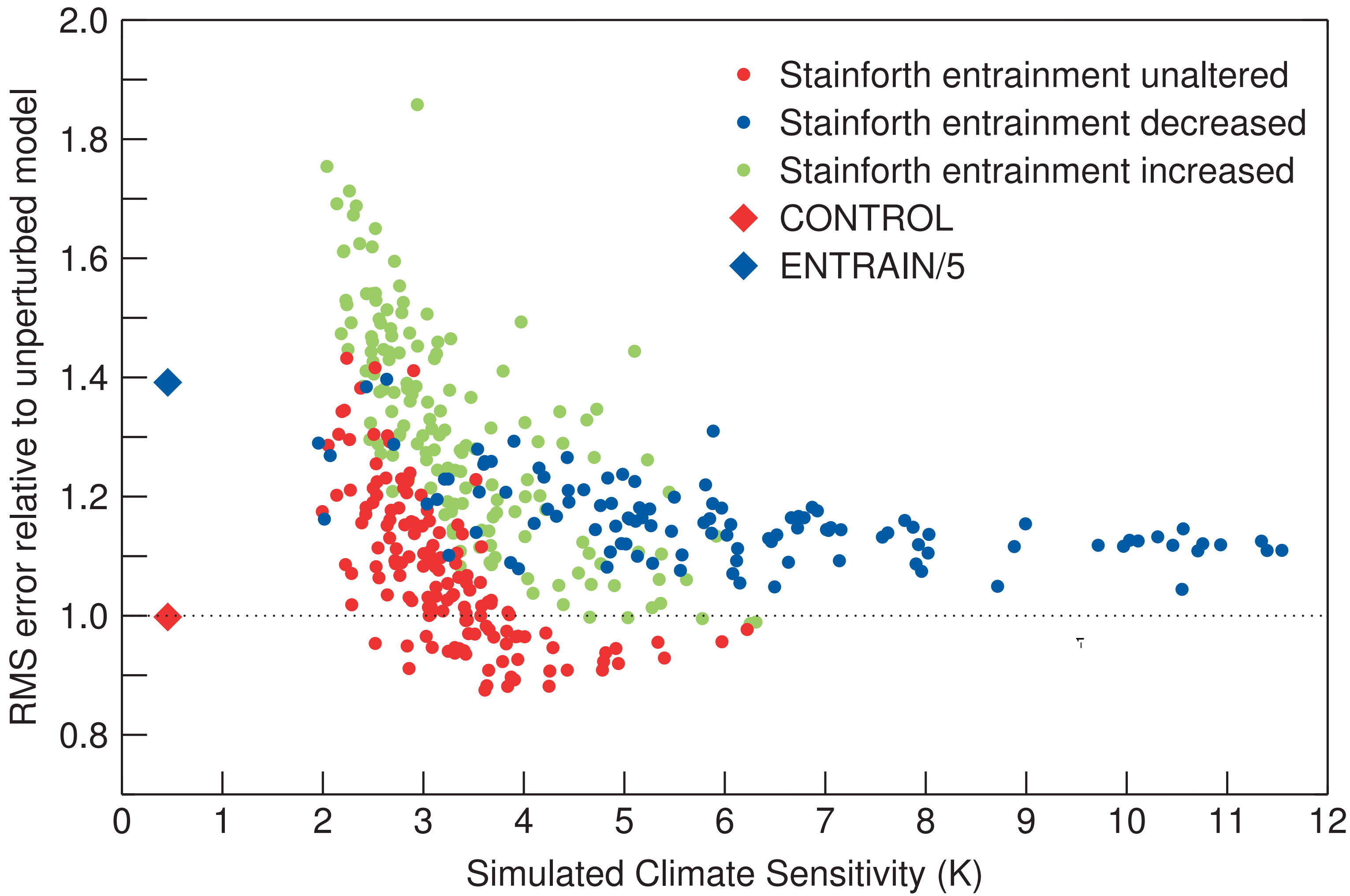
Using data assimilation to improve climate models: Model Error and parameter estimation

Manuel Pulido

Department of Physics, Universidad Nacional del Nordeste, Corrientes, Argentina

Acknowledgments: Juan Ruiz, Guillermo Scheffler, Ted Shepherd, Pierre Tandeo and John Thuburn

Motivation: Parameter estimation



From Stainforth et al, Nature, 2005 (climateprediction.net)

High sensitivity to certain sets of ’realistic parameters’.

Motivation II: Source of errors in climate models

GCMs do not resolve all the motion scales, because of this, they are not able to capture the momentum forcing that is produced by small-scale waves.

There is no simple way to infer this systematic momentum deficit (missing forcing) in a GCM.

If one computes the difference between the true state and the model state, the result is a combination of different sources of errors, recent and past, which once they are generated are advected and interact with other parts of the system.

Is there an objetive way to find the source of model error, i.e., the exact time and position where the momentum errors are produced?

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}_f + M(\mathbf{x}_f) = 0 \quad \rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}_T + M(\mathbf{x}_T) = \mathbf{X}$$

x_f **model state,** M **forecast model,** x_T **true state,** $\mathbf{X}$ **missing unknown forcing.**

Using data assimilation to diagnose 'missing forcing'

4DVar can be used to estimate unknown parameters of a model $\longrightarrow$ the missing forcing.

Assume 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 from t_0 to t_i

$$\mathbf{x}_i = M(\mathbf{x}_0, \boldsymbol{X}, t_i)$$

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, 2005).

The Adjoint Model

The gradient of the cost function is calculated with the adjoint model (Forward model= Reading MAGCM).
The missing forcing is assumed constant within an assimilation window length.

