Development of subgrid parameterizations using ensemble-based data assimilation

Manuel Pulido

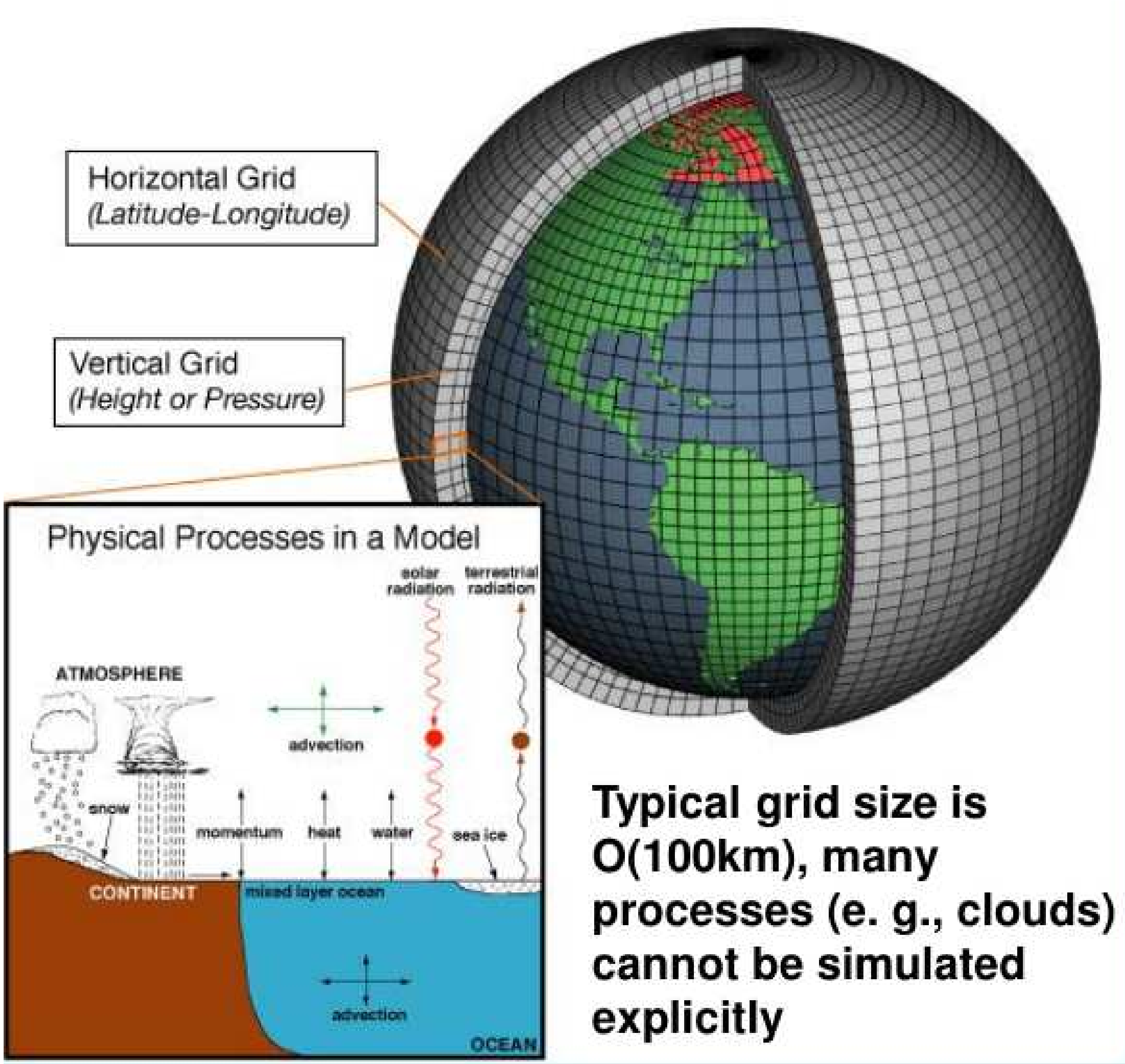
(1) Department of Physics, Universidad Nacional del Nordeste, Corrientes, Argentina

(2) CONICET, Argentina

How to build/improve models from observational data?

In collaboration with: Guillermo Scheffler, Juan Ruiz, Magdalena Lucini and Pierre Tandeo.

Physics in climate models



From Latif

Physics and small-scale dynamical processes in climate models have to be represented by parameterizations: Convection, planetary boundary layer, radiation, turbulence, gravity waves, land surface-atmosphere interactions, etc.

There are a large number of unknown parameters in the physical parameterizations.

Representing “nature” with a two scale model

Two scale ’nature’ model. Large-scale equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}_L^T + M_L(\mathbf{x}_L^T) + M_{LS}(\mathbf{x}_L^T, \mathbf{x}_S^T) = 0$$

where $\mathbf{x}_L^T$ large-scale true model state, M_L large-scale model, M_{LS} interaction model between large and small scales.

Small-scale variable equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}_S^T + M_S(\mathbf{x}_S^T) + M_{SL}(\mathbf{x}_S^T, \mathbf{x}_L^T) = 0$$

Unknown in the forecast model

where $\mathbf{x}_S^T$ small-scale true model state, M_S small-scale model, M_{SL} interaction model between small and large scales.

Subgrid parameterization

The forecast model evolution only represents the large scale variables:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}_L^f + M_L(\mathbf{x}_L^f) = \mathbf{f}(t, \mathbf{x}_L^f)$$

where $\mathbf{x}_L^f$ is the forecast state. M_L is assumed to be “perfect”.

The forcing term is represented through a subgrid parameterization:

$$\mathbf{f}(t, \mathbf{x}_L^f) = Parameterization(t, \mathbf{x}_L^f, \{a_i\})$$

The parameterization should represent the interaction term $-M_{LS}(\mathbf{x}_L, \mathbf{x}_S)$ where $\mathbf{x}_S$ is unknown for the forecast model.

The two-scale separation hypothesis is intrinsic to the sugrid parameterization concept.

Can we improve a climate model using observational information?

Large number of free parameters in climate models

$\longrightarrow$ Observational information can be used to constrain or estimate model parameter values. Ruiz et al. JMSJ 2012.

Some of the current parameterizations have difficulties to represent the observed variability (e.g. Shutts QJ 2015).

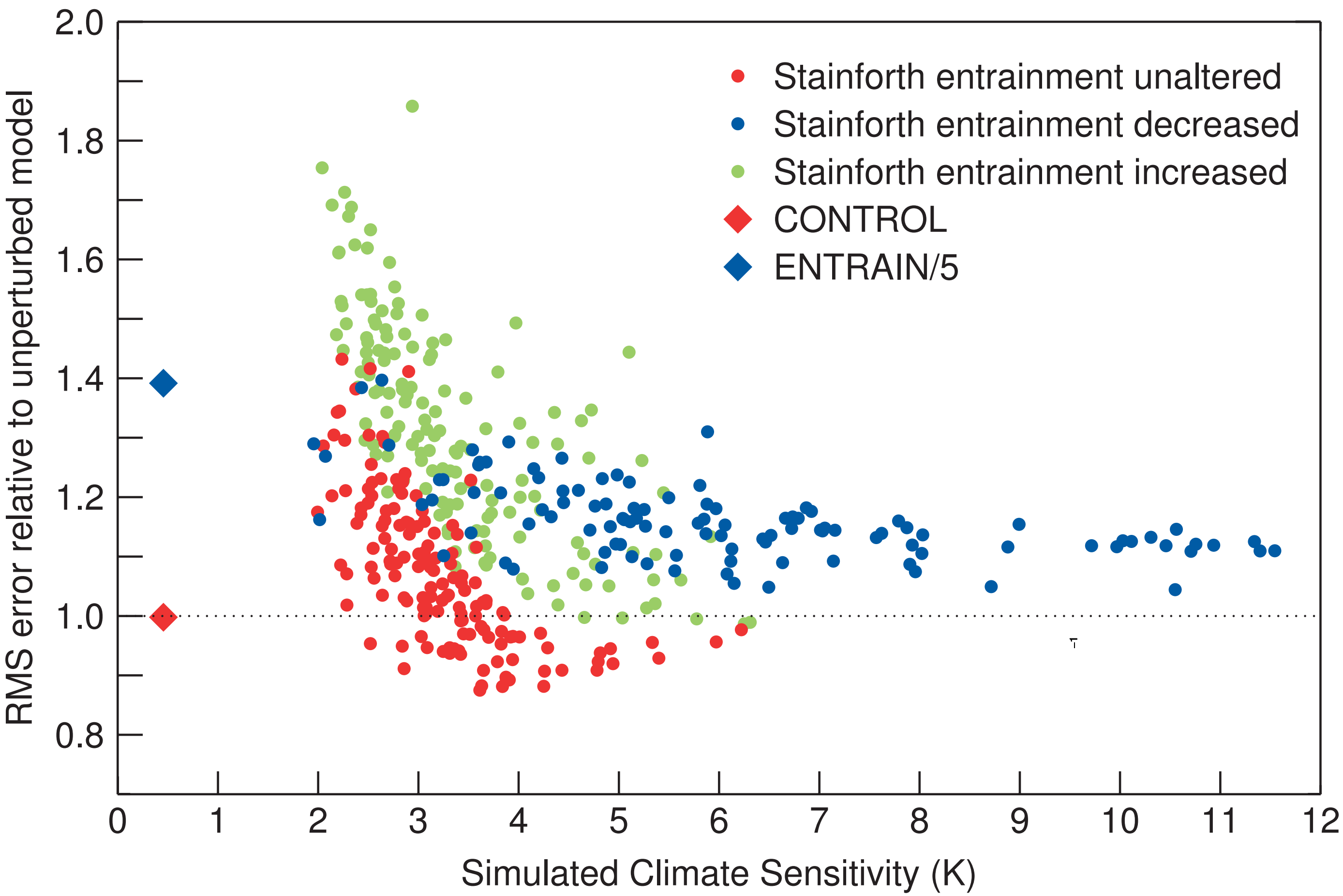
$\longrightarrow$ Observed variability can be used to build stochastic parameterizations.

