A model independent technique to estimate unresolved gravity wave drag: Evaluation in the Lorenz 96 model

Manuel Pulido (1)(2), Guillermo Scheffler (1) and Juan Ruiz (2)(3)

(1) Department of Physics, Universidad Nacional del Nordeste, Corrientes, Argentina

(2) CONICET, Argentina

(3) Universidad de Buenos Aires, Argentina

Outline

1. Missing forcing estimation with 4DVAR
2. Parameter estimation with 4DVAR and Genetic Algorithm
3. Impact of optimal parameters in the model (G. Scheffler’s work)
4. Missing forcing with ENKF
5. Orographic parameter estimation with ENKF

Motivation: Missing 'subgrid' forcing 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 low resolution observations 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.

High resolution observations can estimate momentum fluxes due to both resolved and unresolved waves.

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?

What is the missing forcing?

The model evolution without subgrid-scale effects is represented by:

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

where $\mathbf{x}_f$ **model state and** M **forecast model.**

The model evolution taking into account the missing forcing

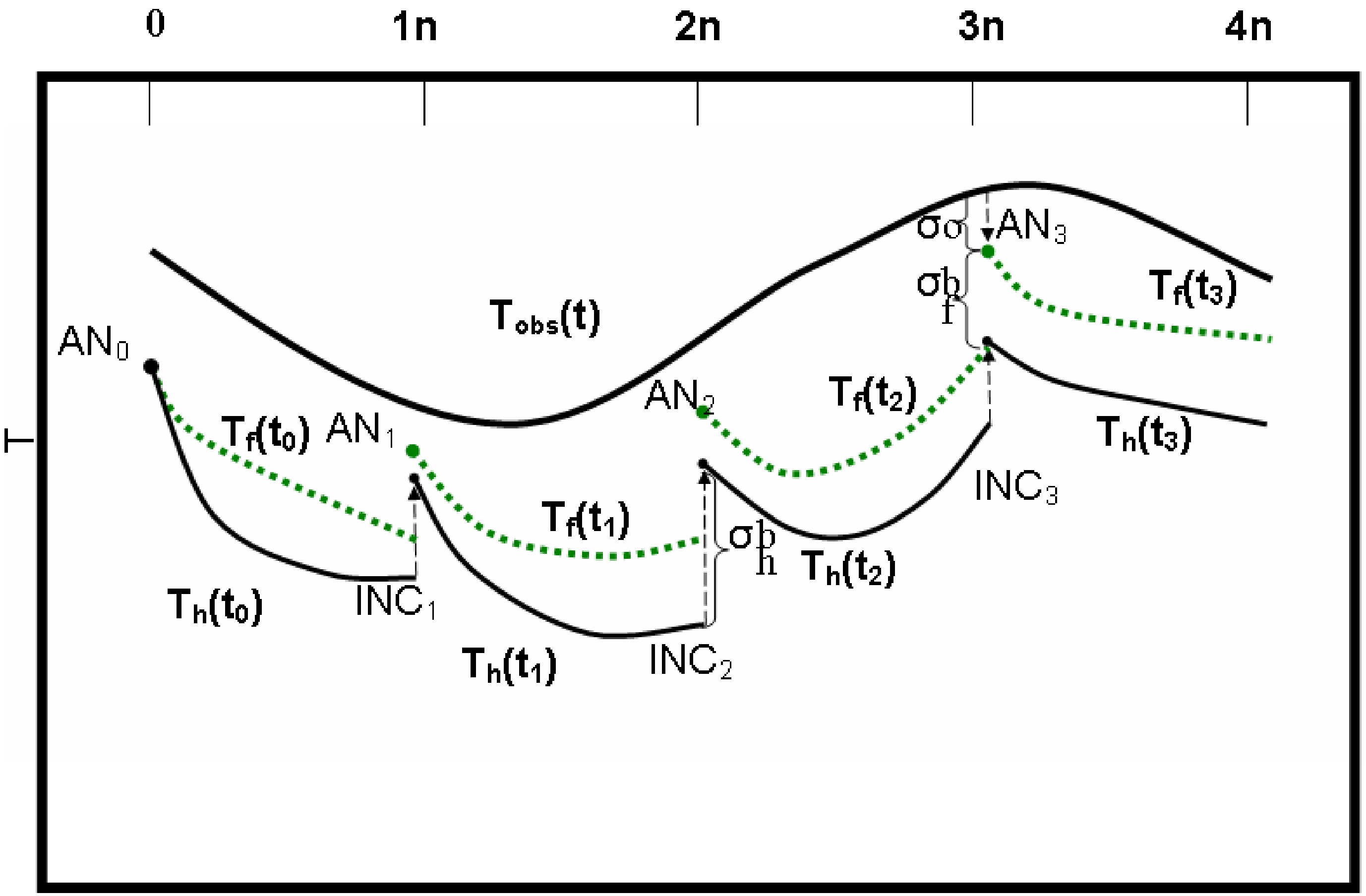
$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}_T + M(\mathbf{x}_T) = \mathbf{X}(t)$$

where $\mathbf{x}_T$ **true state and** $\mathbf{X}(t)$ **missing forcing.**

The missing forcing could be represented through a parameterization, ideally:

$$\mathbf{X}(t) = Parameterization(\mathbf{x}_T)$$

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).**

Continuous model error vs Differences at fixed times .

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 $J = J(\mathbf{x}_0, \boldsymbol{X})$

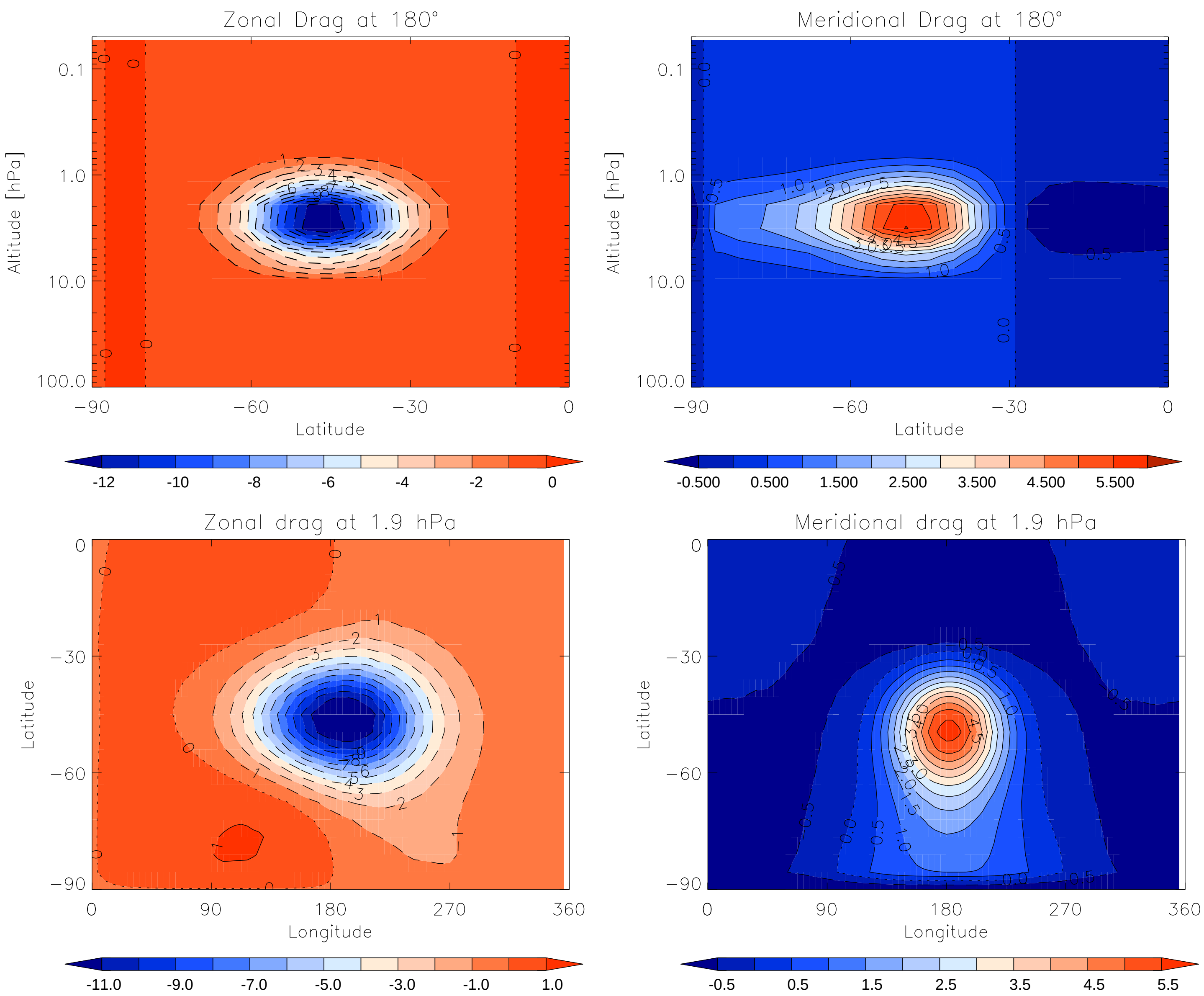
Therefore, if we know $\mathbf{x}_0$ the control space of the cost function is only the field $\mathbf{X}$. The minimum of the cost function gives the ’missing forcing’ (Pulido and Thuburn, 2005).

