Model error and covariance estimation in data assimilation

Alberto Carrassi
acarrassi@ic3.cat

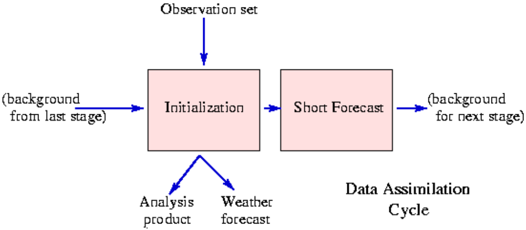
Climate Foreasting Unit - **CFU**
Catalan Institute for Climate Science - **IC3**
Spain

Exploratory Workshop DADA - 16 October 2012
Exploring the Use of Data Assimilation for the Detection and Attribution of Climate Change

Data Assimilation is the entire sequence of operations that, starting from the observations and possibly from a statistical/dynamical knowledge about a system, provides an estimate of its state

The main fields of applications in geophysics are:

- initialize weather prediction
- produce reanalysis
- parameter estimation



Data Assimilation is the entire sequence of operations that, starting from the observations and possibly from a statistical/dynamical knowledge about a system, provides an estimate of its state

Typical sources of informations are:

- observations (synoptic profiles, onboard measurements, remote sensing, etc...)
- background field (climatological, short range forecast)
- evolution dynamics (set of differential equations, numerical model ...)

Data Assimilation is the entire sequence of operations that, starting from the observations and possibly from a statistical/dynamical knowledge about a system, provides an estimate of its state

Typical sources of informations are:

- observations (synoptic profiles, onboard measurements, remote sensing, etc...)
- background field (climatological, short range forecast)
- evolution dynamics (set of differential equations, numerical model ...)

All these information are combined in a statistical fashion to obtain the best-possible estimate the **analysis**

Data Assimilation is the entire sequence of operations that, starting from the observations and possibly from a statistical/dynamical knowledge about a system, provides an estimate of its state

Typical sources of informations are:

- observations (synoptic profiles, onboard measurements, remote sensing, etc...)
- background field (climatological, short range forecast)
- evolution dynamics (set of differential equations, numerical model ...)

All these information are combined in a statistical fashion to obtain the best-possible estimate the **analysis**

But ...

- realistic models are nonlinear and chaotic so that errors amplify rapidly and are subject to a flow-dependent dynamics - an accurate flow-dependent description of the forecast error is crucial
- models are not perfect - incorrect parametrizations of physical processes, numerical discretizations, unresolved scales, etc..

Data Assimilation is the entire sequence of operations that, starting from the observations and possibly from a statistical/dynamical knowledge about a system, provides an estimate of its state

Typical sources of informations are:

- observations (synoptic profiles, onboard measurements, remote sensing, etc...)
- background field (climatological, short range forecast)
- evolution dynamics (set of differential equations, numerical model ...)

All these information are combined in a statistical fashion to obtain the best-possible estimate the **analysis**

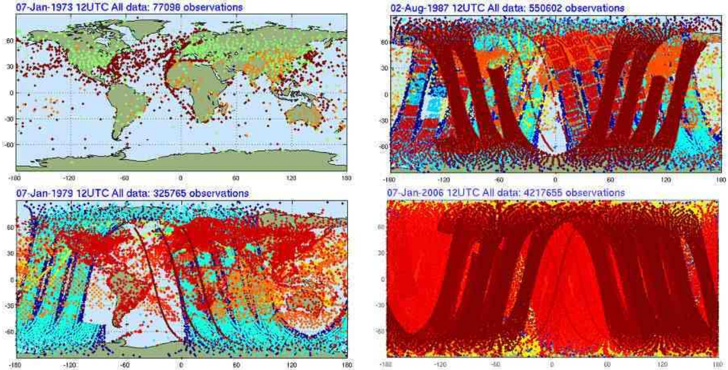
But ...

- realistic models are nonlinear and chaotic so that errors amplify rapidly and are subject to a flow-dependent dynamics - an accurate flow-dependent description of the forecast error is crucial
- models are not perfect - incorrect parametrizations of physical processes, numerical discretizations, unresolved scales, etc..

