Parameter estimation in the presence of model error

Juan J. Ruiz[1,3] and Manuel Pulido[2,3]

[1]Centro de Investigaciones del Mar y de la Atmósfera (CONICET-UBA), Atmospheric Sciences Department, Universidad Nacional de Buenos Aires, Buenos Aires, Argentina [2] Universidad Nacional del Nordeste and CONICET, Corrientes, Argentina.[3] UMI-IFAECI/CNRS

Motivation: The evaluation of parameter estimation using data assimilation techniques is usually performed under the perfect model assumption (i.e. the only source of model error is associated with the unknown parameters). In this work, the impact of other model error sources on the parameter estimation and on analysis quality is explored using a local ensemble transform Kalman filter applied to a simple general circulation model. The forecast skill of the system for different forms of accounting model error in a data assimilation cycle is compared.

Data assimilation and parameter estimation approach:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b \bar{\mathbf{w}}^a$$

$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y}^o - \bar{\mathbf{y}}^b)$$

$$\tilde{\mathbf{P}}^a = [(k-1)\mathbf{I} + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b]^{-1}$$

$$\mathbf{X}^a = \mathbf{X}^b \mathbf{W}^a \quad \mathbf{W}^a = [(k-1)\tilde{\mathbf{P}}^a]^{\frac{1}{2}}$$

The data assimilation technique is based on the LETKF as in Hunt et al. 2007 and Miyoshi et al. 2007. This technique is extended to obtain also the optimal value for certain model parameters.

The analysis space is augmented to include model parameters.

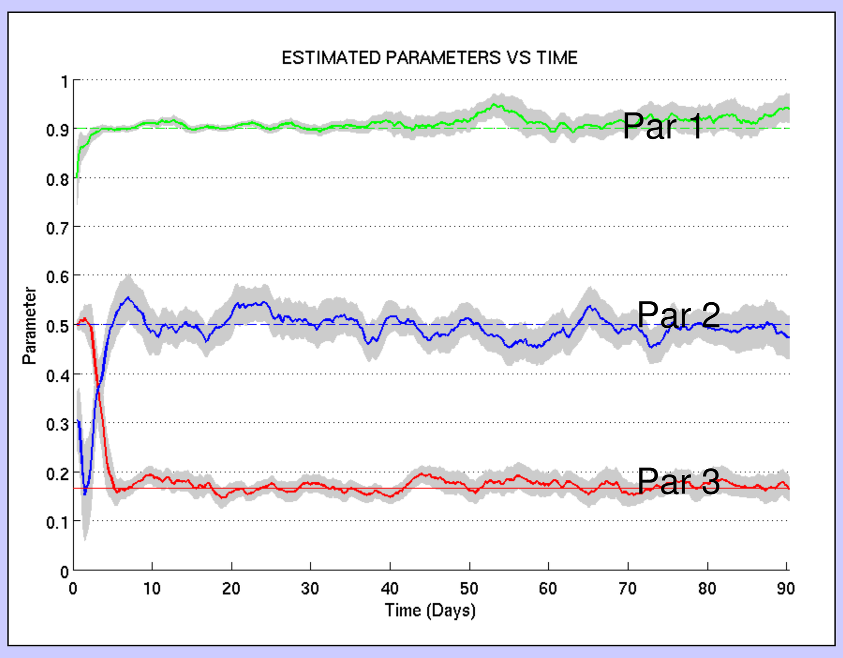
In the case of model state varaibles a localization of the background error covariance and fixed multiplicative inflation are used for their estimation. No localization is used for the estimation of the parameters since parameters are global.

