Efficient parameter estimation for numerical weather prediction models using data assimilation

Juan Ruiz in colaboration with Manuel Pulido,Takemasa Miyoshi and Masaru Kunii

jruiz@cima.fcen.uba.ar

Centro de Investigaciones del Mar y la Atmósfera- CONICET

University of Buenos Aires

Advanced Institute for Computational Science - RIKEN

CONICET

Data Assimilation Team

CIMA

Numerical models have a dynamic core and parameterizations of the unresolved processes.

In both components there are "unknown" parameters that have some impact upon the model performance.

$$\frac{dx}{dt} = f(x, t, p_1) + par(x, t, p_2)$$

Dynamic core

Parameterizations (model physics)

We want obtain better values for different model parameters in order to improve model performance.

We also want to quantify the uncertainty in these parameters.

- **Improve short range forecast and have a better understanding of model errors in short time scales.**

- **Improve the representation of the climate and climate system sensitivity to changes in different forcings (i.e. CO_2 concentrations).**

- **Other applications may include inference about unknown forcings in different components of the system (i.e. pollutan sources), climate change attribution, among others.**

Parameter estimation is expensive: evaluating model sensitivity to the parameters requires several model integrations

Additional challenges are:

- **Model response to the parameter might be non-linear (increased number of required model integrations)**

- **Numercial weather prediction models may have several parameters that needs to be simultaneously estimated**

- **Some parameters represent spatially varying and time dependent forcings. A 2-D distribution of the parameter has to be estimated in this case. (Kang et al. 2011, Pulido and Thuburn 2008, Ito 2010, Bellsky 2014)**

Possible solution: Use data assimilation that are highly efficient methods that combine model outputs and observations to obtain an optimal estimate of the state of the system.

Currently developed data assimilation systems can be extended in order to include the optimization of the model parameters considering them as if they were state variables.

$$s = \begin{Bmatrix} x \\ p \end{Bmatrix}$$

Where S is the “augmented state space” vector which includes the state variables and the model parameters.

This approach has been proposed by Jazwinsky in 1970 as has been used since then in several parameter estimation studies.

In the following experiments the ensemble kalman filter (particularly the LETKF, Hunt et al., 2007) is being used for the simultaneous estimation of the state and some model parameters.

In the ensemble Kalman filter an ensemble of forecast is produced with the model in order to estimate the error covariances among different variables.

Covariances between the model parameters and state variables are estimated using a different set of model parameters for each ensemble member.

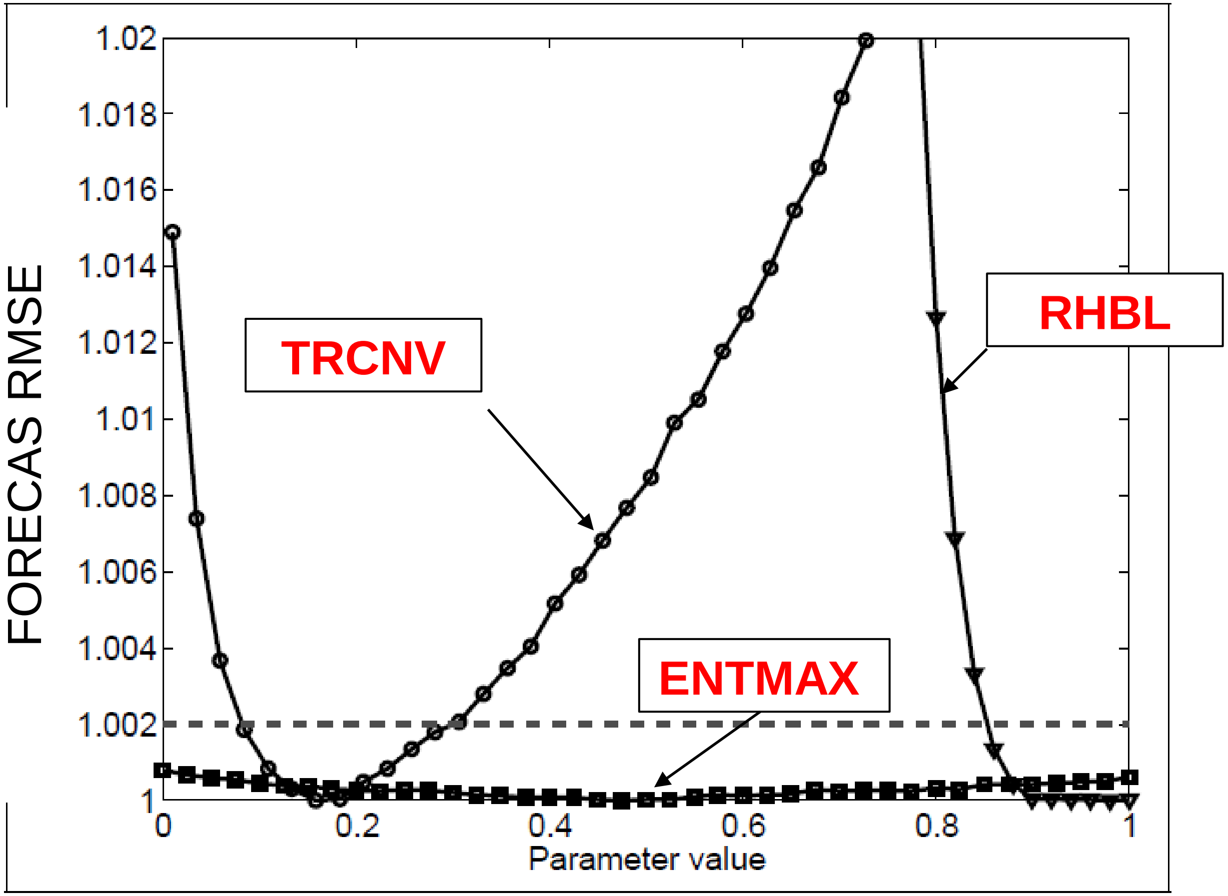
These covariances are the way in which the information provided by the observations can be used to estimate the parameters.

Only minor modifications are required to estimate model parameters in an ensemble based data assimilation system.

When can we estimate the model parameters?

Are parameters identifiable?

Is model sensitive to their changes, is this sensitivity simple, have different parameters the same effect? (Aksoy, 2014)

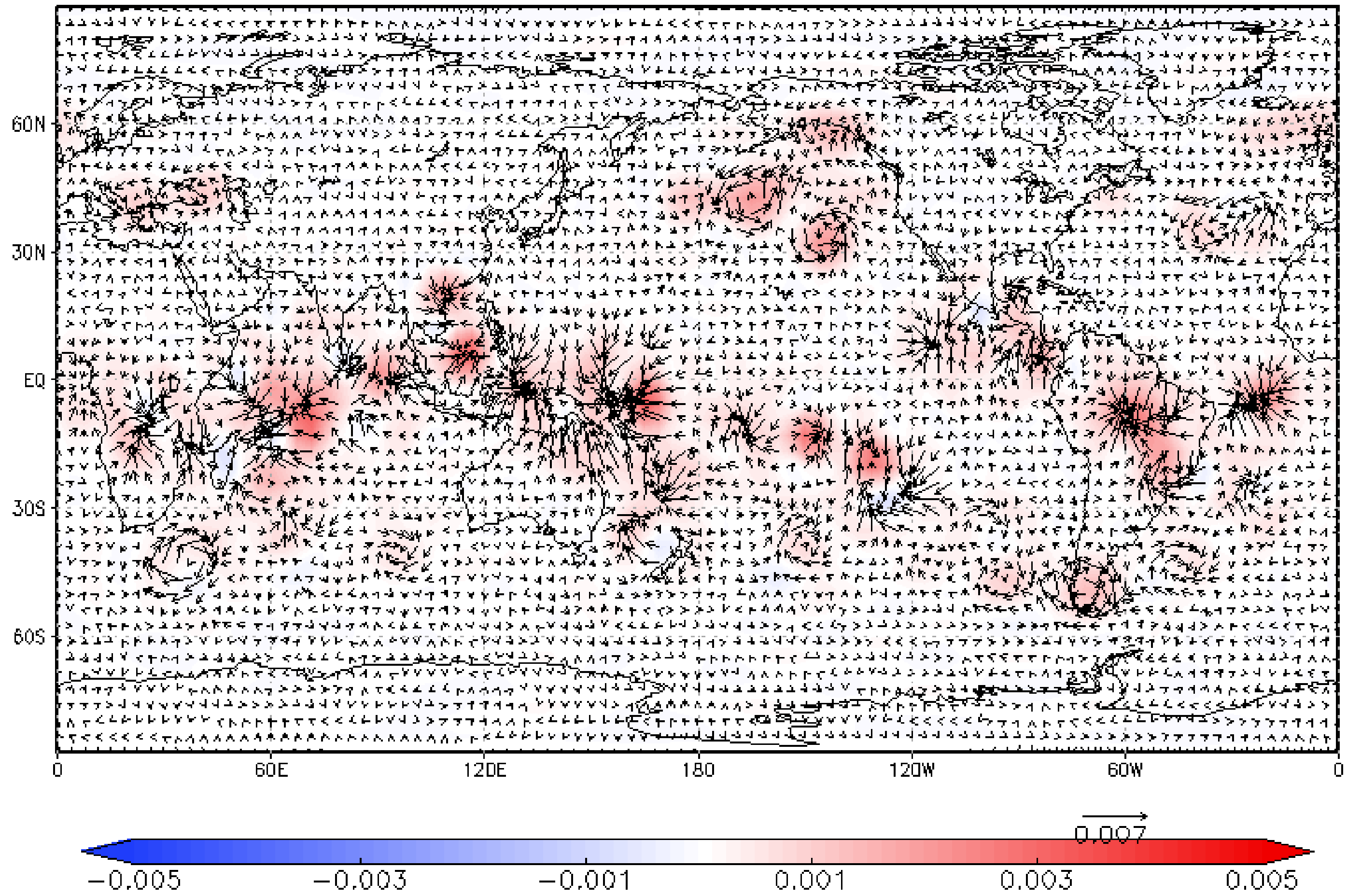


6-hour forecast RMSE as a function of the parameter value for the convective parameter scheme.

Idealized experiments:

How the covarinace between state variables and parameters looks like?

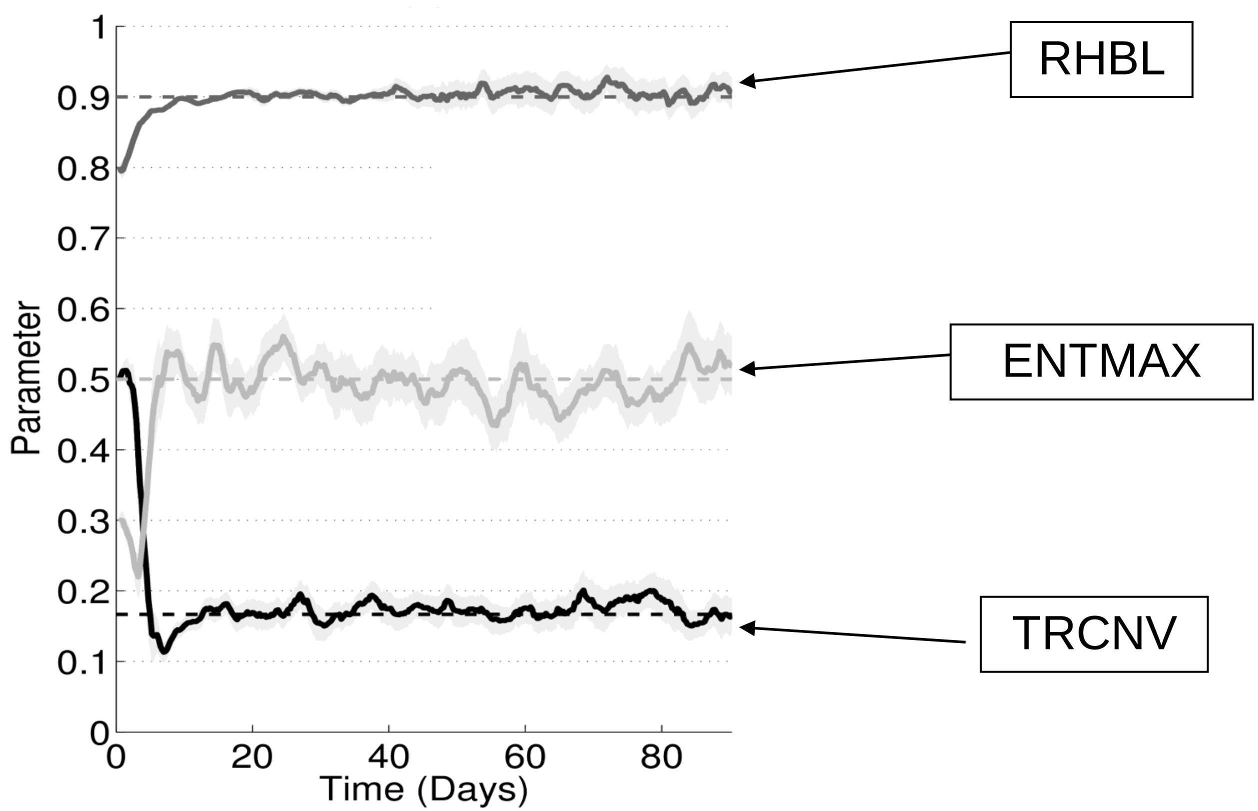
Covariance between TRCNV and middle level temperature and low level winds