1. Tangent linear of the dynamical model:

$$\begin{bmatrix} \delta x^{n+1} \\ \delta X^{n+1} \end{bmatrix} = \begin{bmatrix} M'(x^n) & I \\ 0 & I \end{bmatrix} \begin{bmatrix} \delta x^n \\ \delta X^n \end{bmatrix}$$

2. Adjoint from the tangent linear model

$$\begin{bmatrix} \delta \hat{x}^{n-1} \\ \delta \hat{X}^{n-1} \end{bmatrix} = \begin{bmatrix} M'^T(x^n) & 0 \\ I & I \end{bmatrix} \begin{bmatrix} \delta \hat{x}^n \\ \delta \hat{X}^n \end{bmatrix}$$

3. Finally, the gradient of the cost function is given by

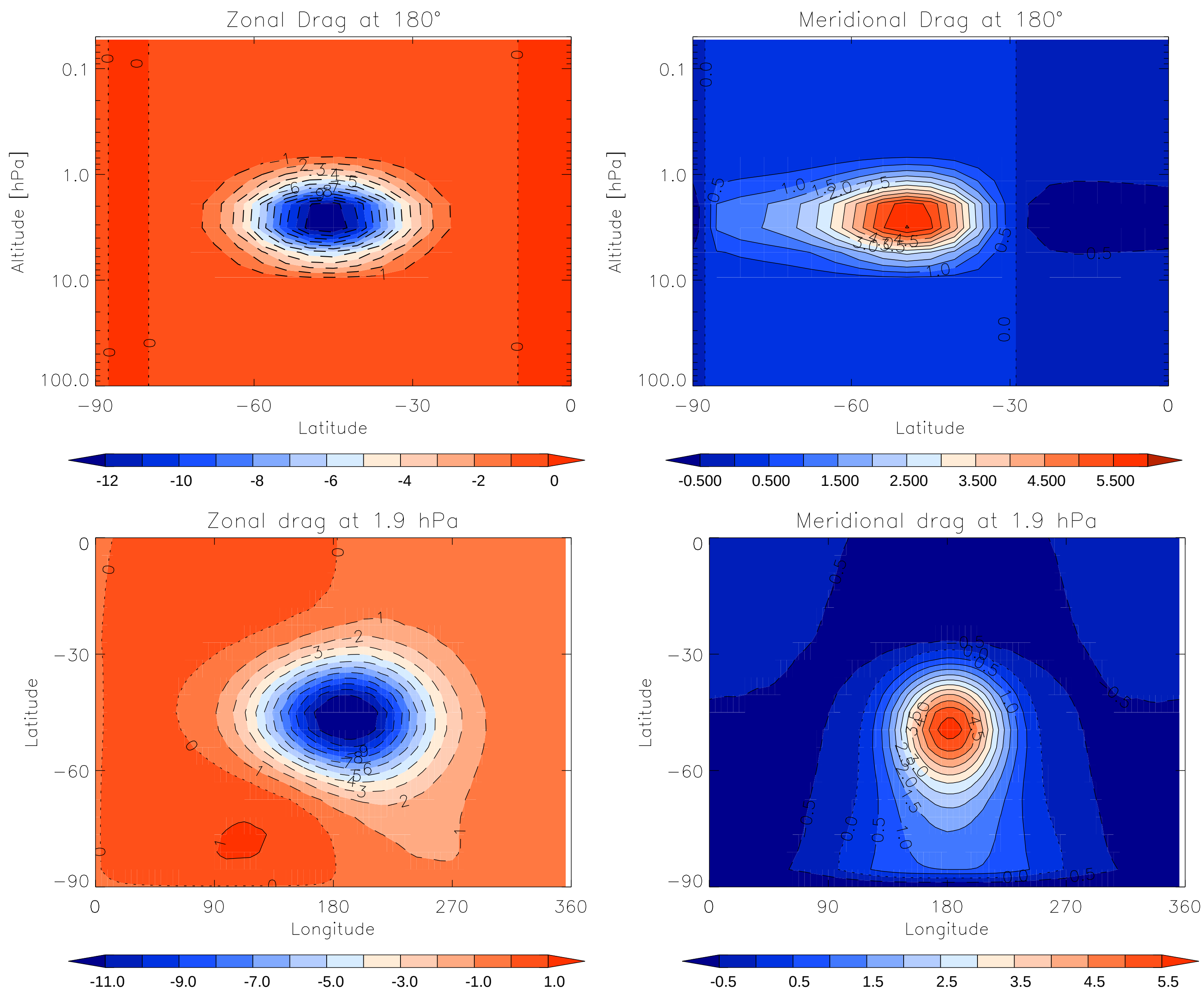
$$\begin{bmatrix} \frac{\partial J}{\partial x^0} \\ \frac{\partial J}{\partial X^0} \end{bmatrix} = \sum_{n=0}^{N-1} \begin{bmatrix} M'^T(x^n) & 0 \\ I & I \end{bmatrix} \cdots \begin{bmatrix} M'^T(x^{N-1}) & 0 \\ I & I \end{bmatrix} \begin{bmatrix} \frac{\partial J}{\partial x_N} \\ \frac{\partial J}{\partial X^N} \end{bmatrix}$$