Since parameters are assumed to be constant within the assimilation window, the parameter ensemble spread does not increase. Therefore, an online estimation method of the parameter ensemble spread is used (Ruiz et al 2012)

Twin experiments with the SPEEDY model (Molteni 2003):

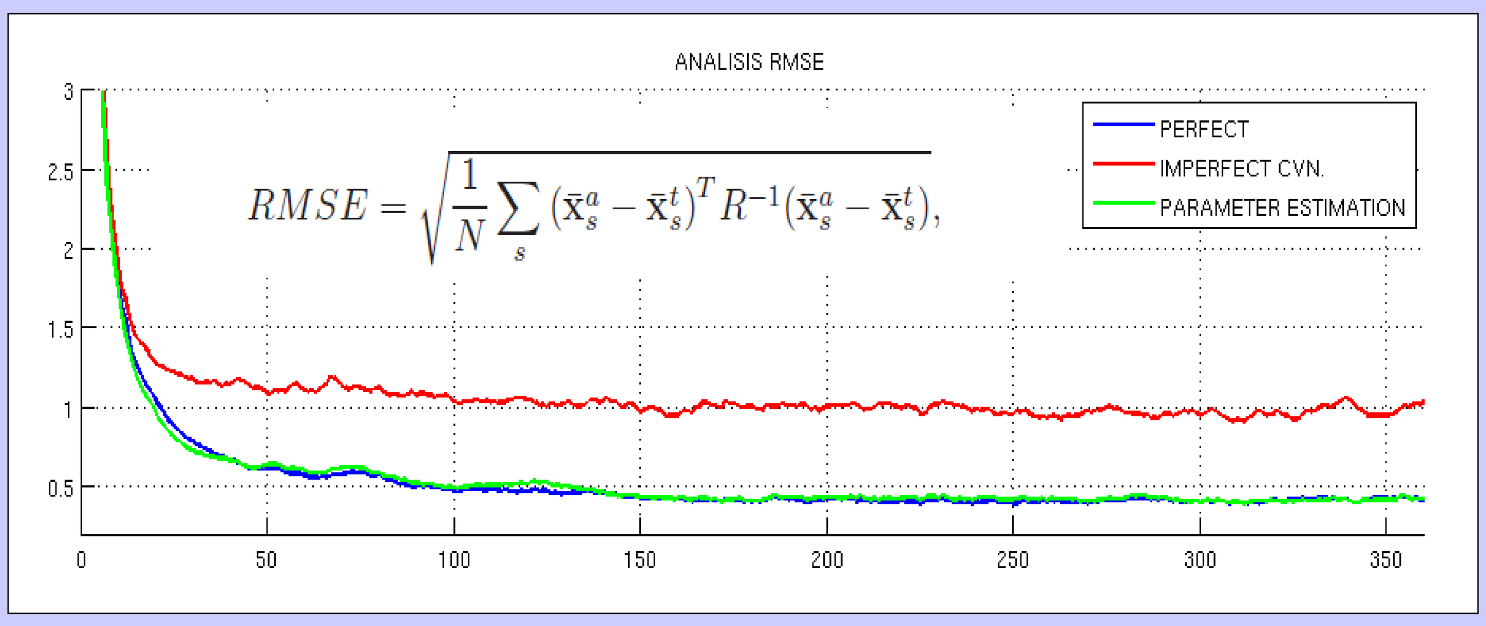
- A long continuous simulation is considered as the true evolution of the system.
- Observations are generated from the true state

Perfect model scenario:

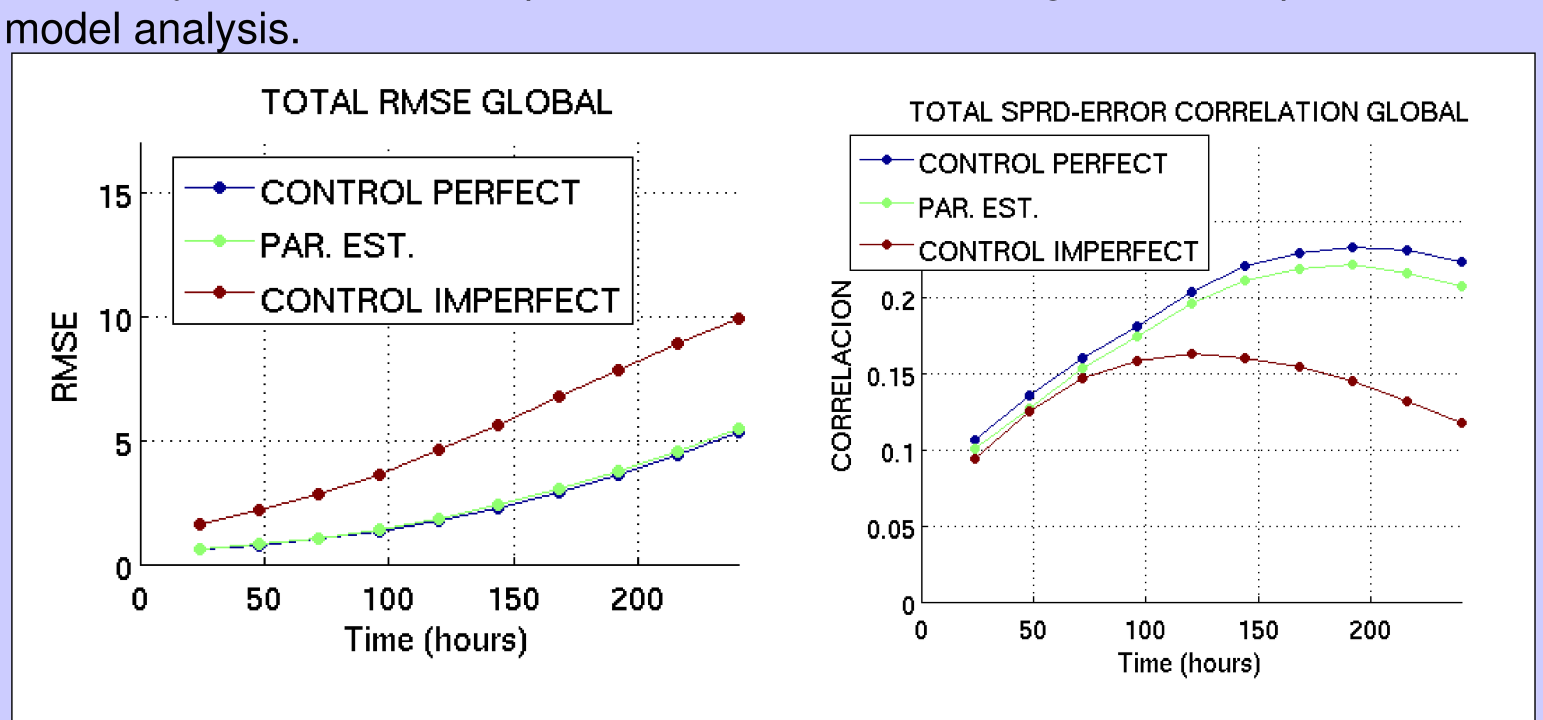


Estimated parameters converge to their true value. (i.e. the value used to generate the true state)

Parameter ensemble spread seems to be consistent with the uncertainty in the estimation of the parameter.



The analysis obtained with parameter estimation is as good as the perfect model analysis.

