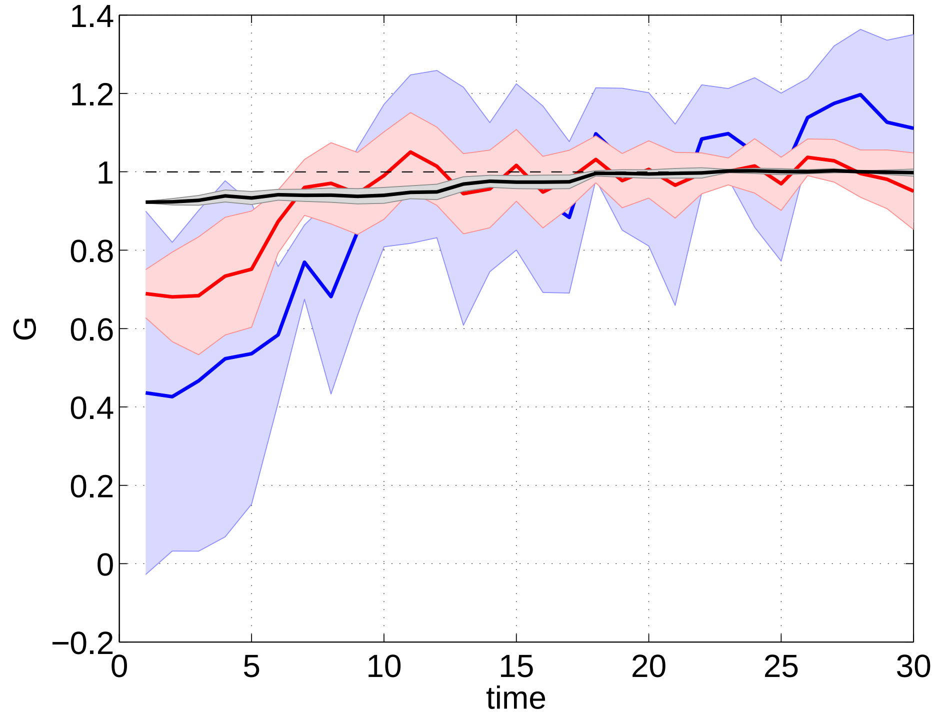
$$\mathcal{H}_k\left(\mathbf{x}^t(t_k)\right) = \mathcal{F}\left(\mathcal{G}\left(\mathbf{x}^t(t_k)\right), \mathbf{Z}_k\right).$$

$\mathcal{F}$ is the orographic scheme. $\mathcal{G}$ maps the parameters to $[-\infty, \infty]$. $\mathbf{Z}_k$ are the forcing fields that change with time (e.g. winds).

EnKF $-->$ Implementation of Pham (2001).

Maximum likelihood error covariance estimation $-->$ Expectation-Maximization algorithm.

Estimation of parameters in a subgrid orographic scheme

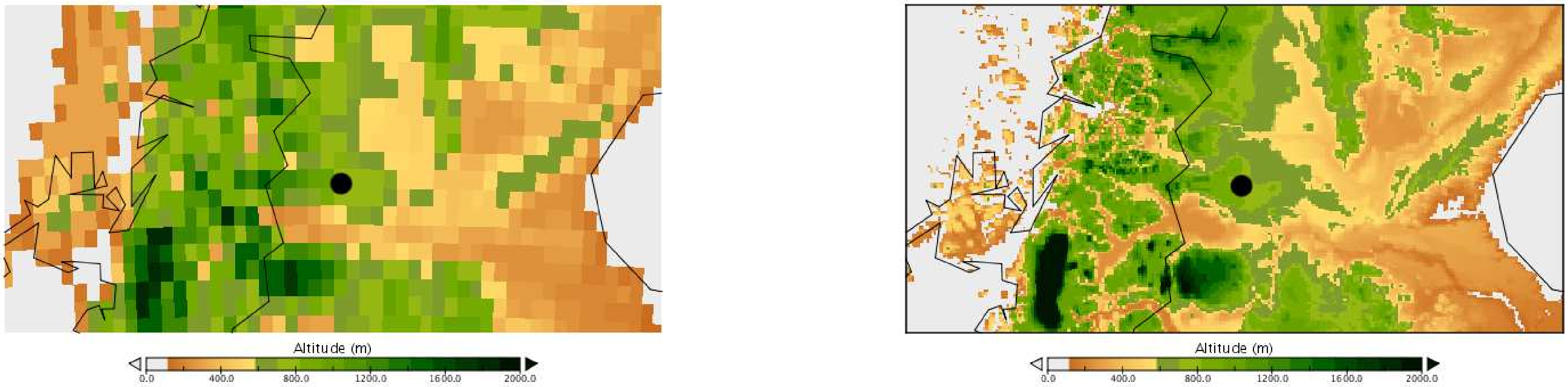


Twin experiment for an offline estimation. Blue (EM Iteration 1), Red (EM it=10) Black (EM it=50).

Should parameters be changed when model resolution is changed?

The resolution of models is often being changed when computer power increases.