Covariance structure is strongly flow dependent depending on where the convective scheme is activated

Idealized experiments:

OSSE with “almost” perfect model

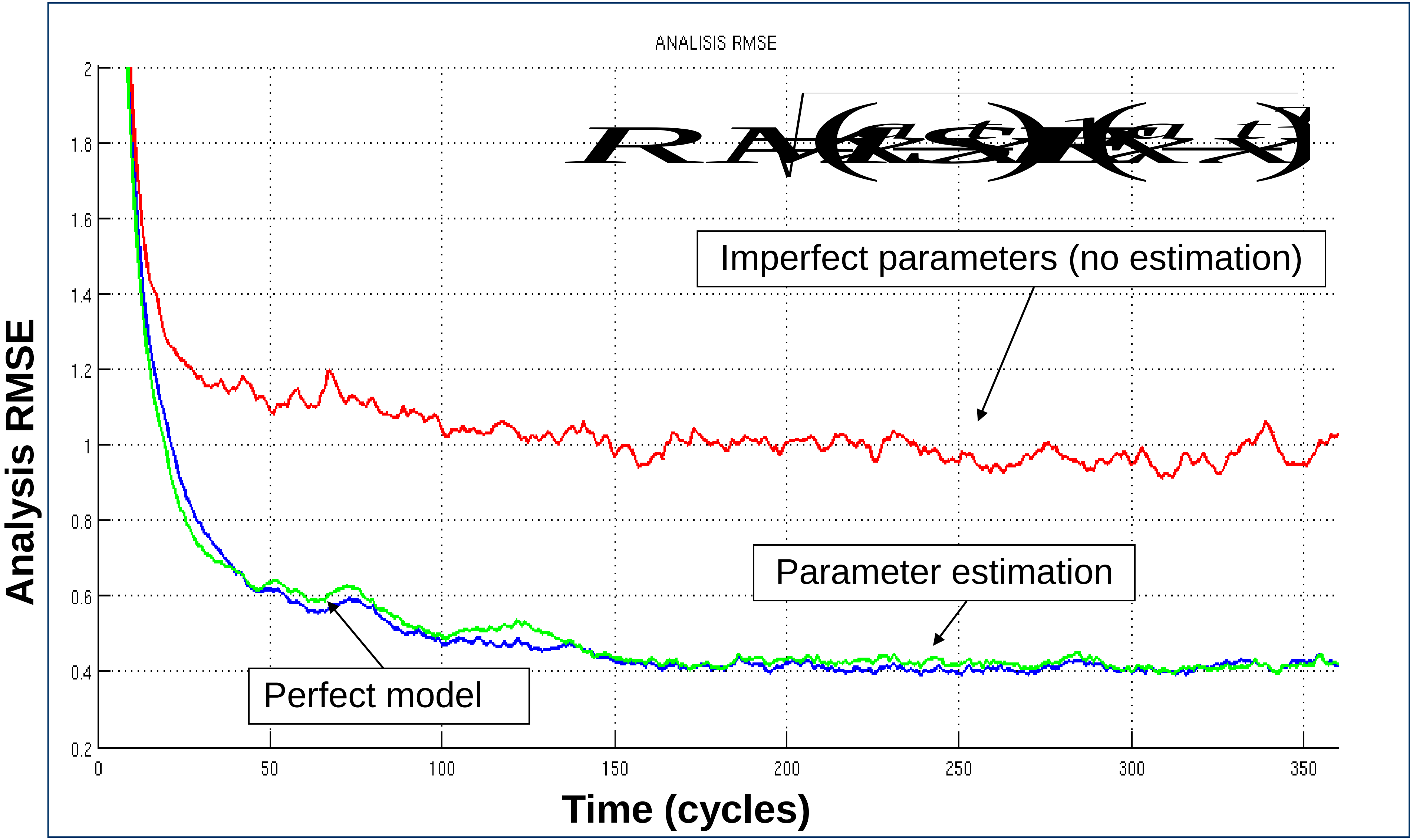


Time evolution of estimated parameters and their uncertainty

Convective schemes parameter are accurately estimated and the spin-up time is around 15 days (including the spin-up of the initial conditions).

Idealized experiments:

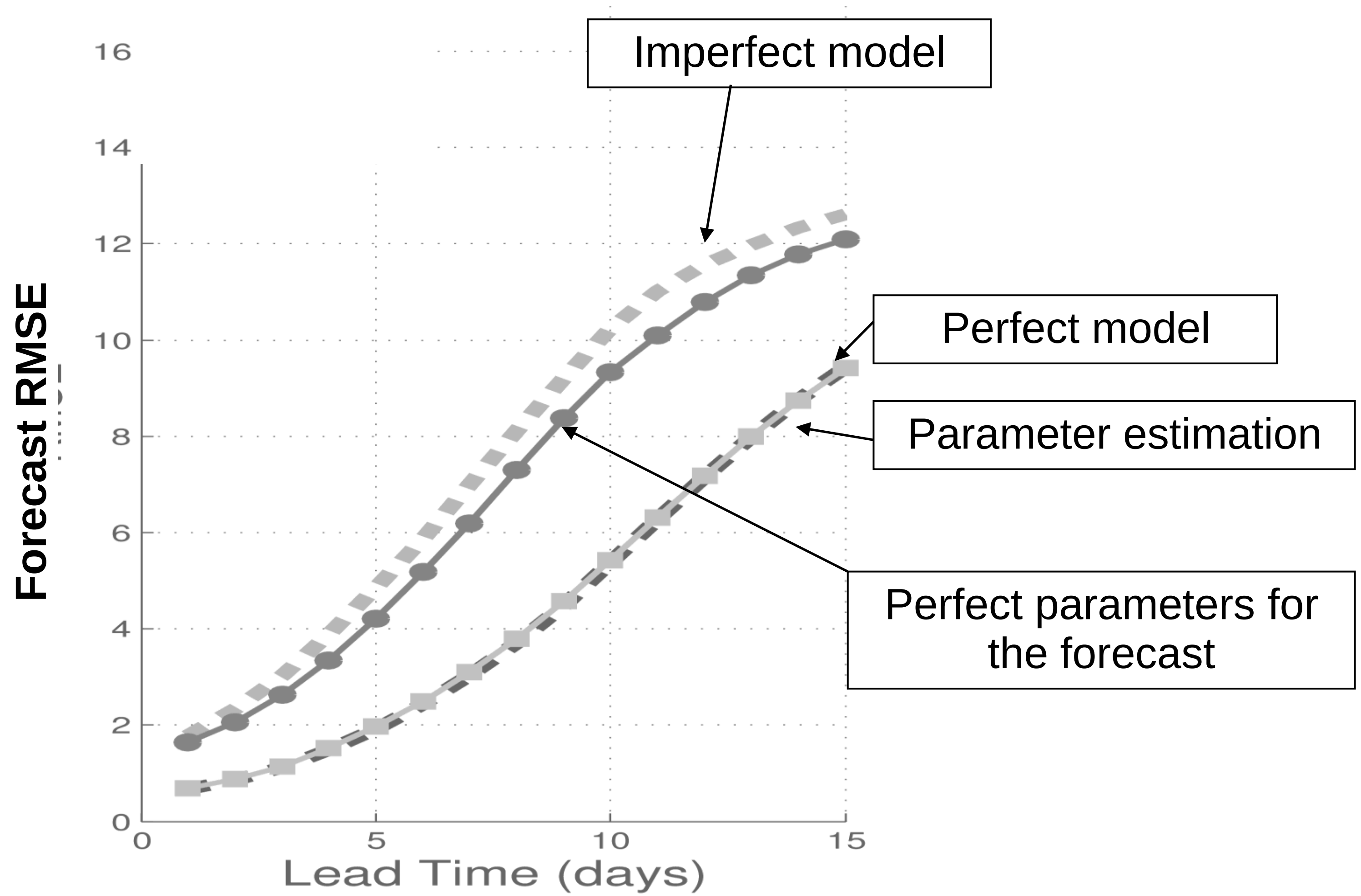
OSSE with “almost” perfect model



Parameter estimation produces a strong improvement in the analysis

Idealized experiments:

OSSE with “almost” perfect model



The positive impact of parameter estimation holds for the medium range forecast.

The stronger impact of parameter estimation is trough the improvement of the initial conditions.

Can we also estimate the uncertainty in the model parameters?

Parameter uncertainty (parameter ensemble spread) is treated in different ways:

- **constant parameter spread (Aksoy et al. 2006)**

- **multiplicative inflation (Koyama and Watanabe 2010, Kang et al. 2011)**

- **additive inflation (Kang et al. 2012)**

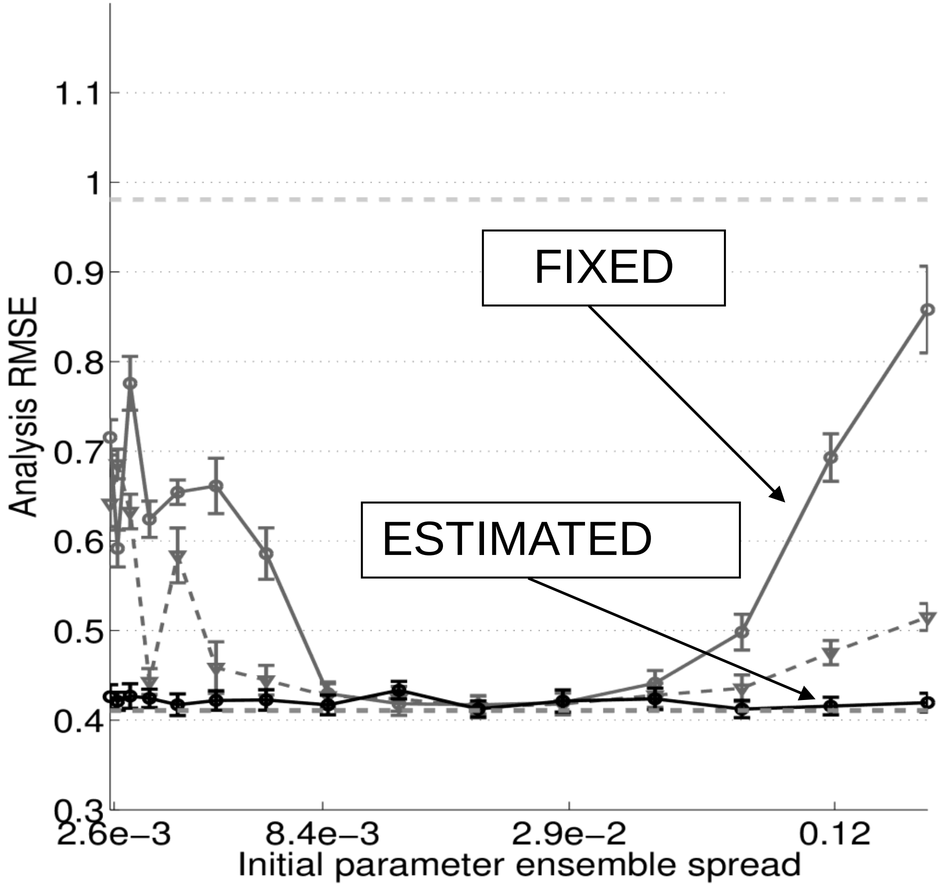
A persistence model is usually assumed for the parameters which contributes to the lack of spread in the parameters

Can we also estimate the uncertainty in the model parameters?

We proposed a simple and cost-free way to estimate the model parameters in the LETKF system.

$$\lambda = \sqrt{\frac{K}{(K-1)tr(\tilde{P}^a)}}$$

Inflate the parameter spread based on the spread in the state variables



Analysis error is sensitive to the proposed parameter ensemble spread

With the new method we do not need to tune the parameter ensemble spread

Can we estimate the model parameters when there are multiple sources of model error?