There are still further limitations in the parameterizations.

$\longrightarrow$ A further step is to improve parameterizations using observations.

Can we determine functional dependencies of the parameterizations to the model state using observations?

Climate sensitivity = parameter sensitivity?



Dots represent different entrainment parameter values (three colors) and initial conditions.

From Stainforth, et al. Nature, 2005.

climateprediction.net

RMSE of climate simulations as a function of the response of the mean global surface temperature produced by a doubling of CO2.

Parameter values may have a large impact, however (long term) RMSE values are similar. Which is better?

Climate vs forecast errors

Finding the source of a systematic error in climate simulations is difficult because feedbacks and compensating errors.

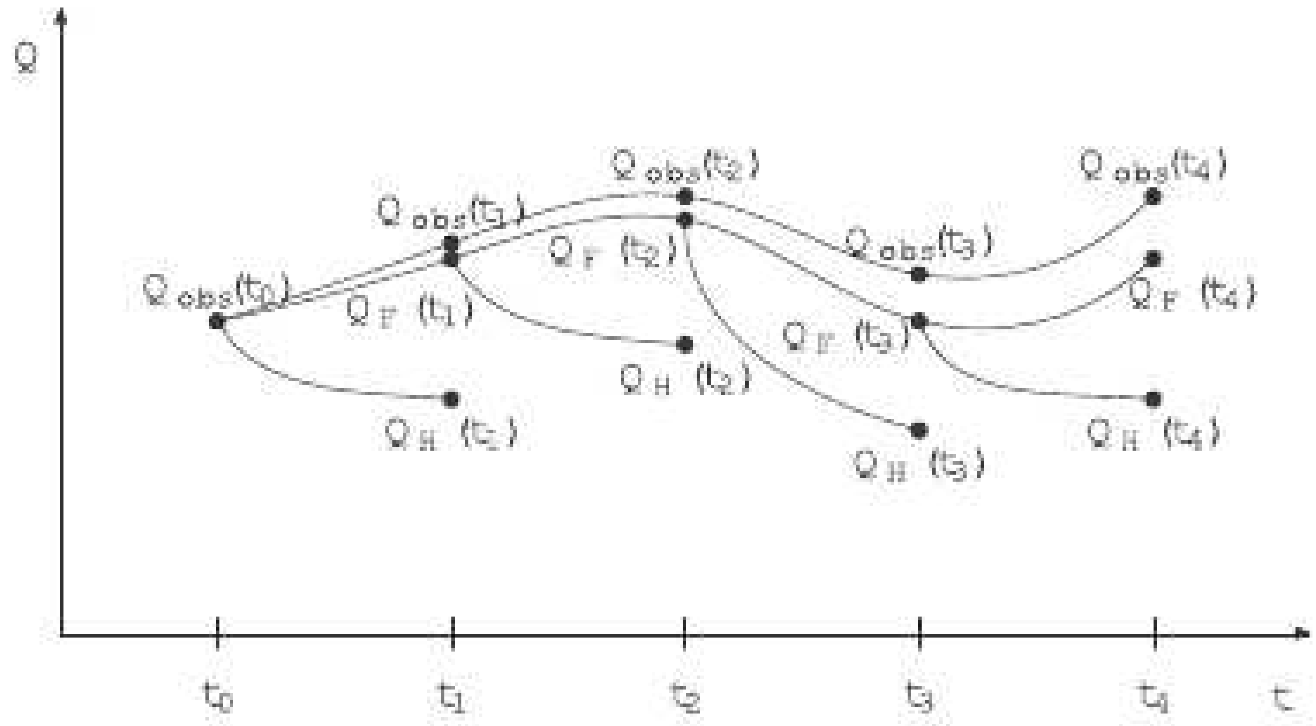
The large-scale state remains close to observations only in the first few days of forecasts.

The errors are likely the results of parameterization errors associated with fast physics instead of slowly evolving feedbacks.

Short-term errors are more easily tracked towards their sources (e.g. Pulido and Thuburn, QJ 2005; Rodwell and Palmer, QJ 2007)

Data assimilation cycles are ideal to identify sources of model error.

Forcing estimation with data assimilation

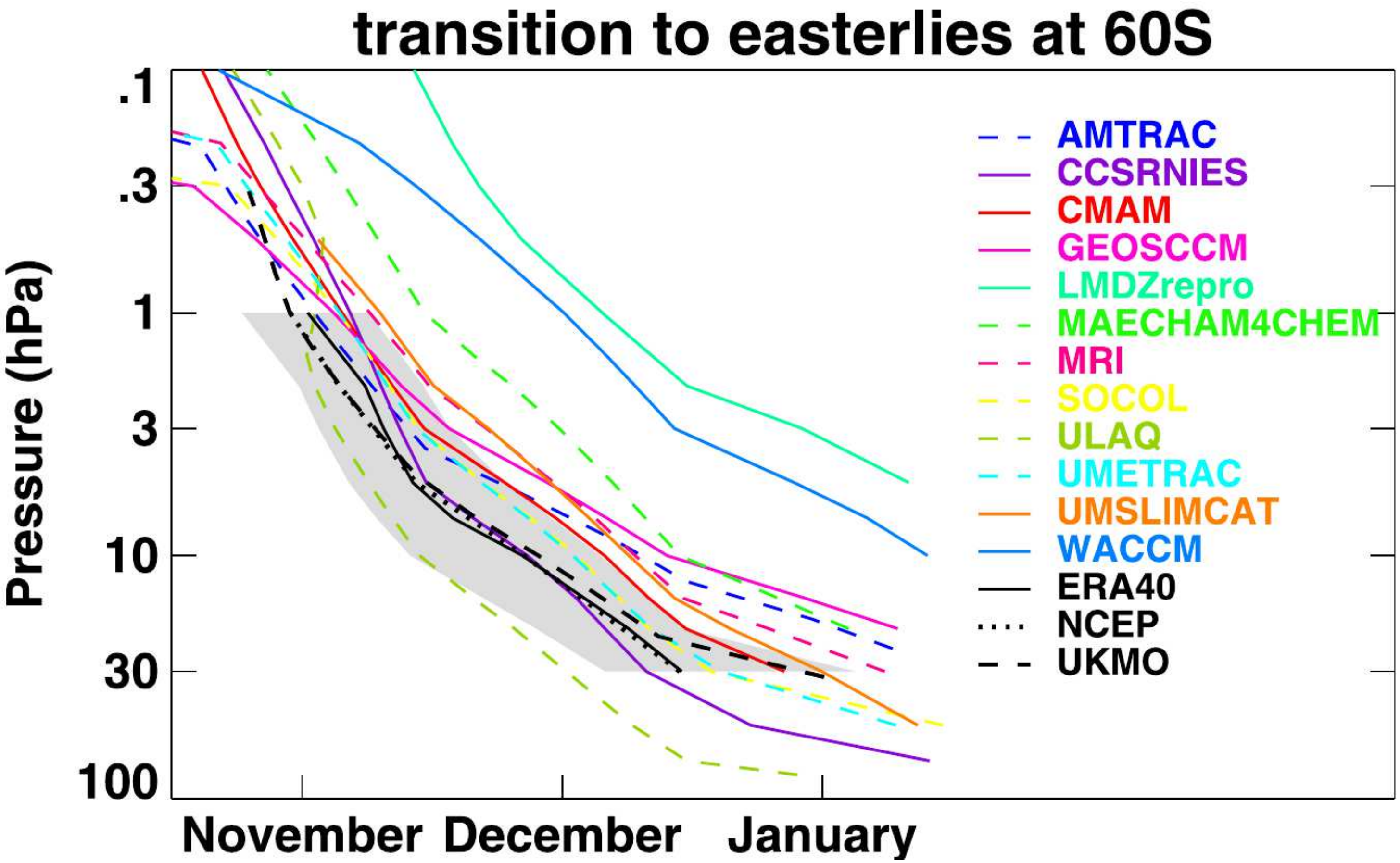


From Pulido, JAS, 2014

Instead of estimating $\mathbf{x}$ with data assimilation we estimate $\mathbf{f}$ or $(\mathbf{x}, \mathbf{f})$. 4-D variational data assimilation or an Ensemble Kalman filter may be used.

