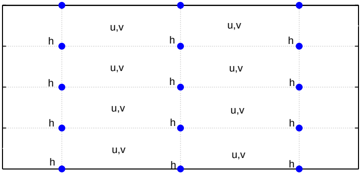
Treatment of model and observation error in an ensemble data assimilation system

Tijana Janjić (Part I: Theory)
and
Andreas Rhodin (Part II)

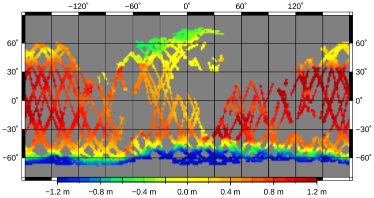
November 25, 2013

Data assimilation algorithm combine forecast and observations to produce the best analysis

$\mathbf{w}_k^b$



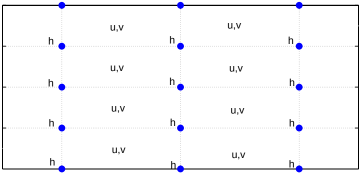
$\mathbf{y}_k^o$



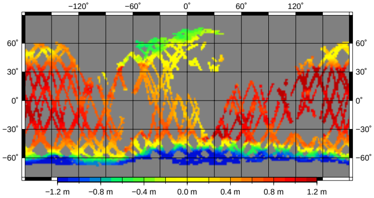
Analysis systems are dependent on appropriate statistics for observation and background errors.

Data assimilation algorithm combine forecast and observations to produce the best analysis

$\mathbf{w}_k^b$



$\mathbf{y}_k^o$



Analysis systems are dependent on appropriate statistics for observation and background errors.

Our goal: Best analysis for a prediction with the numerical model that we are using.

Ensemble Kalman filter methods:

Step 1: Update/Analysis

Combines prior information to obtain an estimate (analysis) $\mathbf{w}_k^a$ of the truth on discrete spatial grid and its error $\mathbf{B}_k^a$.

Ensemble Kalman filter methods:

Step 1: Update/Analysis

Combines prior information to obtain an estimate (analysis) $\mathbf{w}_k^a$ of the truth on discrete spatial grid and its error $\mathbf{B}_k^a$.

Step 2: Resampling

The ensemble Kalman filter requires us to generate a number r of ensemble members $\mathbf{w}_k^{a,i}$, $i = 1, \ldots, N_{ens}$:

$$\mathbf{w}_k^{a,1}, \mathbf{w}_k^{a,2}, \mathbf{w}_k^{a,r-1}, \ldots, \mathbf{w}_k^{a,N_{ens}}$$

Ensemble Kalman filter methods:

Step 1: Update/Analysis

Combines prior information to obtain an estimate (analysis) $\mathbf{w}_k^a$ of the truth on discrete spatial grid and its error $\mathbf{B}_k^a$.

Step 2: Resampling

The ensemble Kalman filter requires us to generate a number r of ensemble members $\mathbf{w}_k^{a,i}$, $i = 1, \ldots, N_{ens}$:

$$\mathbf{w}_k^{a,1}, \mathbf{w}_k^{a,2}, \mathbf{w}_k^{a,r-1}, \ldots, \mathbf{w}_k^{a,N_{ens}}$$

Step 3: Propagation of the ensembles with full nonlinear model

$$\mathbf{w}_k^{a,1}, \mathbf{w}_k^{a,2}, \ldots, \mathbf{w}_k^{a,N_{ens}-1}, \mathbf{w}_k^{a,N_{ens}} \Rightarrow \mathbf{w}_{k+1}^{b,1}, \mathbf{w}_{k+1}^{b,2}, \ldots, \mathbf{w}_{k+1}^{b,N_{ens}-1}, \mathbf{w}_{k+1}^{b,N_{ens}} \Rightarrow \mathbf{w}_{k+1}^{b}, \mathbf{B}_{k+1}^{b}$$