Twin experiments

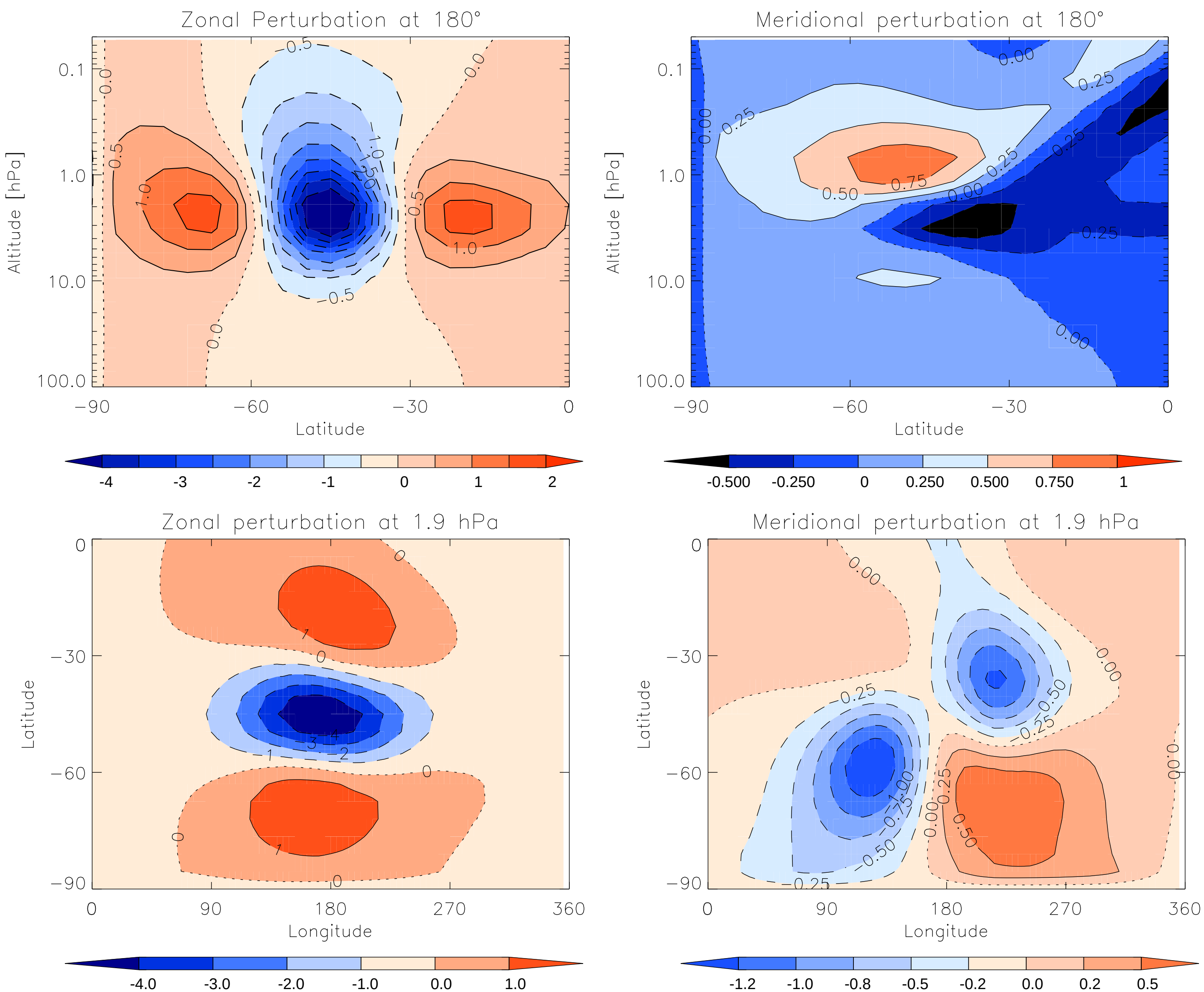
Experiment using Reading MAGCM:

- 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'