Large number of assumptions: Error dominated by a single parameterization. Observations are spatio-temporally dense (to constrain $\mathbf{f}$). Forcing is constant in the ass cycle.

Model bias in the stratosphere



Eyring, et al. JGR, 2006. Climate-Chemistry model assesment.

Butchard et al. JGR, 2011.

Analysis in black (ERA40, NCEP, UKMO).

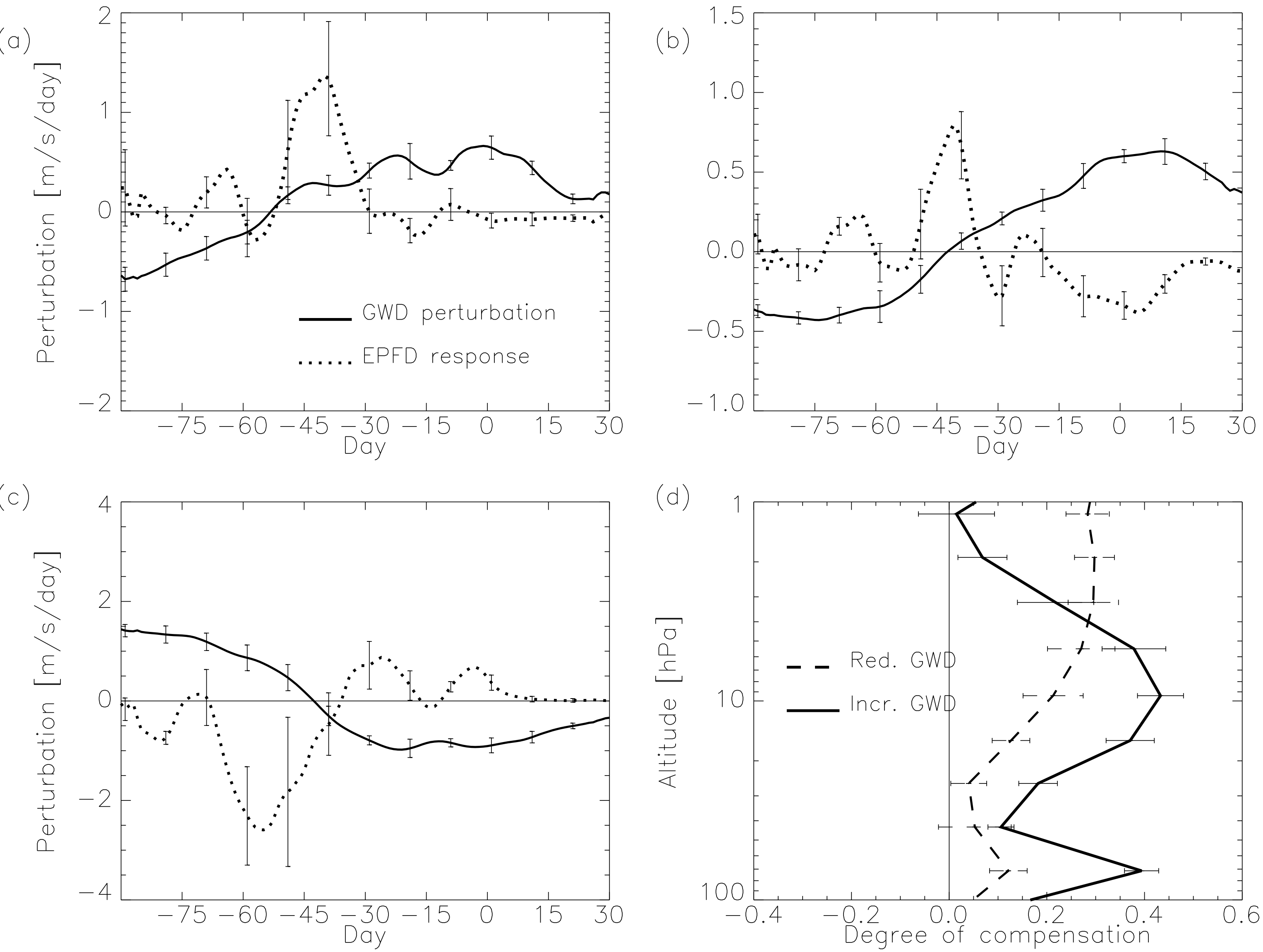
Most of the models have a two-week delay in the polar vortex breakdown compared to observations.

Delay in the polar vortex breakdown in the models

The candidate for the delay is the gravity wave parameterization.

Why is so difficult to manually find good gravity wave parameters?

Resolved-Parameterization Interactions

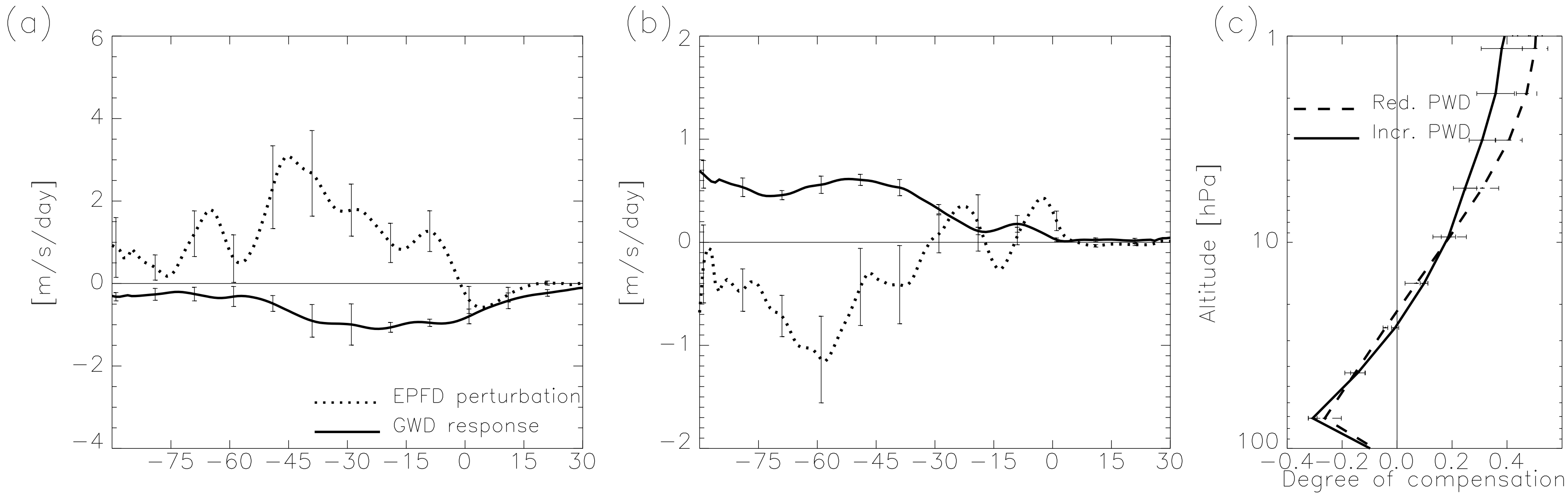


Time series at 2hPa.

Experiments in which we increase and decrease the momentum flux parameter.

From Scheffler and Pulido JAS 2015.