The inputs to the Kalman filter are:

- An initial state at time t_0 and the corresponding covariance matrix $\mathbf{B}_0$
- Observations $\mathbf{y}_k^o$ and observational error covariance $\mathbf{R}_k$ at each analysis time
- Covariance matrix of model error $\mathbf{Q}_k$

We do not need to specify the covariances matrices of background error $\mathbf{B}_k^b$. It is generated and propagated by the filter using full nonlinear dynamics of the model.

$$\mathbf{B}_k^b \approx \frac{1}{N_{ens}-1}\sum_{i=1}^{N_{ens}}[\mathbf{w}_k^{b,i}-\mathbf{w}_k^b][\mathbf{w}_k^{b,i}-\mathbf{w}_k^b]^T.$$

The inputs to the Kalman filter are:

- An initial state at time t_0 and the corresponding covariance matrix $\mathbf{B}_0$
- Observations $\mathbf{y}_k^o$ and observational error covariance $\mathbf{R}_k$ at each analysis time
- Covariance matrix of model error $\mathbf{Q}_k$

We do not need to specify the covariances matrices of background error $\mathbf{B}_k^b$. It is generated and propagated by the filter using full nonlinear dynamics of the model.

$$\mathbf{B}_k^b \approx \frac{1}{N_{ens}-1}\sum_{i=1}^{N_{ens}}[\mathbf{w}_k^{b,i}-\mathbf{w}_k^b][\mathbf{w}_k^{b,i}-\mathbf{w}_k^b]^T.$$

However, we do need to specify $\mathbf{Q}_k$ and $\mathbf{R}_k$ for all k as well as $\mathbf{B}_0$.

Model error

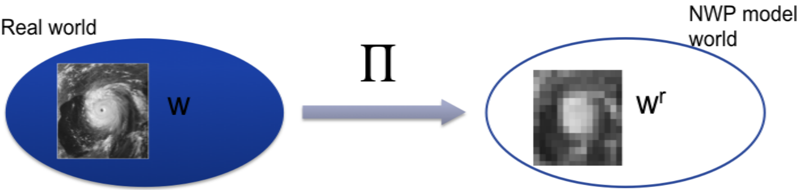
One major contributor to the background error is model error.

Appropriate model error statistics for use in data assimilation algorithms are not known.

What we consider model error:

- accuracy of numerical schemes
- unrepresented subgrid scale processes
- inaccurate forcing and boundary conditions
- representation of orography as well as parametrization uncertainty.

Model Error



from time k to time k+1 atmosphere evolves without us knowing perfectly time propagator, F^c.

from time k to time k+1 numerical model, F, propagates w^r

Model error is the difference: $\Pi F^c(w) - F(w^r).$

Model error

- Model error in DA is assumed to be additive.
- It is defined in the model space, i.e.
  $\mathbf{w}^{r,true} = \mathbf{w}^r + \mathbf{q}$
- Usually $\mathbf{q}$ statistics prescribed (modelled)
- Errors are assumed to be
  - white in time
  - stationary
  - without bias
  - normally distributed.

In ensemble Kalman filter algorithms, full nonlinear numerical model is used to propagate each analysis ensemble member $\mathbf{w}_k^{a,i}$.

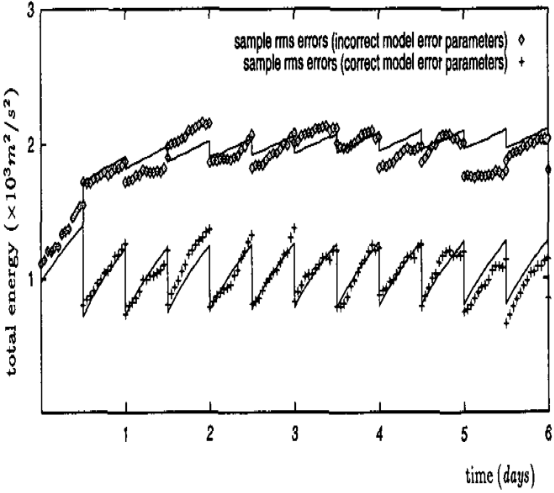
Instead of only propagating the analysis ensemble to obtain the new forecast ensemble, model error $\mathbf{q}_k^i$ can be added to:

$$\mathbf{w}_{k+1}^{b,i} = M_{k+1,k}(\mathbf{w}_k^{a,i}) + \mathbf{q}_{k+1}^i$$

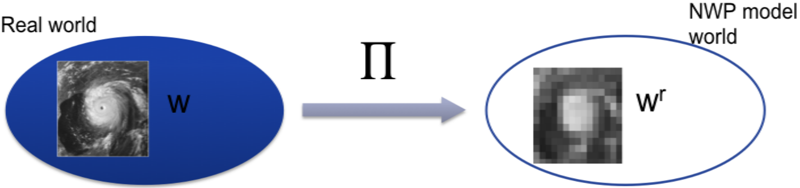
where $\mathbf{q}_k^i$ will be a sample randomly drawn using model error covariance matrix $\mathbf{Q}_k$.

This is done at analysis times!

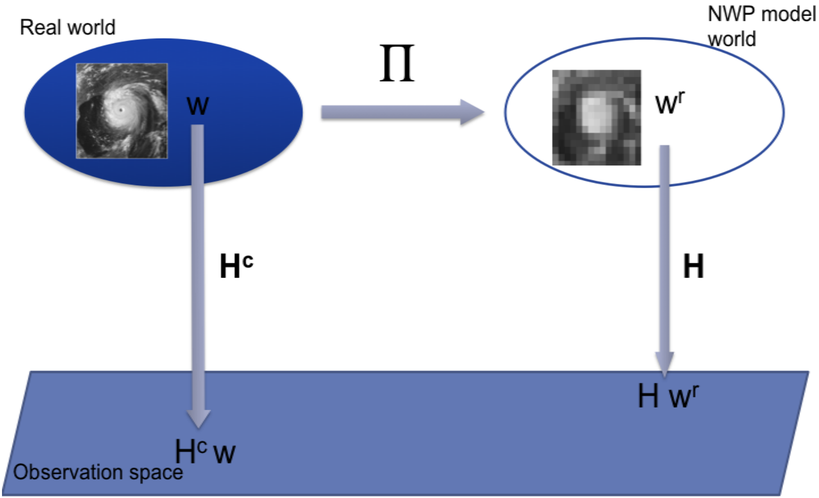
- In case nothing is done about it, the observations even for linear system might degrade the analysis (Dee 1995):



