

Accounting for model error in data assimilation

Alberto Carrassi
a.carrassi@oma.be

Royal Meteorological Institute
Belgium

*Exploring complex dynamics in High dimensional Chaotic Systems:
From Weather Forecasting to Oceanic Flows*

Dresden 25-29 January, 2010



- 1 Data Assimilation - DA overview
- 2 General framework for model error dynamics
 - Modeling error statistics
 - Modeling error statistics for geophysical systems
- 3 Sequential Data Assimilation - EKF
- 4 Variational Data Assimilation
- 5 Conclusion



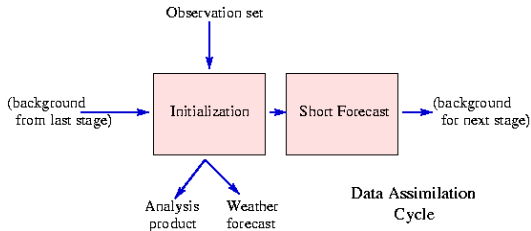
Data Assimilation is the entire sequence of operations that, starting from the observations and possibly from a statistical/dynamical knowledge about a system, provides an estimate of its state



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 (especially for seasonal and climate prediction)



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...)



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:

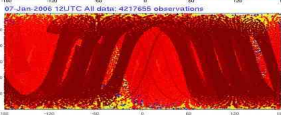
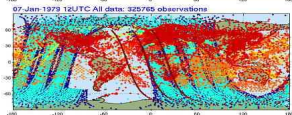
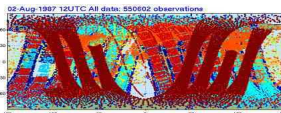
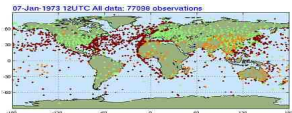
- observations (synoptic profiles, onboard measurements, remote sensing, etc...)
- background field (climatological, short range forecast)



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 ...)



from M.G.Bosilovich (2008)



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

Data assimilation has to deal with:

- initial condition error \implies improved obs coverage - advanced DA algorithms



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

Data assimilation has to deal with:

- initial condition error \implies improved obs coverage - advanced DA algorithms
- **model error** \implies ?

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



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

Data assimilation has to deal with:

- initial condition error \implies improved obs coverage - advanced DA algorithms
- **model error** \implies ?

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:

- 1 large variety of possible error sources



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

Data assimilation has to deal with:

- initial condition error \implies improved obs coverage - advanced DA algorithms
- **model error** \implies ?

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:

- 1 large variety of possible error sources
- 2 the amount of available data insufficient to realistically describe the model error statistics



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

Data assimilation has to deal with:

- initial condition error \implies improved obs coverage - advanced DA algorithms
- **model error** \implies ?

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:

- 1 large variety of possible error sources
- 2 the amount of available data insufficient to realistically describe the model error statistics
- 3 lack of a general framework for model error dynamics and statistics



Formulation

Let assume to have the model:

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

used to describe the true process:

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

$g(\mathbf{y}(t), \lambda')$ represents the dynamics associated to processes not accounted for by the model.



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

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



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

$$\delta \mu = \left. \frac{\partial \mathbf{f}}{\partial \lambda} \right|_{\lambda} \delta \lambda + \gamma \mathbf{g}(\mathbf{y}(t), \lambda')$$

- The model error acts as a deterministic process



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

$$\delta \mu = \left. \frac{\partial \mathbf{f}}{\partial \lambda} \right|_{\lambda} \delta \lambda + \gamma \mathbf{g}(\mathbf{y}(t), \lambda')$$

- The model error acts as a deterministic process
- The important factor controlling the evolution is $\delta \mu(t)$



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

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

- 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 \mathbf{M} , the flow instabilities are expected to influence the model error dynamics



Model error covariance and correlation

Model error covariance

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



Model error covariance and correlation

Model error covariance

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

Model error correlation

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



Model error covariance and correlation

Model error covariance

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

Model error correlation

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

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



Model error covariance and correlation

Model error covariance

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

Model error correlation

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

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

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



Short time approximation

Model error covariance

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



Short time approximation

Model error covariance

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

Model error correlation

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



Short time approximation

Model error covariance

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

Model error correlation

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

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



Short time approximation

Model error covariance

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

Model error correlation

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

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



Short time approximation

Model error covariance

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

Model error correlation

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

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

Model error covariance

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

Model error correlation

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

- 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 $t = t_0$, $\mathbf{Q} = \langle \delta\mu_0 \delta\mu_0^T \rangle$.
- 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 at any time within the short time regime can be computed



Short time approximation

Model error covariance

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

Model error correlation

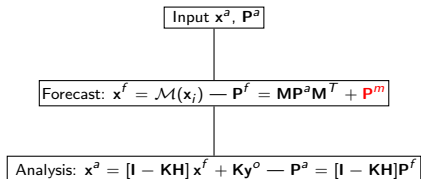
$$\mathbf{P}^m(t_1, t_2) \approx \langle \delta\mu_0 \delta\mu_0^T \rangle (t_1 - t_0)(t_2 - t_0)$$

- 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 $t = t_0$, $\mathbf{Q} = \langle \delta\mu_0 \delta\mu_0^T \rangle$.
- 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 at any time within the short time regime can be computed

Can these short-time approximations be incorporated in data assimilation procedures to account for model error ?