Resolved-Parameterization Interactions

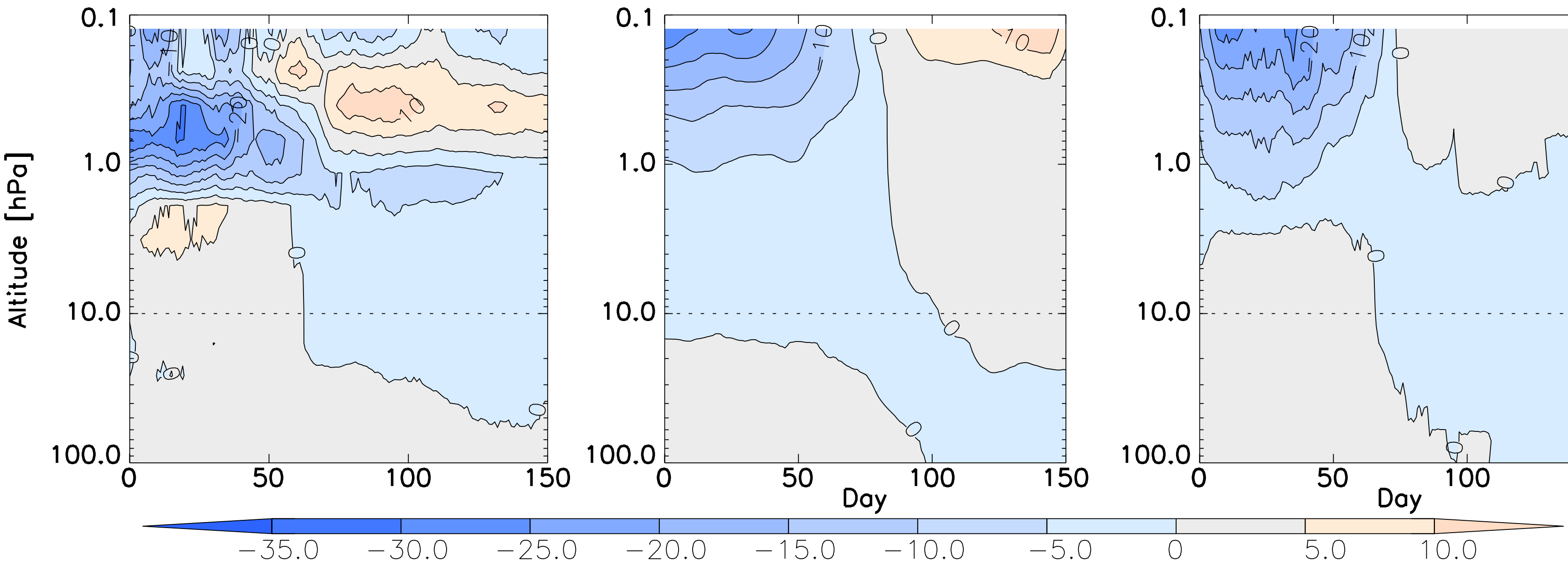


Changes in the parameters are compensated by the resolved scale dynamics. Changes in the resolved scale dynamics are compensated by the parameterization.

Mechanism: gravity wave-planetary wave interactions through mean flow changes (Cohen et al. JAS 2013, JAS 2014) .

Estimated forcing with 4D-Var data assimilation

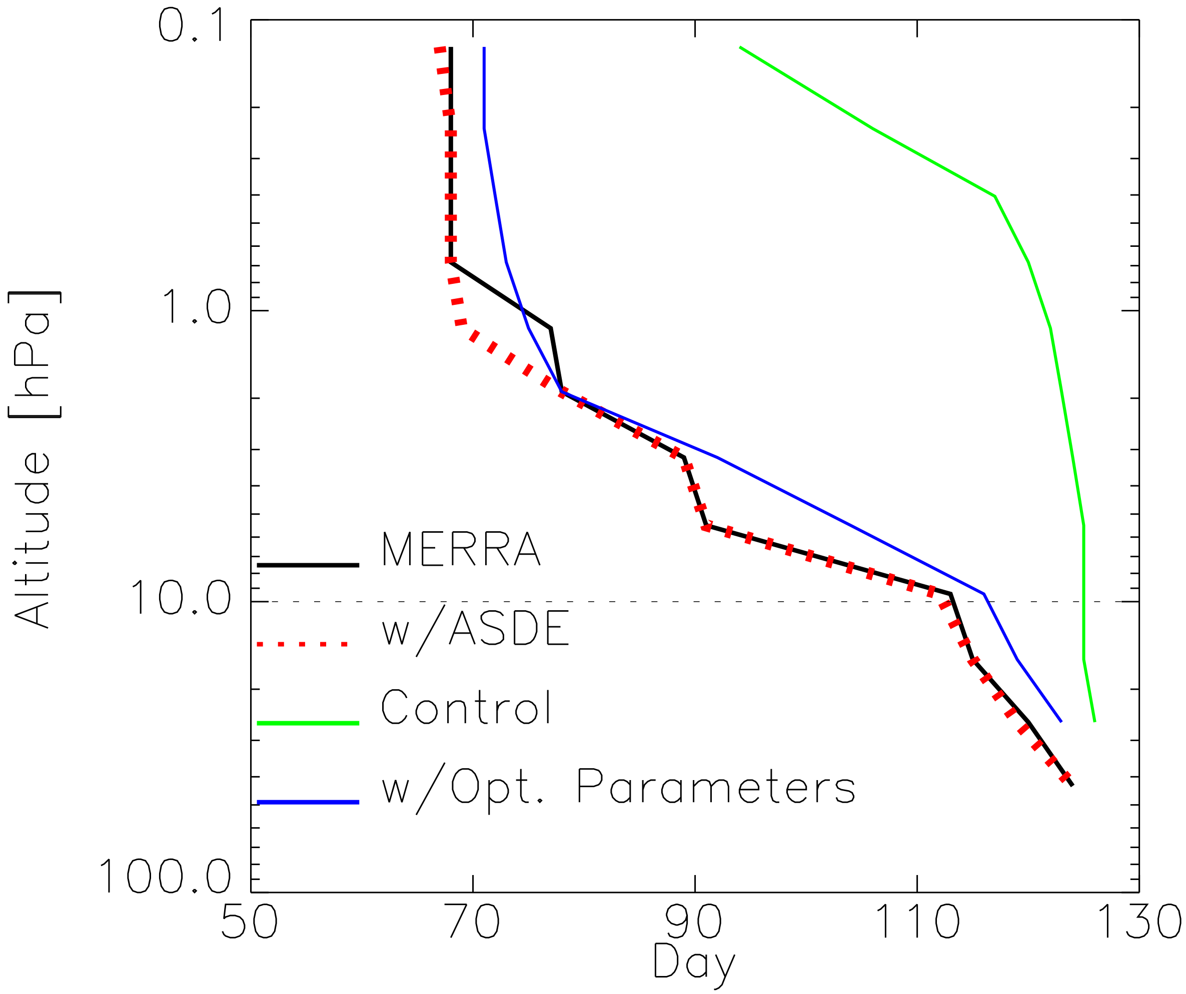
Can data assimilation do a better job instead of manually tuning? Does DA account for resolved-parameterization interactions?



Forcing estimated with 4D-Var. Forcing given by the parameterization. Forcing estimated with optimal parameters.

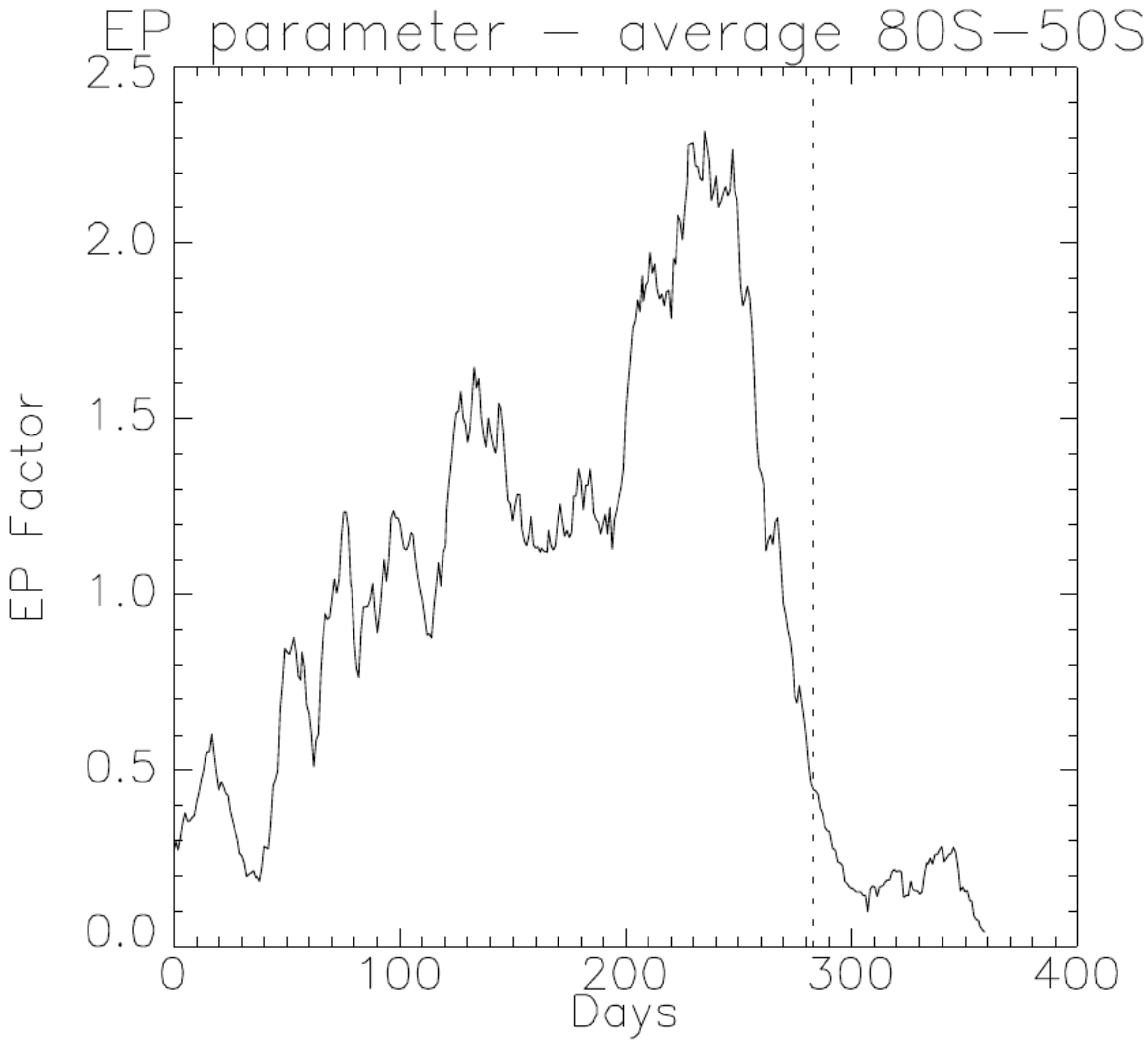
Off-line estimation of the gravity wave parameters using a genetic algorithm (see Pulido et al. QJ 2012).