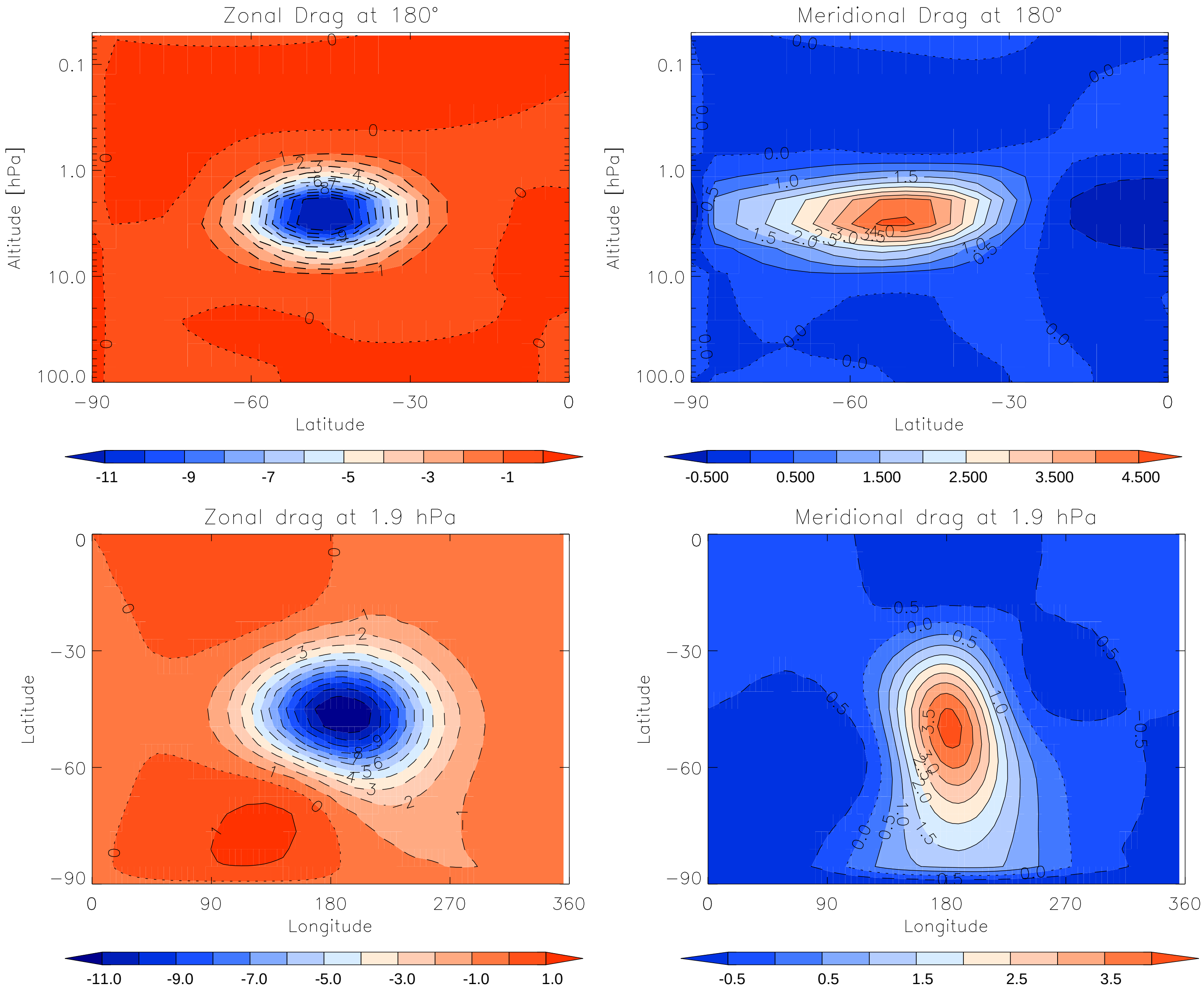


Flow response to the applied forcing at t =1 day.

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, i.e. the forcing.

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).

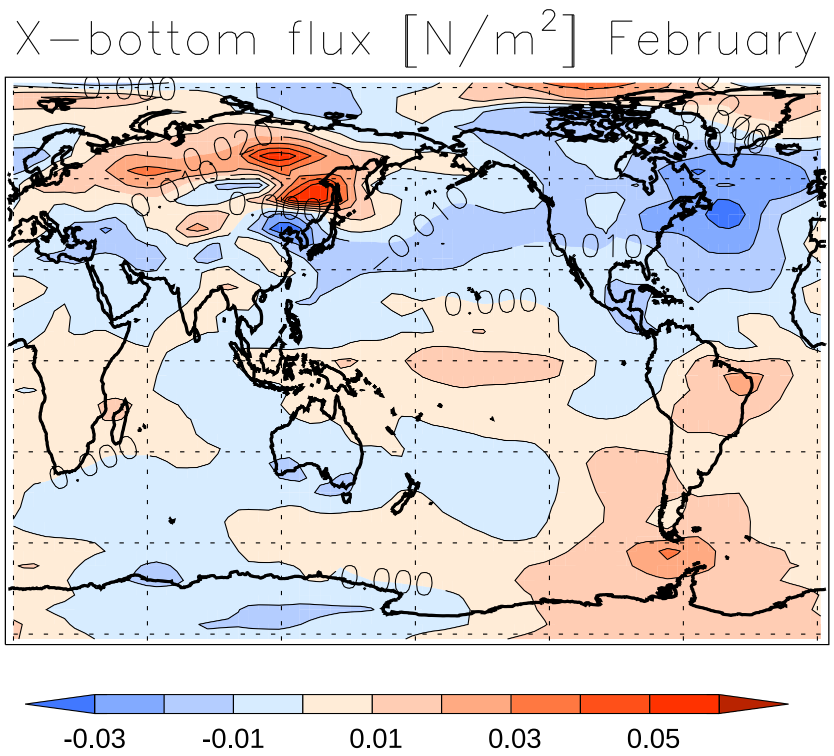
Missing momentum flux: Sources of model error?

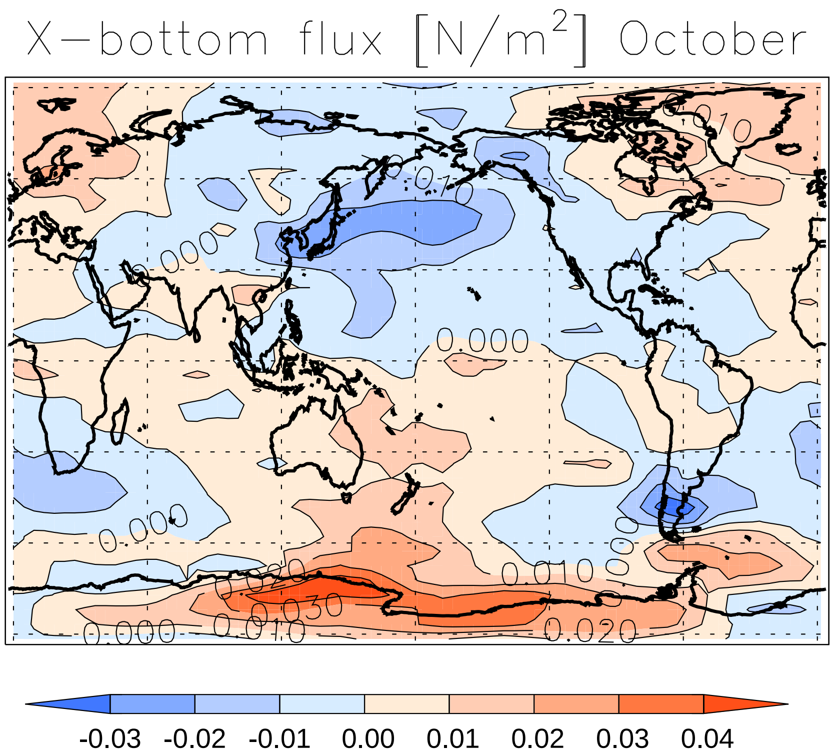
Real experiment: Observations from Met Office analyses.

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.



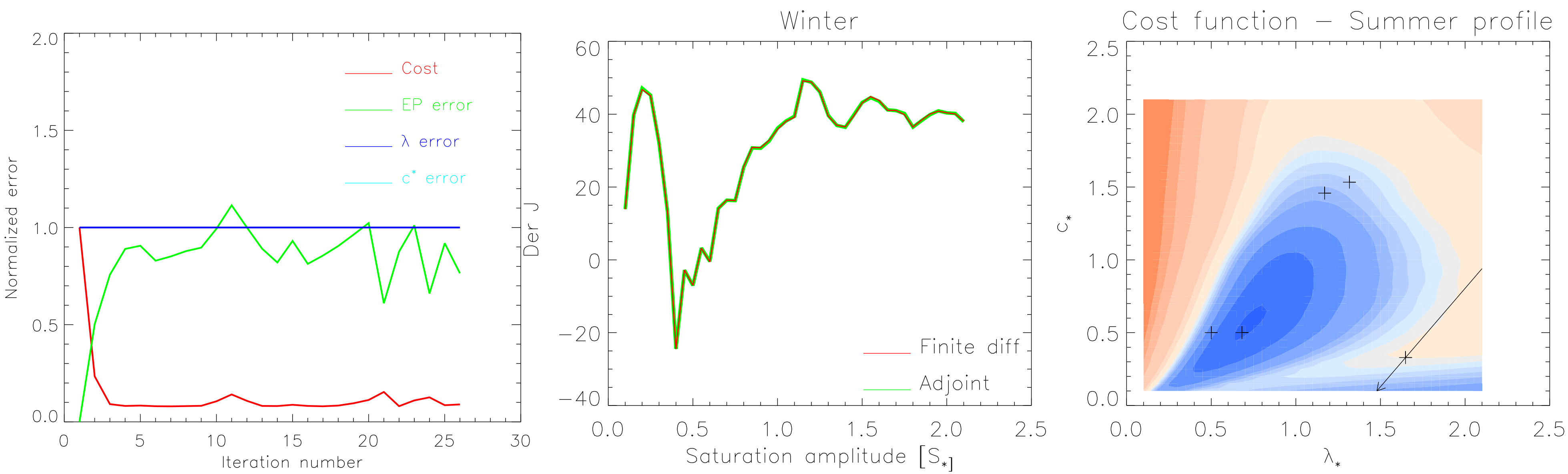


$F_b = \int_{\theta_b}^{\theta_t} \sigma X_x \mathrm{d}\theta$. Pulido and Thuburn, JC (2008).