Extended Kalman Filter in the presence of model error - Deterministic formulation

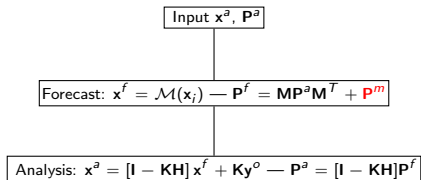


P^m - Model Error Covariance Matrix

Estimate P^m using the short time approximation: $P^m = Q\tau^2$



Extended Kalman Filter in the presence of model error - Deterministic formulation



P^m - Model Error Covariance Matrix

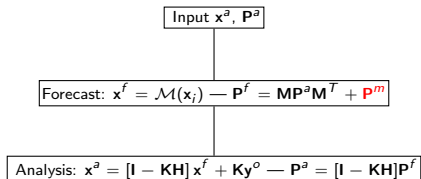
Estimate P^m using the short time approximation: $P^m = Q\tau^2$

...needs to estimate Q ; two solutions analyzed:

- 1 Estimate Q - Statistically based on some a priori information
- 2 Estimate Q - Dynamically (on the fly) using a state/parameter assimilation scheme



Extended Kalman Filter in the presence of model error - Deterministic formulation



P^m - Model Error Covariance Matrix

Estimate P^m using the short time approximation: $P^m = Q\tau^2$

...needs to estimate Q ; two solutions analyzed:

- 1 Estimate Q - Statistically based on some a priori information
- 2 Estimate Q - Dynamically (on the fly) using a state/parameter assimilation scheme

In standard EKF, model error is assumed to be a white-noise and P^m represents the covariance of the process



Observation System Simulation Experiments

- Prototype of nonlinear chaotic dynamics (Lorenz, 1996): $\frac{dx_i}{dt} = \alpha(x_{i+1} - x_{i-2})x_{i-1} - \beta x_i + F \quad 1 \leq i \leq 36$



Observation System Simulation Experiments

- Prototype of nonlinear chaotic dynamics (Lorenz, 1996): $\frac{dx_i}{dt} = \alpha(x_{i+1} - x_{i-2})x_{i-1} - \beta x_i + F \quad 1 \leq i \leq 36$
- Homogeneous network of 18 noisy observations of the state



Observation System Simulation Experiments

- Prototype of nonlinear chaotic dynamics (Lorenz, 1996): $\frac{dx_i}{dt} = \alpha(x_{i+1} - x_{i-2})x_{i-1} - \beta x_i + F \quad 1 \leq i \leq 36$
- Homogeneous network of 18 noisy observations of the state
- Random parametric error (Gaussian sample) $\delta\lambda \in \mathcal{N}(C_I/C_{II}, \sigma_\lambda)$.



Observation System Simulation Experiments

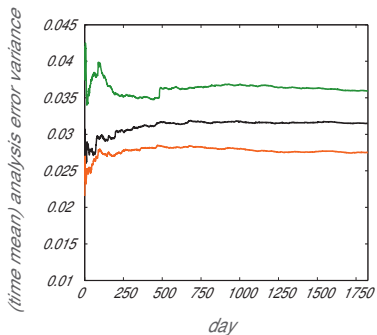
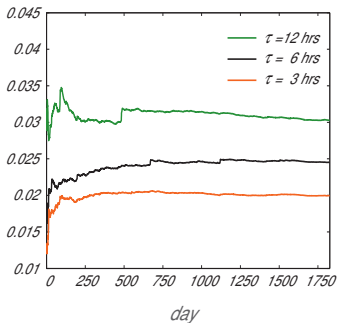
- Prototype of nonlinear chaotic dynamics (Lorenz, 1996): $\frac{dx_i}{dt} = \alpha(x_{i+1} - x_{i-2})x_{i-1} - \beta x_i + F \quad 1 \leq i \leq 36$
- Homogeneous network of 18 noisy observations of the state
- Random parametric error (Gaussian sample) $\delta\lambda \in \mathcal{N}(C_I/C_{II}, \sigma_\lambda)$.