Final warming date



See Scheffler's Poster for more details.

Limitations in the parameterization.



Should the parameterization be switched off?

Some aspects of the estimated forcing can not be reproduced by the parameterization.

Can we go further?

Using EnKF to estimate the forcing in a “toy” model

Nature state: Lorenz 96 two scale model.

Forecast state: Lorenz 96 one scale model (only large-scale variables).

Since small-scale motions are not modeled, $\mathbf{f}$ is unknown. We want to estimate the $\mathbf{f}$ term using data assimilation.

This requires only the modelling and observations of large-scale variables $\mathbf{x}_L$.

Two metodologies:

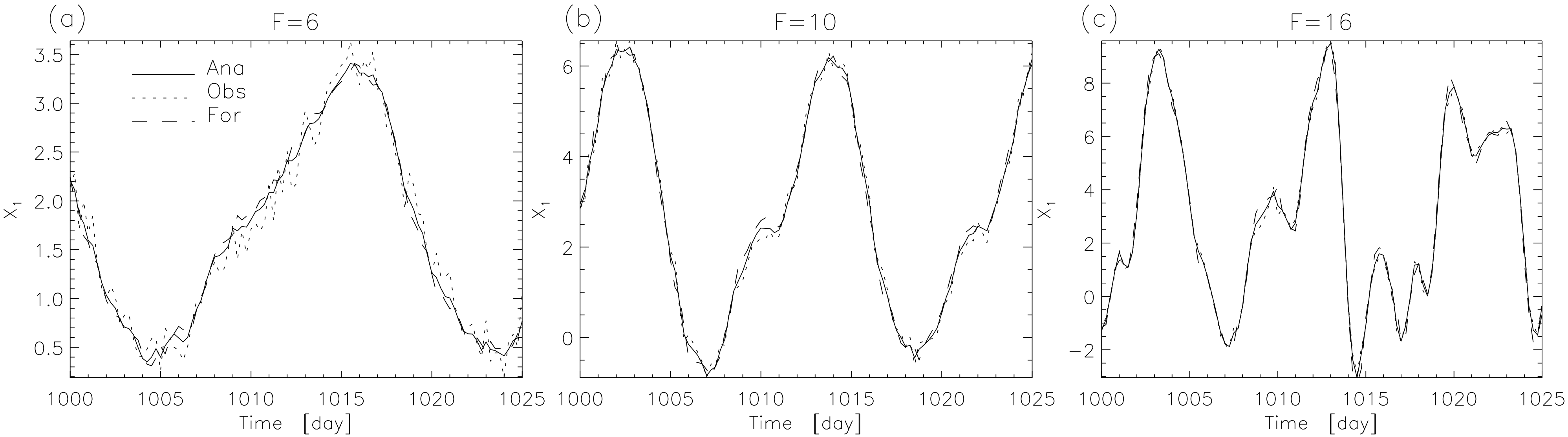
Offline estimation. Apart from the state variables, the forcing variables are included in the state to be estimated (Augmented state). Having the pairs x_n^a, f_n^a for each time, offline linear regression gives the polynomial function that best fits them.

Online estimation. we assume a priori the forcing is a polynomial function and the polynomial coefficients are augmented to the state.

Assimilation details

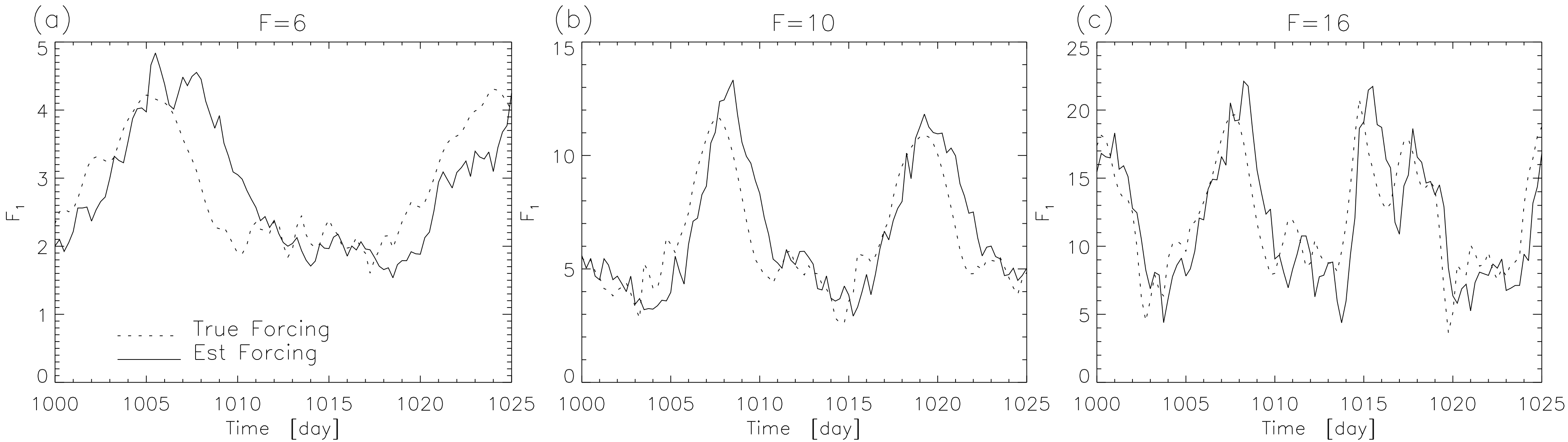
- The LETKF algorithm is used (Hunt et al. 2007).
- Persistence model for the forcing and parameters:
$\mathbf{f}^f(t+1) = \mathbf{f}^a(t)$, $\mathbf{a}^f(t+1) = \mathbf{a}^a(t)$
- All large-scale variables are observed at each cycle.
- No localization is used (some experiments were conducted with localization, a low impact was found for an $N = 8$ and $K = 30$ experiment).
- Adaptive independent inflation factors for the model variables and the forcing variables (Miyoshi, MWR, 2009).

Results off-line. Analysis state



Evolution of the analysis state (x_1^a), observations (x_1^o) and forecast state (x_1^f) from day 1000 to 1025 for case (a) $F = 6$, (b) $F = 10$, and (c) $F = 16$.