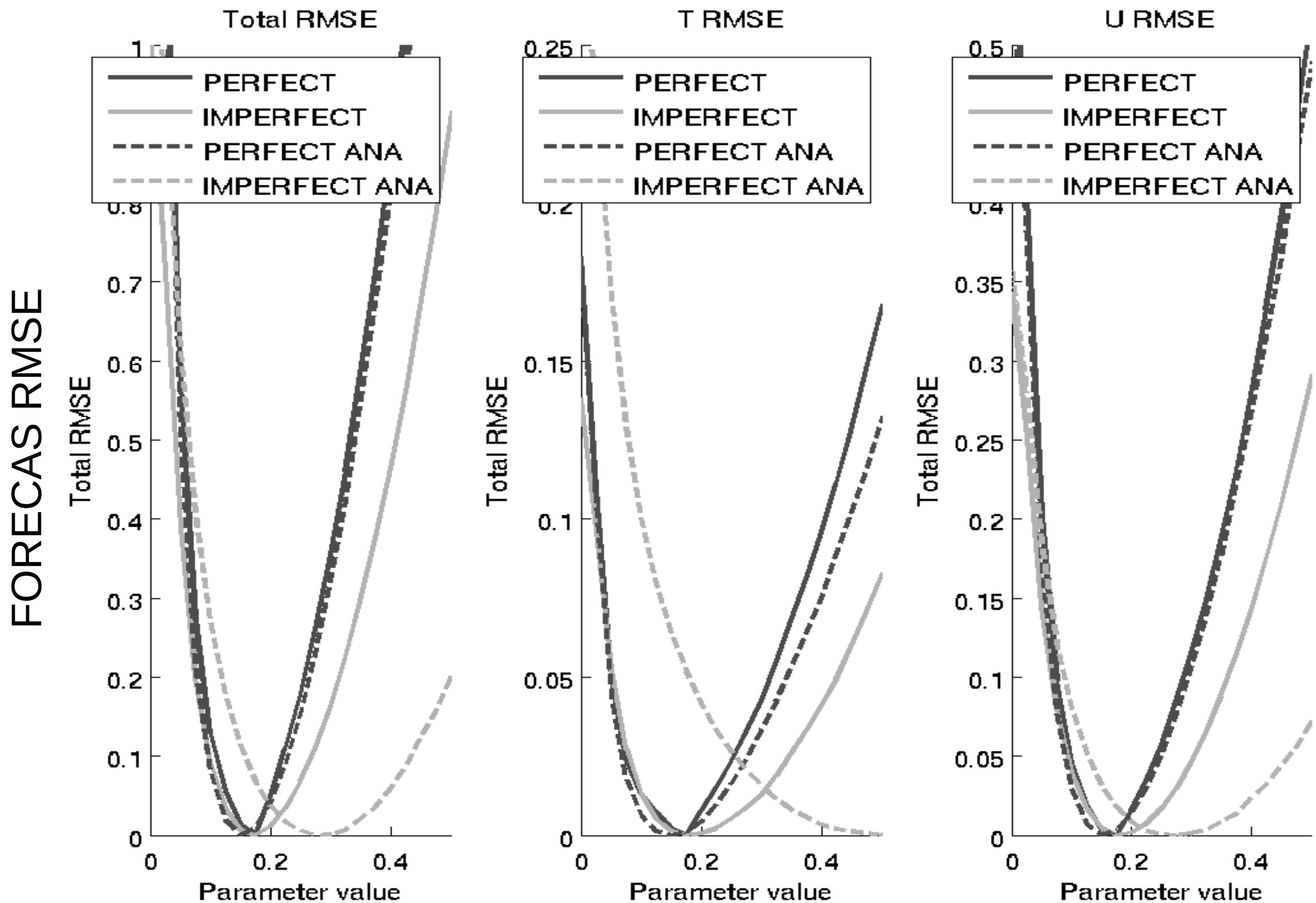
In the experiments presented so far, the model was (almost) perfect.
All model error is due to the uncertainty in the convective scheme parameters.

What happens when those are not the only source of errors and when the model error cannot be completely corrected by the tuning of some parameters (as in real applications)?

Can we combine parameter estimation with other ways to deal with model error in data assimilation methods?

Idealized experiments:

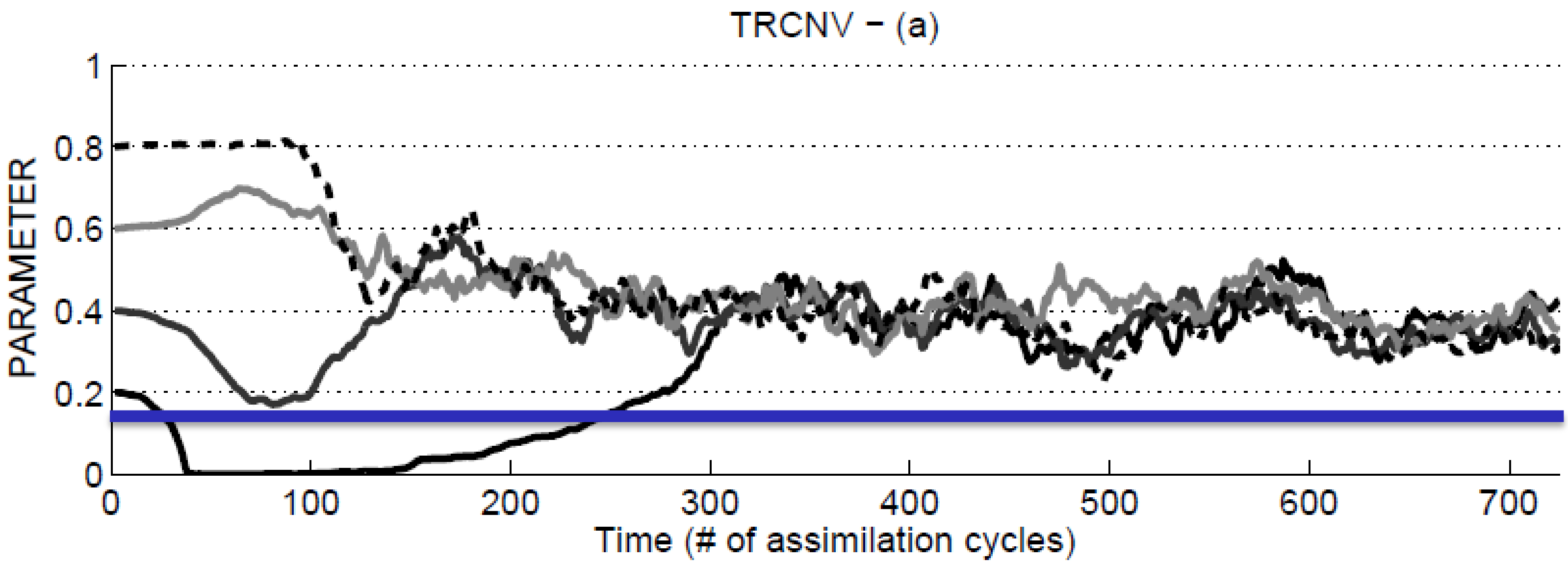
Model sensitivity in the presence of model error



Errors in the initial conditions can affect the model response to changes in the parameters.
Optimal parameter might depend on the variable considered (i.e. available observations). (Schirber et al. 2013)

Idealized experiments:

Parameter estimation with an imperfect model



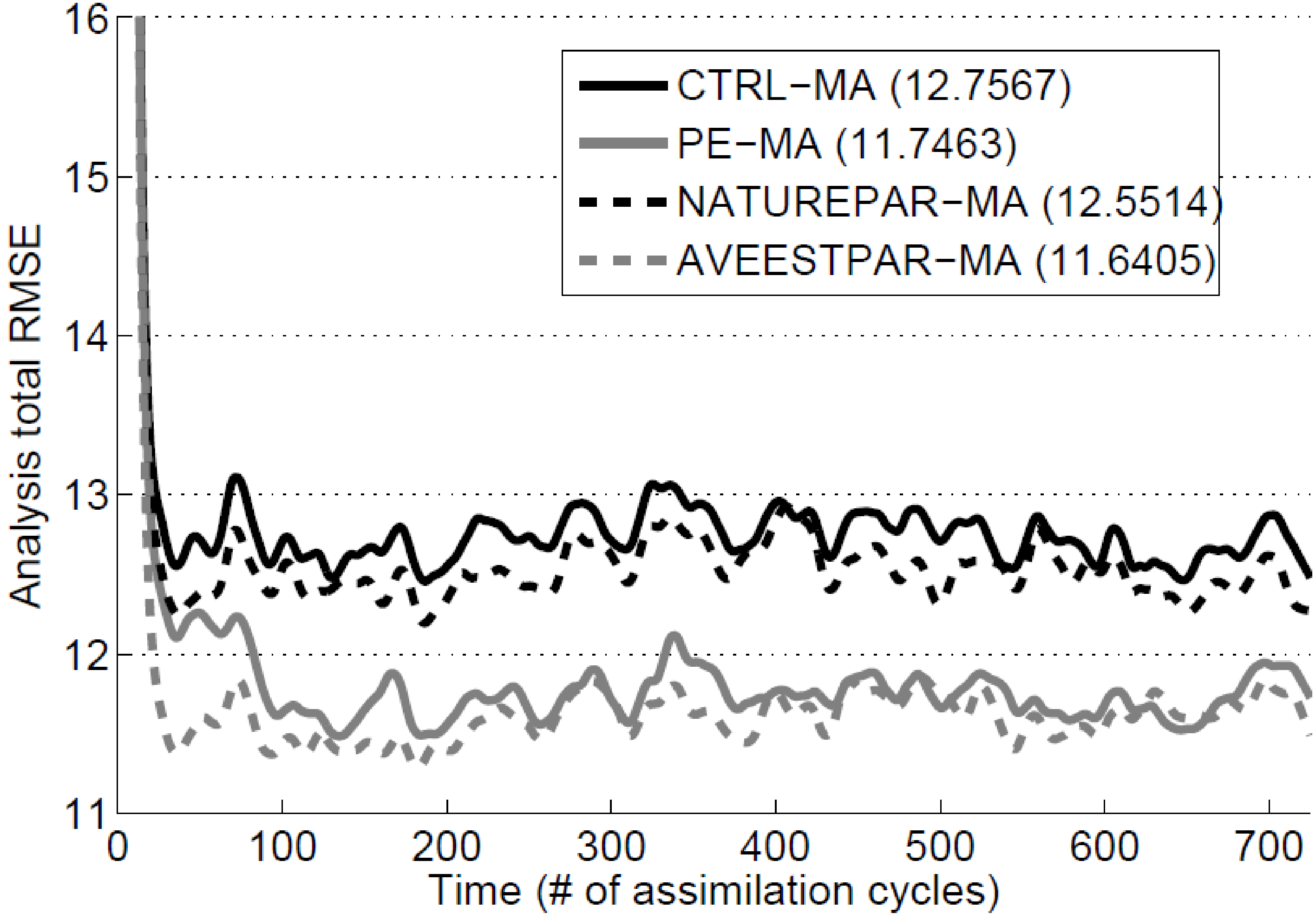
The estimated parameters do not converge to the parameter value used in the nature run.

But we still have convergence of the estimated parameter values.

Estimated parameters shows an increased variability in time.

Idealized experiments:

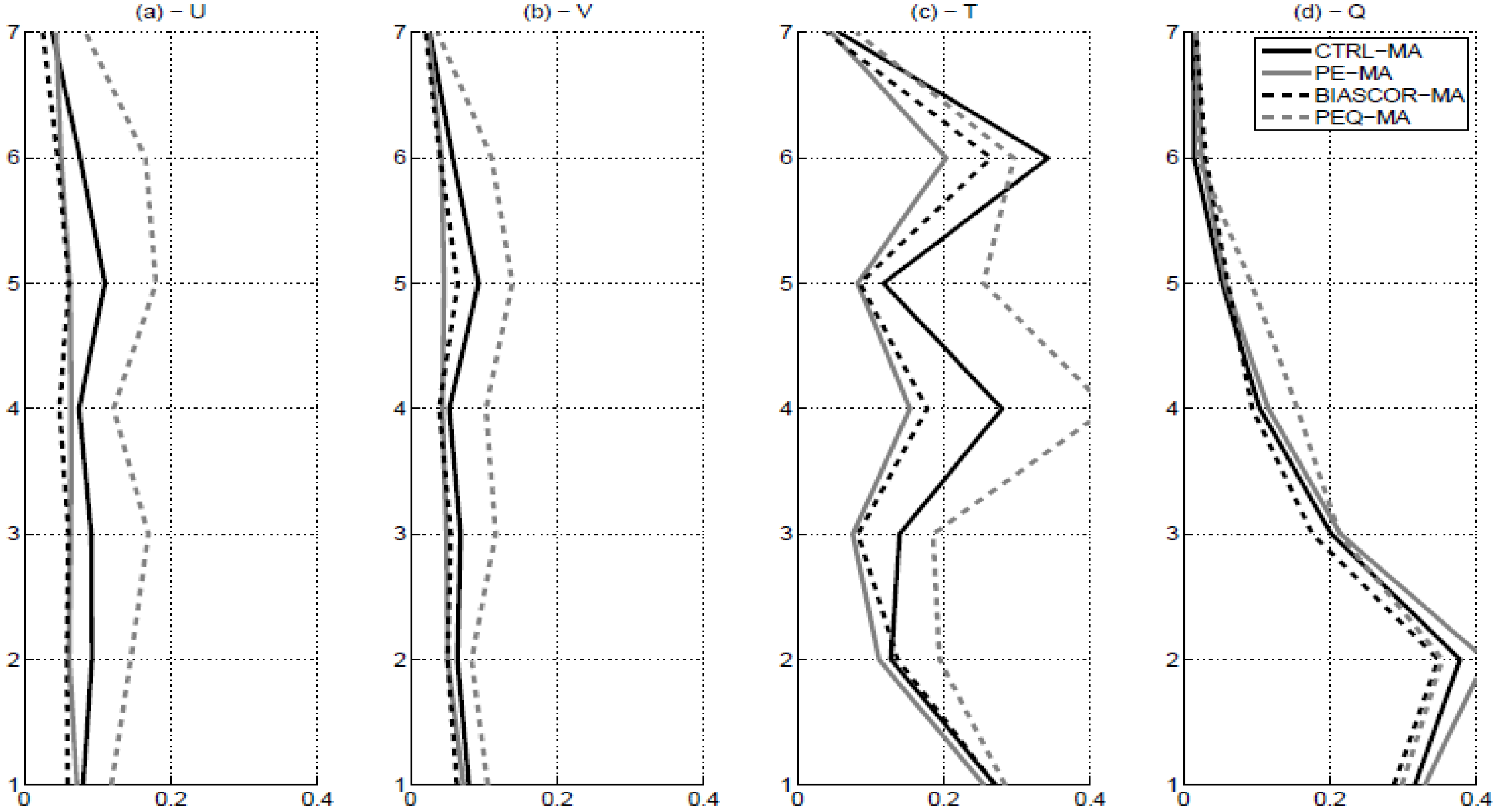
Parameter estimation with an imperfect model
Are estimated parameters useful?



