Hybrid Variational/Ensemble Data Assimilation

Zhiquan Liu (liuz@ucar.edu)
NCAR/MMM

Acknowledgements: D. M. Barker, M. Demirtas, A. P. Mizzi
Outline

• Motivation and differently proposed hybrid DA

• Elements of hybrid DA

• Preliminary results

• Introduction to practice
Why Hybrid?

• 3D-Var uses static ("climate") BE

\[ J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} [y - H(x)]^T R^{-1} [y - H(x)] \]

• 4D-Var implicitly uses flow-dependent information, but still starts from static BE

\[ J_j(\delta x_j) = \frac{1}{2} (\delta x_j - \delta x_j^b)^T B^{-1} (\delta x_j - \delta x_j^b) \]

\[ + \frac{1}{2} \sum_{k=0}^{K} (H_{j,k} M_{j,k} \delta x_j - d_{j,k})^T R^{-1} (H_{j,k} M_{j,k} \delta x_j - d_{j,k}) \]

• Hybrid: using flow-dependent background error information from ensemble in a variational DA system
T Analysis increments from a single T obs

1K difference, 1K error

3DVAR

Hybrid (64 members)
What is the Hybrid DA?

• Combine ensemble and variational DA together

• Ensemble mean is analyzed by a variational algorithm (i.e., minimize a cost function). It combines the 3DVAR “climate” background error covariance and “error of the day” from ensemble.

• A system for updating ensemble
  – Could be (independent) ensemble forecasts already available from NWP centers
  – Could be an EnKF-based DA system
  – Could be an ETKF-based ensemble system
Hamill and Snyder, 2000

- 3DVAR cost function

\[
J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} [H(x) - y]^T R^{-1} [H(x) - y]
\]

- Idea: replace B by a weighted sum of 3DVAR B and the ensemble covariance

\[
B = \alpha_1 B_1 + \alpha_2 B_2, \quad \alpha_1 = 1 - \alpha_2
\]
Lorenc, 2003

- Ensemble covariance is included in the 3DVAR cost function through augmentation of control variables (Lorenc, 2003)

\[ J(x, \alpha) = \beta_1 J_b + \beta_2 J_e + J_o \]

= \beta_1 \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \beta_2 \frac{1}{2} \alpha^T A^{-1} \alpha + \frac{1}{2} [y - H(x + x_e)]^T R^{-1} [y - H(x + x_e)]

- This is implemented in WRFDA (Wang et al., 2008)

- It is mathematically equivalent to Hamill and Snyder (2000).
Advantages of the Hybrid DA?

• Hybrid DA system can be more robust than a pure EnKF-based DA

  – For some observations type, e.g., radiances, localization is not well defined in observation space, bias correction issues

  – Localization is in model space in a variational framework.

  – For small-size ensemble since can adjust amount of 3DVAR and ensemble covariances.
Elements of Hybrid DA

• Ensemble forecasts: *WRF-ensemble forecasts*

• Ensemble Transform Kalman Filter (ETKF):
  • Update forecast/background ensemble perturbations to analysis ensemble perturbations

• A Variational DA to update ensemble mean.
Cycling WRF/WRFDA/ETKF System (Hybrid DA)

Ensemble Forecast -> Ensemble Perturbations -> Updated Ensemble Perturbations

\[ \begin{align*}
X_1^a & \rightarrow X_1^f \\
X_2^a & \rightarrow X_2^f \\
& \cdots \\
X_N^a & \rightarrow X_N^f \\
\end{align*} \]

\[ \begin{align*}
\delta X_1^f & \rightarrow H(x_1^f), \sigma_o \\
\delta X_2^f & \rightarrow H(x_2^f), \sigma_o \\
& \cdots \\
\delta X_N^f & \rightarrow H(x_N^f), \sigma_o \\
\end{align*} \]

\[ \begin{align*}
\delta X_1^a & \rightarrow X_1^a \\
\delta X_2^a & \rightarrow X_2^a \\
& \cdots \\
\delta X_N^a & \rightarrow X_N^a \\
\end{align*} \]

Ensemble Mean of forecasts \[ \bar{X}^f \]

Ensemble Mean analysis \[ \bar{X}^a \]

3/4D-Var

\[ y^o \]
Hybrid DA: Variational Part

Ensemble Perturbations (extra input)

$\delta x_1^f$

$\delta x_2^f$

$\delta x_N^f$

$\bar{x}^f$

3/4D-Var

$\bar{x}^a$

$y^o$

Ensemble Mean of forecasts

Ensemble Mean analysis

2011 Summer WRFDA tutorial
Hybrid 3DVAR formulation

- Ensemble covariance is included in the 3DVAR cost function through augmentation of control variables (Lorenc, 2003; Wang et al., 2008)

\[
J(x, \alpha) = \beta_1 J_b + \beta_2 J_e + J_o
\]

\[
= \beta_1 \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \beta_2 \frac{1}{2} \alpha^T A^{-1} \alpha + \frac{1}{2} [y - H(x + x_e)]^T R^{-1} [y - H(x + x_e)]
\]

\[
\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1, \quad x_e = \frac{1}{\sqrt{N} - 1} \sum_{i=1}^{N} \alpha_i \cdot x_i', \text{ where } x_i' \text{ is the ensemble perturbation for the member } i.
\]