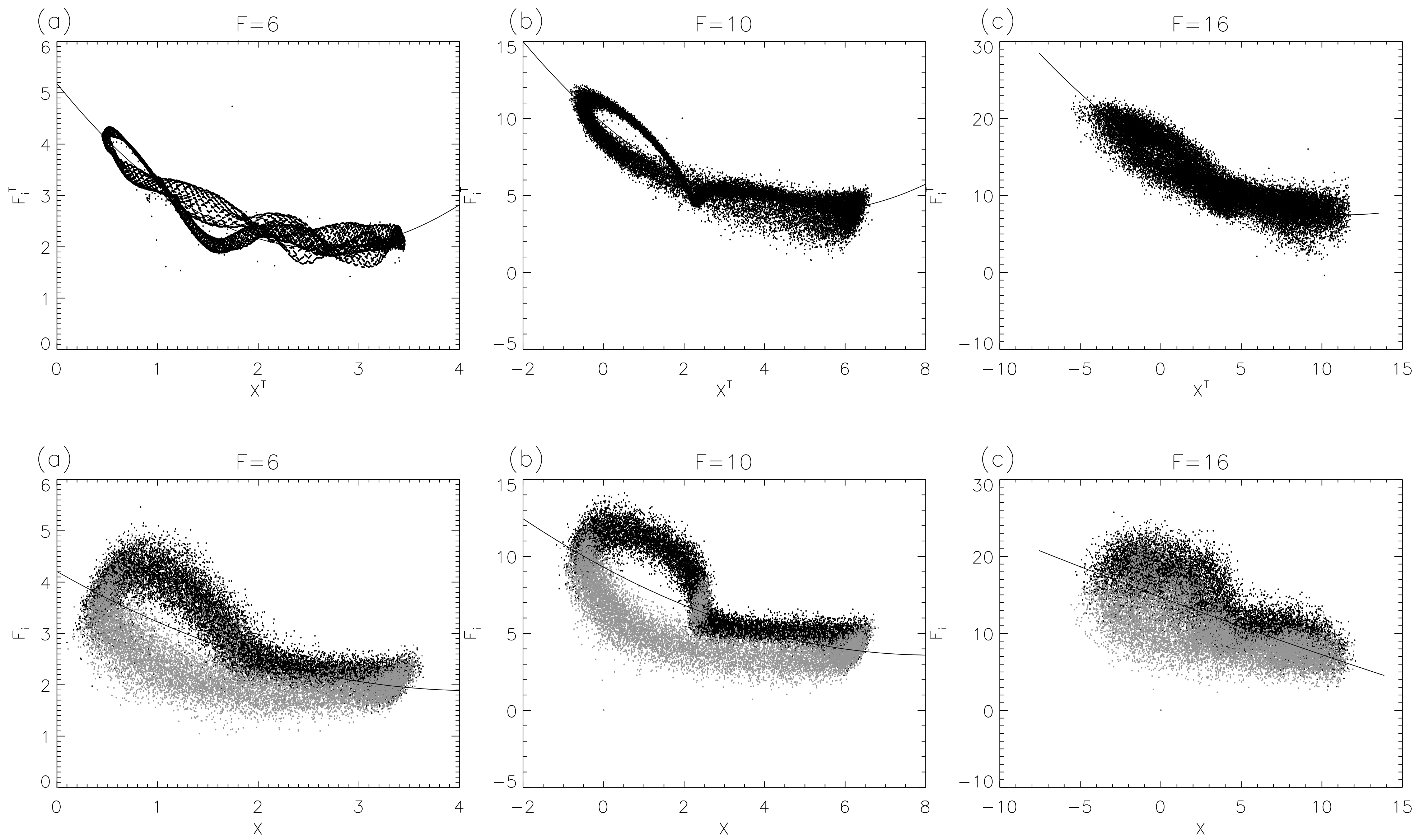
Results off-line. Estimated forcing



Evolution of the estimated forcing (F_1^a), and the true forcing (F_1^T) from day 1000 to 1025 for case (a) $F = 6$, (b) $F = 10$, and (c) $F = 16$.

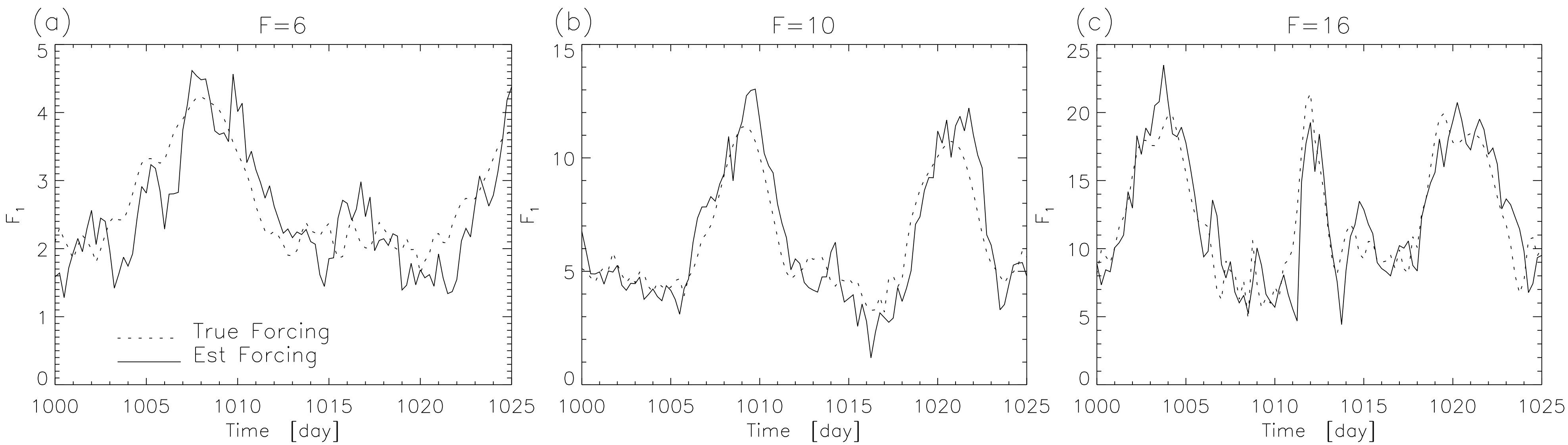
Systematic lag between the true and estimated forcing.

Is the X-F relationship accurate?



Scatterplot of the true small-scale forcing as a function of the true state, and of the estimated forcing. Points with $\frac{\mathrm{d}X}{\mathrm{d}t} < 0$ are represented in gray.

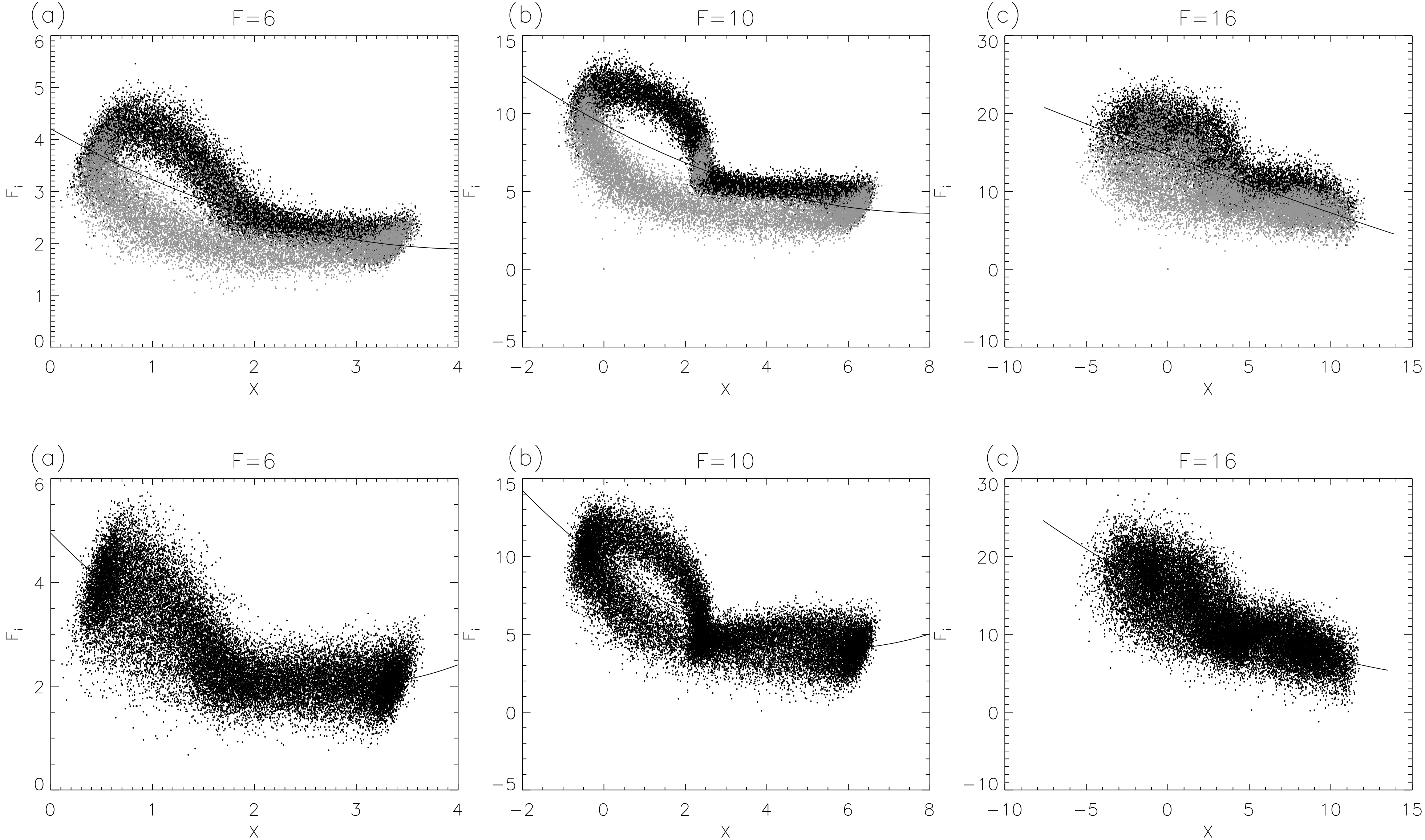
Impact of the inflation factor



Two fixed inflation factors (one for the model state, and one for the forcing) may reduce time lag to the expenses of noise.