Estimated parameters improves the analysis even more than the nature run parameters. We can call this parameters optimal in the sense that they reduce the analysis RMSE.

Idealized experiments:

Parameter estimation with an imperfect model
Are estimated parameters useful?

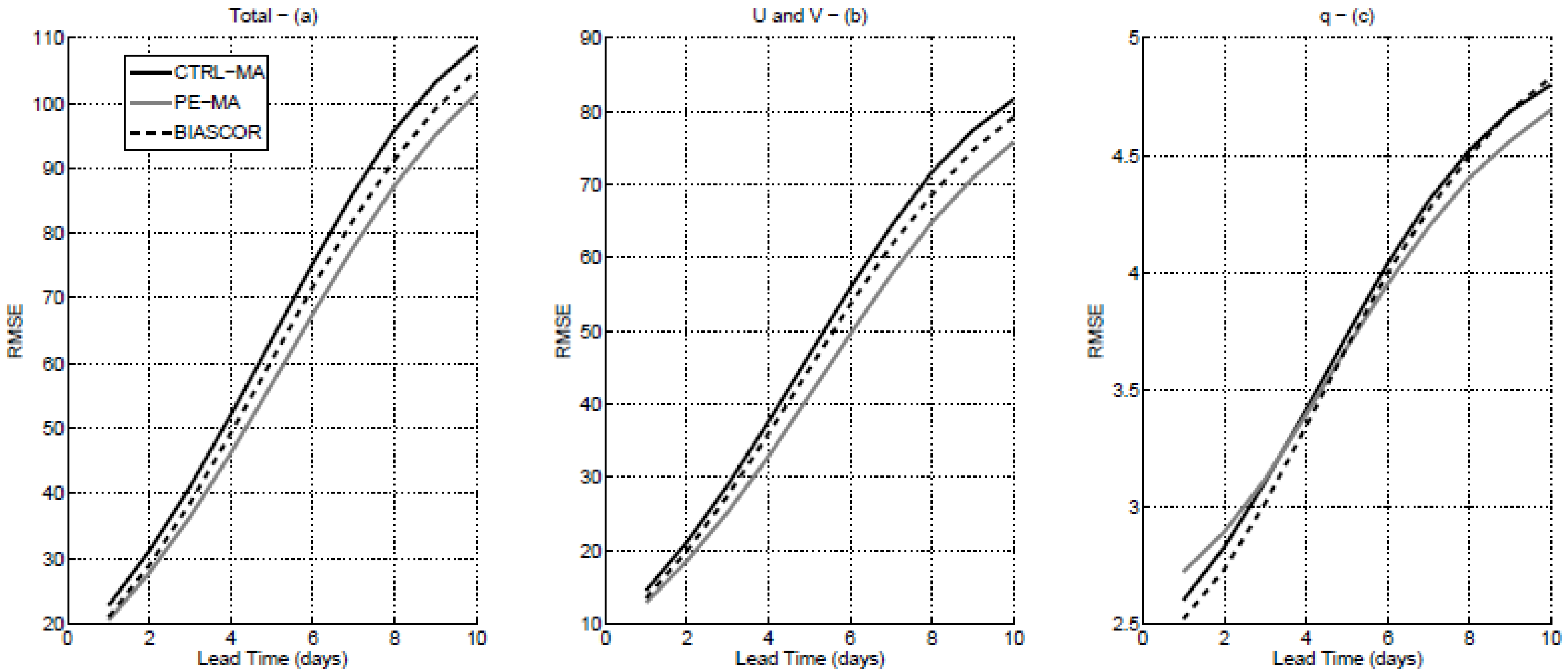


Some variables are improved, some others are degraded

Properly selecting the observations we can improve specific variables

Idealized experiments:

Parameter estimation with an imperfect model
Are estimated parameters useful?

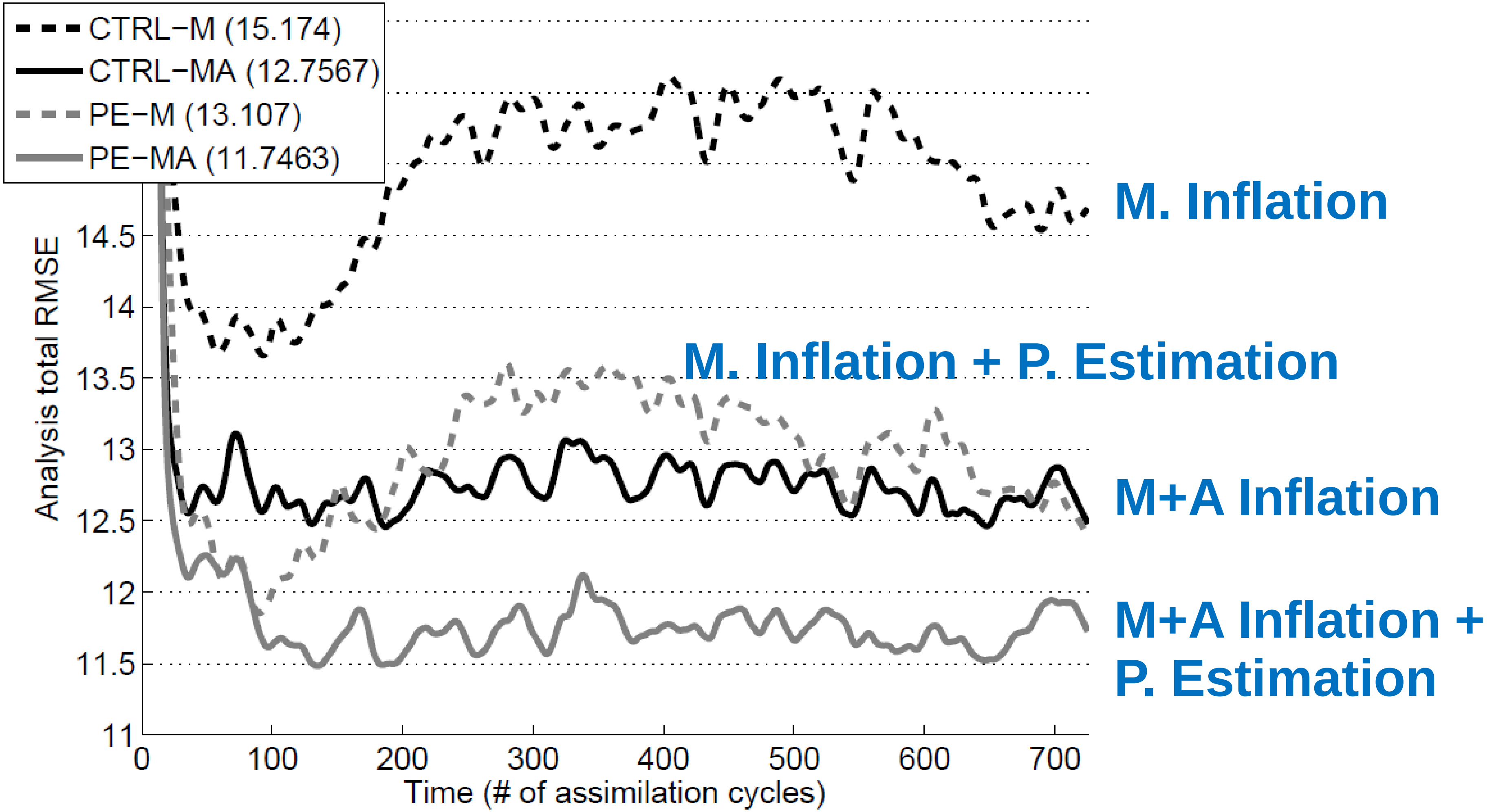


Some variables are improved, some others are degraded

But the overall impact is positive

Idealized experiments:

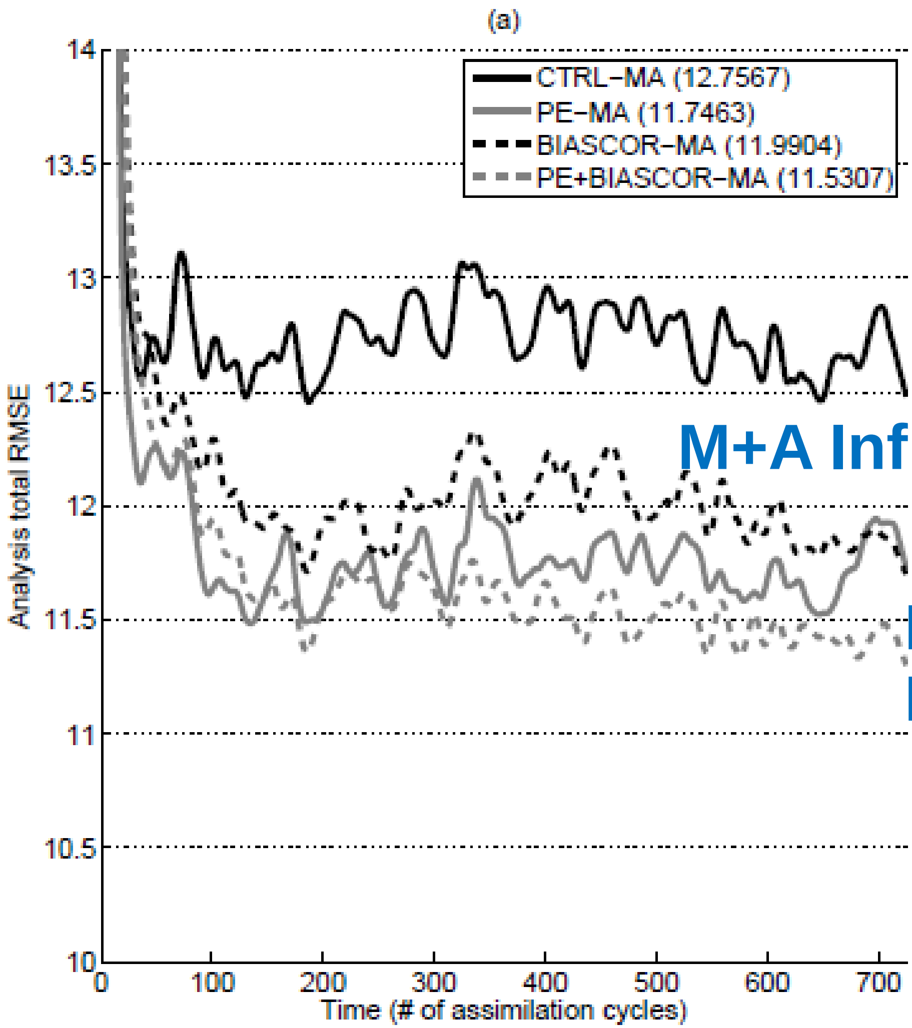
Can we combine parameter estimation with other model error treatment techniques?



Parameter estimation combined with each of these techniques produces further reduction in the analysis RMSE.

Idealized experiments:

Can we combine parameter estimation with other model error treatment techniques?



M+A Inflation

M+A Inflation+Bias

M+A Inflation + P. Estimation

M+A Inflation + P. Estimation + Bias

Combination of parameter estimation and bias estimation produce only marginal further improvement.

Can parameter estimation work in a real world application?

In real world applications there are many different sources of model error that interact with each other in a complex way.

Experiment: Motivated by previous works by Ito et al. 2010 and Kang et al. 2012 who showed that surface fluxes or their associated parameters can be estimated we implement parameter estimation for the improvement of surface fluxes in the WRF-LETKF system.

