Hybrid Variational-Ensemble Data Assimilation at NCEP

Daryl Kleist
NOAA/NWS/NCEP/EMC

with acknowledgements to Kayo Ide, Dave Parrish, Jeff Whitaker, John Derber, Russ Treadon, Wan-Shu Wu, Jacob Carley, and Mingjing Tong

The 4th Annual PSU-UMD DA Workshop – 13 December 2012
Outline

• Introduction
  – (Brief) background on hybrid data assimilation

• Hybrid Var/Ens at NCEP: Present and Future

• Research Developments (toward improving global hybrid)
  – 4DEnsVar (and hybridization)
  – Scale-dependent weights
Kalman Filter in Variational Setting

Forecast Step
\[
\begin{align*}
x^b_{t+1} &= M(x^a_t) \\
B_{KF} &= MA_{KF} M^T + Q \\
x^a &= x^b + K (y^o - Hx^b) \\
K &= B_{KF} H^T (R + HB_{KF} H^T)^{-1} \\
A_{KF} &= (I - KH) B_{KF}
\end{align*}
\]

Analysis

- Analysis step in variational framework (cost function)
  \[
  J_{KF}(x') = \frac{1}{2} (x')^T B_{KF}^{-1} (x') + \frac{1}{2} (Hx' - y')^T R^{-1} (Hx' - y')
  \]
- \( B_{KF} \): Time evolving background error covariance
  \[
  A_{KF}^{-1} = B_{KF}^{-1} + H^T R^{-1} H
  \]

Extended Kalman Filter
Incorporate ensemble perturbations directly into variational cost function through extended control variable

- Lorenc (2003), Buehner (2005), Wang et. al. (2007), etc.

\[ J(x'_f, \alpha) = \beta_f \frac{1}{2}(x'_f)^T B_f^{-1}(x'_f) + \beta_e \frac{1}{2} \sum_{n=1}^{N} (\alpha^n)^T L^{-1}(\alpha^n) + \frac{1}{2}(Hx'_t - y')^T R^{-1}(Hx'_t - y') \]

\[ x'_t = x'_f + \sum_{n=1}^{N} (\alpha^n \circ x_e^n) \]

\( b_f \) & \( b_e \): weighting coefficients for fixed and ensemble covariance respectively
\( x'_t \): (total increment) sum of increment from fixed/static \( B \) (\( x'_f \)) and ensemble \( B \)
\( a_k \): extended control variable; \( X^e_k \):ensemble perturbations
- analogous to the weights in the LETKF formulation
\( L \): correlation matrix [effectively the localization of ensemble perturbations]
Single Temperature Observation

\[ b_f^{-1} = 0.0 \]

\[ b_f^{-1} = 0.5 \]
Outline

• Introduction
  – (Brief) background on hybrid data assimilation

• Hybrid Var/Ens at NCEP: Present and Future

• Research Developments (toward improving global hybrid)
  – 4DEnsVar (and hybridization)
  – Scale-dependent weights
NOAA’s NWS Model Production Suite

Global Forecast System

- Climate
  - CFS
  - MOM3

- Hurricane
  - GFDL
  - HWRF

Regional NAM

NMM-B

Oceans

- HYCOM
- WaveWatch III

Dispersion

- ARL/HYSPLIT

Severe Weather

- WRF NMM/ARW
- Workstation WRF

Air Quality

- NAM/CMAQ

*Rapid Update for Aviation

Global Data Assimilation

1.7B Obs/Day Satellites 99.9%

Regional Data Assimilation

North American Ensemble Forecast System

GFS, Canadian Global Model

Short-Range Ensemble Forecast

WRF: ARW, NMM

NMM-B

NOAH Land Surface Model

- NOAH Land Surface Model

Coupled
• Package included other changes
  • NPP ATMS (MW): 7 months after launch!!
    • This is by far the fastest we have ever begun assimilating data from a new satellite sensor after launch
  • GPS RO Bending Angle replaced Refractivity

• Summary of pre-implementation retrospective testing
  • Improved Tropical winds
  • Improved mid-latitude forecasts
  • Fewer Dropouts
  • Improved fits to observations of forecasts
  • Some improvement in NA precip. in winter
  • Increased bias in NA precip. – decreased rain/no rain skill in summer (Improved by land surface bug fix)
  • Overall significant improvement of GFS forecasts
TC Track Error Reduction

HYBRID TEST

GFS OPERATIONAL

NHC/JTWC OFFICIAL
500 hPa AC

Northern Hemisphere

Southern Hemisphere

AC differences outside of outline bars are significant at the 95% confidence level.
NAM vs NAM parallels upper air stats vs raobs

Ops NAM = Solid ; NAMB (with Physics changes) = Dashed ; NAMX (with physics changes and using global EnKF in GSI) = Dash-Dot