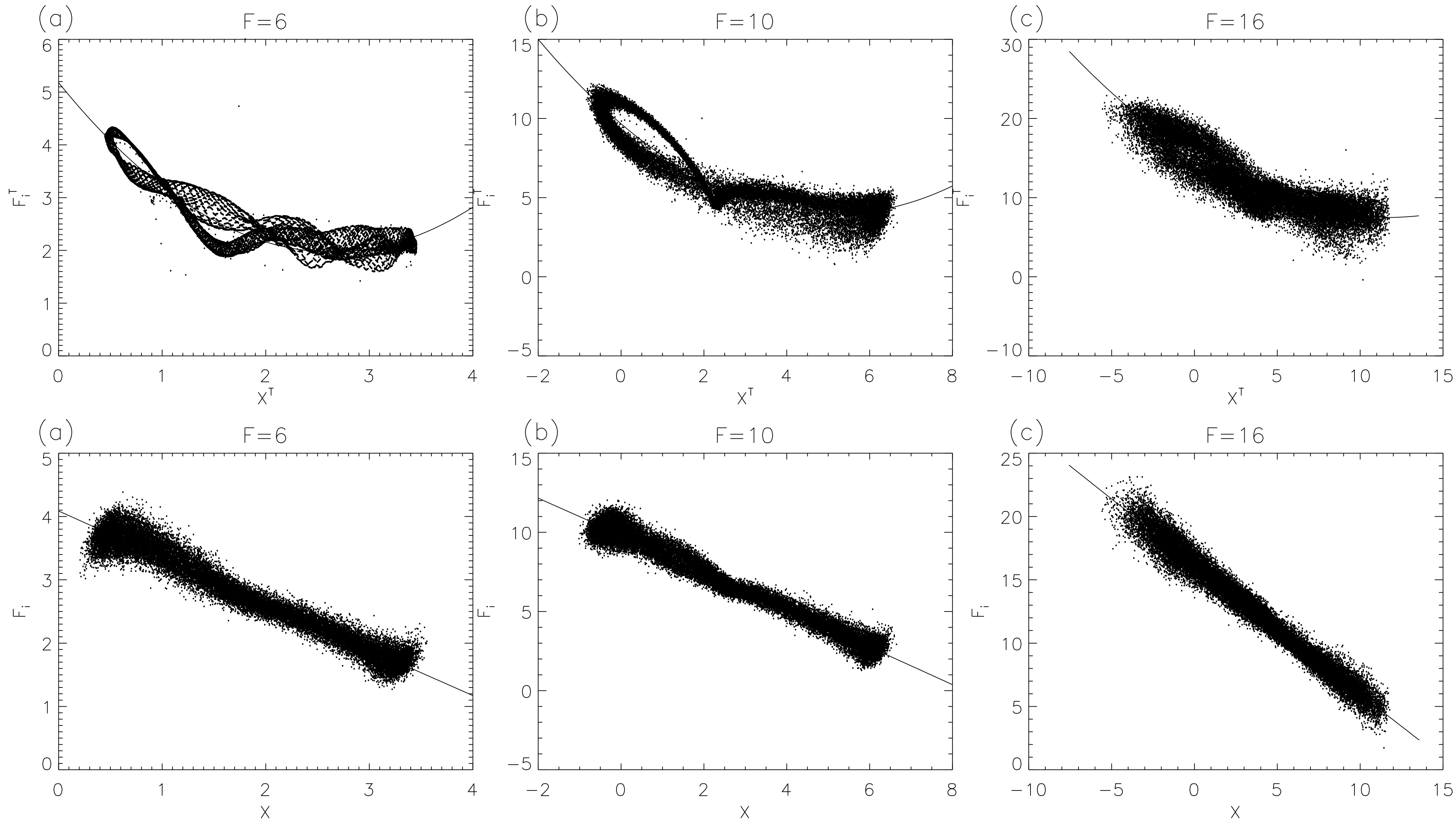
Typical “good” inflation factors: 1.05 for the model state, 1.5-2.2 for the forcing.

X-F scatterplot



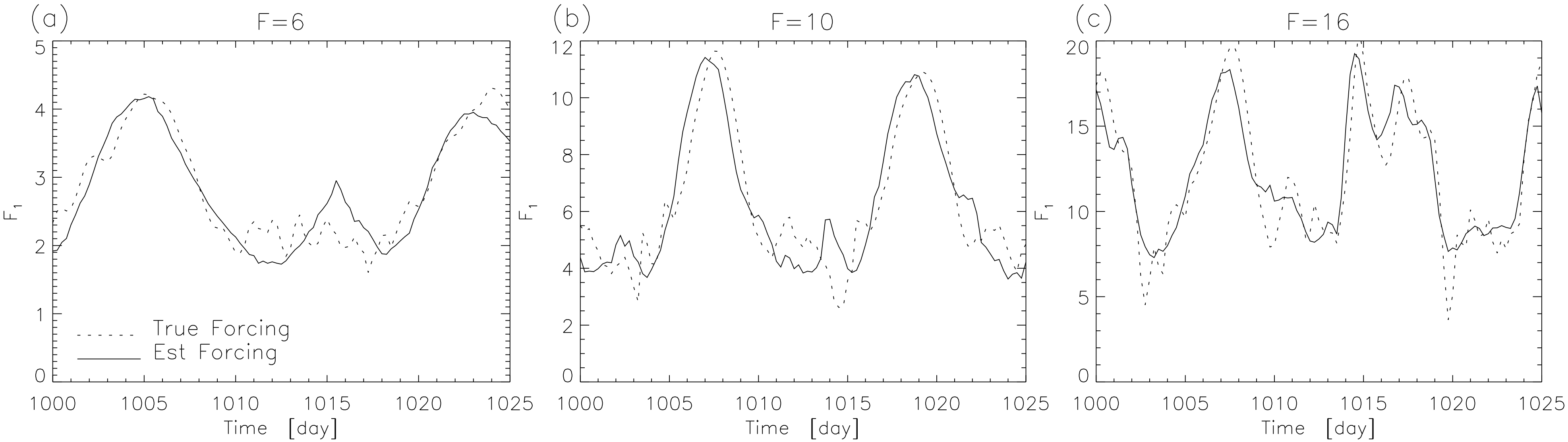
Scatterplot for the experiment with adaptive inflation factors, and for the one with fixed inflation factors.

Scatterplots on-line linear estimation



Scatterplot of the true small-scale forcing as a function of the true state, and of the online linear estimation as a function of the estimated state.

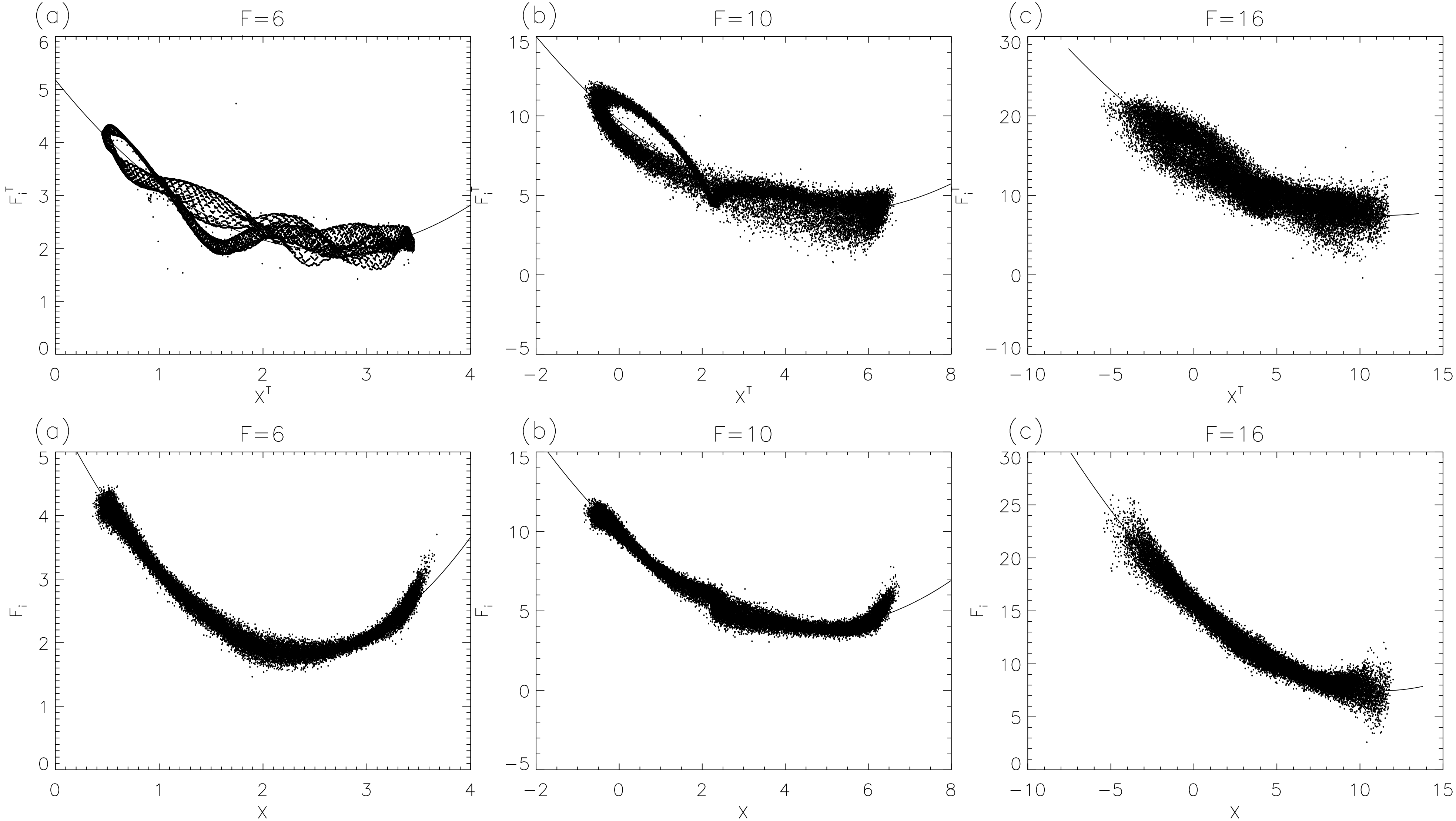
On-line estimation. Quadratic forcing



Evolution of the estimated forcing (F_1^a) for the quadratic online estimation, and the true forcing (F_1^T) from day 1000 to 1025 for case (a) $F = 6$, (b) $F = 10$, and (c) $F = 16$.

