Deterministic Treatment of the Model Error in Data Assimilation

Alberto Carrassi[1] and Stéphane Vannitsem[2]

(1): Institut Català de Ciències del Clima (IC3), Barcelona, Spain (Contact: alberto.carrassi@ic3.cat)

(2): Institut Royal Météorologique de Belgique (IRM), Brussels, Belgium

1 Introduction

The prediction problem in geophysical fluid dynamics typically relies on two complementary elements: the model and the data. The sequence of operations that merges model and data to obtain a possibly improved estimate of the flows state is known as data assimilation. Nevertheless the treatment of model error in data assimilation procedures is still done, in most instances, following simple assumptions such as the absence of time correlation.

Fundamental problems making difficult an adequate treatment of model error in data assimilation:

- large variety of possible error sources (incorrect parametrizations of physical processes, numerical discretizations, unresolved scales, etc..)
- the amount of available data insufficient to realistically describe the model error statistics
- lack of a general framework for model error dynamics

OBJECTIVES

1. Identifying some general laws for the evolution of the model error dynamics (with suitable *application-oriented* approximations)
2. Use of these dynamical laws to prescribe the model error statistics required by DA algorithms

2 Formulation

Let assume that the model and Nature are given by:

$$\textit{Model}\quad \frac{d\mathbf{x}(t)}{dt}=f(\mathbf{x},\lambda)\qquad \textit{Nature}\quad \frac{d\hat{\mathbf{x}}(t)}{dt}=\hat{f}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda')+\epsilon\hat{g}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda'),\quad \frac{d\hat{\mathbf{y}}(t)}{dt}=\hat{h}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda')$$

- $\hat{g}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda')$ represents the dynamics associated to extra processes not accounted for by the model;
- $\hat{h}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda')$ - unresolved scale

2.1 CASE I - Parametric Error

ASSUMPTIONS:

- the model resolves all the relevant scales $\Rightarrow \hat{h}=0$ and $f=\hat{f}$
- error in the parameter $\delta\lambda\neq 0$
- set $\epsilon=\gamma\delta\lambda$

Linear Error $\quad \delta\mathbf{x}(t)\approx\mathbf{M}_{t,t_0}\delta\mathbf{x}_0+\int_{t_0}^{t}d\tau\mathbf{M}_{t,\tau}\delta\mu(\tau)=\delta\mathbf{x}^{ic}(t)+\delta\mathbf{x}^{m}(t)$

- The model error acts as a deterministic process
- The important factor controlling the evolution is $\delta\mu(t)=\delta\mu=[\frac{\partial\mathbf{f}}{\partial\lambda}|_{\lambda}+\gamma g(\mathbf{y}(t),\lambda')]\delta\lambda$
- In view of the presence of the propagator M, the flow instabilities are expected to influence the model error dynamics

Model error covariance $\quad \mathbf{P}^m(t)=\int_{t_0}^{t}d\tau\int_{t_0}^{t}d\tau'\mathbf{M}_{t,\tau}<(\delta\mu(\tau))(\delta\mu(\tau'))^T)>\mathbf{M}_{t,\tau'}^T$

Model error correlation $\quad \mathbf{P}^m(t_1,t_2)=\int_{t_0}^{t_1}d\tau\int_{t_0}^{t_2}d\tau'\mathbf{M}_{t_1,\tau}<\delta\mu(\tau)\delta\mu(\tau')^T>\mathbf{M}_{t_2,\tau'}^T$

These equations are NOT suitable for realistic geophysical applications - Some approximation is required

SHORT TIME APPROXIMATION - CASE I

Approx. Model error Cov. $\quad \mathbf{P}^m(t)\approx<\delta\mu_0\delta\mu_0^T>(t-t_0)^2$

Approx. Model error Corr. $\quad \mathbf{P}^m(t_1,t_2)\approx<\delta\mu_0\delta\mu_0^T>(t_1-t_0)(t_2-t_0)$