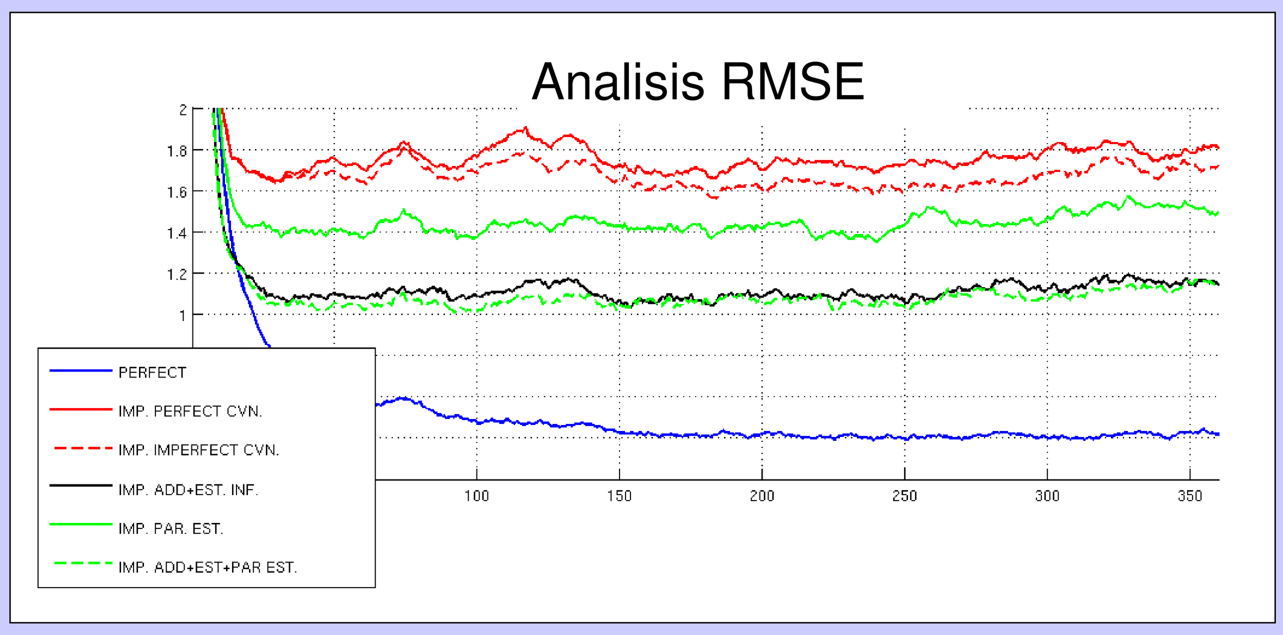


10 day ensemble forecast are started from each experiment.
Short to medium range ensemble forecasts show a large improvement with parameter estimation.
The RMSE for the forecast generated by the model with estimated parameters is almost as good as in the case of the one generated by the perfect model.
The error-spread relationship improves when parameters are estimated, because of the reduction of model error.

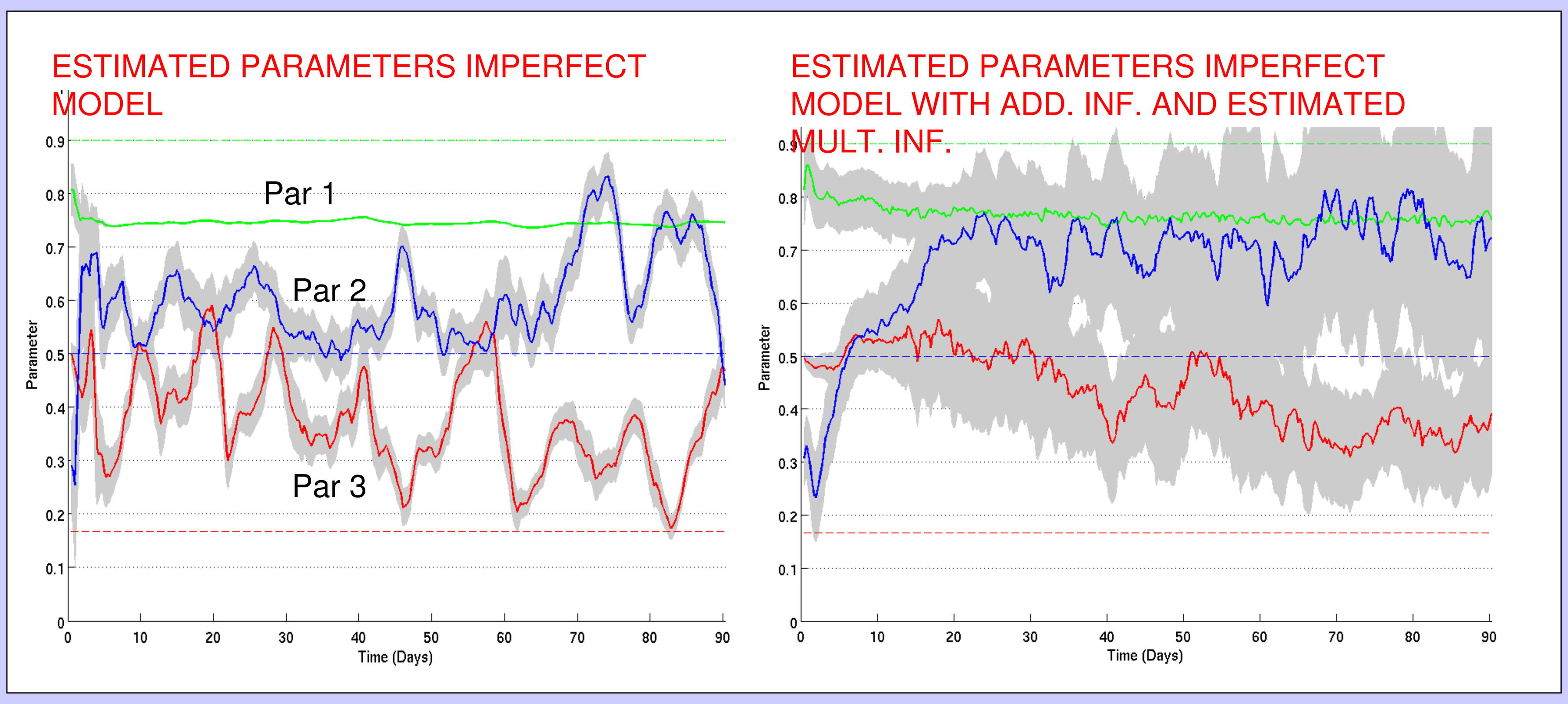
Imperfect model scenario:

In these experiments the model used in the data assimilation cycle is not perfect. This imperfection represent all the error sources that cannot be corrected by tuning the model parameters.

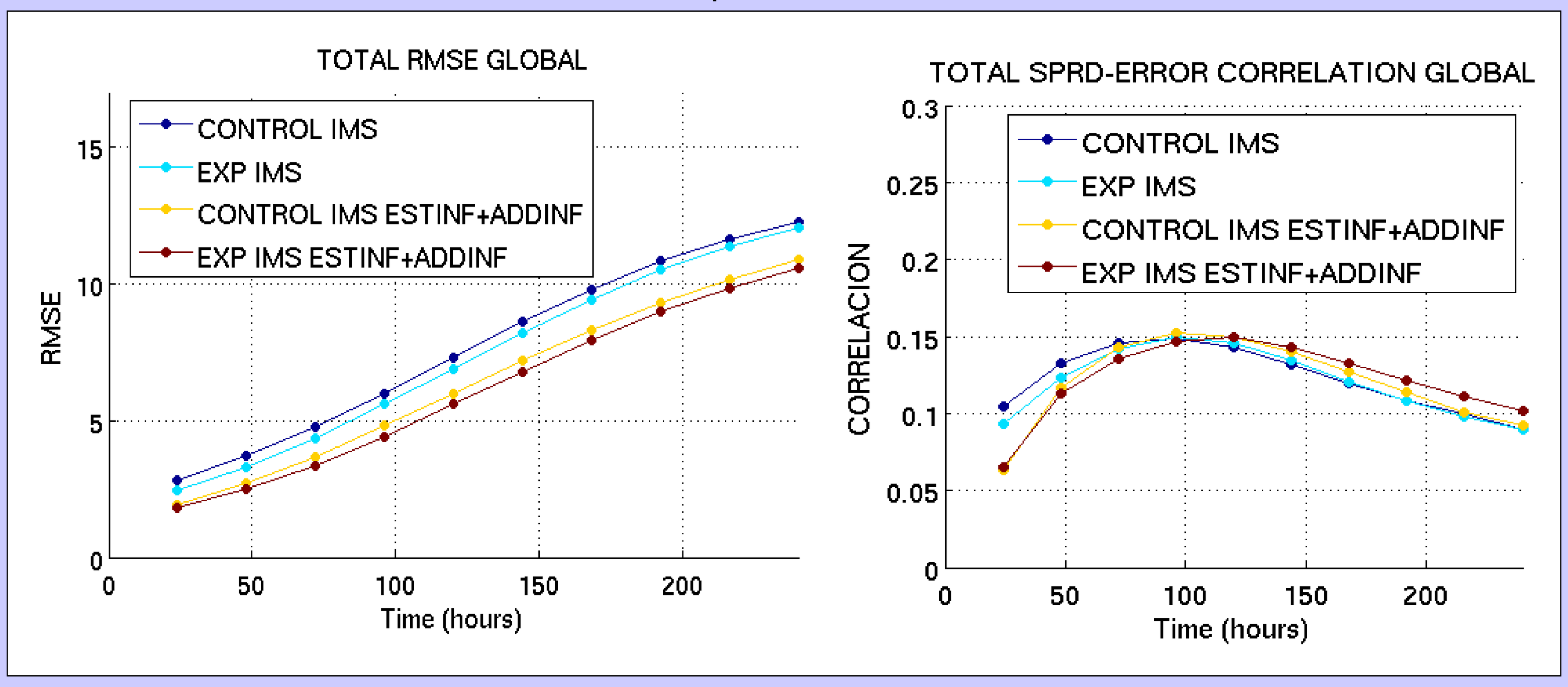
Parameter estimation is combined with methods to account for model error (Li et al. 2009 MWR). Two experiments are conducted:
- LETKF + ADD. INF. + ESTIMATED MULTIPLICATIVE INFLATION
- LETKF + ADD. INF. + ESTIMATED MULTIPLICATIVE INFLATION + PARAMETER ESTIMATION.