Twin experiments

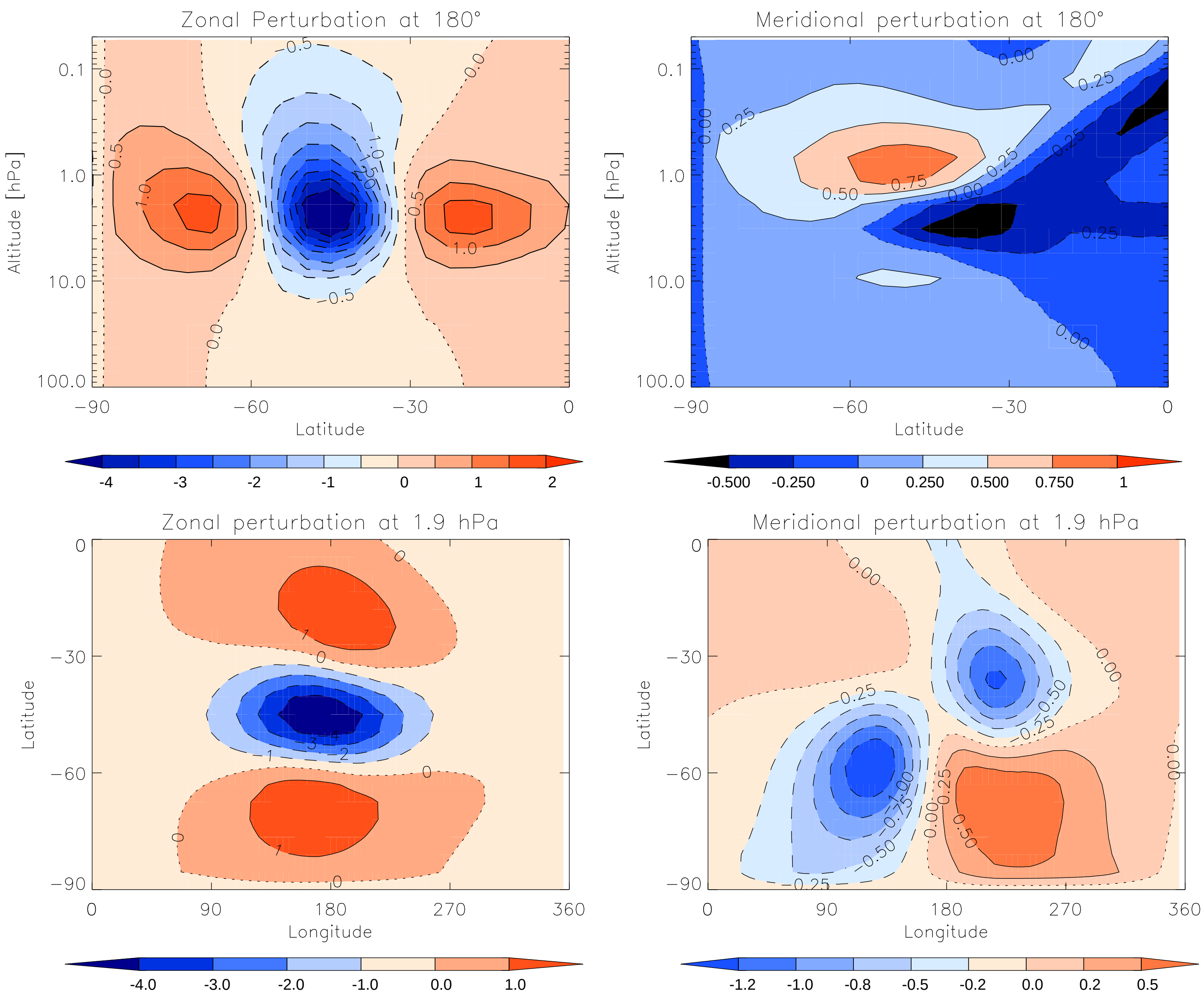
Experiment using Reading GCM:

- Gaussian forcing used as the prescribed forcing for the twin experiments.

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



Flow response. 'The observations'

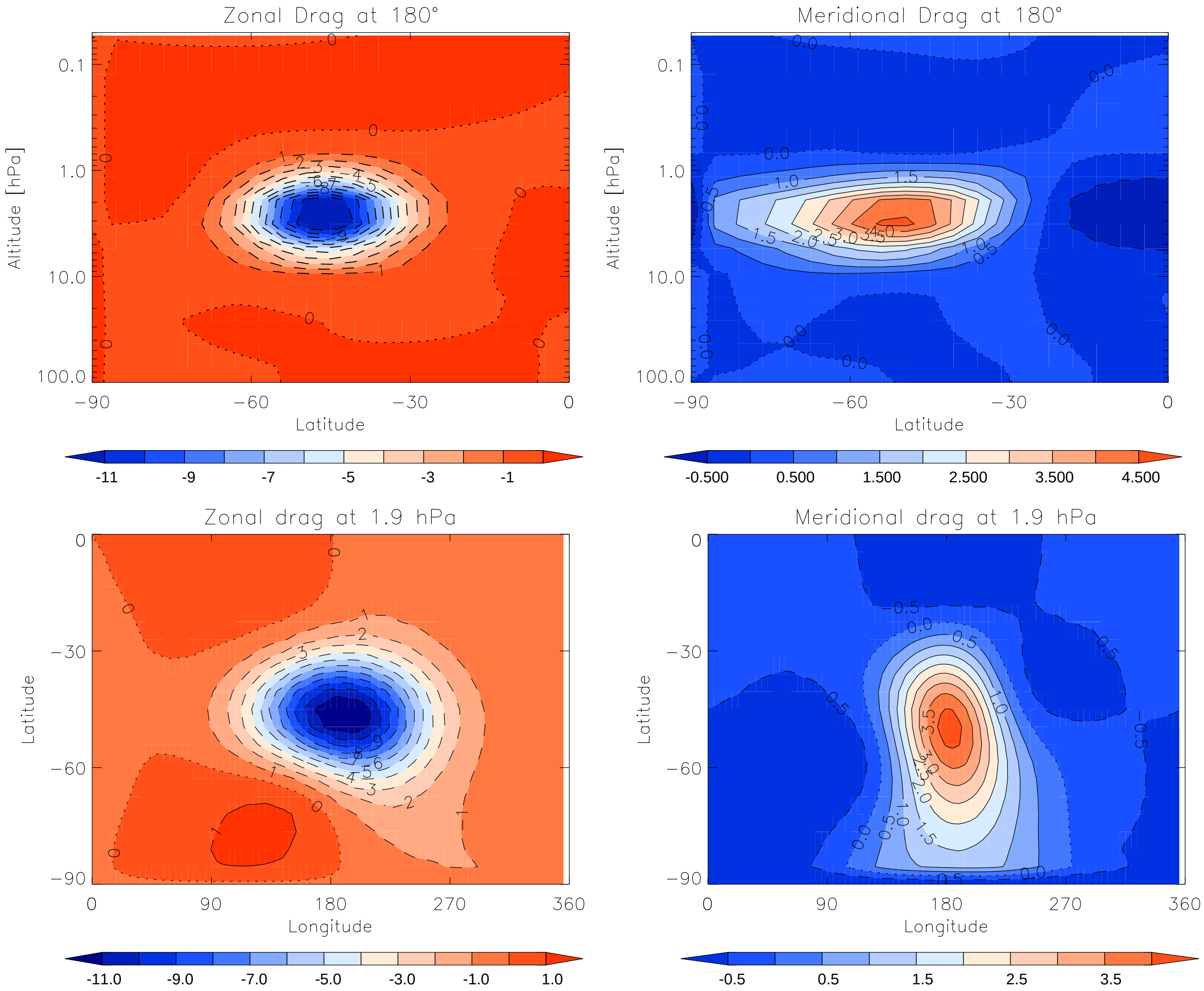


Flow response to the applied forcing at $t = 1$ day.