Data Assimilation is the entire sequence of operations that, starting from the observations and possibly from a statistical/dynamical knowledge about a system, provides an estimate of its state

In the last decades the accuracy of initial conditions has improved:

- observational network has been enlarged and refined (a major contribution came from remote sensing measurements)
- there has been a flourishing of data assimilation techniques aimed at a flow dependent description of the forecast error (KF-like algorithms, Monte Carlo and Deterministic filters, AUS, 4DVar, ...)



controlling errors: **what about model error ?**

In the past, model error has been considered small with respect to the (growth of) initial condition error, and thus often neglected

Nowadays model error is recognized as a main source of uncertainty in NWP, seasonal and climate prediction

controlling errors: **what about model error ?**

In the past, model error has been considered small with respect to the (growth of) initial condition error, and thus often neglected

Nowadays model error is recognized as a main source of uncertainty in NWP, seasonal and climate prediction

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..)

controlling errors: **what about model error ?**

In the past, model error has been considered small with respect to the (growth of) initial condition error, and thus often neglected

Nowadays model error is recognized as a main source of uncertainty in NWP, seasonal and climate prediction

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

controlling errors: **what about model error ?**

In the past, model error has been considered small with respect to the (growth of) initial condition error, and thus often neglected

Nowadays model error is recognized as a main source of uncertainty in NWP, seasonal and climate prediction

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

controlling errors: **what about model error ?**

In the past, model error has been considered small with respect to the (growth of) initial condition error, and thus often neglected

Nowadays model error is recognized as a main source of uncertainty in NWP, seasonal and climate prediction

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

controlling errors: **what about model error ?**

In the past, model error has been considered small with respect to the (growth of) initial condition error, and thus often neglected

Nowadays model error is recognized as a main source of uncertainty in NWP, seasonal and climate prediction

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

The posing of the problem

Let assume to have the model:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}, \lambda)$$

used to describe the true process:

$$\frac{d\hat{\mathbf{x}}(t)}{dt} = \hat{f}(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \lambda^{'}) + \epsilon\hat{g}(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \lambda^{'})$$
$$\frac{d\hat{\mathbf{y}}(t)}{dt} = \hat{h}(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \lambda^{'})$$

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

CASE I - Parametric Error

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

Estimation error evolution in the linear approximation

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

where

$$\delta\mu = [\frac{\partial\mathbf{f}}{\partial\lambda}|_{\lambda} + \gamma g(\mathbf{y}(t), \lambda^{'})]\delta\lambda$$

CASE I - Parametric Error

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

Estimation error evolution in the linear approximation

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

where

$$\delta\mu = [\frac{\partial\mathbf{f}}{\partial\lambda}|_{\lambda} + \gamma g(\mathbf{y}(t), \lambda^{'})]\delta\lambda$$

- The model error acts as a deterministic process

CASE I - Parametric Error

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

Estimation error evolution in the linear approximation

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

where

$$\delta\mu = [\frac{\partial\mathbf{f}}{\partial\lambda}|_{\lambda} + \gamma g(\mathbf{y}(t), \lambda^{'})]\delta\lambda$$

- The model error acts as a deterministic process
- The important factor controlling the evolution is $\delta\mu(t)$
- In view of the presence of the propagator M, the flow instabilities are expected to influence the model error dynamics

Model error covariance and correlation - CASE I

Model error covariance

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

Model error covariance and correlation - CASE I

Model error covariance

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

Model error correlation

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

Model error covariance and correlation - CASE I

Model error covariance

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

Model error correlation

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

These covariance and correlations are exactly what we need in DA !

Model error covariance and correlation - CASE I

Model error covariance

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

Model error correlation

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

These covariance and correlations are exactly what we need in DA !