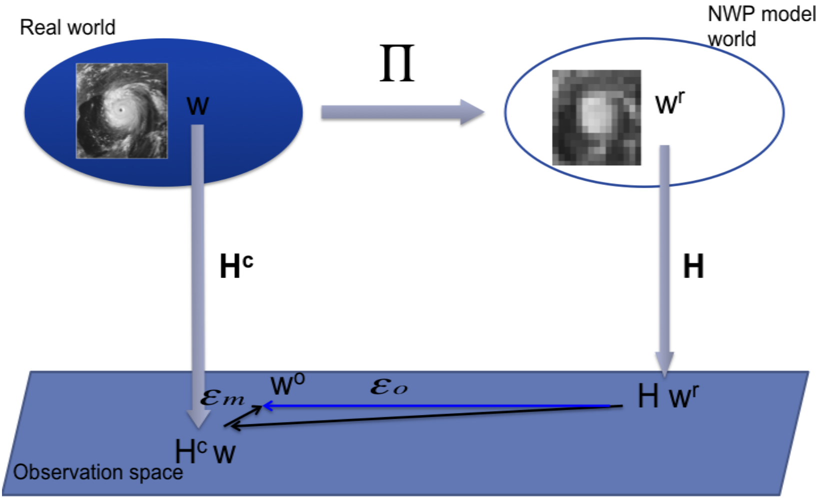
Observation Error



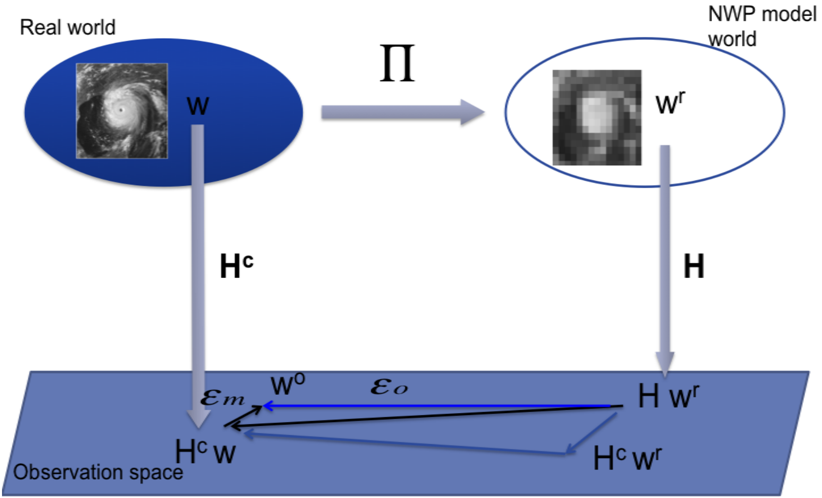
Observation Error



Observation Error



Observation Error



Observation error

ϵ_k^o consists of measurement error and representativeness error. It can be divided into three parts:

$$\epsilon_k^o = \epsilon_k' + \epsilon_k'' + \epsilon_k^m$$

where

$$\begin{aligned} \epsilon_k' &\equiv \mathbf{H}_k^c \mathbf{w}(\cdot, t_k) - \mathbf{H}_k^c \mathbf{\Pi} \mathbf{w}(\cdot, t_k) \\ &= \mathbf{H}_k^c (\mathbf{I} - \mathbf{\Pi}) \mathbf{w}(\cdot, t_k) \end{aligned}$$

ϵ_k' – will be called *error due to unresolved scales*.

$$\begin{aligned} \epsilon_k'' &\equiv \mathbf{H}_k^c \mathbf{\Pi} \mathbf{w}(\cdot, t_k) - \mathbf{H}_k \mathbf{\Pi} \mathbf{w}(\cdot, t_k) \\ &= [\mathbf{H}_k^c - \mathbf{H}_k] \mathbf{\Pi} \mathbf{w}(\cdot, t_k) \end{aligned}$$

ϵ_k'' – will be called *forward interpolation error*.

Representativeness error (Lorenc 1986; Daley 1993; Cohn 1997)

- representativeness error introduces spatial correlations in the observational error
- and it is state and time dependent (Janjic 2001, Janjic and Cohn 2006)

Representativeness error (Lorenc 1986; Daley 1993; Cohn 1997)

- representativeness error introduces spatial correlations in the observational error
- and it is state and time dependent (Janjic 2001, Janjic and Cohn 2006)
- difficult to estimate
- important for optimal use of observations, since it tell us how observations are to be provided to best adopt to model resolution
- for variable model resolutions needs to be scale adaptive
- it depends on the observation type

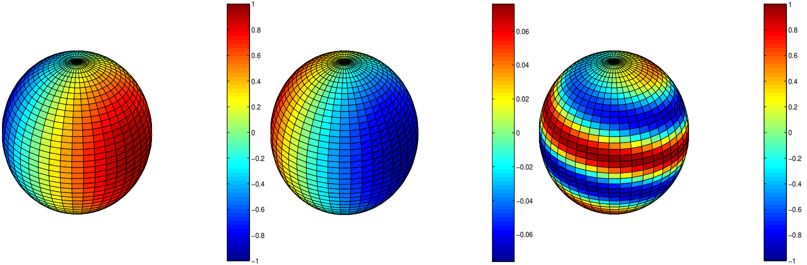
Representativeness error (Lorenc 1986; Daley 1993; Cohn 1997)