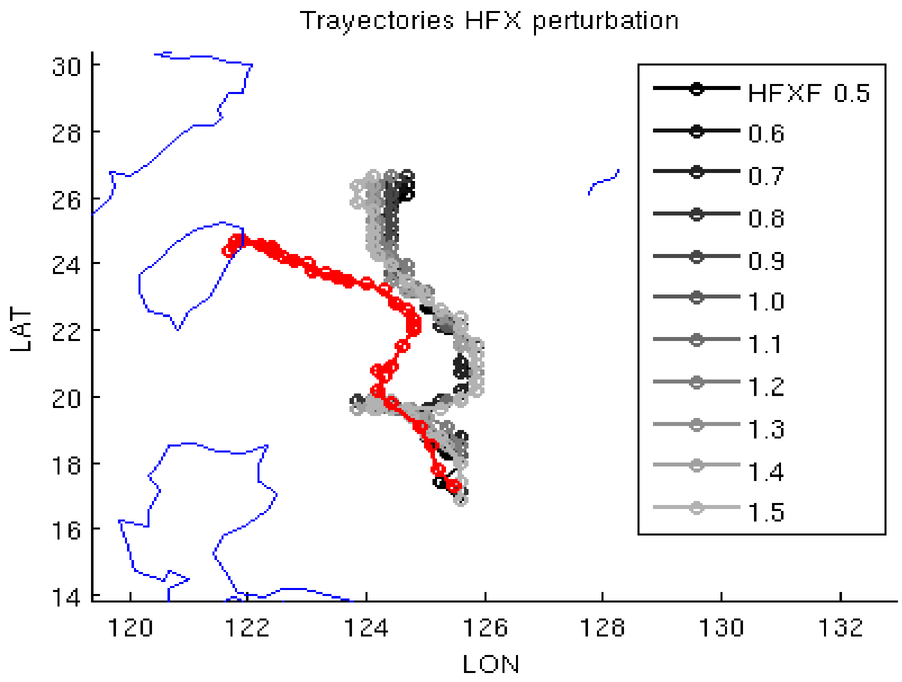
$$F_q = F_q^{WRF} \alpha$$

Simple parameter estimation approach, a multiplicative correction factor is introduced and is estimated using the LETKF-WRF system.

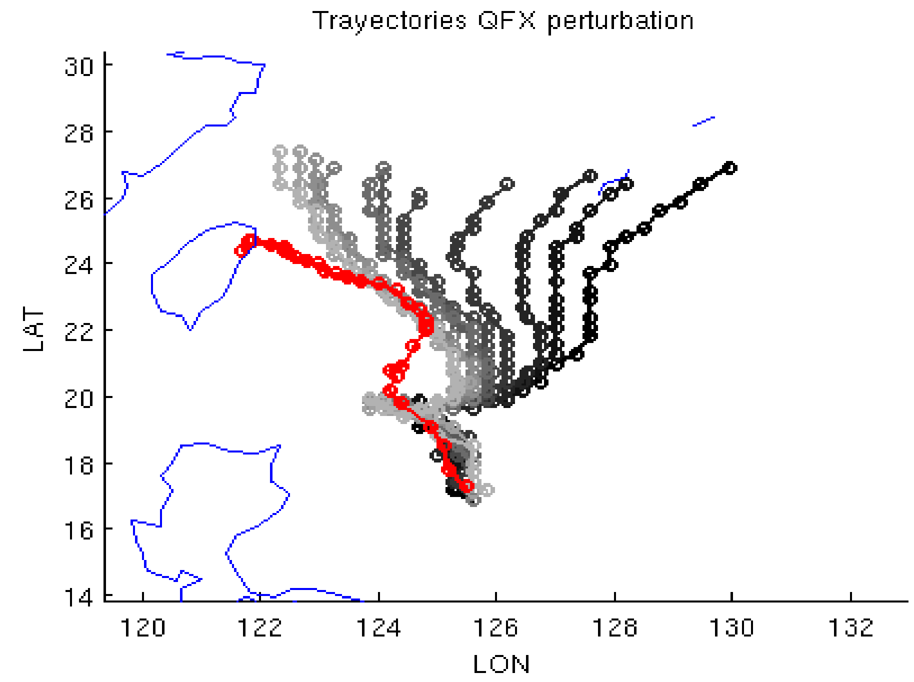
Real world experiments:

TC Sinlaku (2008)

Less sensitive (heat exchange)



More sensitive (latent heat exchange)



Given the stronger impact of latent heat fluxes we test the methodology focusing on these fluxes.

Ruiz , Miyoshi and Kunii (2014, in preparation)

Real world experiments:

Four parameter estimation experiments has been conducted:

0D parameters with vertical localization: **Parameters are considered global constants and only near surface observations are used to estimate them.**

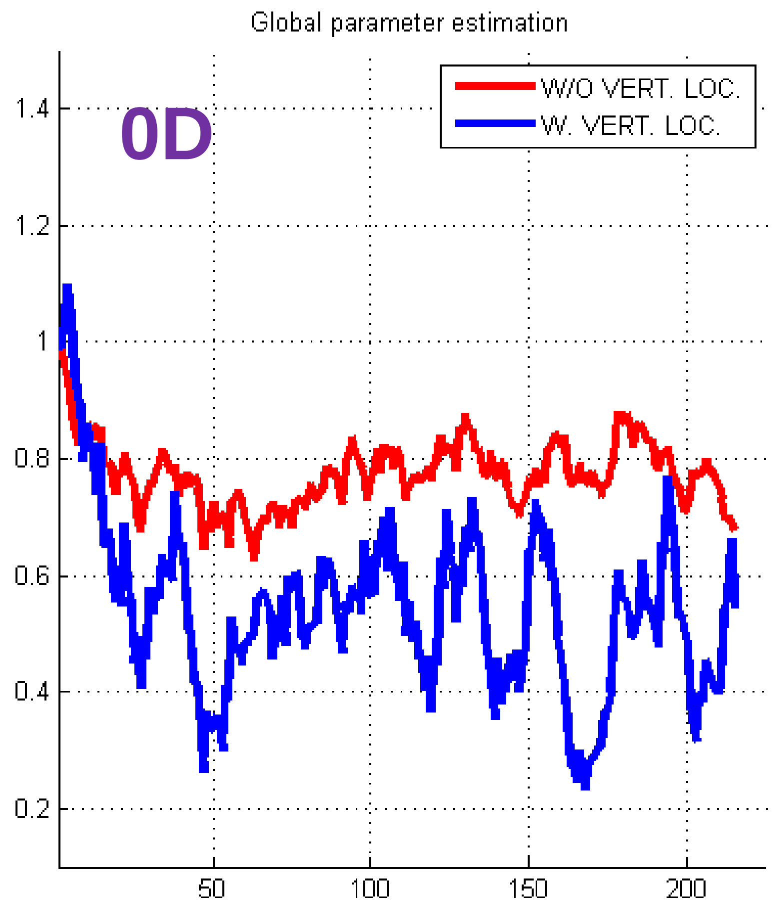
0D parameters without vertical localization: **All the observations in the vertical column are used to estimate the parameters.**

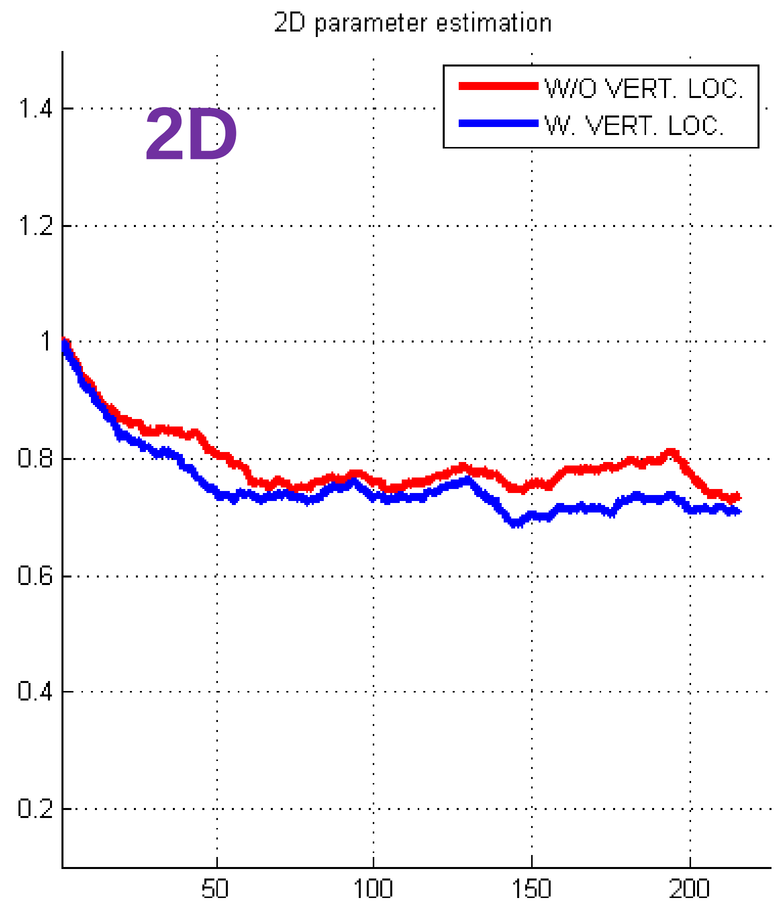
2D parameters with vertical localizations: **Parameters are a function of latitude and longitude only near surface observations are used to estimate the parameters.**

2D parameters without vertical localization: **Idem as before but removing the vertical localization in the estimation of the parameters.**

Real world experiments:

Estimated model parameters as a function of time



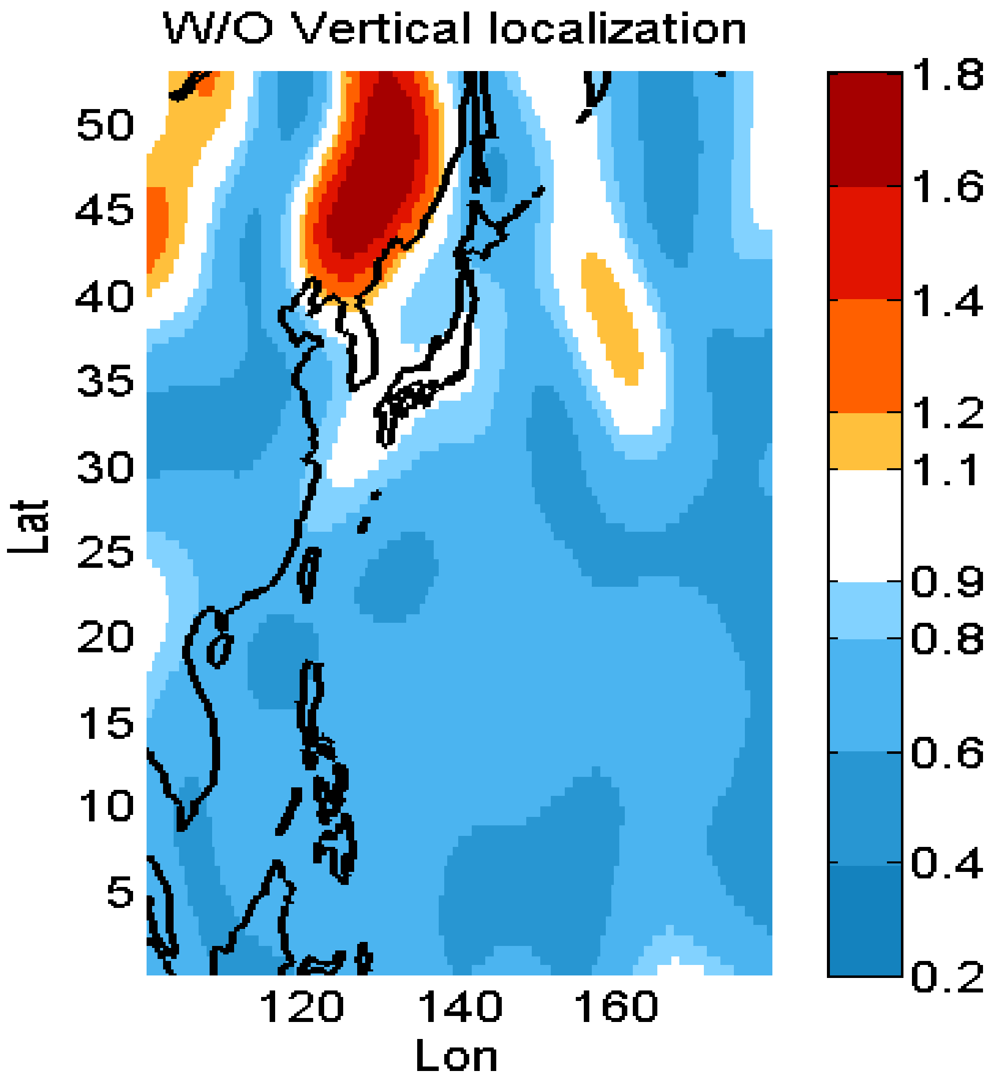


Estimated parameters are below one for all the experiments reducing latent heat fluxes.