Parameter estimation produces a large improvement in the analysis when model error is not taken into account. And produce a small improvement when additive and multiplicative inflation are tuned to account for model error in LETKF.



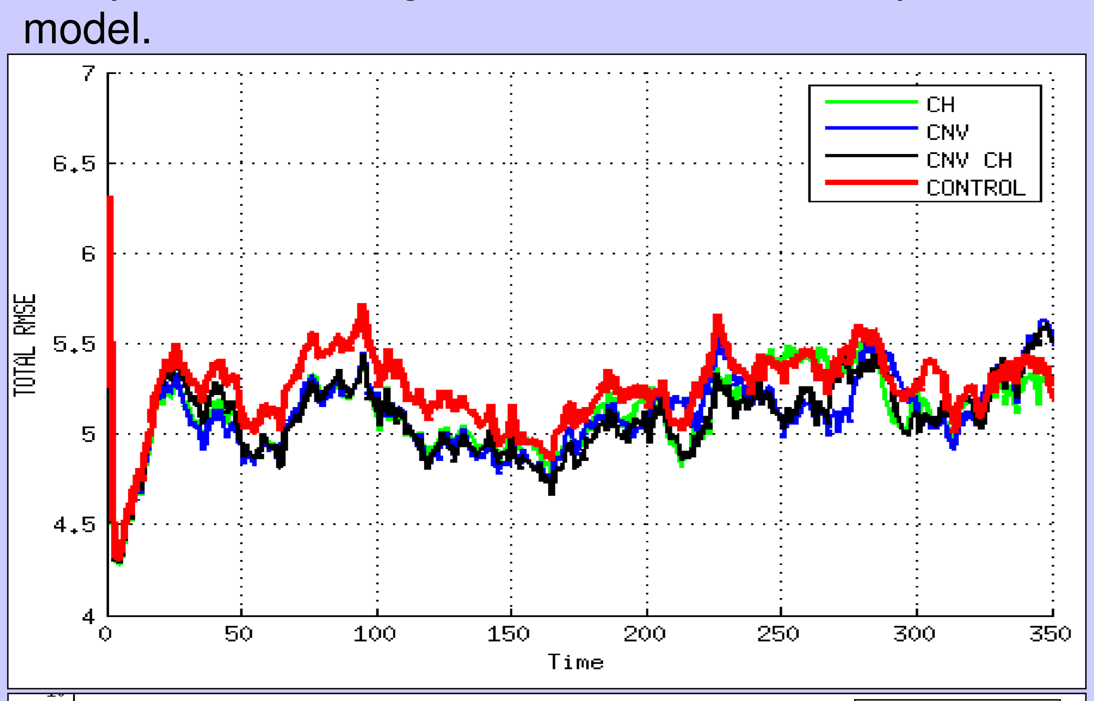
When the model is not perfect the estimated parameters do not converge to the "true" value anymore. In some cases (parameter 1) they converge to an "optimal" value. In some other cases there is no convergence and the value of the parameter oscillates in time (as if the optimal value is strongly state dependent). When ADD. INF.+ ESTIMATED MULT. INF. Is used the time oscillation of the estimated parameters is reduced.