This could be interpreted as the model error: $\mathbf{X} = [\mathbf{u_F}(1\mathrm{d}) - \mathbf{u_{F=0}}(1\mathrm{d})]/1\mathrm{d}$

This is the effect of model error but not the source of model error.

Estimated missing forcing with 4DVar



Estimated forcing after 25 minimisation iterations without a priori information.

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 2005).

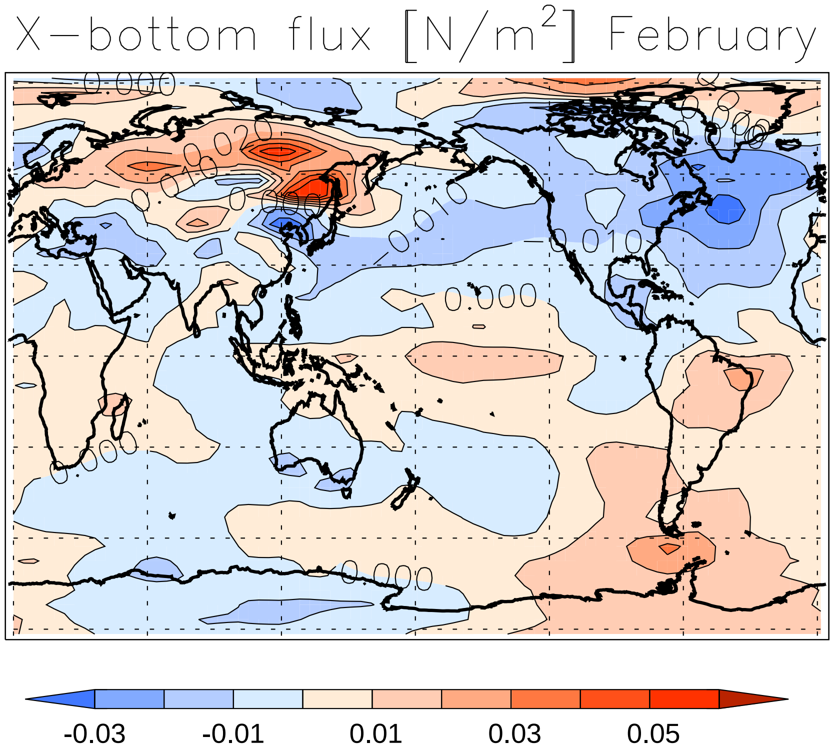
Missing momentum flux: Sources of model error?

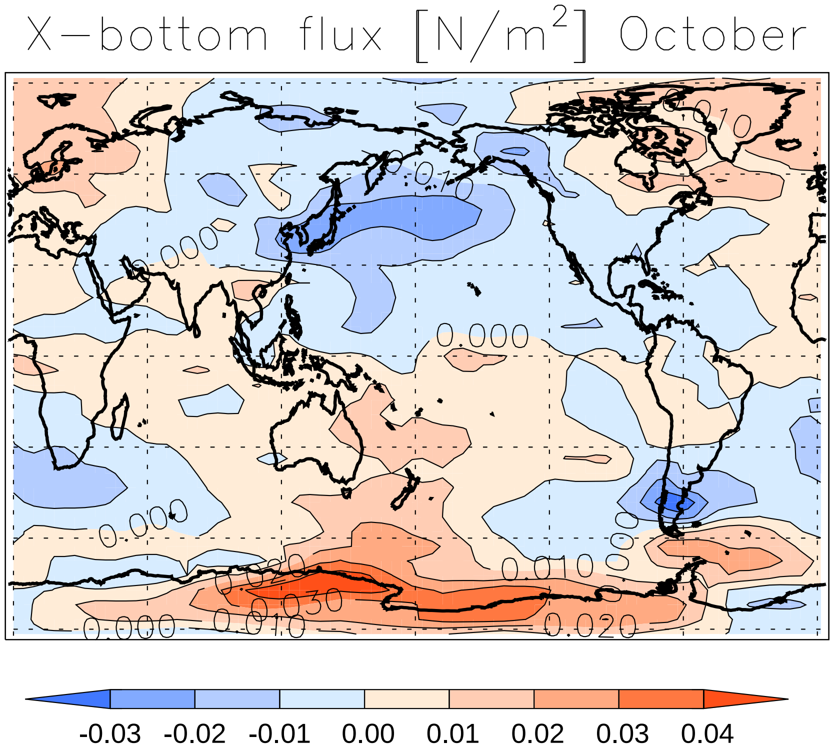
Real experiment: Observations from Met Office analyses.

Model: Reading MAGCM without GW parameterizations.

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

Control space: Curl of the forcing only.



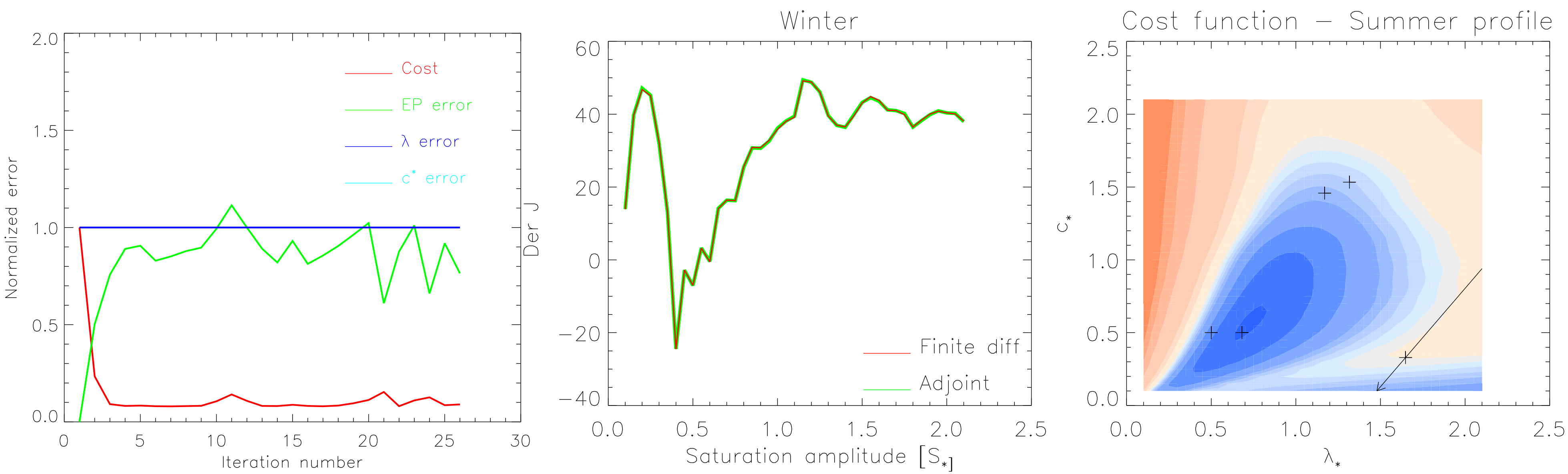


$F_b = \int_{\theta_b}^{\theta_t} \sigma X_x \mathrm{d}\theta$. Pulido and Thuburn, JC (2008).