The extended control variable \( \alpha = (\alpha_1, \alpha_2, ..., \alpha_N) \) has dimension of \( M(\text{dimension of } x) \times N \) (ensemble size)

The matrix \( A \) plays the role for ensemble covariance localization.

\[
A = \begin{pmatrix}
S \\
\vdots \\
S
\end{pmatrix}, \quad S = \langle \alpha_i \rangle (\alpha_i)^T
\]
Hybrid 3DVAR formulation

• Equivalently can write in another form (Wang et al., 2008)

\[
J(x, \alpha) = \frac{1}{2} (x + x_e - x_b)^T \left( \frac{1}{\beta_1} B + \frac{1}{\beta_2} P^e \circ S \right)^{-1} (x + x_e - x_b)
\]

\[
+ \frac{1}{2} [y - H(x + x_e)]^T R^{-1} [y - H(x + x_e)]
\]

\[
P^e = \frac{1}{N-1} (x')(x')^T \text{ is the sample ensemble covariances.}
\]

• This explains why S is for localization.

• This is also equivalent to Hamill and Snyder (2000).
Hybrid DA: Ensemble Part (ETKF-based)

\[ \delta x_1^f \rightarrow H(x_1^f, \sigma_i) \rightarrow \delta x_1^a \]

\[ \delta x_2^f \rightarrow H(x_2^f, \sigma_o) \rightarrow \delta x_2^a \]

\[ \delta x_N^f \rightarrow H(x_N^f, \sigma_o) \rightarrow \delta x_N^a \]

\[ \text{3/4D-Var} \]

\[ y^o \]
ETKF formulation

- The ETKF (Wang et al. 2007) finds the transformation matrix $T$ to update forecast/background perturbations to analysis perturbations

$$\delta x^a = \delta x^f T$$

$$T = rE(\rho \lambda + I)^{-1/2} E^T$$

- Where $E$ and $\lambda$ contain eigenvectors and eigenvalues of a NxN (N is ensemble size) matrix

$$[H(\delta x^f)]^T R^{-1} [H(\delta x^f)] / (N - 1)$$

$$H(\delta x_k^f) = H(x_k^f) - H(x_k^f)$$
Inflation

• $r$ inflation factor, $\rho$ accounts for the fraction of the forecast-error variance projected onto the ensemble subspace.
  
  – Both factors are adaptively calculated for each DA cycle by using innovation statistics.

• Inflation is to ensure that on average the background error variance estimated from the spread of ensembles is consistent with innovation statistics, i.e.,

$$d^T R^{-1} d \approx \text{trace}\left( \frac{1}{N-1} \sum_{i=1}^{N} \left[ H(x_i) - H(\bar{x}) \right] R^{-1} \left[ H(x_i) - H(\bar{x}) \right] + I \right)$$
Pros and Cons of ETKF

• Desirable aspects:
  – ETKF is fast (computations are done in model ensemble perturbation subspace).
  – It directly updates perturbations.

• Less desirable aspects:
  – Not localized, therefore it does not represent sampling error efficiently. It may need very high inflation factors.

• Alternatives for ensemble part
  – EnKF, Perturbed obs, LETKF
Old results (need to update in the future)

- Ensemble size: 10
- Cycle frequency: 3 hours
- Observations: GTS conventional observations
- Deterministic ICs/BCs: Down-scaled GFS forecasts
- Ensemble ICs/BCs: Produced by adding spatially correlated Gaussian noise to GFS forecasts.
- Horizontal resolution: 45km
- Number of vertical levels: 57
- Model top: 50 hPa
Ensemble spread: 500 hPa height (m) std. dev.

WRF t+3 valid at 2007081900

Modest inflations factors used

Higher inflations factors used

2011 Summer WRFDA tutorial
Ensemble Mean and Std. Deviation (spread)
Inflation Factors
Prior Ensemble Mean

Failed ETKF (Hurricane Dean)
Restarted for those cases

Ensemble Spread
Hybrid gives better RMSE scores for wind compared to 3D-VAR.
Hybrid gives better RMSE scores for wind compared to 3D-VAR.
Introduction of Hybrid practice session

- **Computation:**
  - Computing ensemble mean.
  - Extracting ensemble perturbations (EP).
  - Running WRFDA in "hybrid" mode.
  - Displaying results for: ens_mean, std_dev, ensemble perturbations, hybrid increments, cost function and, etc.
  - If time permits, tailor your own test by changing hybrid settings; testing different values of “je_factor” and “alpha_corr_scale” parameters.

- **Scripts to use:**
  - Some NCL scripts to display results.
  - **Ensemble generation part not included in current practice**
Brief information for the chosen case

Ensemble size: 10

Domain info:
- time_step=240,
- e_we=122,
- e_sn=110,
- e_vert=42,
- dx=45000,
- dy=45000,

Input data provided (courtesy of JME Group):
- WRF ensemble forecasts valid at 2006102800
- Observation data (ob.ascii) for 2006102800
- 3D-VAR “be.dat” file
References