Great job in the peaks. Variability of the small-forcing is not well captured.

Scatterplots on-line quadratic estimation



Scatterplot of the true small-scale forcing as a function of the true state, and of the online quadratic estimation as a function of estimated state.

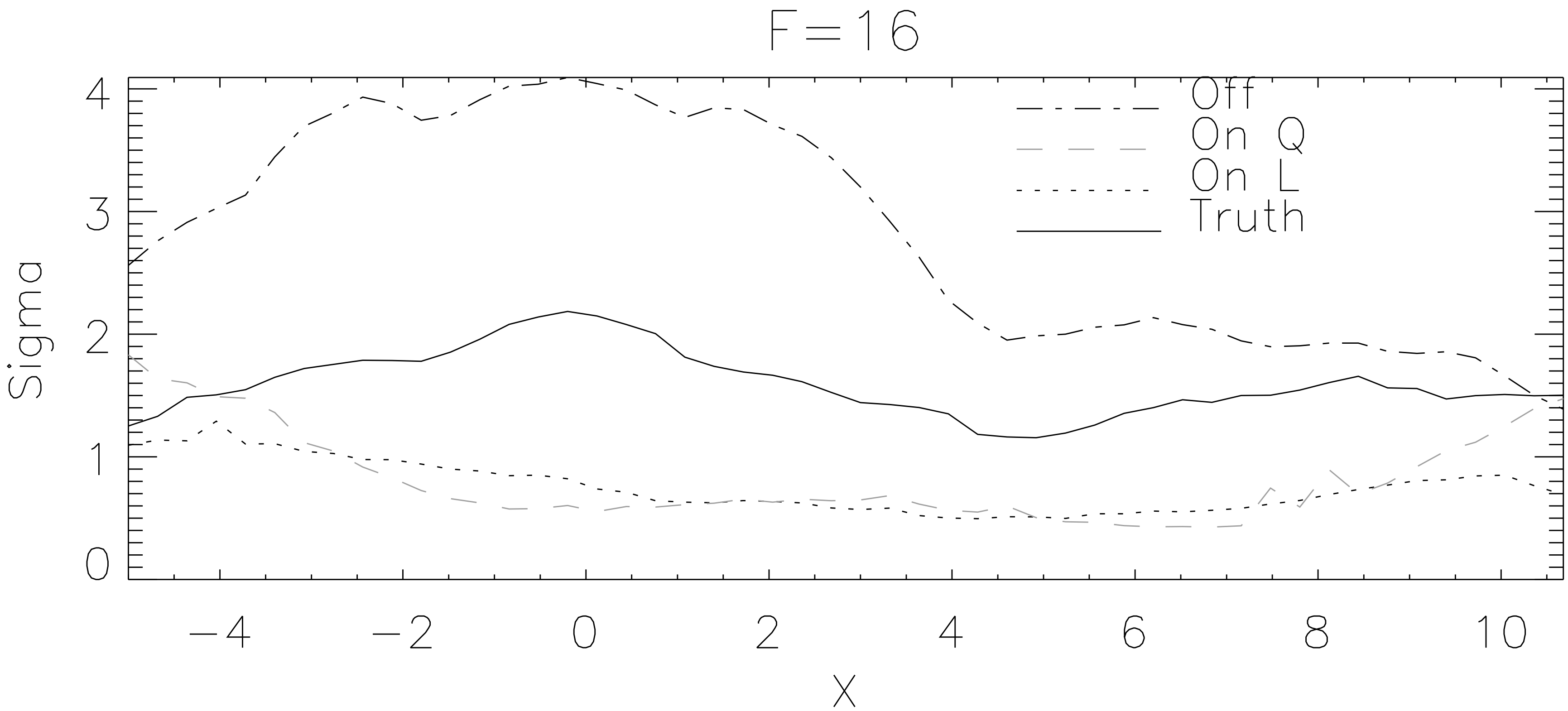
Mean polynomial coefficients

Coef/Case	$F = 6$		$F = 10$		$F = 16$	
	F^{off}	F^{on}	F^{off}	F^{on}	F^{off}	F^{on}
a_0	5	9	0.1	3	0.9	3
a_1	18	32	11	13	21	12
a_2	26	47	20	18	56	19

Relative error (%) in the time mean coefficients of the forcing.

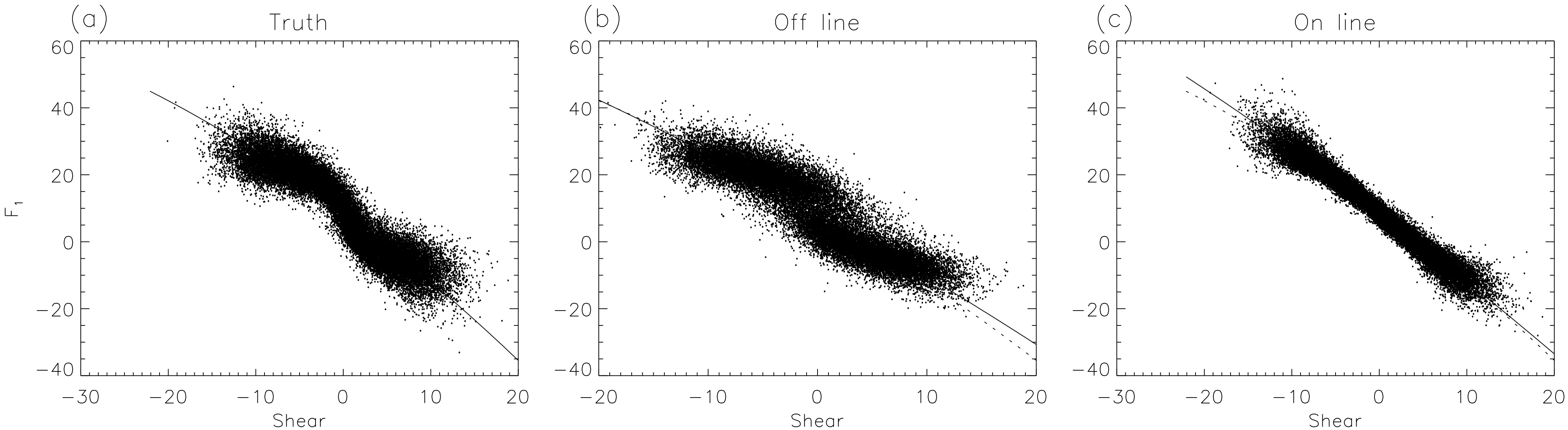
Dispersion of the forcing

The dispersion of the forcing around the deterministic value $\sigma^2(x_n) = \frac{1}{I-1}\sum_i^I \left[f_n(t_i) - \sum_{j=0}^2 a_j(x_n(t_i))^j\right]^2$ may be used to set a stochastic term in the parameterization (e.g. first order autoregression, Wilks QJ 2005).



Nonlocal dependencies of the forcing

The forcing of the small-scale variables in the TRUE model depends on the derivative (instead of the local variable). $d_t X_i^s + M_S(X_i^s) + \alpha(X_{i+1}^L - X_{i-1}^L) = 0$



The forcing term in the parameterization (off and online) is assumed to be:

$F_n = \sum_{i=0}^{2} a_i (X_{n+1}^f - X_{n-1}^f)^i$

Conclusions

- Forcing estimation can be used without apriori information about the functional dependences of the forcing.

- The inflaction factor plays a mayor role in parameter estimation. It does influence the forcing estimation (time lag in the forcing).

- Offline estimation may be useful to determine the functional dependences of the parameterization.

- Once the functional dependences of the forcing are known, online estimation may give slightly better accuracy in parameter values.

- Neither the online nor the offline estimation appears to give an accurate overview of the stochastic characteristics of the small-scale subgrid effects.