Horizontal and vertical localization has an impact on the value of the estimated parameters.

Real world experiments:

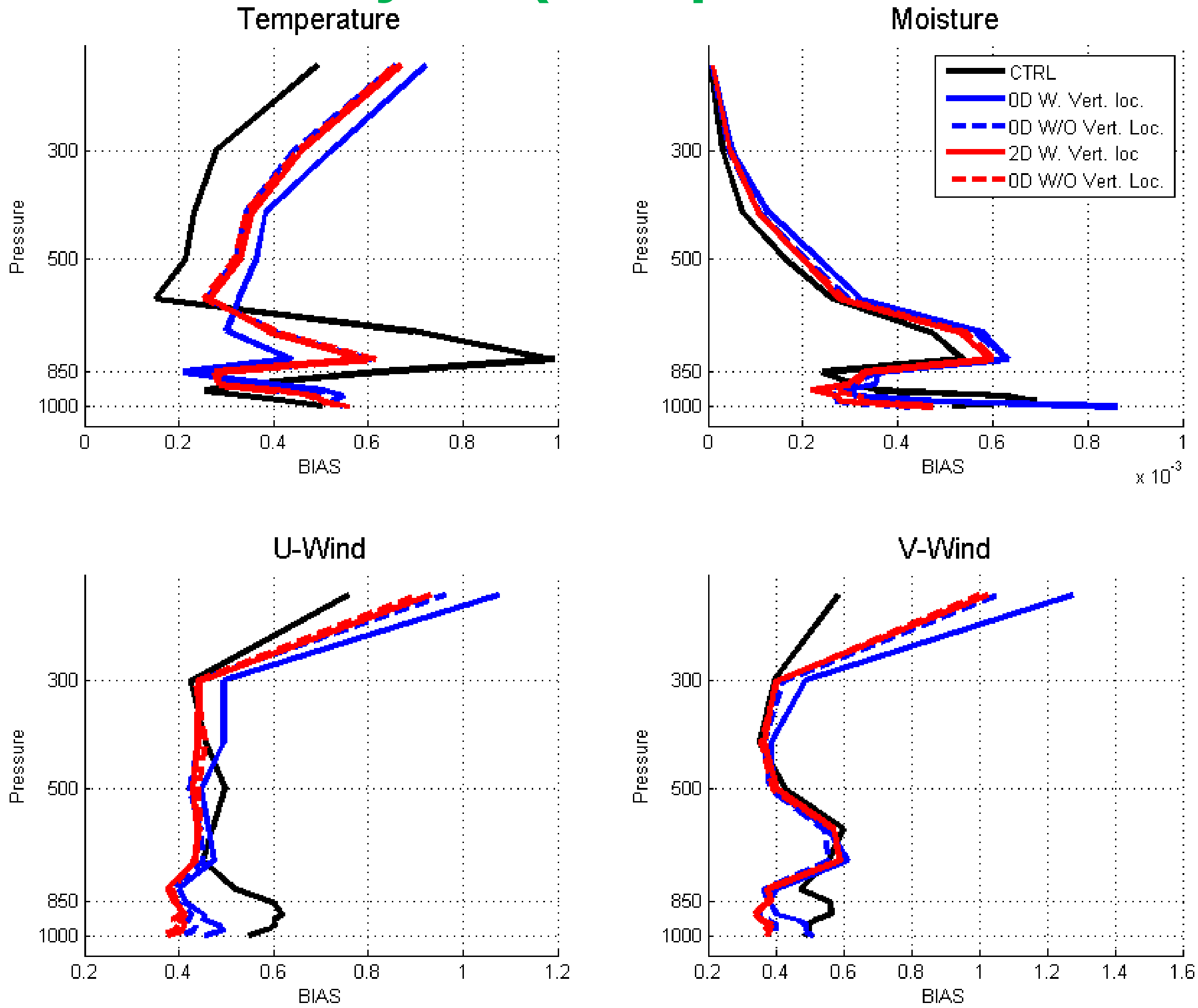
Estimated model parameters



Horizontal distribution is quite homogeneous particularly over the tropical ocean where the model sensitivity to the parameter is stronger.

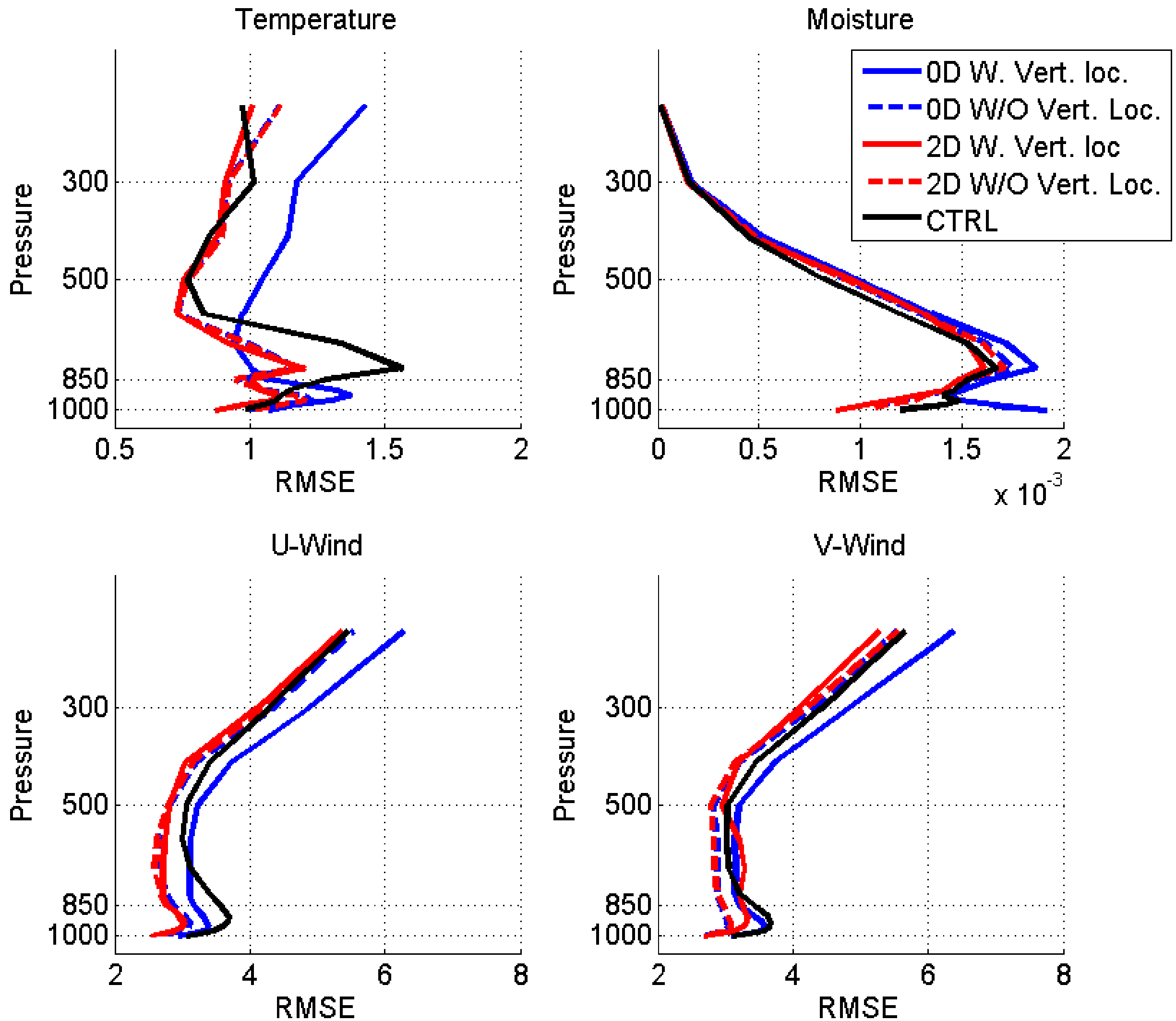
Real world experiments:

Impact upon the analysis (compared with GDAS)



Low level biases are removed in almost all variables.
Upper level biases are increased.

Real world experiments:
Impact upon the forecast (compared with GDAS)

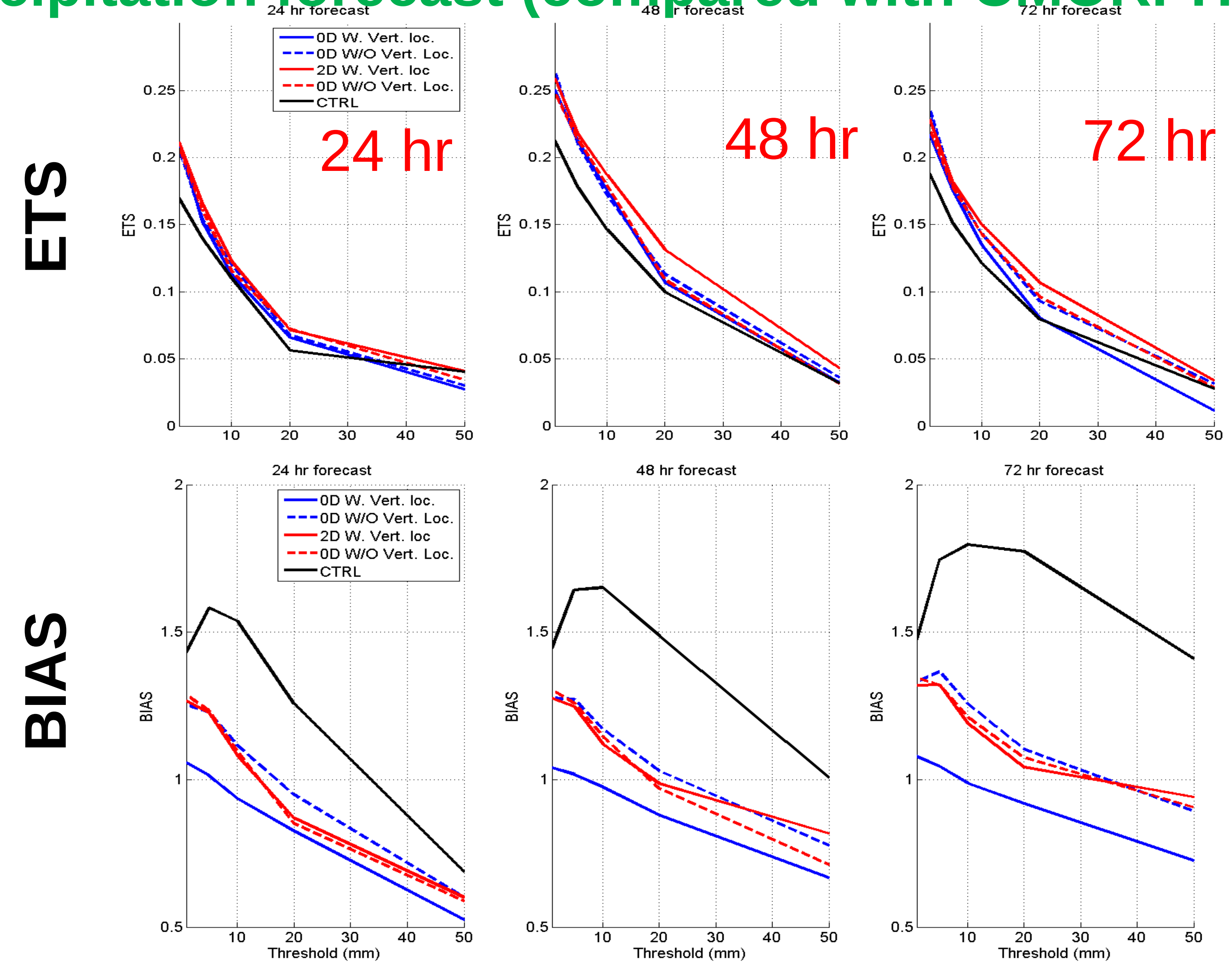


40-member ensemble forecast up to 72-hr lead time.

For the 72 hr forecast most parameter estimation experiments shows an improvement for T, U and V. Q is improved at low levels but degraded at upper levels.

