Some thoughts on hybrid approach to data assimilation

Monika Krysta, Xudong Sun, Jin Lee, Susan Rennie, Peter Steinle
Variational DA in an incremental formulation

Search for an optimal state of the atmosphere defined as the solution to the minimization problem

\[ J(\delta w) = \frac{1}{2} (\delta w - \delta w^b)^T B^{-1} (\delta w - \delta w^b) + \frac{1}{2} (y - y^0)^T (R)^{-1} (y - y^0) \]

Focus on the background error covariance matrix

- regularises an ill-posed problem
- weights a contribution of the background state relative to the information provided by observations
- spreads out spatially information carried by observations
- imposes physical balances on the analysed fields
Construction of $B$

- For an underlying discretised problem it is a very large object
- Control variable transform
  - $B$ is factorised into $B = UU^T$
  - $U$ translates model variables $\delta w$ into control variables $\nu$ and is defined via a sequence of transforms

$$\delta w = U_p U_h U_v \nu$$

$$J(\nu) = \frac{1}{2} \nu^T \nu + \frac{1}{2} (y - y_0)^T R^{-1} (y - y_0)$$

- $U$ derived empirically from statistical properties of a training set
- UM control variables:
  $$\psi, \chi, A p, \mu$$
Training set

- Based on an ensemble prediction system
- MOGREPS with ETKF-based perturbations
- Departures of T6 forecasts from the ensemble mean constitute proxies of the background error
- 44 members for base time every 3.25 days in August 2015 selected months with 00Z, 06Z, 12Z, 18Z considered twice
- 352 elements in a training set
Can we get an insight into $B$?

- **Directly**
  - vertical profiles of standard deviations and correlations degree of explained variances for statistical balances
  - horizontal lengthscales for the horizontal transform
  - empirical eigenvalues and eigenvectors for vertical transform

- **Indirectly – influence it exerts on the analyses**
  - single observation tests
  - sections of the $B$ matrix (columns) via an action of a tailored control vector
  - sampling of $B$ matrix by acting on a random sample drawn from a distribution with an identity covariance matrix
Standard deviations
Vertical profiles averaged over the domain

$U_{inc}$

$V_{inc}$

$PSI_{inc}$

$CHI_{inc}$
Standard deviations
Vertical profiles averaged over the domain

\[ p^A_{inc} = p^H - p^G_v \]

\[ p^G_v = B^H_v (B^G_v)^{-1} p^G \]
aP explained by hP

\[ 1 - \frac{\text{var}(p^A_{\text{inc}})}{\text{var}(p^H_{\text{inc}})} \]
Vertical correlations

\[ \text{PSI}_{\text{inc}} \]

\[ \text{CHI}_{\text{inc}} \]

\[ \text{P}_{\text{inc}}^A \]

\[ \text{MU}_{\text{inc}} \]
Horizontal lengthscales

$\text{PSI}_{\text{inc}}$

$\text{CHI}_{\text{inc}}$

$\rho^A_{\text{inc}}$

$\text{MU}_{\text{inc}}$
Training set

- Based on an ensemble prediction system
- MOGREPS with ETKF-based perturbations
- Departures of T6 forecasts from the ensemble mean constitute proxies of the background error
- 44 members for base time every 3.25 days in August 2015 selected months with 00Z, 06Z, 12Z, 18Z considered twice
- 352 elements in a training set
- Hybrid 4D-Var instead 4D-Var runs to build the training set
- Increase the number of elements in the training set
- Select a different time period
Outlook

- Compare with the climatological B currently used in operations

- Assess if the MOGREPS-based training set results in a viable background error covariance matrix

- Transfer, with care, experience gathered in a global to a regional model

- Test a hybrid 4D-Var with the MOGREPS-based climatological B in a regional model
Thank you...
Sensitivity to ensemble size
Lengthscales
352/176 members for Aug 2015
Sensitivity to ensemble size
Vertical correlations
352/176 members for Aug 2015
Old B hybrid (aa-670)
Hybrid with new B (u-ah826)
Two training sets

- **August 2015**
  - 4D-Var
    - Departures of T6 forecasts with respect to the ensemble mean constitute proxies of the background error
    - MOGREPS ETKF-based ensemble comprising 44 members for base time every 3.25 days in the selected months with 00Z, 06Z, 12Z, 18Z considered twice
      - 352 states
    - Suboptimal resolution of n216 replaced with n320 which is the UM resolution
      - Increase the number of elements in the training set

- **February 2016**
  - Hybrid 4D-Var
Illustration of standard deviations
Illustration of lengthscales