A further step: Offline parameter estimation

Can GW parameterizations with optimum parameters reproduce the estimated missing forcing? We use the Scinocca (2002) parameterization implemented operationally in the Canadian GCM and the ECMWF model.

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 forcing profile and $\mathbf{x} = Sch(E_*, \lambda_*, S_*)$ is the one resulting from the GW scheme.

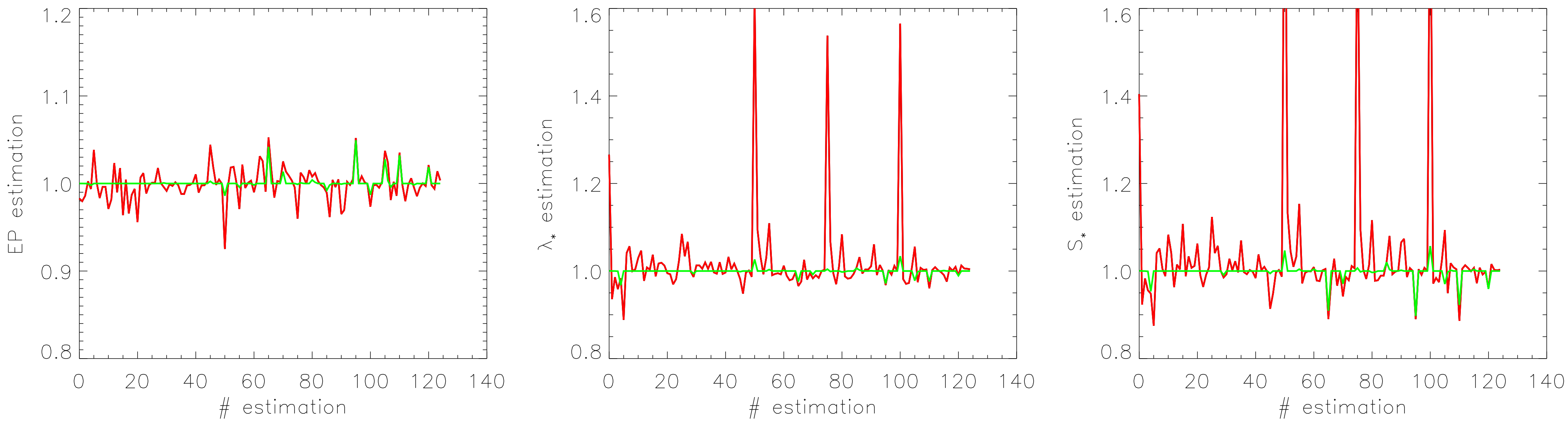


Parameterizations are highly non-linear and ill-conditioned NOT suitable for variational data assimilation.

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 (about 20000 parameterization evaluations).

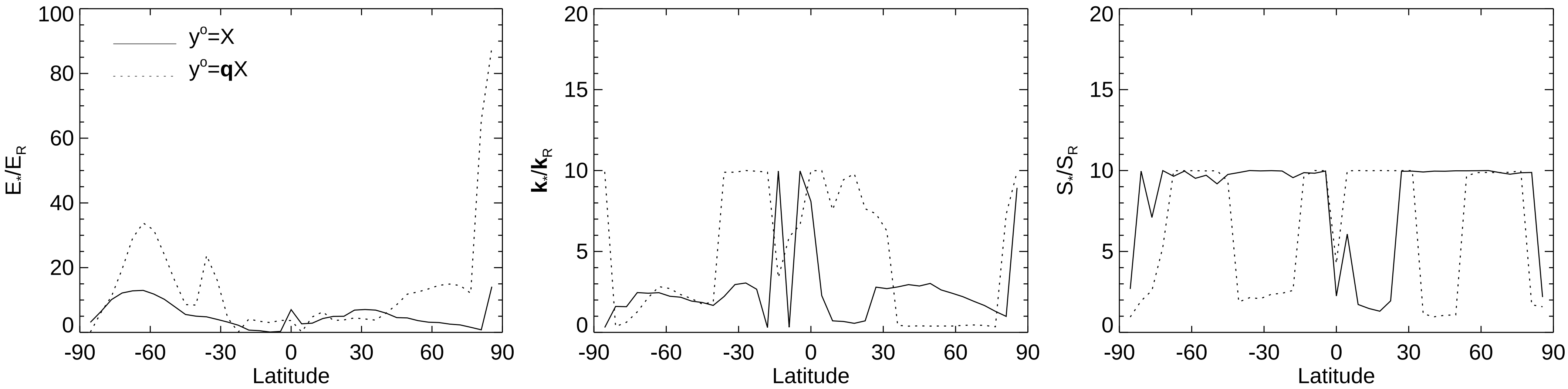


All the experiments converge toward the true parameters.

Estimated parameters

Zonal wind and temperature is taken from Met Office analysis.