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.

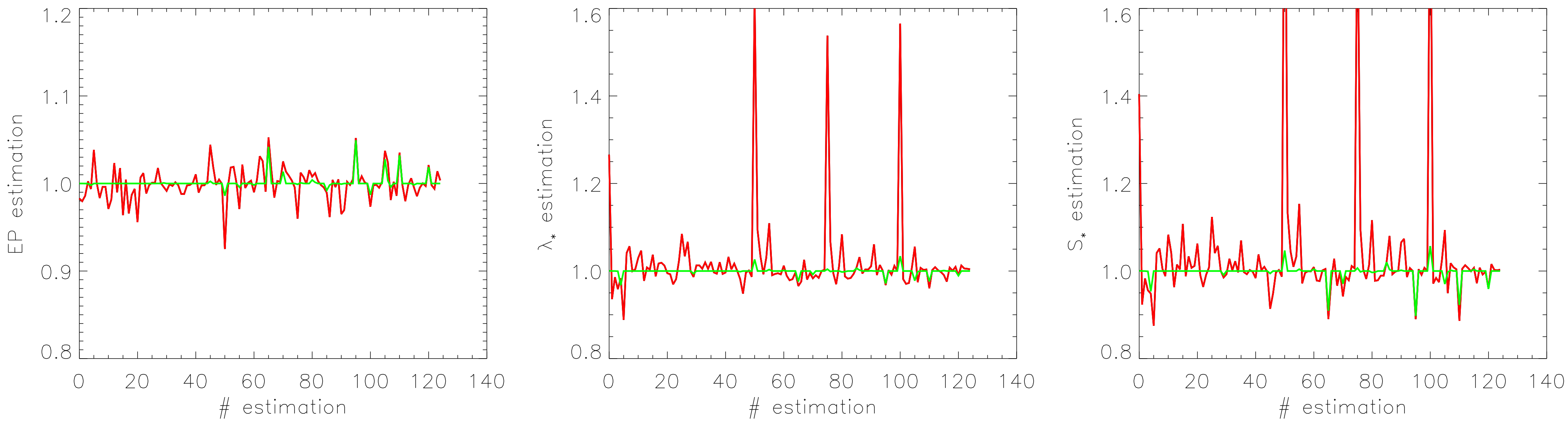


Parameterizations are highly non-linear and ill-conditioned NOT suitable for variational data assimilation.

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).

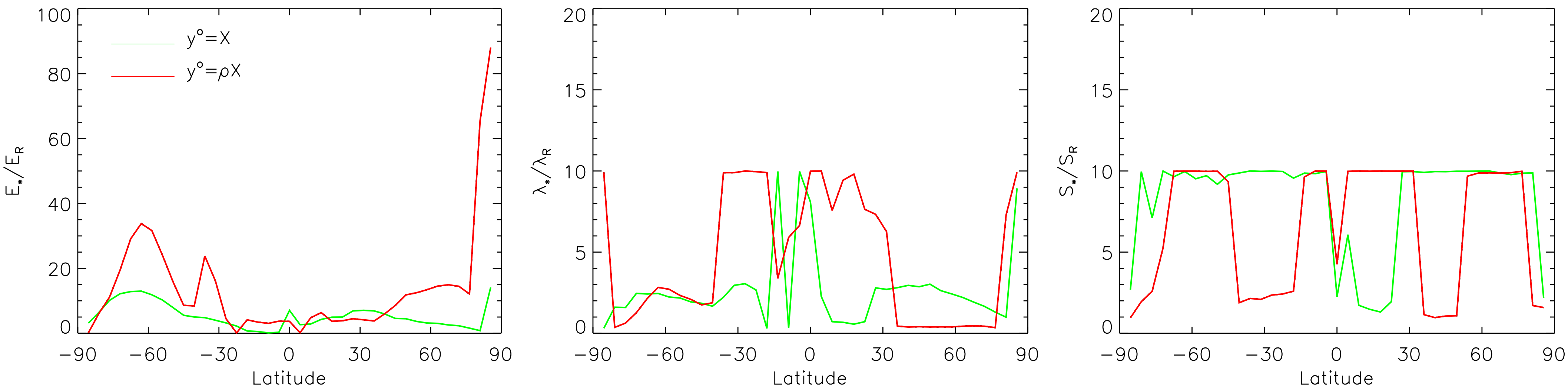


All the experiments converge toward the true parameters.

Estimated parameters

Zonal wind and temperature is taken from Met Office analysis.

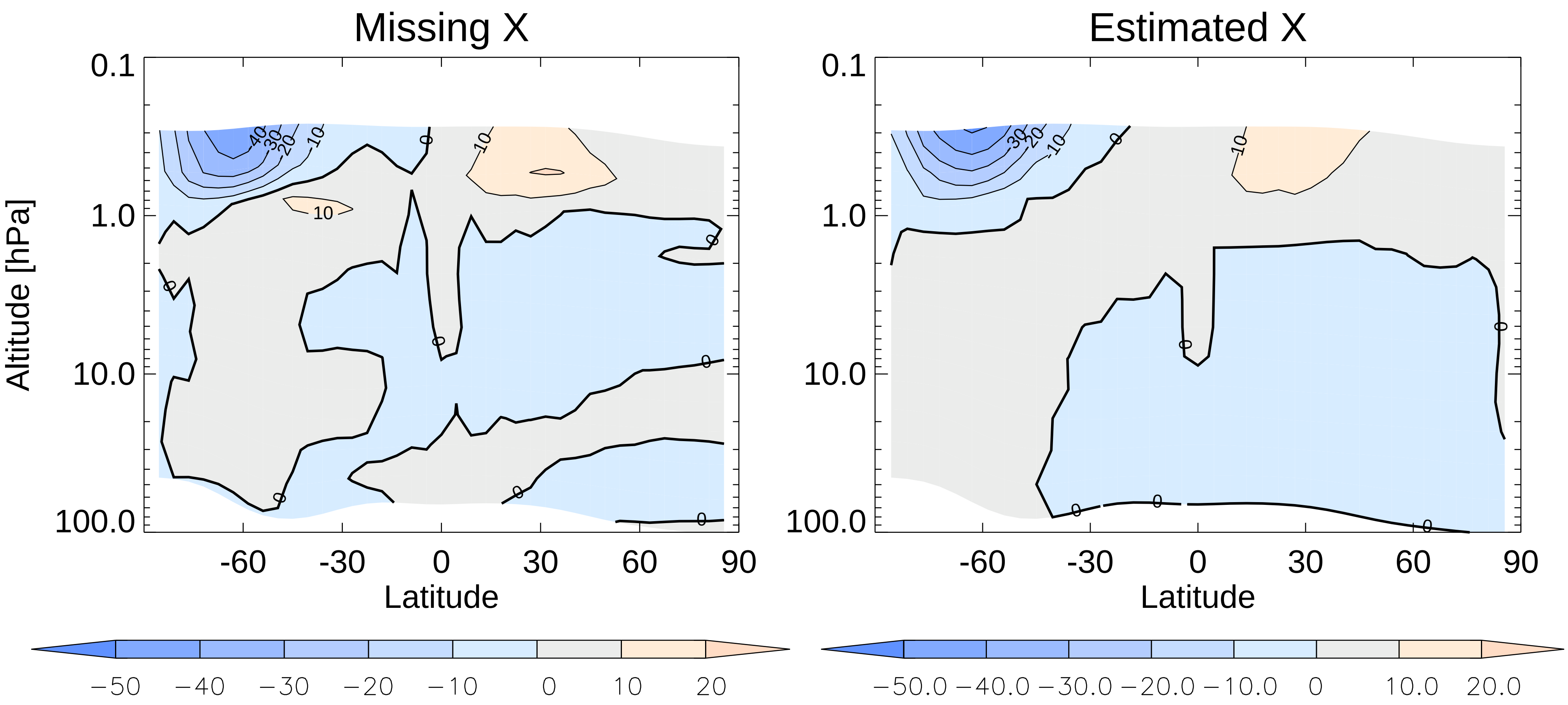
The missing forcing estimated with the ASDE-4DVar technique (Pulido and Thuburn, JC 2008) for July 2002 is used as the “observations” $\mathbf{y}$.