Real world experiments:

Precipitation forecast (compared with CMORPH)

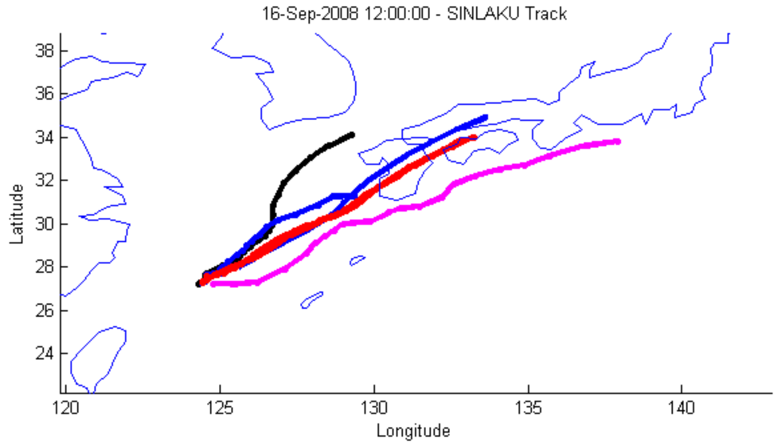


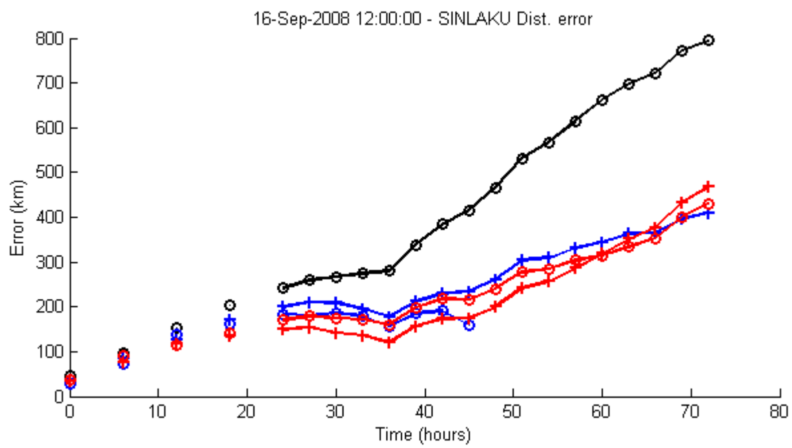
Precipitation forecast improved ETS. Precipitation frequency decreases .

Real world experiments:

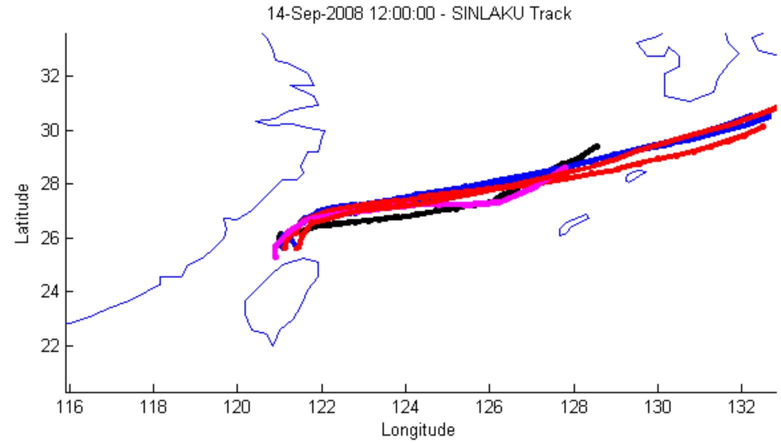
Impact upon TC forecast

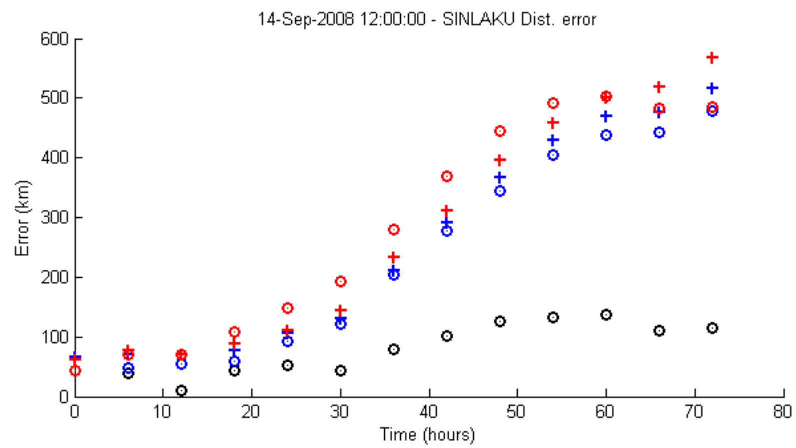
Forecast improved





Forecast degrade

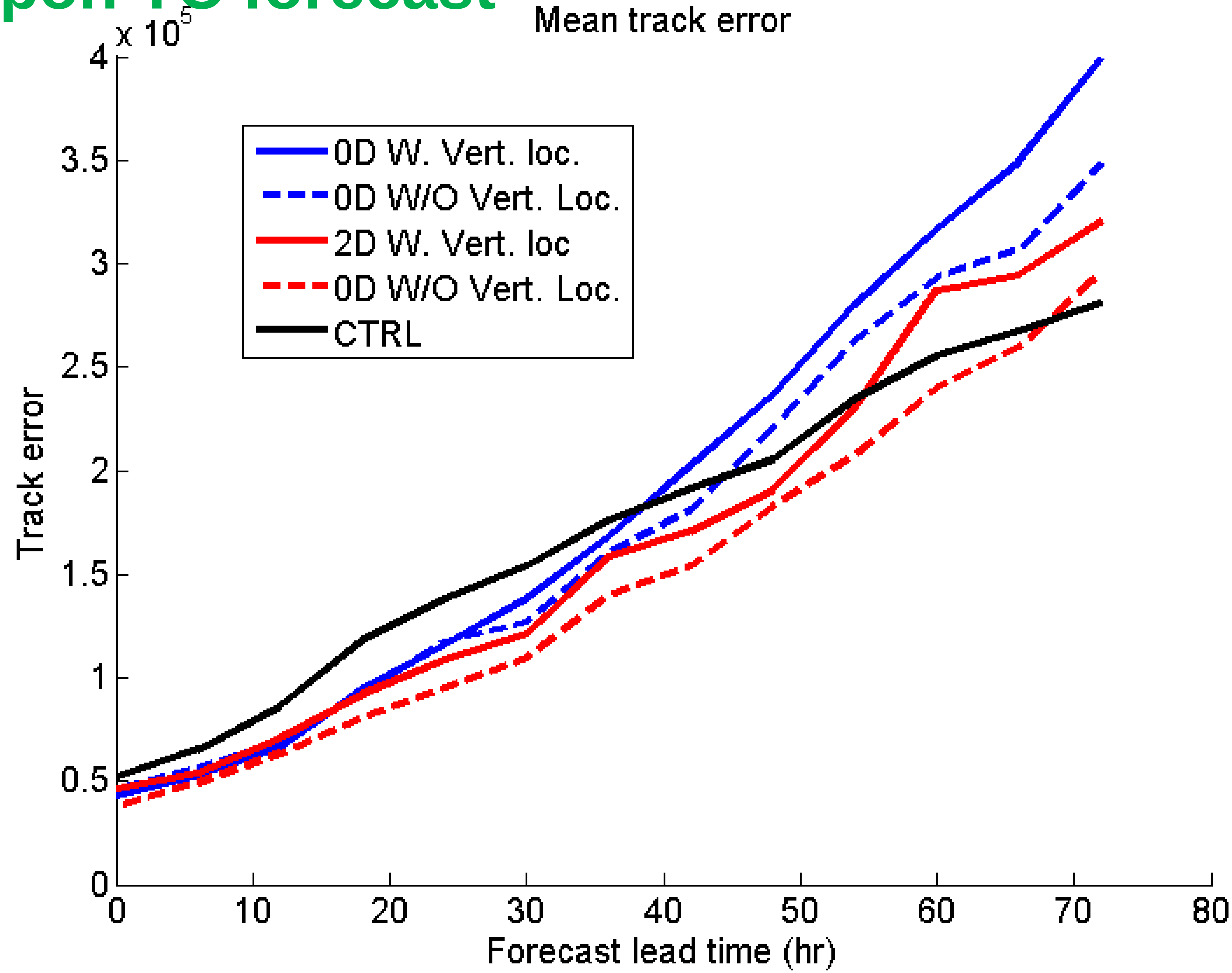




Some cases shows a consistent improvement while others shows a consistent degradation...

Real world experiments:

Impact upon TC forecast



The mean track error is better for the 2D parameter estimation experiments.

The sample is too small to have robust results.

Real world experiments:

Parameters are successfully estimated using the LETKF-WRF system. In all the experiments parameters indicate that moisture surface fluxes are too strong and are possible responsible for a too much moisture at low levels.

Although small, parameter estimation impact upon the forecast is positive for most variables. The impact upon the TC forecast is still unclear although results suggest that estimated parameters can potentially improve TC forecasting.

Localization has an impact upon the estimated parameters. Best results has been obtained with 2D parameters.

One final thought:

Parameter estimation is not a way for model improvement, this ultimate goal can only be reached by the improvement of our understanding of the physical processes and to its application to the development of more realistic parametrizations of the unresolved-scale processes (Jakob, 2010).

However, model error and errors arising from parametrization will be there for a long time. Data assimilation based parameter estimation can provide an efficient way to optimize multiple parameters from different parametrizations in an ever changing model scenario.

Thank you!!

Can we find parameter values that improves the model climatology?

In a long climatological run the model sensitivity to the parameters might change as a result of the interaction of different sources of model error in a longer time scale.

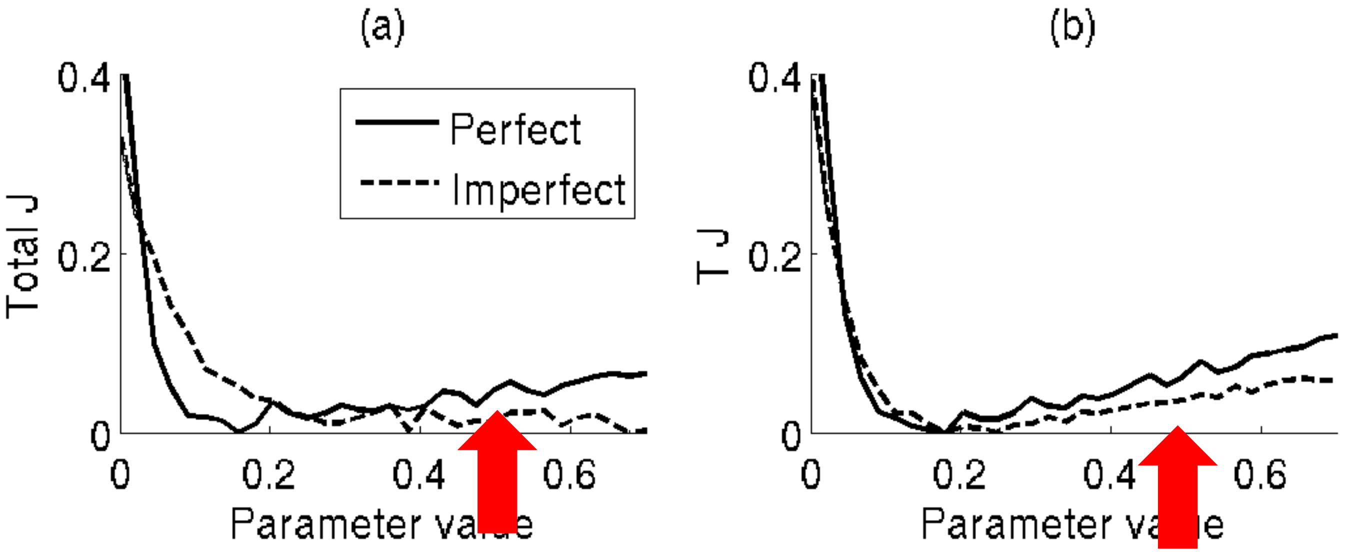
Annan 2005 proposed a method to improve the model climatology but performing the assimilation in the long time scale.

Rodwell and Palmer (2007) found that model biases in the first forecast time steps are associated with model biases in longer time scales.

Schirber et al. 2013 showed that model climatology can be improved by the optimization of model parameters using a data assimilation cycle.