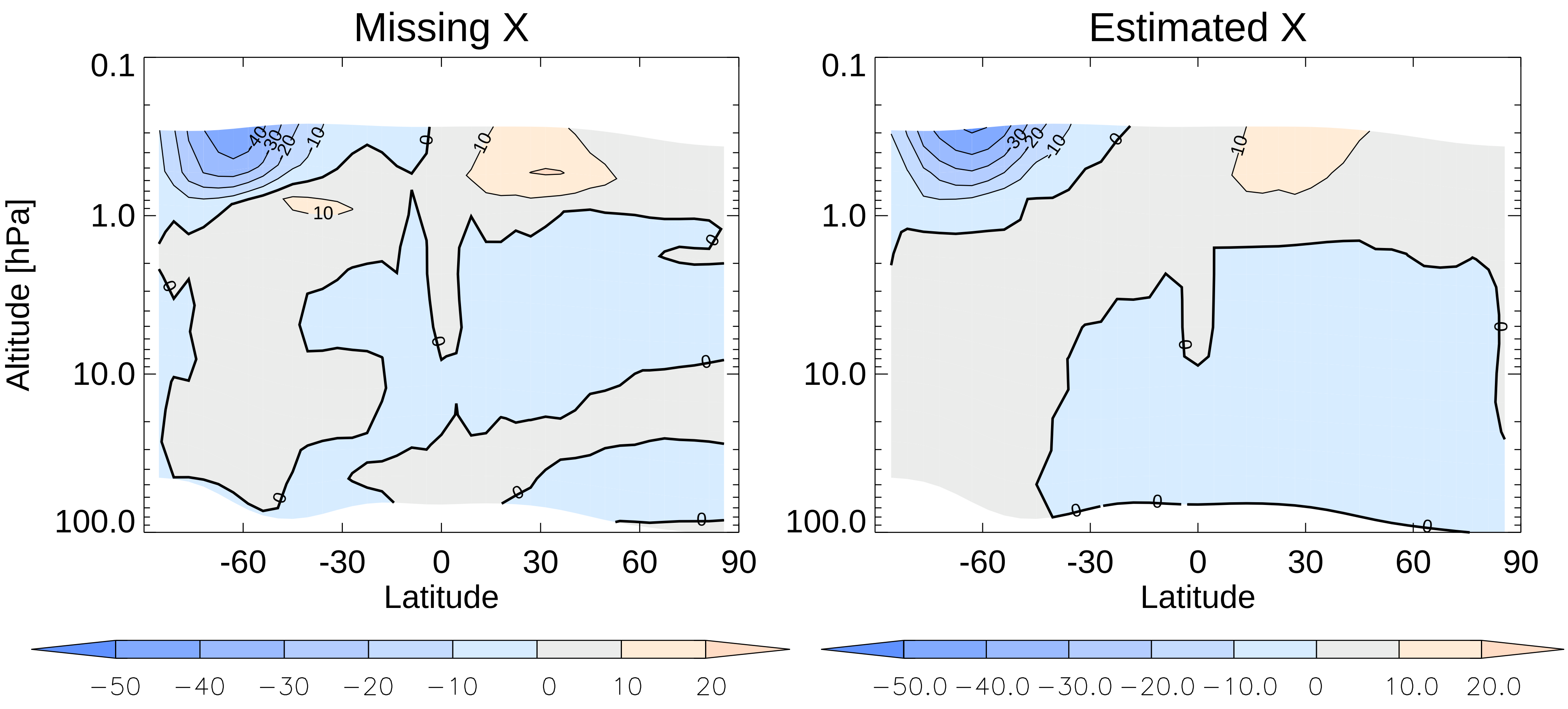
The missing forcing estimated with the ASDE-4DVar technique (Pulido and Thuburn, JC 2008) for July 2002 is used as the “observations” $\mathbf{y}$.



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

Parameters E_*, λ_* agree rather well with high-resolution mesurements.

Estimated and "parameterized" forcing

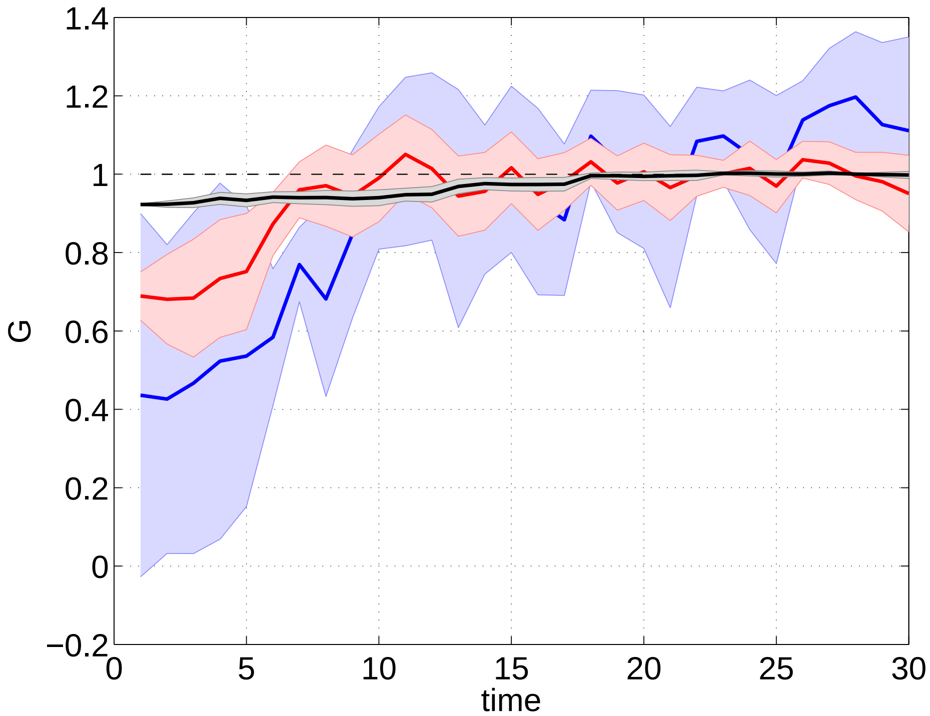


Missing forcing (momentum flux divergence) from observations and the estimated forcing using GW Scinocca scheme with optimum parameters (right panel).

Ensemble-based data assimilation: Parameter estimation

Optimization of the subgrid orographic parameterization (Lott 1998, operational in ECMWF, LMD-Z).

Technique: EnKF + Maximum likelihood error covariance.