- The model error covariance and correlation evolve quadratically in the short-time
- The main factor determining this evolution is the covariance of $\delta\mu$ at $t=t_0$, $\mathbf{Q}=<\delta\mu_0\delta\mu_0^T>$
- The covariance $\mathbf{Q}$ embeds the information on the model error through $\delta\lambda$ and the functional dependence of the dynamics on the parameters
- Once **Q** is known, $\mathbf{P}^m$ can be computed at any time within the short time regime

2.2 CASE II - Error due to unresolved scales

ASSUMPTIONS:

- the model does not describe the scale given by $\hat{h}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda')$
- assume correct parameter, $\delta\lambda=0$, and set $\epsilon=0$

Error in the resolved scale $\quad \delta\mathbf{x}(t)=\mathbf{x}(t)-\hat{\mathbf{x}}(t)=\delta\mathbf{x}_0+\int_{t_0}^{t}d\tau(f(\mathbf{x},\lambda)-\hat{f}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda))$

Error covariance in the resolved scale

$$\mathbf{P}(t)=<\delta\mathbf{x}_0\delta\mathbf{x}_0^T>+\int_{t_0}^{t}d\tau\int_{t_0}^{t}d\tau'<[f(\mathbf{x},\lambda)-\hat{f}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda)][f(\mathbf{x},\lambda)-\hat{f}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda)]^T>$$

- the correlation between i.c. and model error neglected (standard hyp. in DA)
- the important factor controlling the evolution is the difference between the velocity fields $f(\mathbf{x},\lambda)-\hat{f}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda)$

SHORT TIME APPROXIMATION - CASE II

- the contribution $f(\mathbf{x},\lambda)-\hat{f}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda)$ is treated as a deterministic process
- the short time evolution of $\mathbf{P}(t)$ reads:

$$\mathbf{P}(t)\approx<\delta\mathbf{x}_0\delta\mathbf{x}_0^T>+<[f(\mathbf{x},\lambda)-\hat{f}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda)][f(\mathbf{x},\lambda)-\hat{f}(\hat{\mathbf{x}},\hat{\mathbf{y}},\lambda)]^T>t^2+O(3)$$

Can we incorporate the short-time approximation for the model error covariance in the context of DA procedures ?

Specific goals:

1. **Computation of the model error covariance in the sequential data assimilation - EKF**
2. **Computation of the model error correlations in the weak-constraint 4DVar**

3 Results

3.1 Sequential Data Assimilation - EKF

- EKF Forecast Error Covariance Update: $\quad \mathbf{P}^f=\mathbf{M}\mathbf{P}^a\mathbf{M}^T+\mathbf{P}^m$
- $\mathbf{P}^m$ - **Model Error Covariance Matrix**

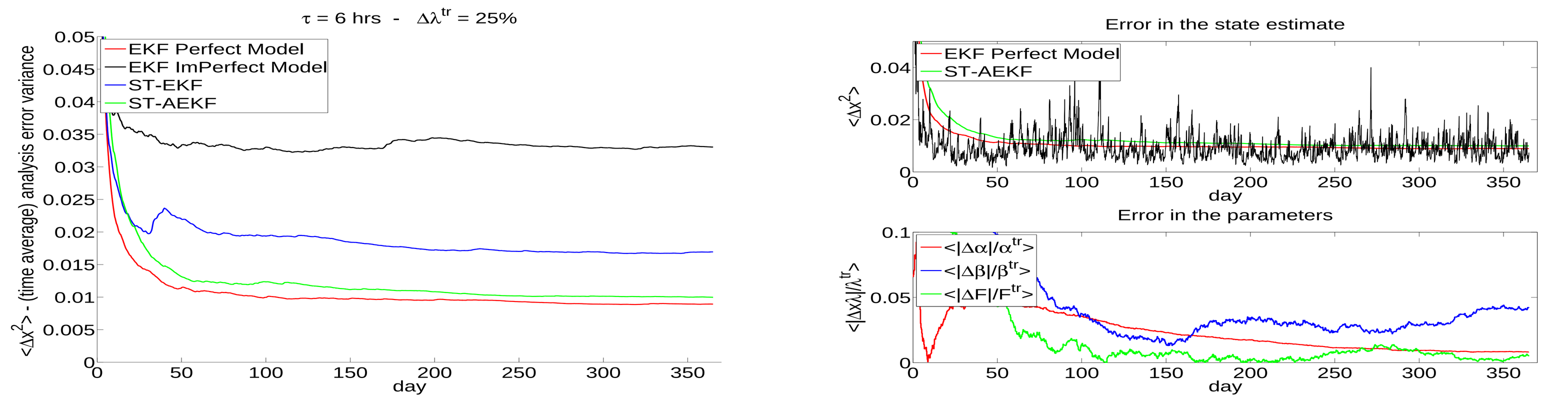
Estimate $\mathbf{P}^m$ using the short time approximation