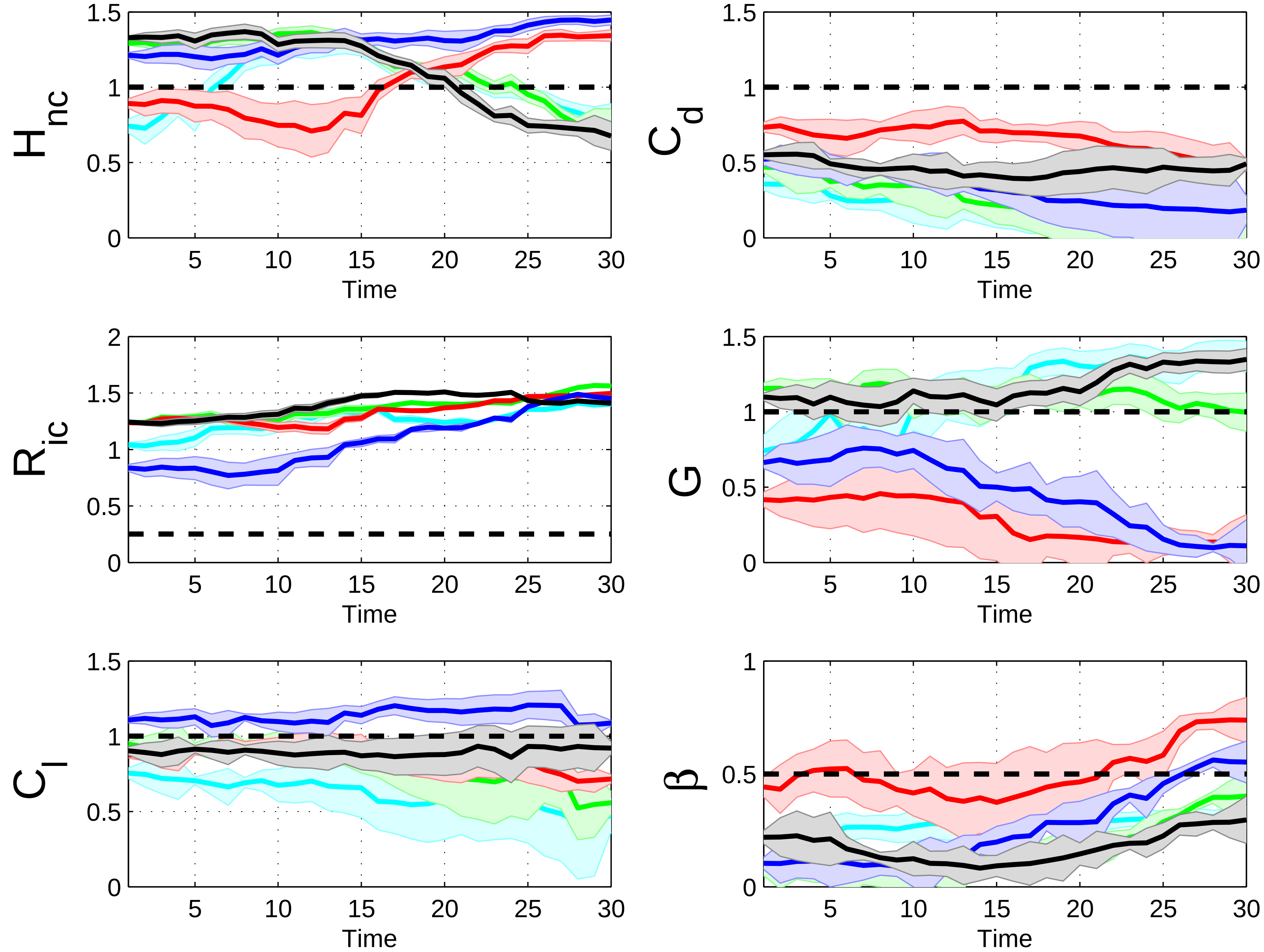
Are the “optimal” parameters in one model configuration still optimal for other configuration?



The orographic scheme, $\mathcal{F}$, changes with model resolution changes, i.e. the representation of mountain-ridge orientation, anisotropy and elevation of the subgrid orography change.

Assuming that the “observations”, i.e. small-scale momentum flux divergence, remain the same when the resolution is changed; the EnKF-EM technique can be used to determine the optimal parameters in the higher resolution.

Parameter optimization when model resolution is increased



Estimation of the parameters using different initial seeds for 25 EM iterations.

Conclusions

Estimating the source of missing momentum:

- **Variational data assimilation may be used to estimate the missing force for a given climate model.**
- **The 4DVar technique appears to give robust results with very good convergence.**
- **It is able to estimate the 'launch' momentum flux**

Estimating parameters of GW schemes:

- **Variational data assimilation needs a good first guess for estimating parameters of physical parameterizations, since the sensitivity far from the minimum is usually nonlinear.**
- **A genetic algorithm works well for this low dimension problem.**
- **Ensemble Kalman filtering also needs a good a priori knowledge to converge. EM may be used to estimate the statistical parameters.**