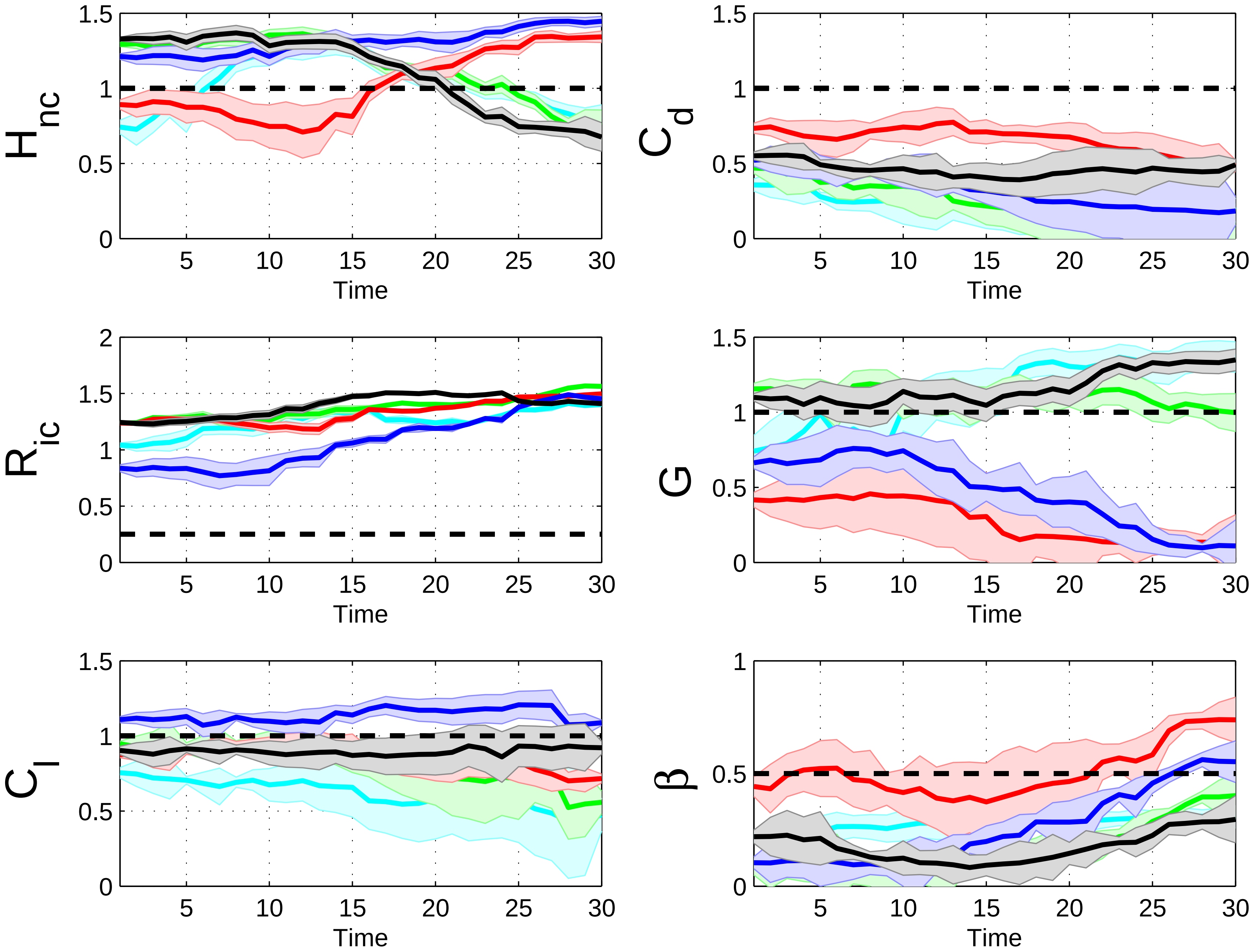


Twin experiment for an offline estimation. Blue (Iteration 1), Red (it=10) Black (it=50).

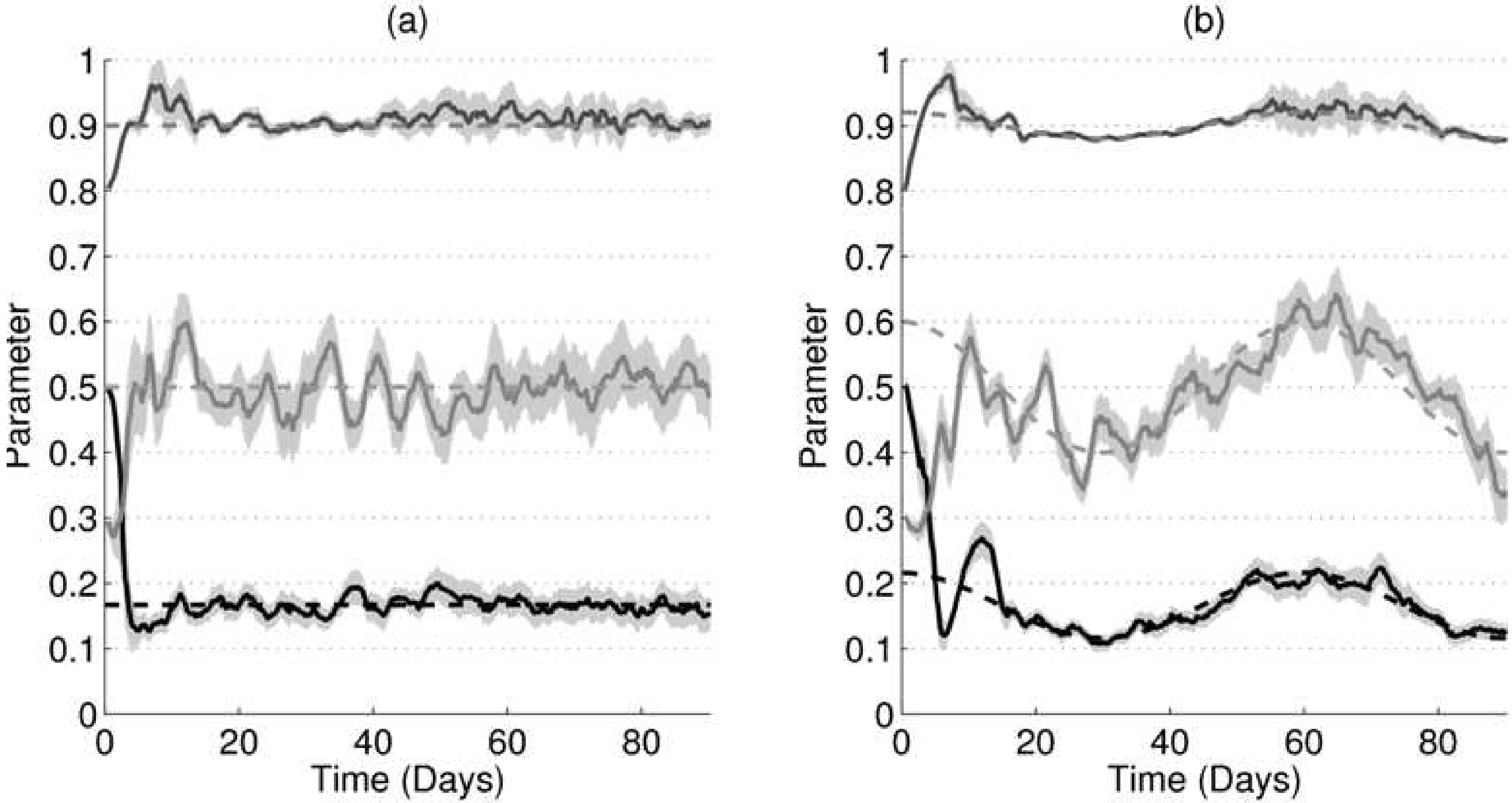
Note that model is time-independent but the forcing terms (u,v,T) change with time.