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

Short time approximation - CASE I

Model error covariance

$$\mathbf{P}^m(t) \approx < \delta\mu_0 \delta\mu_0^T > (t - t_0)^2$$

Short time approximation - CASE I

Model error covariance

$$\mathbf{P}^m(t) \approx < \delta\mu_0 \delta\mu_0^T > (t - t_0)^2$$

Model error correlation

$$\mathbf{P}^m(t_1, t_2) \approx < \delta\mu_0 \delta\mu_0^T > (t_1 - t_0)(t_2 - t_0)$$

Short time approximation - CASE I

Model error covariance

$$\mathbf{P}^m(t) \approx <\delta\mu_0\delta\mu_0^T> (t-t_0)^2$$

Model error correlation

$$\mathbf{P}^m(t_1,t_2) \approx <\delta\mu_0\delta\mu_0^T> (t_1-t_0)(t_2-t_0)$$

- The model error covariance and correlation evolve quadratically in the short-time.

Short time approximation - CASE I

Model error covariance

$$\mathbf{P}^m(t) \approx < \delta\mu_0 \delta\mu_0^T > (t - t_0)^2$$

Model error correlation

$$\mathbf{P}^m(t_1, t_2) \approx < \delta\mu_0 \delta\mu_0^T > (t_1 - t_0)(t_2 - t_0)$$

- 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 $t = t_0$, $\mathbf{Q} = < \delta\mu_0 \delta\mu_0^T >$.

Short time approximation - CASE I

Model error covariance

$$\mathbf{P}^m(t) \approx < \delta\mu_0 \delta\mu_0^T > (t - t_0)^2$$

Model error correlation

$$\mathbf{P}^m(t_1, t_2) \approx < \delta\mu_0 \delta\mu_0^T > (t_1 - t_0)(t_2 - t_0)$$

- 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.

Short time approximation - CASE I

Model error covariance

$$\mathbf{P}^m(t) \approx < \delta\mu_0 \delta\mu_0^T > (t - t_0)^2$$

Model error correlation

$$\mathbf{P}^m(t_1, t_2) \approx < \delta\mu_0 \delta\mu_0^T > (t_1 - t_0)(t_2 - t_0)$$

- 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 $\mathbf{Q}$ is known, $\mathbf{P}^m$ can be computed at any time within the short time regime

CASE II - Error due to unresolved scales

- 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 $\epsilon = 0$

Estimation error evolution in the resolved scale

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

CASE II - Error due to unresolved scales

- 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 $\epsilon = 0$

Estimation error evolution in the resolved scale

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

Evolution of the estimation error covariance in the resolved scale

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

CASE II - Error due to unresolved scales

- 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 $\epsilon = 0$

Estimation error evolution in the resolved scale

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

Evolution of the estimation 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)

CASE II - Error due to unresolved scales

- 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 $\epsilon = 0$

Estimation error evolution in the resolved scale

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

Evolution of the estimation 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:

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

DA in the presence of model error

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**

Extended Kalman Filter (EKF) in the presence of model error

$$\boxed{\text{Input } \mathbf{x}^a,\ \mathbf{P}^a}$$

$$\boxed{\text{Forecast: } \mathbf{x}^f = \mathcal{M}(\mathbf{x}^a)\ \&\ \mathbf{P}^f = \mathbf{M}\mathbf{P}^a\mathbf{M}^T + \mathbf{P}^m}$$

$$\boxed{\text{Analysis Update: } \mathbf{x}^a = [\mathbf{I} - \mathbf{K}\mathbf{H}]\,\mathbf{x}^f + \mathbf{K}\mathbf{y}^o\ \&\ \mathbf{P}^a = [\mathbf{I} - \mathbf{K}\mathbf{H}]\mathbf{P}^f}$$

$\mathbf{P}^m$ - Model Error Covariance Matrix

Extended Kalman Filter (EKF) in the presence of model error

Input $\mathbf{x}^a$, $\mathbf{P}^a$

Forecast: $\mathbf{x}^f = \mathcal{M}(\mathbf{x}^a)$ & $\mathbf{P}^f = \mathbf{M}\mathbf{P}^a\mathbf{M}^T + \mathbf{P}^m$

Analysis Update: $\mathbf{x}^a = [\mathbf{I} - \mathbf{K}\mathbf{H}]\,\mathbf{x}^f + \mathbf{K}\mathbf{y}^o$ & $\mathbf{P}^a = [\mathbf{I} - \mathbf{K}\mathbf{H}]\mathbf{P}^f$