Idealized experiments:

Long therm model sensitivity to changes in the parameter



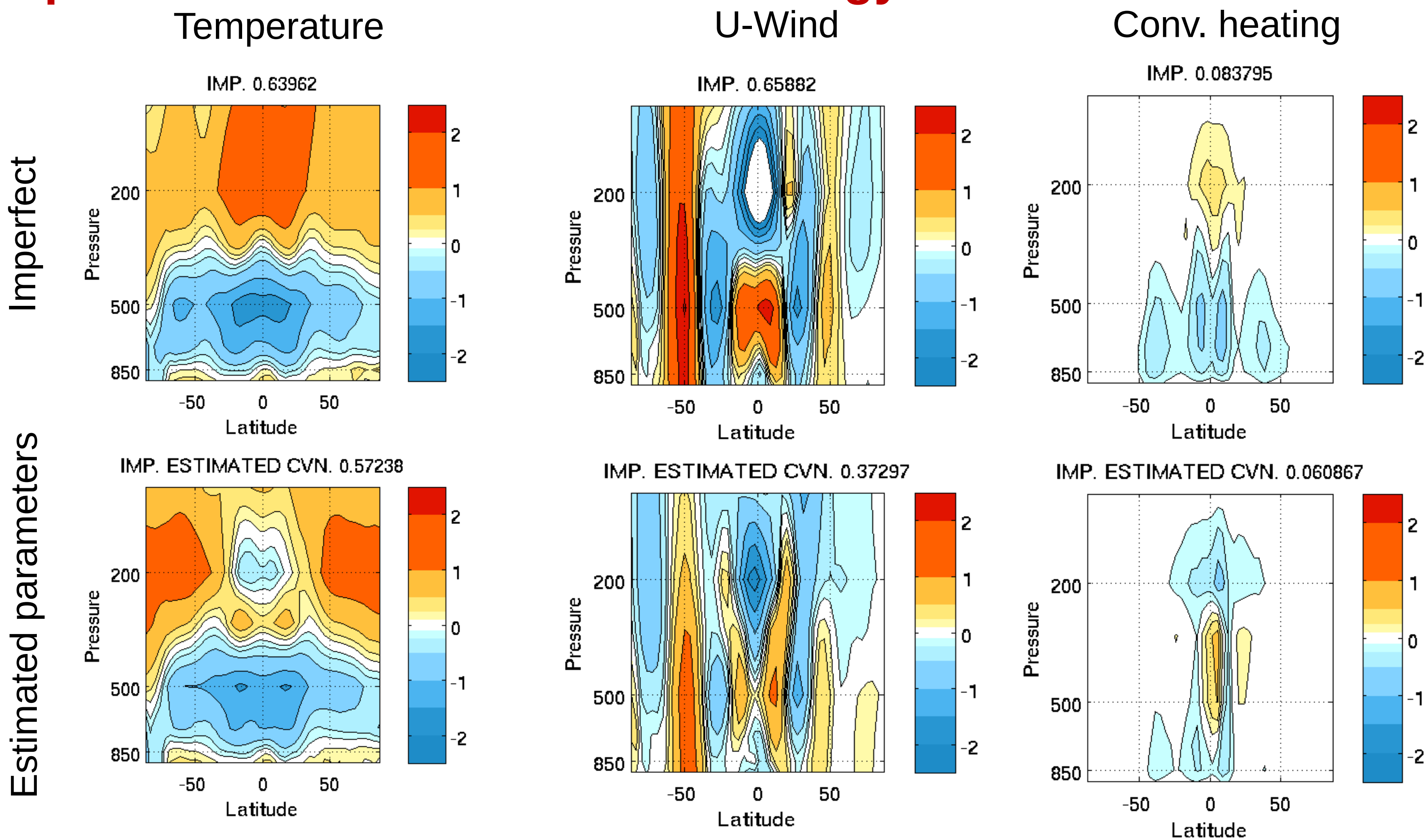
RMSE of the model climatology as a function of TRCNV value.

The parameter that produces the best representation of the climatology depends on the considered variable. (Schirber et al 2013)

The estimated parameters fall within the region of weak model sensitivity and low total RMSE.

Idealized experiments:

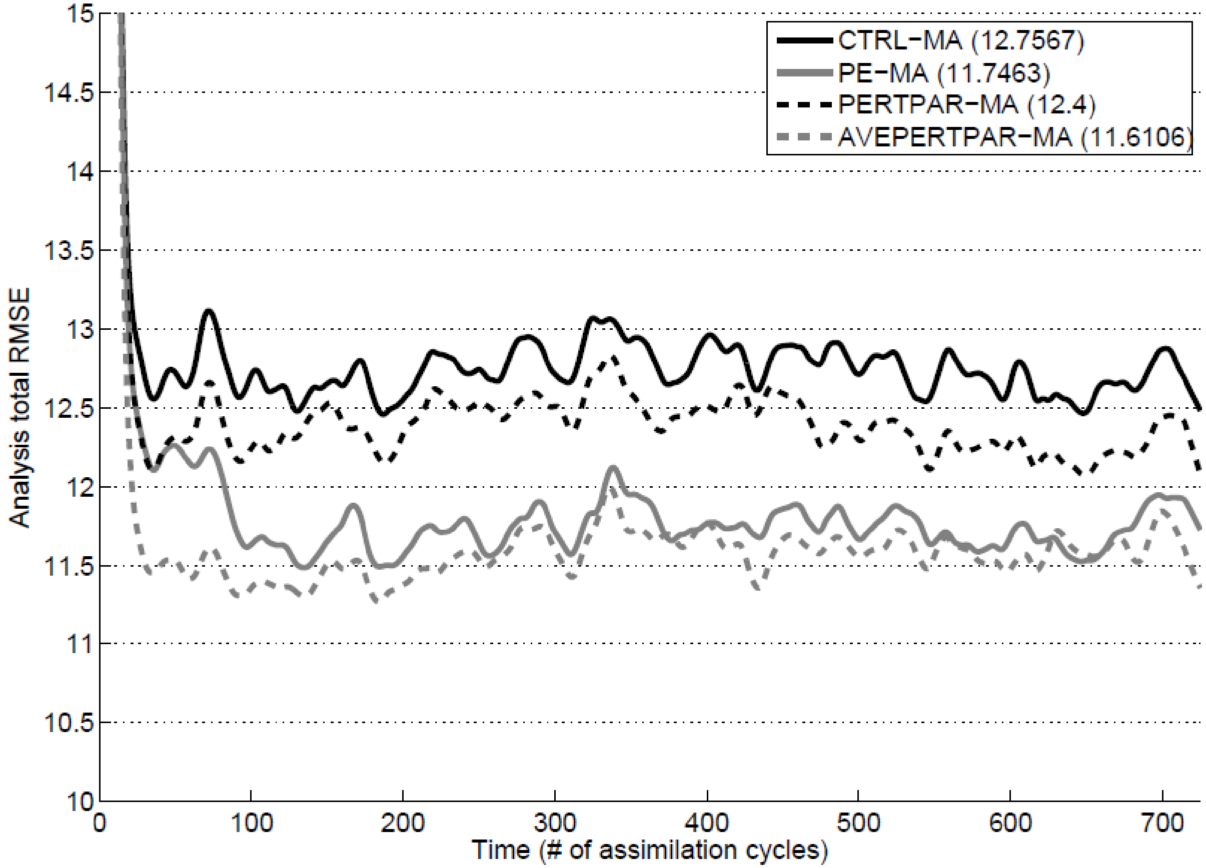
Improvement in the model climatology



Most variables are improved when the estimated parameters are used.

Idealized experiments:

Where does the improvement come from?

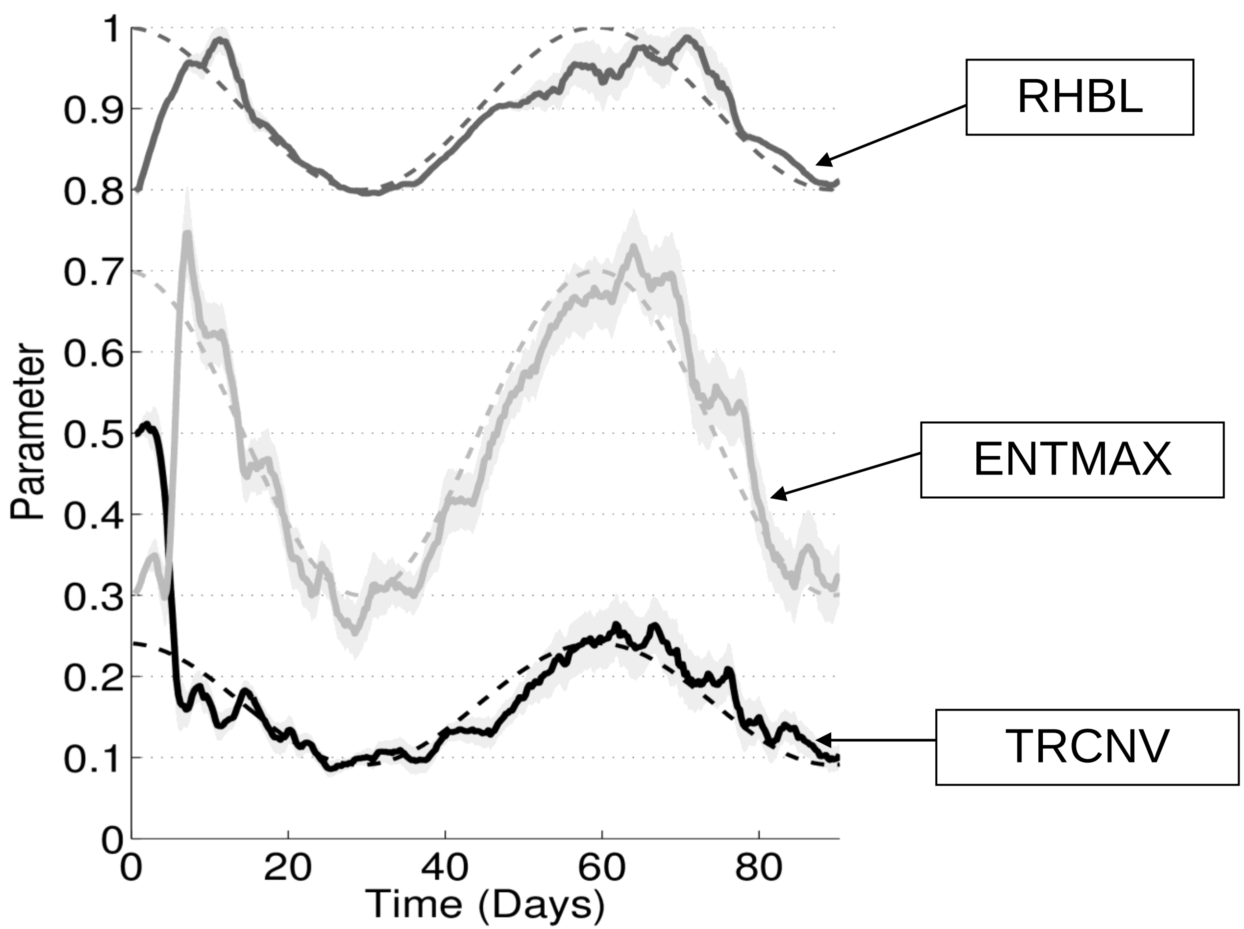


Perturbing the parameters increases the ensemble spread. It also introduces state dependent perturbations that contains information about model errors.

Most of the improvement is produced by the update of the parameter values.

Idealized experiments:

OSSE with “almost” perfect model and time dependent parameters



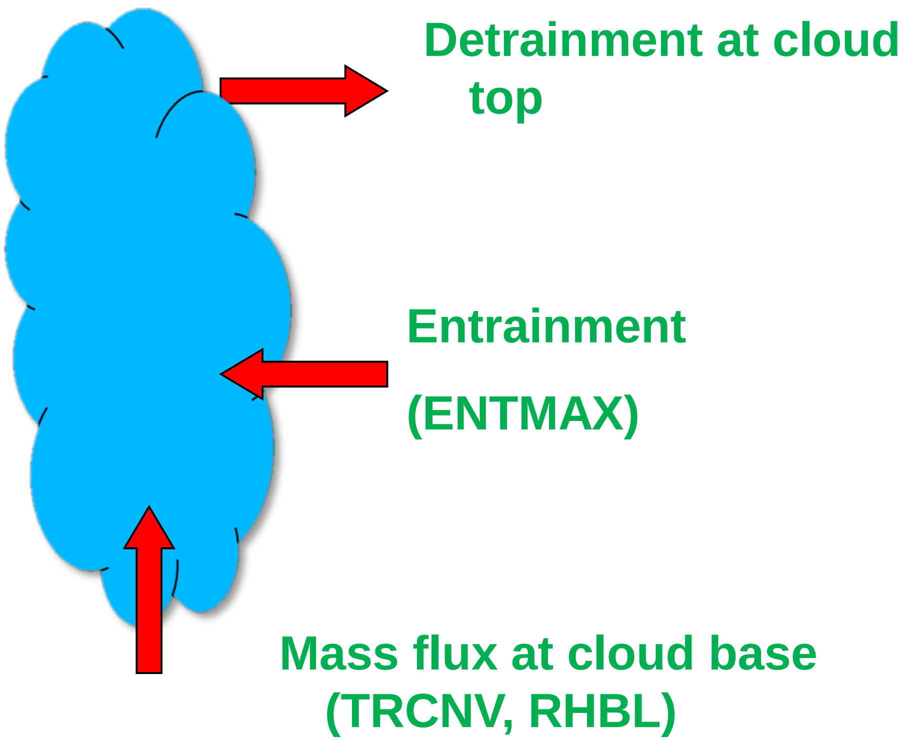
Time evolution of estimated parameters and their uncertainty

Time dependent parameters are well estimated. However a small lag is present in the estimated parameters.

Idealized experiments:

Convective scheme parameter estimation in the T30L7 SPEEDY model (Molteni 2003).

Three parameters are most important: RHBL, ENTMAX and TRCNV.



Convection produce a high impact in the system

It is also challenging since it is intermittent in time and space.