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

Parameter λ_* appears to agree in midlatitude with measurements.

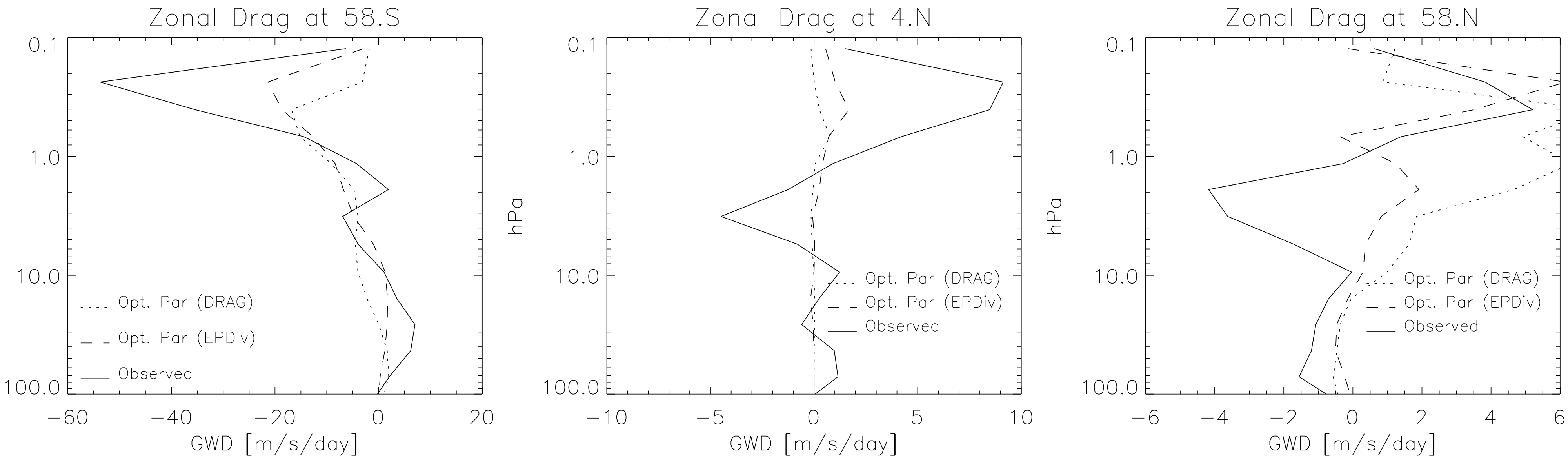
Estimated and "parameterized" forcing



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

Impact of optimal parameters in the model

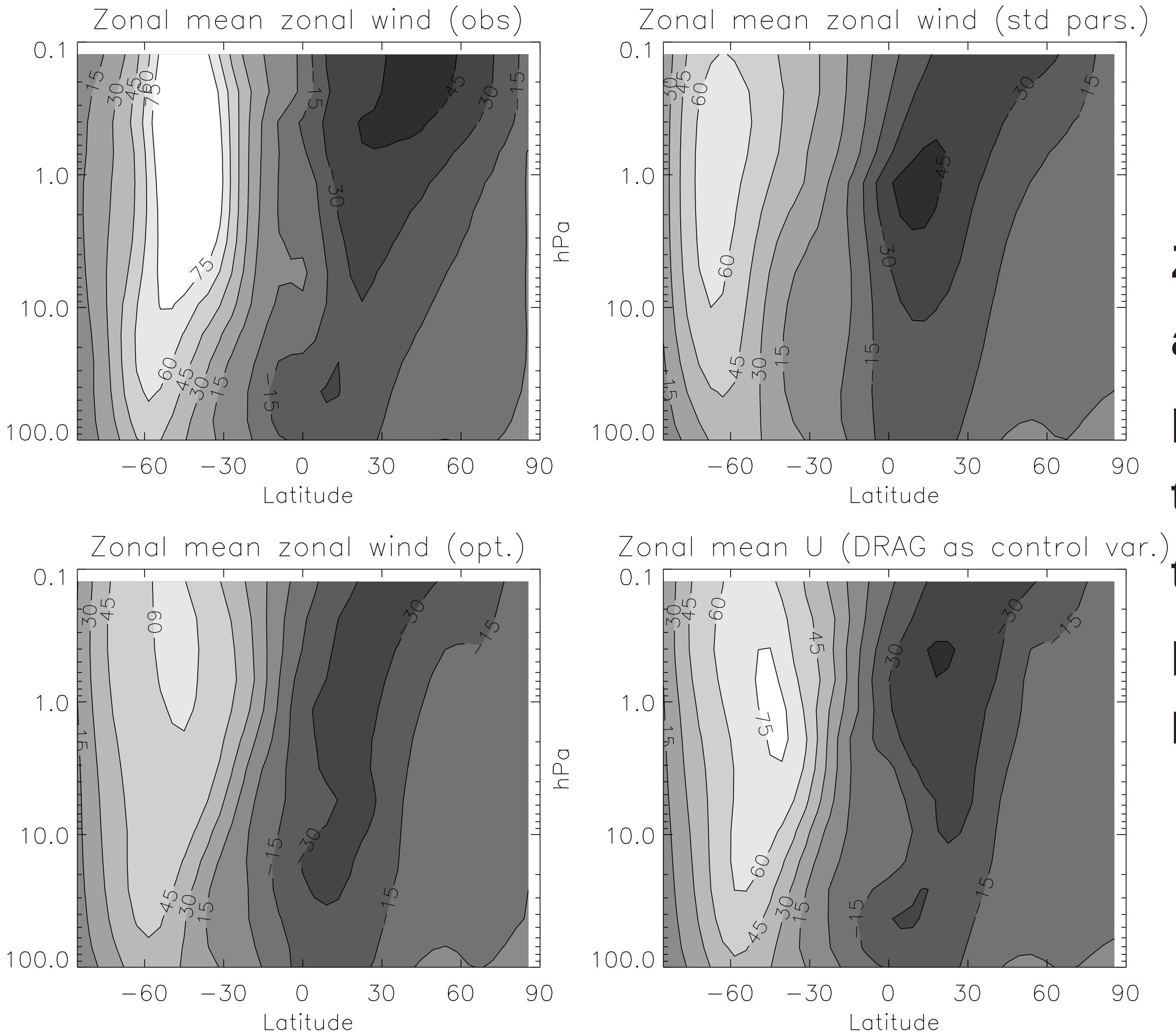
On-line simulation with the model using the optimal parameters: 60 days (June/July 2001).