Observation System Simulation Experiments

- Prototype of nonlinear chaotic dynamics (Lorenz, 1996): $\frac{dx_i}{dt} = \alpha(x_{i+1} - x_{i-2})x_{i-1} - \beta x_i + F \quad 1 \leq i \leq 36$
- Homogeneous network of 18 noisy observations of the state
- Random parametric error (Gaussian sample) $\delta\lambda \in \mathcal{N}(C_I/C_{II}, \sigma_\lambda)$.

\mathbf{Q} estimated a-priori - average on a sample of system realizations $\mathbf{Q} = \langle \delta\mu(\mathbf{x})\delta\mu(\mathbf{x})^T \rangle$

white noise EKF*deterministic EKF*

Results EKF - Case (2) \mathbf{Q} online estimate

\mathbf{Q} estimated online by assimilating observations - State Augmented formulation

- augmented system $\mathbf{z} = (\mathcal{M}(\mathbf{x}), \mathcal{F}(\lambda))^T$



Results EKF - Case (2) \mathbf{Q} online estimate \mathbf{Q} estimated online by assimilating observations - 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)



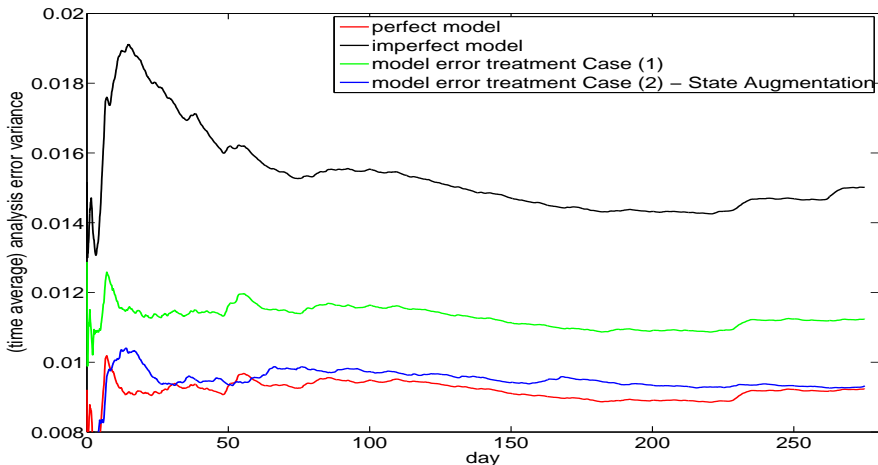
Results EKF - Case (2) \mathbf{Q} online estimate \mathbf{Q} estimated online by assimilating observations - State Augmented formulation

- 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)
- the updated parametric error covariance, $\mathbf{P}_{\lambda}^a = \langle \delta\lambda\delta\lambda^T \rangle$, is then used to update $\mathbf{Q} \Rightarrow \mathbf{P}^m$

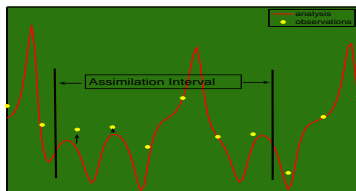


Results EKF - Case (2) \mathbf{Q} online estimate \mathbf{Q} estimated online by assimilating observations - State Augmented formulation

- 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)
- the updated parametric error covariance, $\mathbf{P}_{\lambda}^a = \langle \delta\lambda\delta\lambda^T \rangle$, is then used to update $\mathbf{Q} \Rightarrow \mathbf{P}^m$

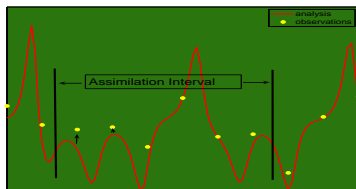


4DVar in the presence of model error - Weak Constraint deterministic formulation



- assimilate observations distributed over the time window τ

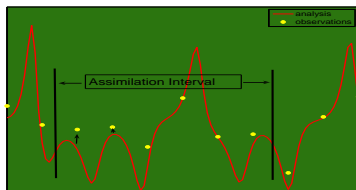
4DVar in the presence of model error - Weak Constraint deterministic formulation



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

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

4DVar in the presence of model error - Weak Constraint deterministic formulation



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

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

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

- In the standard weak-constraint 4DVar, model error is assumed to be white in time \Rightarrow Only the model error covariances \mathbf{P}_t^m need to be estimated



- Lorenz 3-variable (1963) system



Results weak-constraint 4DVar

- Lorenz 3-variable (1963) system
- Assimilation interval $\tau = 8$ time-steps, Obs frequency $\Delta t_{obs} = 2$ time-steps