$\mathbf{P}^m$ - Model Error Covariance Matrix

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

Extended Kalman Filter in the presence of parametric error

CASE I - Parametric Error

$$\mathbf{P}^m \approx < \delta\mu_0 \delta\mu_0^T > \tau^2 = \mathbf{Q}\tau^2$$

...needs to estimate $\mathbf{Q}$

Extended Kalman Filter in the presence of parametric error

CASE I - Parametric Error

$$\mathbf{P}^m \approx < \delta\mu_0 \delta\mu_0^T > \tau^2 = \mathbf{Q}\tau^2$$

...needs to estimate **Q**

Two solutions proposed:

1. Statistically based on a priori information – Short Time EKF (**ST-EKF**)

Extended Kalman Filter in the presence of parametric error

CASE I - Parametric Error

$$\mathbf{P}^m \approx < \delta\mu_0 \delta\mu_0^T > \tau^2 = \mathbf{Q}\tau^2$$

...needs to estimate $\mathbf{Q}$

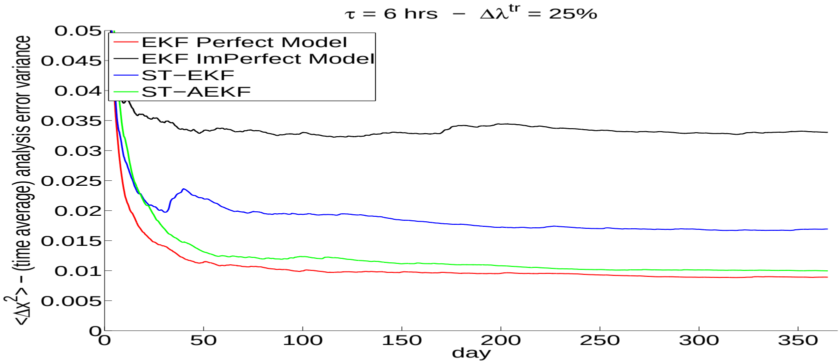
Two solutions proposed:

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**)

Parametric Error - Numerical Analysis with ST-EKF and ST-AEKF

Carrassi, Vannitsem, Nicolis (2008) QJRMS and Carrassi & Vannitsem (2011) QJRMS

- Prototype of nonlinear chaotic dynamics (Lorenz, 1996): $\frac{dx_i}{dt} = \alpha(x_{i+1} - x_{i-2})x_{i-1} - \beta x_i + F \qquad 1 \leq i \leq 36$
- ST-EKF - **Q** estimate statistically and then kept fixed along the assimilation cycle
- ST-AEKF - **Q** estimated online by measuring system's observables - State Augmented formulation
  - augmented system $\mathbf{z} = (\mathcal{M}(\mathbf{x}), \mathcal{F}(\lambda))^T$
  - at analysis time the state and parameters are estimated along with their associated uncertainty (covariances) and cross correlations
  - the updated parametric error covariance, $\mathbf{P}^a_\lambda = < \delta\lambda\delta\lambda^T >$, is then used to update $\mathbf{Q} \Rightarrow \mathbf{P}^m$

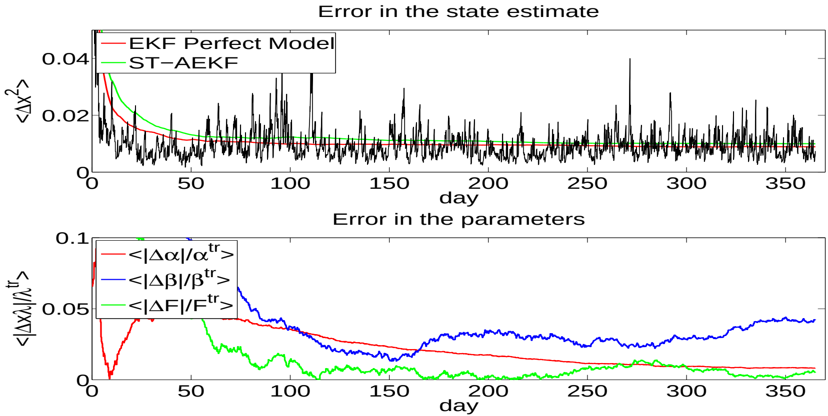


Parametric Error - Numerical Analysis with ST-EKF and ST-AEKF

Carrassi, Vannitsem, Nicolis (2008) QJRMS and Carrassi & Vannitsem (2011) QJRMS

Q estimated online by measuring system's observables - State Augmented formulation