Parameter estimation always results in a better forecast (lower RMSE), this reflects the improvement in the model and in the initial conditions. However the improvement is smaller than in the perfect model scenario. The spread-error relationship is usually better when parameter are not estimated for short lead times. The use of additive inflation and multiplicative inflation improves the skill of the forecast but also reduces the error-spread relationship (probably because it introduces random noise in the ensemble perturbations).

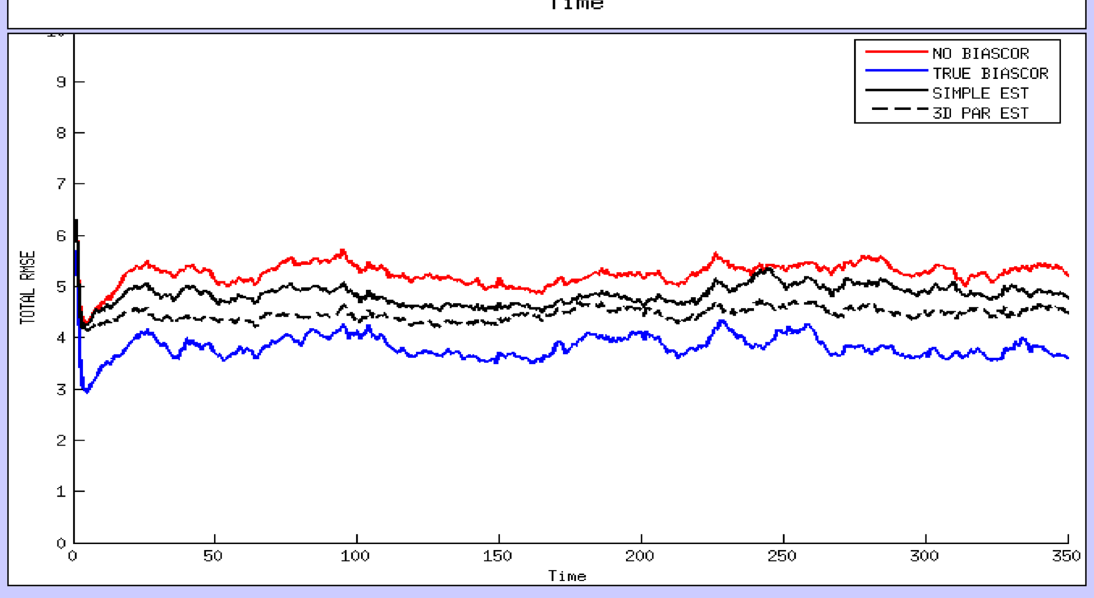
Preliminary experiments with NNR:

In these experiments NCEP Reanalysis are used as the true (as in Li et al 2009 MWR). Noisy observations generated from the reanalysis dataset are assimilated in the SPEEDY model.



The sensitivity of the analysis RMSE to the estimation of the convective scheme parameters and the surface heat exchange coefficient is smaller than in the previous experiments.

Maybe because only a small part of the total model error is related to parameter uncertainty in this case.



Bias estimation experiments:

-Simplified Dee and Da Silva.

-Estimation of bias as a 3D parameter (based on model error statistics)

-Forecast correction using the true forecast bias

Conclusions

Parameter estimation in an imperfect model scenario do not (usually) converge to the "true" parameters. However parameter estimation might produce an "optimal" set of parameters that reduce analysis error and improves the model thus improving the forecast.
Parameter estimation can be easily combined with other model error schemes resulting in a further improvement of the analysis and in an increase of the forecast skill.
This suggest that parameter estimation can improve the model in real data applications, however the estimated parameter will be extremely model dependent as they will depend upon the behavior of other sources of model error.

Aknowledgements: This research was partly financed by ANPICT 2007-411 and ANPICT 2010-1986. Thanks to Takemasa Miyoshi for providing the LETKF code for the SPEEDY model.