Thanks to Eric Rogers and Wan-Shu Wu
RAP hybrid DA using global ensemble

RAP GSI-hybrid vs. RAP GSI-3dvar
upper-air verification

+ 6 h forecast

RMS Error

28 Nov – 3 Dec 2012
Assimilation of NOAA-P3 Tail Doppler Radar (TDR) Data using GSI hybrid method for HWRF

- HWRF Model: 3 domains with 0.18-0.06-0.02 degree (27-9-3 km) horizontal resolutions, 43 vertical levels with model top at 50 hPa, with ocean coupling
- TC environment cold start from GDAS forecast, TC vortex cycled from HWRF forecast
- GSI hybrid analysis using GFS EnKF ensemble
  - 80% of background error covariance from ensemble B.
  - Horizontal localization 150 km, vertical localization 10 model levels for weak storm and 20 model levels for strong storm
- Conventional data plus TDR data
- Modified gross error check, re-tuned observation error and rejected data dump with very small data coverage for TDR data
- 19 TDR missions for Hurricane Isaac, Leslie and Sandy
GSI + NMMB (currently operational NAM code, 1.33 km grid spacing)
Assimilate both dBZ and velocity over a 1 hour period at 5 min. intervals
Followed by 1hr forecast

Experiments:
3DVar
Single, deterministic simulation
Hybrid ensemble-3DVar (32 ensemble members)
3 configurations
25%, 50%, 75% static B contribution (Hyb B25, Hyb B50, Hyb B75)
Control (CTL)
No data assimilation
Single, deterministic forecast
Outline

• Introduction
  – (Brief) background on hybrid data assimilation

• Hybrid Var/Ens at NCEP: Present and Future

• Research Developments (toward improving global hybrid)
  – 4DEnsVar (and hybridization)
  – Scale-dependent weights
As in Buehner (2010), the H-4DVAR_AD cost function can be modified to solve for the ensemble control variable (without static contribution)

\[ J(\alpha) = \frac{1}{2} \sum_{n=1}^{N} (\alpha^n)^T L^{-1}(\alpha^n) + \frac{1}{2} \sum_{k=1}^{K} (H_k x'_k - y'_k)^T R_k^{-1}(H_k x'_k - y'_k) \]

Where the 4D increment is prescribed exclusively through linear combinations of the 4D ensemble perturbations

\[ x'_k = \sum_{n=1}^{N} (\alpha^n \circ (x^n_c)_k) \]

Here, the control variables (ensemble weights) are assumed to be valid throughout the assimilation window (analogous to the 4D-LETKF without temporal localization). Note that the need for the computationally expensive linear and adjoint models in the minimization is conveniently avoided.
The 4DENSV cost function can be easily expanded to include a static contribution

$$J(x', \alpha) = \beta_f \frac{1}{2} (x_f')^T B_f^{-1} (x_f') + \beta_e \frac{1}{2} \sum_{n=1}^{N} (\alpha^n)^T L^{-1} (\alpha^n) +$$

$$\frac{1}{2} \sum_{k=1}^{K} (H_k x_k' - y_k')^T R_k^{-1} (H_k x_k' - y_k')$$

Where the 4D increment is prescribed exclusively through linear combinations of the 4D ensemble perturbations plus static contribution

$$x_k' = x_f' + \sum_{n=1}^{N} (\alpha^n \circ (x_e)_k^n)$$

Here, the static contribution is considered time-invariant (i.e. from 3DVAR-FGAT). Weighting parameters exist just as in the other hybrid variants.
Single Observation (-3h) Example for 4D Variants

4DVAR (B-NMC)

4DENSV

H-4DVAR_AD

$\beta_i^{-1} = 0.25$

H-4DENSV

$\beta_i^{-1} = 0.25$
Time Evolution of Increment

Solution at beginning of window same to within round-off (because observation is taken at that time, and same weighting parameters used)

Evolution of increment qualitatively similar between dynamic and ensemble specification

** Current linear and adjoint models in GSI are computationally unfeasible for use in 4DVAR other than simple single observation testing at low resolution
OSSE-based comparison of 3DHYB and H-4DENS

Something in the 4D experiments is resulting in more moisture in the analysis, triggering more convective precipitation
Scale-Dependence Motivation
(Courtesy: Tom Hamill)

(1) Generally more power at all wavenumbers relative to ETR.
(2) Overestimate of power (i.e., amplitude of perturbations) at small scales. Likely this is attributable to inappropriate analysis increments due to the use of smaller-than-ideal ensemble size (n=80) in the EnKF, and still-crude methods (covariance localization) for filtering usable signal from sampling noise.
Spectrum of Dual-Resolution Increment with SD-weighting
Initial scale-dependent tests in
and OSSE
• NCEP successfully implemented hybrid variational-ensemble algorithm into GDAS

• NCEP aggressively pursuing application of hybrid to other systems
  – Mesoscale (NAM), HWRF, Rapid Refresh (and HRRR follow on), storm scale ensemble
  – Future Reanalysis
    • Have already run preliminary tests for 1981-1983 periods, attempting to capture QBO transitions (a notoriously difficult problem for reduced observing system periods)

• Extensions to the GDAS hybrid are ongoing, including 4DEnsVar (with a planned future implementation, much like the UKMO and Canadians)
6th WMO Symposium on DA

• NCEP will be hosting the next WMO DA Symposium from 7-11 October 2013 at NCWCP

• Call for papers will be out in a month or two

• Register early, as space will be limited

• Will potentially be looking for volunteers to help out, especially during week of the symposium