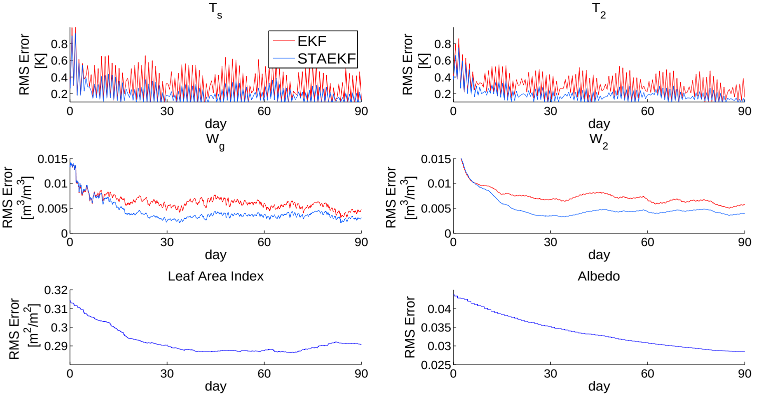
- simultaneous estimate of the three parameters
- results averaged over an ensemble (O(100)) of experiments



Parametric Error - Numerical Analysis with ST-EKF and ST-AEKF

Carrassi, Hamdi, Vannitsem, Termonia (2012) ASL

- Land Surface model ISBA (Mahfouf and Noilhan, 1996)
- State Variables: soil temperature (T_s and T_2) and moisture content (w_g and w_2).
- Observations of screen-level variables (temperature and humidity at 2 meter)
- Parametric error in the *Leaf Area Index* (LAI) and *Albedo*
- Comparison between EKF and ST-AEKF



Error due to unresolved scales – ST-EKF

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

...needs to estimate the statistics of the vel. fields discrepancy.

Solution proposed:

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

$$f-\hat{f} = \frac{d\mathbf{x}}{dt} - \frac{d\hat{\mathbf{x}}}{dt} \approx \frac{\mathbf{x}_r^f(t+\tau_r) - \mathbf{x}_r^a(t)}{\tau_r} - \frac{\mathbf{x}_r^a(t+\tau_r) - \mathbf{x}_r^a(t)}{\tau_r} = \frac{\delta \mathbf{x}_r^a}{\tau_r} \Rightarrow$$

$$\mathbf{P}^m(t) \approx < \delta \mathbf{x}_r^a \delta {\mathbf{x}_r^a}^T > \frac{\tau^2}{\tau_r^2}$$

- τ_r reanalysis assimilation interval
- τ current assimilation interval

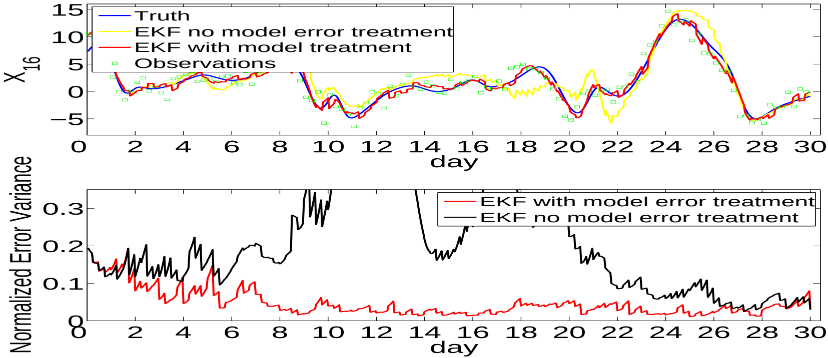
Error due to unresolved scales – ST-EKF

Carrassi & Vannitsem (2011) IJBC

- Lorenz (1996) with two scales (*large scale* - **x**; *small scale* - **y**)
- 12 regular observations of the large scale, **x**, only