Drag profiles given by the parameterization using the optimal parameters, standard parameters and the estimated drag. Day 10/7/2001

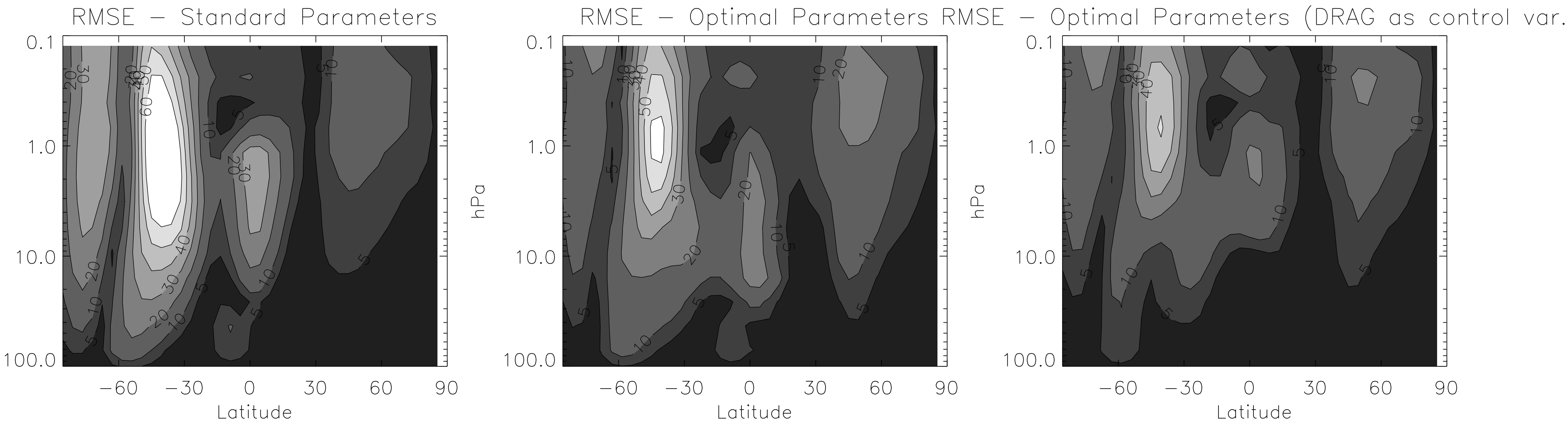
G. Scheffler's work

Zonal wind with optimal parameters



Zonal wind from the analysis, given by the parameterization using the standard parameters and the optimal parameters (EPDIV and DRAG).

Model error comparison



RMSE between the model with standard parameters and MO analysis, and the one with optimal parameters (EPDIV and DRAG).

G. Scheffler's work. Preliminary results.

4DVar works really well to estimate GWDrag, however it is model dependent. If we want to estimate GWD 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.

Ensemble-based data assimilation: Forcing estimation

Evaluation of Ensemble Kalman filtering to estimate GW drag.

Advantage: Model independent technique.

Disavantage: Some tuning/development is required.

Implementation: Local Ensemble Transform Kalman Filter (LETKF) Hunt, et al Physica D 2007

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathrm{K}_k(\mathbf{y}_k^o - H(\mathbf{x}_k^f))$$

where $\mathrm{K}_k = \mathrm{P}_k^f \mathrm{H}_k^T (\mathrm{H}\mathrm{P}_k^f \mathrm{H}^T + \mathrm{R}_k)^{-1}$**.**

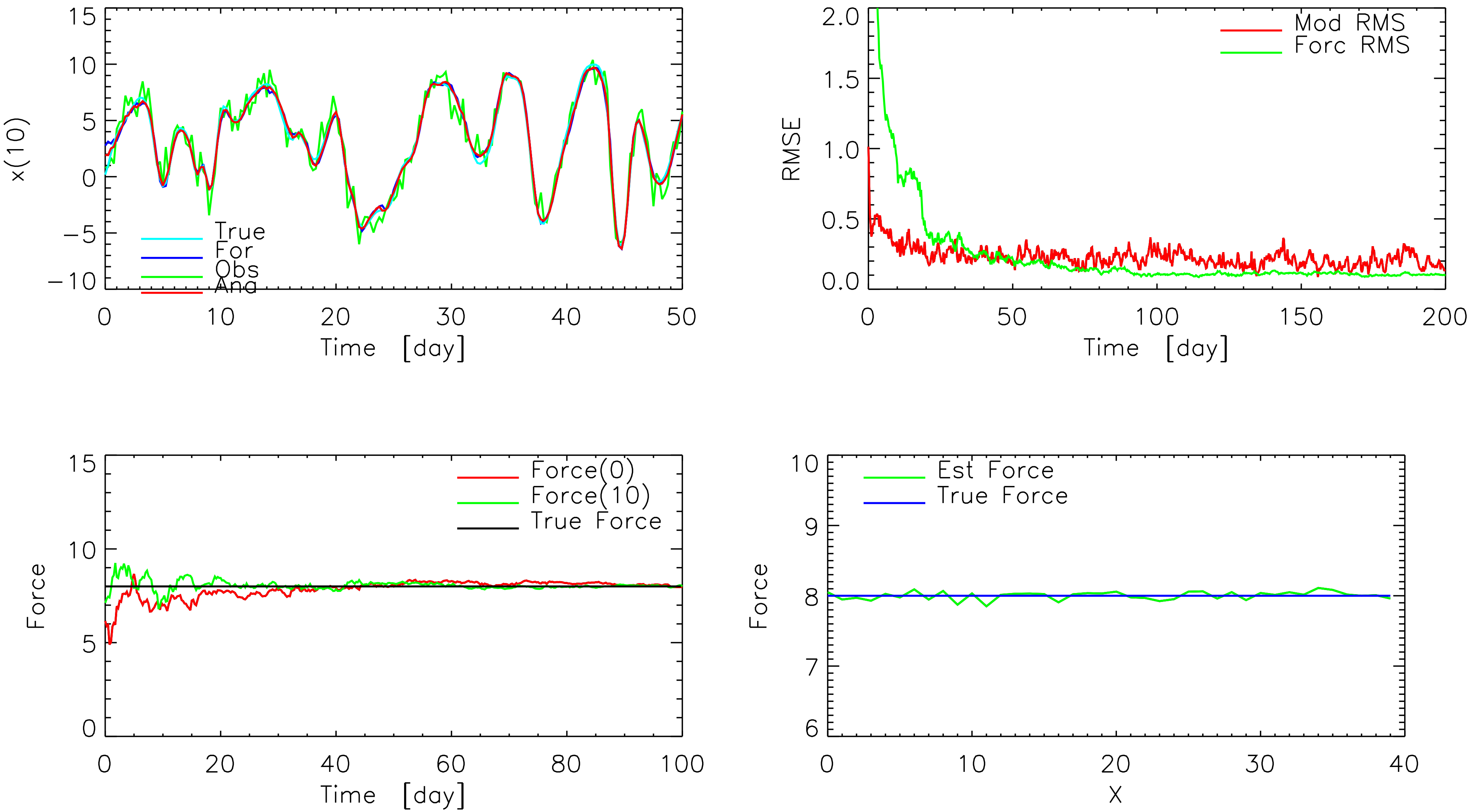
Test model: Lorenz'96 dynamical system (nonlinear chaotic model).

