Hybrid Variational/Ensemble Data Assimilation

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

• Background

• Hybrid formulation in a variational framework

• Some results

• Introduction to hybrid practice
Motivation of Hybrid DA

• 3D-Var uses static ("climate") BE

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

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

\[ J(\delta x) = \frac{1}{2} \delta x^T B^{-1} \delta x + \frac{1}{2} \sum_{i=1}^{I} [HM_i \delta x - d_i]^T R^{-1} [HM_i \delta x - d_i] \]

• Hybrid uses flow-dependent background error covariance from forecast ensemble perturbation in a variational DA system
What is the Hybrid DA?

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

• Hybrid algorithm (again in a variational framework) itself usually does not generate ensemble analyses.

• Need a separate system to update ensemble
  – Could be ensemble forecasts already available from NWP centers
  – Could be an Ensemble Kalman Filter-based DA system
  – Or multiple model/physics ensemble

• Ensemble needs to be good to well represent “error of the day”
single observation tests

Potential temperature increment, 21\textsuperscript{st} model level

Pure EnKF

3DVAR

Hybrid-full ensemble

Hybrid 50/50

Average increment of T (K)

-0.28 -0.24 -0.2 -0.16 -0.12 -0.08 -0.04 0 0.04 0.08 0.12 0.16 0.2 0.24 0.28
Hybrid formulation (1)  
(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 static \( B_s \) and the ensemble \( B_e \)

\[
B = a_s B_s + a_e B_e, \quad a_s = 1 - a_e
\]

  – Has been demonstrated on a simple model.
  – Difficult to implement for large NWP model.
Hybrid formulation (2): used in WRFDA (Lorenc, 2003)

• Ensemble covariance is included in the 3DVAR cost function through augmentation of control variables

\[
J(x, \alpha) = \beta_s \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \beta_e \frac{1}{2} \sum_{i=1}^{N} \alpha_i^T C^{-1} \alpha_i \\
+ \frac{1}{2}[y - H(x + x_e')]^T R^{-1}[y - H(x + x_e')]
\]

\[x_e' = \sum_{i=1}^{N} \alpha_i \circ x_i', \text{ where } x_i' \text{ is the ensemble perturbation for the ensemble member } i.\]

\[\circ \text{ denote element-wise product. } \alpha_i \text{ is in effect the ensemble weight.}\]

\[C: \text{ correlation matrix (effectively localization of ensemble perturbations)}\]

• In practical implementation, \(\alpha_i\) can be reduced to horizontal 2D fields (i.e., use same weight in different vertical levels) to save computing cost.

• \(\beta_s\) and \(\beta_e\) (\(1/\beta_s + 1/\beta_e = 1\)) can be tuned to have different weight between static and ensemble part.
Hybrid formulation (3)

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

\[
J(x, \alpha) = \frac{1}{2} (x + x_e - x_b)^T \left( \frac{1}{\beta_s} B + \frac{1}{\beta_e} B_e \circ C \right)^{-1} (x + x_e - x_b) \\
+ \frac{1}{2} [y - H(x + x_e)]^T R^{-1} [y - H(x + x_e)]
\]

• \( C \) is “localization” matrix
Hybrid DA data flow

Ensemble Perturbations (extra input for hybrid)

For cycling data assimilation/forecast Experiment, need a mechanism to update ensemble.
EnKF-based Ensemble Generation

- EnKF with perturbed observations

- EnKF without perturbed observations
  - All based on square-root filter
  - Ensemble Transformed Kalman Filter (ETKF)
  - Ensemble Adjustment Kalman Filter (EAKF)
  - Ensemble Square-Root Filter (EnSRF)

- Most implementation assimilates obs sequentially (i.e., one by one, or box by box)
  - can be parallelized

More information was given in 2012 slides.
Advantages of the Hybrid DA

• Hybrid localization is in model space while EnKF localization is usually in observation space.

• For some observations type, e.g., radiances, localization is not well defined in observation space

• Easier to make use of existing radiance VarBC in hybrid

• For small-size ensemble, use of static B could be beneficial to have a higher-rank covariance.
Paula case: 0600 UTC 10 October 2010 to 1200 UTC 15 October 2010;
Background: 15km interpolated from GFS data;
Resolution: 718x 373 (15km) and 43 levels;
Observations: GTS and TAMDAR;
Cycle frequency: 6 hours;
Background error: CV5;
Time widows: 2 hours;

TAMDAR: a new Tropospheric Airborne Meteorological Data Reporting (TAMDAR) observing system that has been developed by AirDat company.
Experiments:

CYC1: assimilate GTS and TAMDAR with Hybrid (w/ TAMDAR H);

CYC2: same to CYC1, but no TAMDAR (w/o TAMDAR H)

CYC3: assimilate GTS and TAMDAR with standard 3DVAR (Deterministic WRFDA)
inflation and fraction factor
Forecast Verification: RMSE

+12hr

+24hr
Track Forecast Verification (+24hr)
Hybrid practice

- **Computation steps:**
  - Computing ensemble mean (gen_be_ensmean.exe).
  - Extracting ensemble perturbations (gen_be_ep2.exe).
  - Running WRFDA in “hybrid” mode (da_wrfvar.exe).
  - Displaying results for: ens_mean, std_dev, ensemble perturbations, hybrid increments, cost function
  - If time permits, play with different namelist settings: “je_factor” and “alpha_corr_scale”.

- **Scripts to use:**
  - Some NCL scripts to display results.

- **Ensemble generation part not included in current practice**
Namelist for WRFDA in hybrid mode

&wrfvar7
je_factor=2,     # half/half for Jb and Je term (tunable parameter)

&wrfvar16
alphacv_method=2,       # ensemble part is in model space (u,v,t,q,ps)
ensdim_alpha=10,
alpha_corr_type=3,  # 1=Exponential; 2=SOAR; 3=Gaussian
alpha_corr_scale=750.,  # correlation scale in km (tunable parameter)
alpha_std_dev=1.,
alpha_vertloc=true,  (use program “gen_be_vertloc.exe 42” to generate file)
Dual-Resolution hybrid (V3.6)


Doing Hybrid-Analysis at 15km d02 grid but with ensemble perturbation input from 45km d01 grid
Dual-resolution cost-function

• High-resolution (HR) variables:
  – $x_1, B, H, \delta x$

• Low-resolution (LR) variables:
  – $a, A, D$

$$J(x_1,a) = \frac{\beta_1}{2} (x_1)^T B^{-1} x_1 + \frac{\beta_2}{2} a^T A^{-1} a + \frac{1}{2} (d - H\delta x)^T R^{-1} (d - H\delta x)$$

This term requires interpolation from low to high resolution.
Intermediate domain

- WRFDA directly reads in d01 ensembles, then cut to d02 size (making use of WRF model nest namelist setting)
References