CASE I - Parametric Error $\Rightarrow \mathbf{P}^m\approx<\delta\mu_0\delta\mu_0^T>\tau^2=\mathbf{Q}\tau^2$

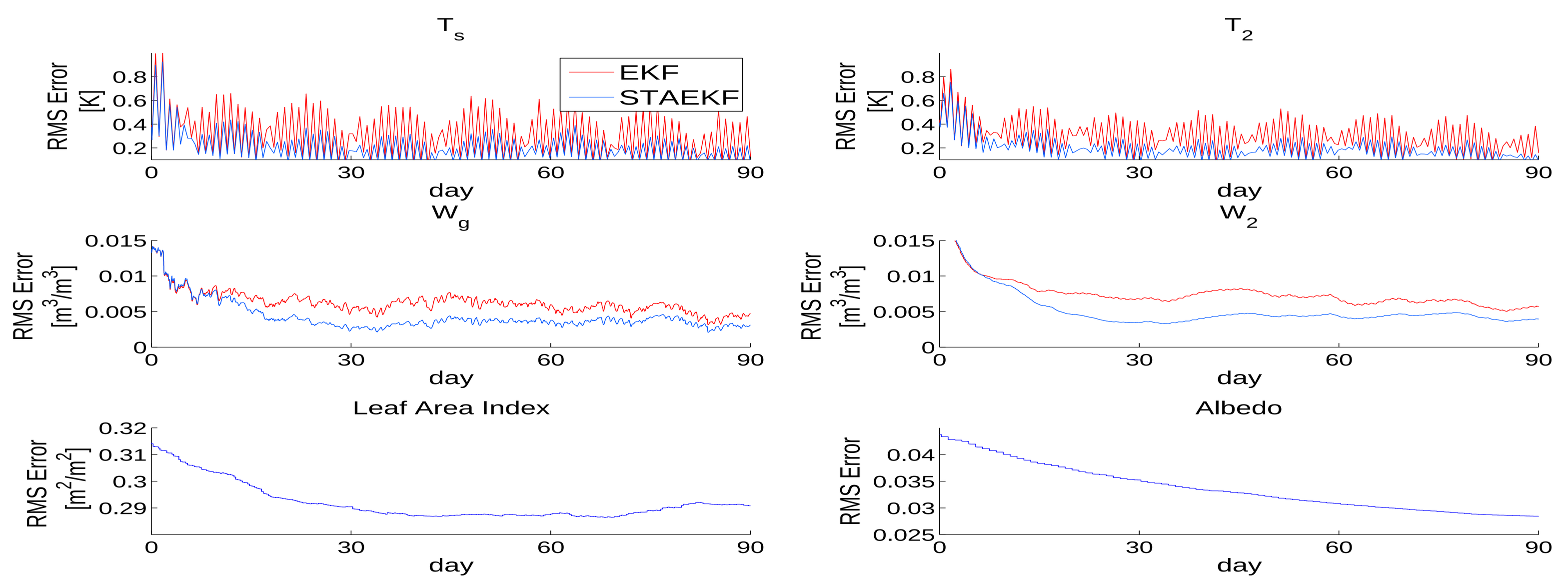
Two solutions proposed to estimate $\mathbf{Q}$:

1. Statistically based on a priori information – Short Time EKF (**ST-EKF**)
2. Dynamically (on the fly) using a state/parameter estimation approach – Short Time Augmented EKF (**ST-AEKF**)

Experiments with Lorenz (1996) 36-variable. State and parameter estimation - Carrassi *et al.* (2008) and Carrassi and Vannitsem (2011)a



Experiments with Land Surface model ISBA (Mahfouf and Noilhan, 1996) State and parameter estimation - Carrassi *et al.* (2012)



CASE II - Error of Unresolved Scales $\Rightarrow \mathbf{P}^m\approx<(f-\hat{f})(f-\hat{f})^T>\tau^2$

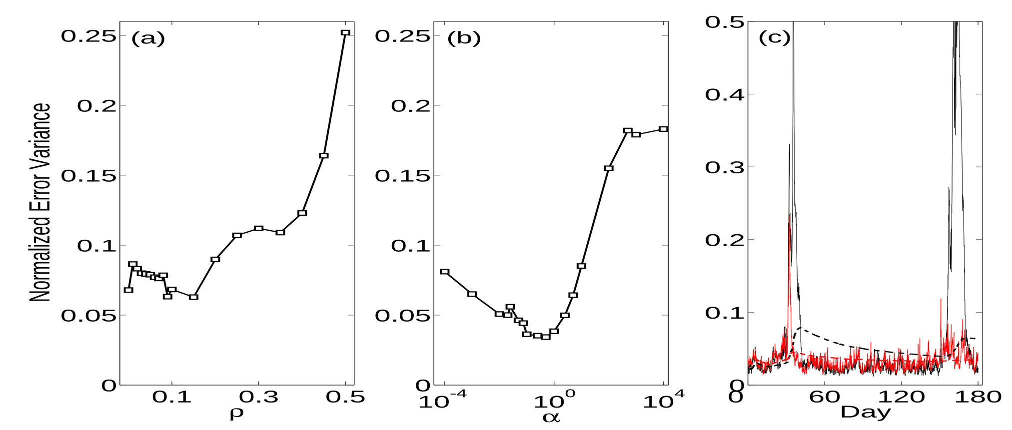
Solution proposed to estimate the statistics of the error in the advection:

- Use of the analysis increments of a reanalysis data-set :

$$f-\hat{f}=\frac{d\mathbf{x}}{dt}-\frac{d\hat{\mathbf{x}}}{dt}\approx\frac{\mathbf{x}_r^f(t+\tau_r)-\mathbf{x}_r^a(t)}{\tau_r}-\frac{\mathbf{x}_r^a(t+\tau_r)-\mathbf{x}_r^a(t)}{\tau_r}=\frac{\delta\mathbf{x}_r^a}{\tau_r}\Rightarrow\mathbf{P}^m(t)\approx<\delta\mathbf{x}_r^a\delta\mathbf{x}_r^{aT}>\frac{\tau^2}{\tau_r^2}$$

- τ_r reanalysis assimilation interval - τ current assimilation interval

Experiments with Lorenz (1996) model. 360 variables - two scales. - Comparison with EKF using multiplicative inflation. - Carrassi and Vannitsem (2011)b