$$\frac{\mathrm{dX}_k}{\mathrm{d}t} = -\mathrm{X}_{k-1}\left(\mathrm{X}_{k-2} - \mathrm{X}_{k+1}\right) - \mathrm{X}_k + \mathrm{F}_k(t), \qquad k = 1, \cdots, K$$

X_k **model variables,** $\mathrm{F}_k(t)$ **forcing terms.**

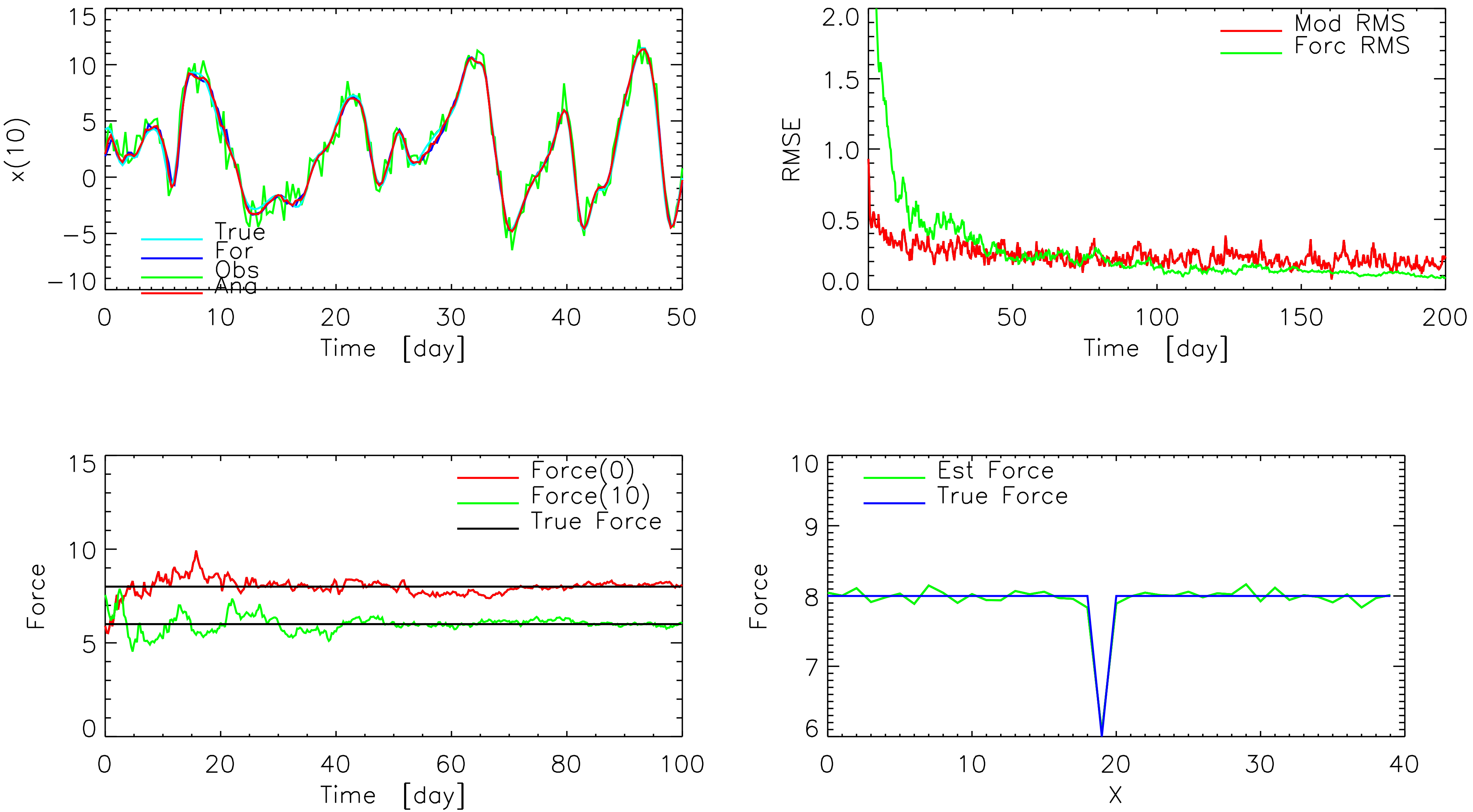
Estimation of the forcing with ETKF

Twin Experiment: True Constant forcing=8, Initial guess (mean forcing)=6, Initial spread=2.



Estimation with ETKF: A localized forcing feature

Twin Experiment: True Constant forcing=8 except $F_T(20) = 6$, Initial guess (mean forcing)=6, Initial spread=2.

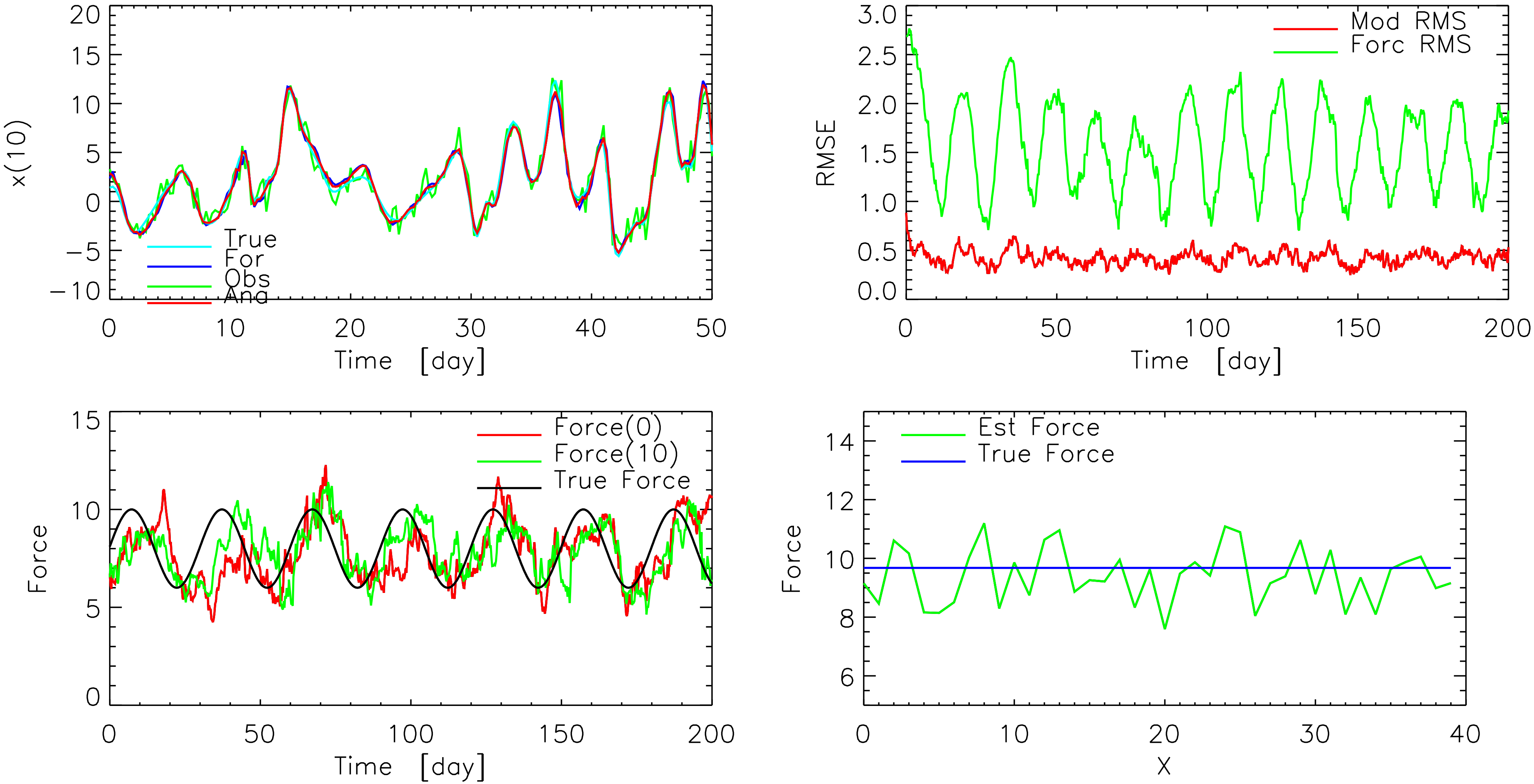


Estimation with ETKF: Time dependent

Weak point for EnKF techniques: slow convergence in time.