- representativeness error introduces spatial correlations in the observational error
- and it is state and time dependent (Janjic 2001, Janjic and Cohn 2006)
- difficult to estimate
- important for optimal use of observations, since it tell us how observations are to be provided to best adopt to model resolution
- for variable model resolutions needs to be scale adaptive
- it depends on the observation type
- Example: for 40×40 km Radiosonde/Dropsonde wind observation, observational error < 0.5 m/s and Assigned error: $2 - 3$ m/s

Data assimilation approach for dealing with observation and model error statistics

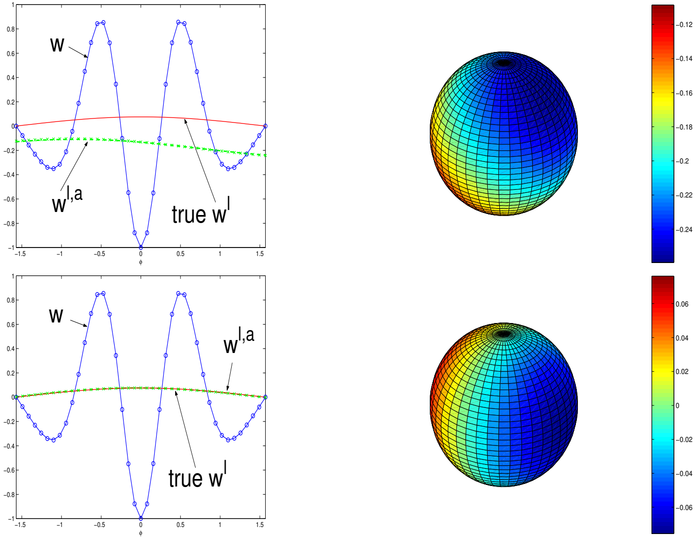
- parametrize the covariances $\mathbf{Q} = \mathbf{V}\mathbf{Q}(\alpha)\mathbf{V}$ and estimate missing parameters from observation minus forecast statistics (Dee 1995, Dee and Arlindo 1997) $v_k = w_k^o - H(w)$, which are assumed $\mathbf{H}\mathbf{P}^{\mathbf{f}}\mathbf{H}^{\mathbf{T}} + \mathbf{R}$
- or from two observing system (Dee and Arlindo 1997) that simultaneously measure the same quantity.
- similar for the observation error statistics (Desrozier et al. 2005)
- or Jung et al. MWR 2012 for assimilation of radar data during forecast perturb shape parameters in the ensemble plus multiplicative and additive inflation (every 5 min u, v, θ perturbed with noise). Noise had std 0.5, 0.5, 0.5 with SM scheme and with DM 0.75,0.75,0.75.

Example: representativeness error



Initial state

Left: The state $w^r(\lambda, \varphi, t)$ which is being estimated.
Right: The full state $w(\lambda, \varphi, t)$ from which the observations are taken.



Cross section at $\lambda = \pi$ and estimate of $w^r(\lambda, \varphi, t)$.

Upper: $\mathbf{H}^c_{2k}[\mathbf{H}^c_{1k} W^{uu}_k(\cdot,\cdot)]^T$ approximated by zero.

Lower: $\mathbf{H}^c_{2k}[\mathbf{H}^c_{1k} W^{uu}_k(\cdot,\cdot)]^T$ approximated by adaptive method.

Conclusion

- Data assimilation algorithms require us to specify the statistical properties of the observation and model error.
- Both of these errors depend on the state of the atmosphere.
- Since we are searching for the best estimate for the scales that our model can represent,
- the unresolved scales are part of the model error as well as observation error.
- Not include them can have a significant impact on accuracy of the analysis.