Tandeo and Pulido (2012) in preparation.

Should parameters be changed when model resolution is changed?



Convective parameters. Online estimation



Convective parameter estimation (RHBL, ENTMAX and TRCNV) using the Ensemble Transform Kalman Filter in the SPEEDY GCM (online estimation).

Ruiz, Pulido and Miyoshi 2012, submitted to JMSJ.

Conclusions

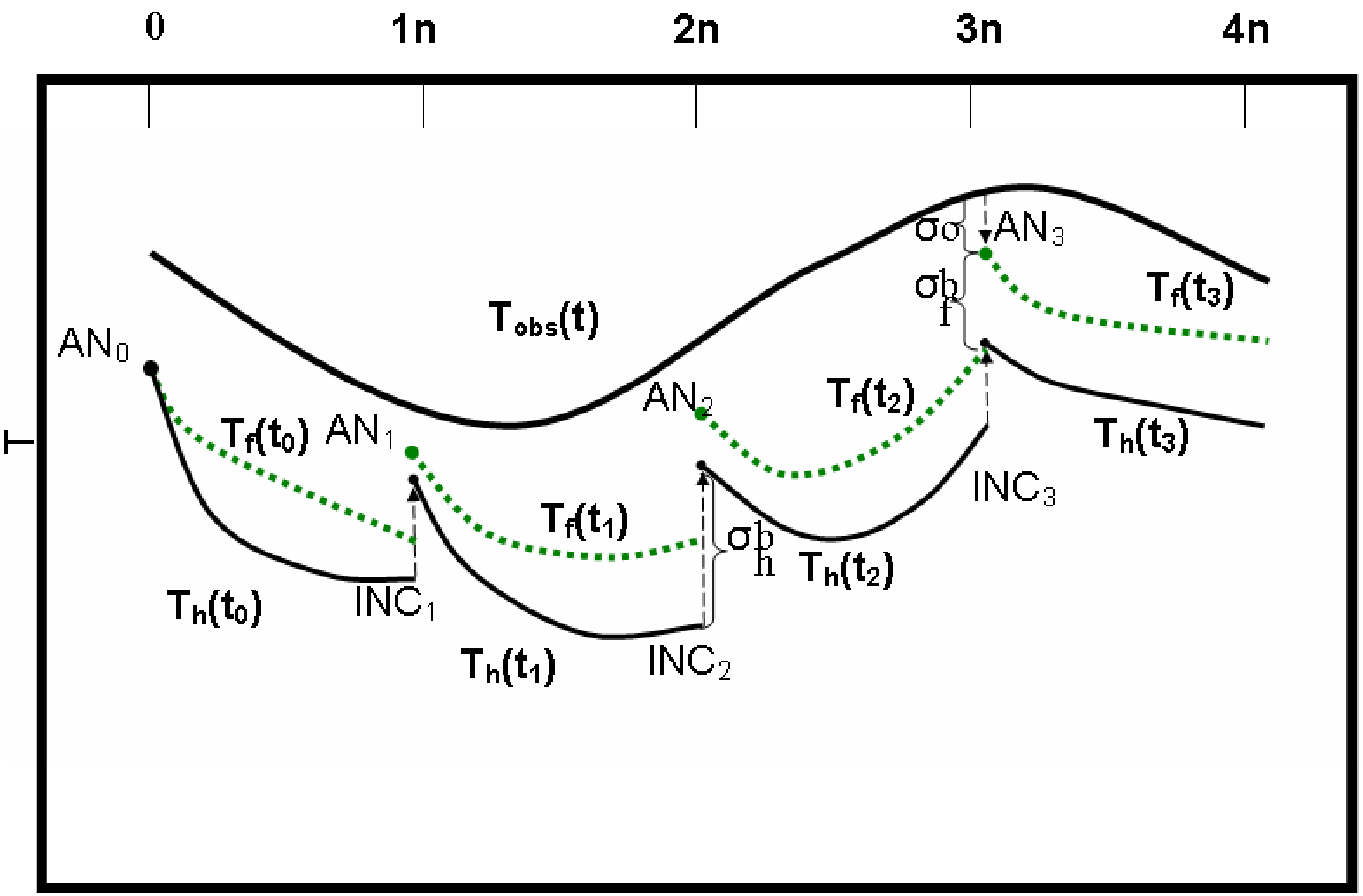
Estimating the sources of model errors:

- Variational data assimilation may be used to estimate the missing force.
- The 4DVar technique appears to give robust results with very good convergence.
- The information “missing forcing” is useful to improve parameterizations

Parameter estimation:

- Variational data assimilation may be not useful for estimating parameters of physical parameterizations, since the sensitivity is usually nonlinear.
- A genetic algorithm and Ensamble Kalman filtering + Maximum likelihood error covariance estimation works well for off-line estimations.
- Ensemble transform Kalman filtering works for on-line estimations (only evaluated with twin experiments)

Model error in a DA cycle



Adapted from Rodwell and Palmer QJ 2007. $T_h(t)$ **evolution of the "homogeneus" model,** $T_f(t)$ **evolution of the model with the "continous model error" forcing term (green line, Pulido and Thuburn QJ 2005).**