Inflation factor is critical. Best results with Ruiz, et al. (2012) submitted to JMSJ

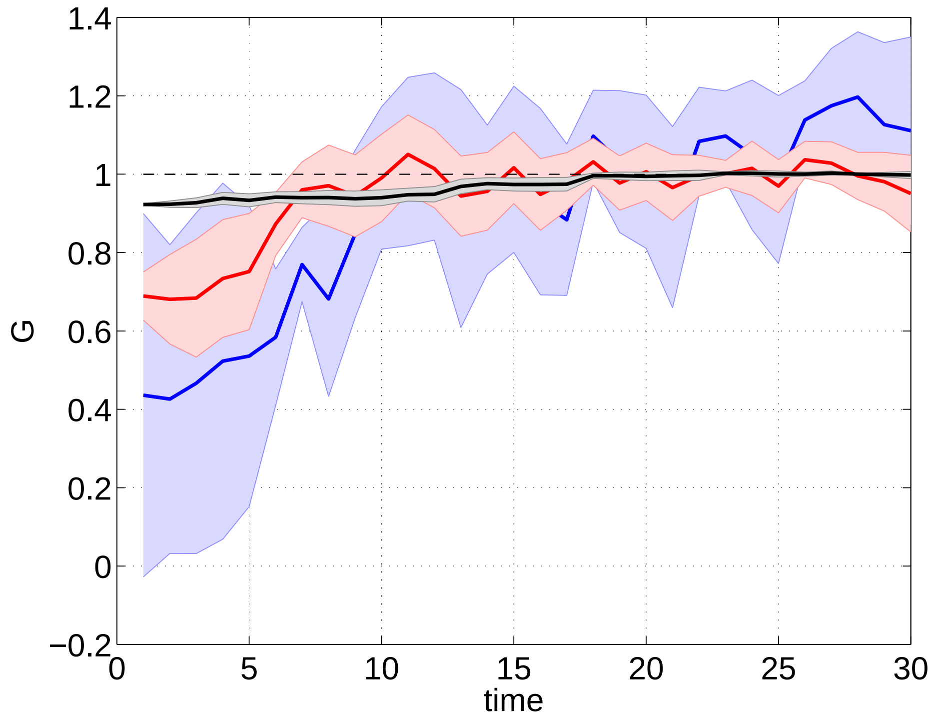
Twin Experiment: True Constant forcing=8 with a sinusoidal 30 day period.



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.

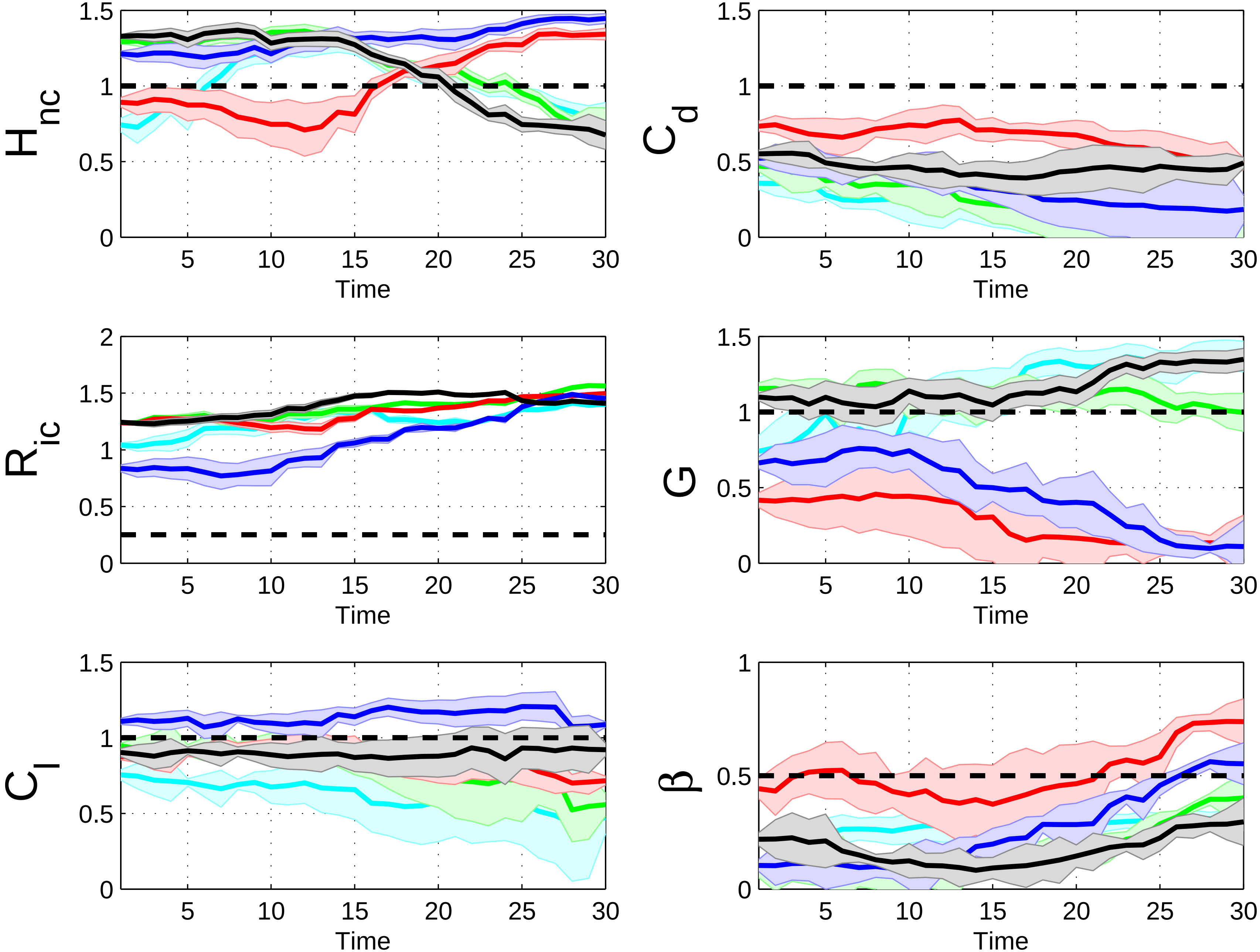


Twin experiment for an offline estimation. Blue (Iteration 1), Red (it=10) Black (it=50).

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?



Conclusions

Estimating the missing forcing = source 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 + simple model with twin experiments)

Conclusions

Estimation of missing forcing with ENKF.

- **Great potential since it is model independent.**
- **Works well for persistent forcing and localized forcing.**
- **Do not capture well time dependences even for clever inflation factors.**
- **Possible solutions (future work): Running in Place, Maximum Likelihood Cov Estimation**