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F - \frac{hc}{b}\sum_{j=1}^{10} y_{j,i}, \qquad i = \{1, \ldots, 36\}$$

$$\frac{dy_{j,i}}{dt} = -cby_{j+1,i}(y_{j+2,i} - y_{j-1,i}) - cy_{j,i} + \frac{hc}{b}x_i, \qquad j = \{1, \ldots, 10\}$$

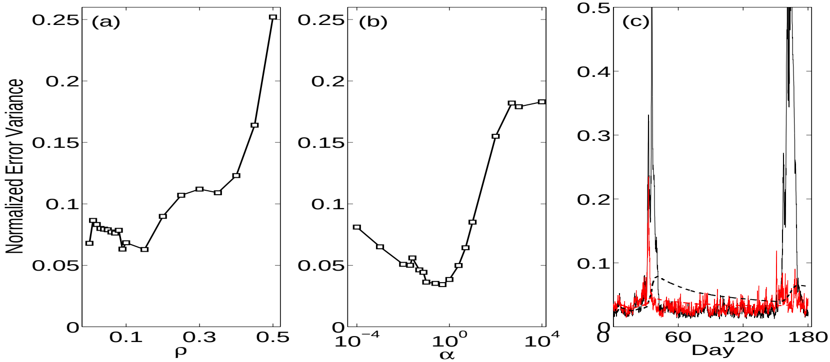


Error due to unresolved scales – ST-EKF

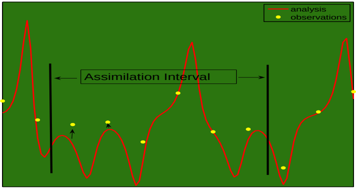
Carrassi & Vannitsem (2011) IJBC

Comparison with the EKF employing the inflation of the $\mathbf{P}^f$ as a tool to account for model error

- (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)

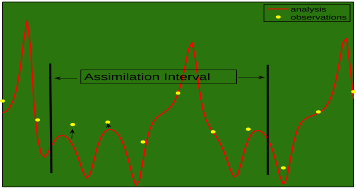


4DVar in the presence of model error - Short Time Weak Constraint 4DVar



- assimilate observations distributed over the time window τ

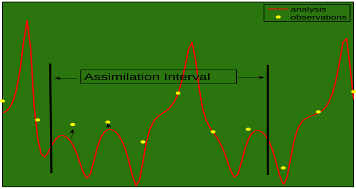
4DVar in the presence of model error - Short Time Weak Constraint 4DVar



- assimilate observations distributed over the time window τ
- analysis state as the minimum of a cost-function:

$$2J = \int_0^\tau \int_0^\tau (\delta \mathbf{x}_{t_1}^m)^T (\mathbf{P}^m)^{-1}_{t_1 t_2} (\delta \mathbf{x}_{t_2}^m) dt_1 dt_2 + \sum_{k=1}^{M} \epsilon_{\mathbf{k}}{}^T \mathbf{R}_k^{-1} \epsilon_{\mathbf{k}} + \epsilon_b^T \mathbf{B}^{-1} \epsilon_b$$

4DVar in the presence of model error - Short Time Weak Constraint 4DVar



- assimilate observations distributed over the time window τ
- analysis state as the minimum of a cost-function:

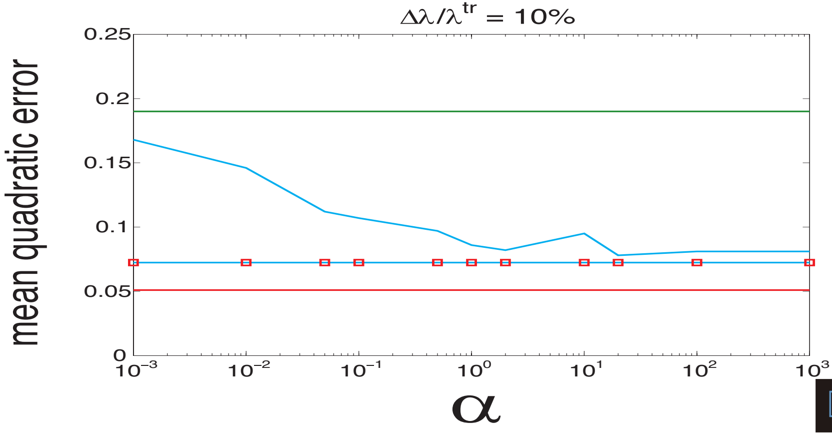
$$2J = \int_0^\tau \int_0^\tau (\delta \mathbf{x}_{t_1}^m)^T (\mathbf{P}^m)^{-1}_{t_1 t_2} (\delta \mathbf{x}_{t_2}^m) dt_1 dt_2 + \sum_{k=1}^{M} \epsilon_{\mathbf{k}}{}^T \mathbf{R}_k^{-1} \epsilon_{\mathbf{k}} + \epsilon_b^T \mathbf{B}^{-1} \epsilon_b$$

Estimate model error covariances/correlations using $\mathbf{P}(t_1, t_2) \approx \mathbf{Q}(t_1 - t_0)(t_2 - t_0)$

Results weak-constraint 4DVar

Carrassi and Vannitsem, 2010 (MWR)

- Lorenz 3-variable (1963) system
- Assimilation interval $\tau = 8$ time-steps, Obs frequency $\Delta t_{obs} = 2$ time-steps
- **Strong-constraint** - **Short-time weak constraint 4DVar** - **Weak constraint 4DVar with uncorrelated model error: with $P^m_t = \alpha \mathbf{B}$ (blue) or $P^m_t = \mathbf{Q}(t - t_0)^2$ (blue with red marks)**



Conclusions and Perspectives

- the proposed formulations gave encouraging results in the framework of both sequential and variational assimilation
- treating the model error as a deterministic process makes possible to derive short-time approximations for the error covariance suitable for DA applications
- the estimation of model error covariances is based on fundamental features rather than estimated using ad-hoc procedures
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals

Conclusions and Perspectives

- the proposed formulations gave encouraging results in the framework of both sequential and variational assimilation
- treating the model error as a deterministic process makes possible to derive short-time approximations for the error covariance suitable for DA applications
- the estimation of model error covariances is based on fundamental features rather than estimated using ad-hoc procedures
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals

Future directions:

- simultaneous treatment of parametric and unresolved scales error

Conclusions and Perspectives

- the proposed formulations gave encouraging results in the framework of both sequential and variational assimilation
- treating the model error as a deterministic process makes possible to derive short-time approximations for the error covariance suitable for DA applications
- the estimation of model error covariances is based on fundamental features rather than estimated using ad-hoc procedures
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals

Future directions:

- simultaneous treatment of parametric and unresolved scales error
- application to more realistic model (and model error) - observational scenarios (surface data assimilation, use of adaptive control variable ...)

Conclusions and Perspectives

- the proposed formulations gave encouraging results in the framework of both sequential and variational assimilation
- treating the model error as a deterministic process makes possible to derive short-time approximations for the error covariance suitable for DA applications
- the estimation of model error covariances is based on fundamental features rather than estimated using ad-hoc procedures
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals

Future directions:

- simultaneous treatment of parametric and unresolved scales error
- application to more realistic model (and model error) - observational scenarios (surface data assimilation, use of adaptive control variable ...)
- application of the state augmentation formulation for the state and parameter estimation for seasonal and climate predictions

Conclusions and Perspectives

- the proposed formulations gave encouraging results in the framework of both sequential and variational assimilation
- treating the model error as a deterministic process makes possible to derive short-time approximations for the error covariance suitable for DA applications
- the estimation of model error covariances is based on fundamental features rather than estimated using ad-hoc procedures
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals

Future directions:

- simultaneous treatment of parametric and unresolved scales error
- application to more realistic model (and model error) - observational scenarios (surface data assimilation, use of adaptive control variable ...)
- application of the state augmentation formulation for the state and parameter estimation for seasonal and climate predictions

thanks