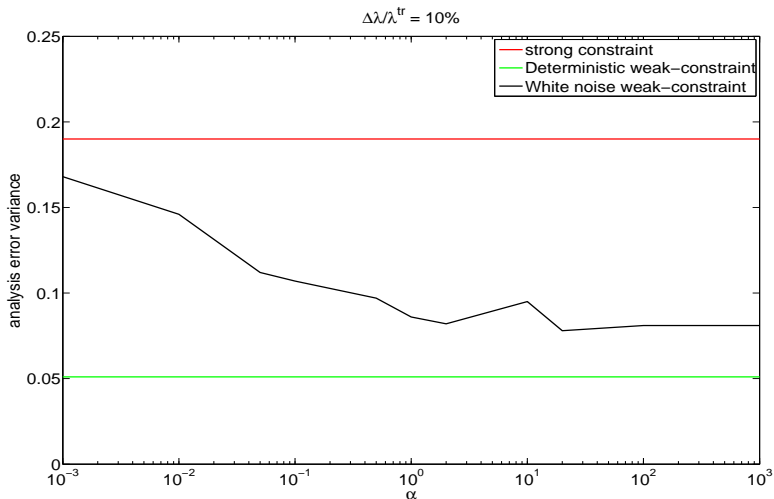


- Lorenz 3-variable (1963) system
- Assimilation interval $\tau = 8$ time-steps, Obs frequency $\Delta t_{obs} = 2$ time-steps
- Random parametric error (Gaussian sample) $\delta\lambda \in \mathcal{N}(0, \Delta\lambda/\lambda)$



Results weak-constraint 4DVar

- Lorenz 3-variable (1963) system
- Assimilation interval $\tau = 8$ time-steps, Obs frequency $\Delta t_{obs} = 2$ time-steps
- Random parametric error (Gaussian sample) $\delta\lambda \in \mathcal{N}(0, \Delta\lambda/\lambda)$



Conclusions and Perspectives

- the deterministic treatment of model error in DA systematically outperformed the standard white noise approach



Conclusions and Perspectives

- the deterministic treatment of model error in DA systematically outperformed the standard white noise approach
- it allows for the estimation of the model error covariance



Conclusions and Perspectives

- the deterministic treatment of model error in DA systematically outperformed the standard white noise approach
- it allows for the estimation of the model error covariance
- in conjunction with the state augmentation formulation it provides a state-dependent estimate of the model error covariance and overcomes the need for a-priori information on model deficiencies



Conclusions and Perspectives

- the deterministic treatment of model error in DA systematically outperformed the standard white noise approach
- it allows for the estimation of the model error covariance
- in conjunction with the state augmentation formulation it provides a state-dependent estimate of the model error covariance and overcomes the need for a-priori information on model deficiencies
- it provides a way to introduce the model error correlations in the weak-constraint variational assimilation



Conclusions and Perspectives

- the deterministic treatment of model error in DA systematically outperformed the standard white noise approach
- it allows for the estimation of the model error covariance
- in conjunction with the state augmentation formulation it provides a state-dependent estimate of the model error covariance and overcomes the need for a-priori information on model deficiencies
- it provides a way to introduce the model error correlations in the weak-constraint variational assimilation
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals



Conclusions and Perspectives

- the deterministic treatment of model error in DA systematically outperformed the standard white noise approach
- it allows for the estimation of the model error covariance
- in conjunction with the state augmentation formulation it provides a state-dependent estimate of the model error covariance and overcomes the need for a-priori information on model deficiencies
- it provides a way to introduce the model error correlations in the weak-constraint variational assimilation
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals

Future directions:

- application to more realistic model and observational scenarios



Conclusions and Perspectives

- the deterministic treatment of model error in DA systematically outperformed the standard white noise approach
- it allows for the estimation of the model error covariance
- in conjunction with the state augmentation formulation it provides a state-dependent estimate of the model error covariance and overcomes the need for a-priori information on model deficiencies
- it provides a way to introduce the model error correlations in the weak-constraint variational assimilation
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals

Future directions:

- application to more realistic model and observational scenarios
- application of the deterministic approach with the state augmentation formulation for the state and parameter estimation for seasonal and climate predictions



Conclusions and Perspectives

- the deterministic treatment of model error in DA systematically outperformed the standard white noise approach
- it allows for the estimation of the model error covariance
- in conjunction with the state augmentation formulation it provides a state-dependent estimate of the model error covariance and overcomes the need for a-priori information on model deficiencies
- it provides a way to introduce the model error correlations in the weak-constraint variational assimilation
- the model error statistics are easily adaptable to different observational frequencies and/or assimilation intervals

Future directions:

- application to more realistic model and observational scenarios
- application of the deterministic approach with the state augmentation formulation for the state and parameter estimation for seasonal and climate predictions

thanks