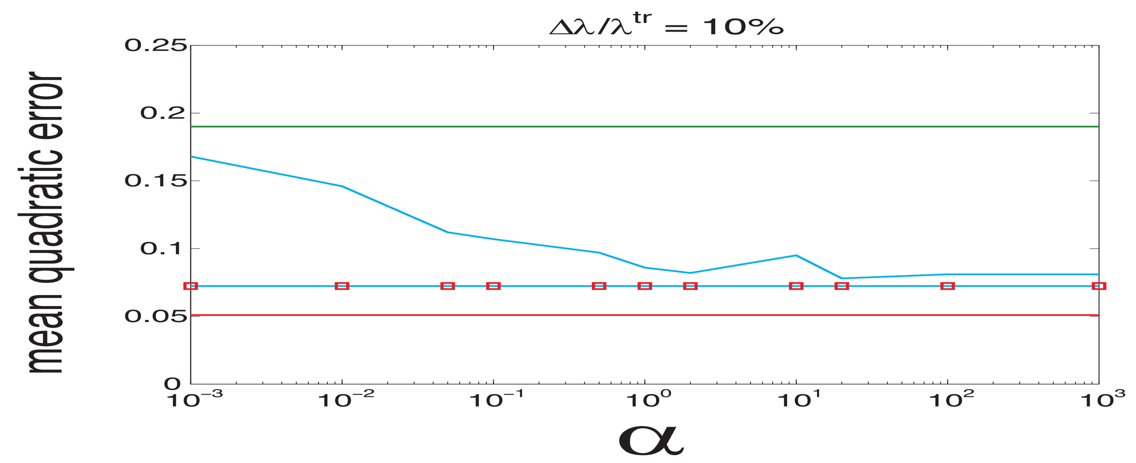


- (a) - EKF; Inflation procedure on the $\mathbf{P}^f\rightarrow(1+\rho)\mathbf{P}^f$
- (b) - ST-EKF; Tuning of $\mathbf{P}^m\rightarrow\alpha\mathbf{P}^m$ ($\mathbf{P}^m$ estimated statistically and then kept fixed)
- (c) - Analysis Error Comparison ST-EKF ($\alpha=0.5$ red line) and EKF ($\rho=0.09$ black line)

3.2 Variational Data Assimilation - 4DVar

- analysis state as the minimum of a cost-function: $2J=\int_0^{\tau}\int_0^{\tau}(\delta\mathbf{x}_{t_1}^m)^T(\mathbf{P}^m)^{-1}_{t_1t_2}(\delta\mathbf{x}_{t_2}^m)dt_1dt_2+\sum_{k=1}^{M}\epsilon_k^T\mathbf{R}_k^{-1}\epsilon_k+\epsilon_b^T\mathbf{B}^{-1}\epsilon_b$
- **Estimate model error covariances/correlations using $\mathbf{P}(t_1,t_2)\approx\mathbf{Q}(t_1-t_0)(t_2-t_0)$**

Experiments with Lorenz (1963) model - Carrassi and Vannitsem (2010)



- **Strong-constraint**
- **Short-time weak constraint 4DVar**
- **Weak constraint 4DVar with uncorrelated model error**: $P_t^m=\alpha\mathbf{B}$ (blue)
- **Weak constraint 4DVar with uncorrelated model error**: $P_t^m=\mathbf{Q}(t-t_0)^2$ (blue with red marks)

****

AC was financed through the IEF Marie Curie Project INCLIDA of the FP7

****

Key References

Carrassi A., S. Vannitsem and C. Nicolis, 2008: Model error and sequential data assimilation. A deterministic formulation. *Q. J. R. Meteorol. Soc.*, **134**: 1297-1313

Carrassi A. and S. Vannitsem, 2010: Accounting for model error in variational data assimilation. A deterministic formulation. *Mon. Wea. Rev.*, **138**, 3369–3386

Carrassi A. and S. Vannitsem, 2011(a): State and parameter estimation with the extended Kalman filter. An alternative formulation of the model error dynamics. *Q. J. R. Meteorol. Soc.*, **137**, 437–451

Carrassi A. and S. Vannitsem, 2011(b): Treatment of the error due to unresolved scales in sequential data assimilation. *I. J. B. C.*, **21**, 3619-3626

Carrassi, A., R. Hamdi, S. Vannitsem and P. Termonia, 2012: Short-time Augmented Extended Kalman filter for soil analysis. A feasibility study. *Atm. Sci. Lett.*. DOI: 10.1002/asl.394

Lorenz E.N. 1963: Deterministic non-periodic flows. *J. Atmos. Sci.*, **20**, 130¿141.

Lorenz E.N. 1996: Predictability: A problem partly solved" *Proc. Seminar on Predictability - ECMWF* **1**, pp. 1-18