GSI-based hybrid variation-ensemble data assimilation for the NCEP GFS

Daryl T. Kleist

GSI Community Workshop
28 June 2011

Collaborators: Jeff Whitaker*, Dave Parrish*, Kayo Ide, John Derber, Russ Treadon, Xuguang Wang
Variational Data Assimilation

\[ J_{\text{Var}}(x') = \frac{1}{2} (x')^T B_{\text{Var}}^{-1} (x') + \frac{1}{2} (y'_o - Hx')^T R^{-1} (y'_o - Hx') + J_c \]

\( J \): Penalty (Fit to background + Fit to observations + Constraints)
\( x' \): Analysis increment \((x_a - x_b)\); where \(x_b\) is a background
\( B_{\text{Var}} \): Background error covariance
\( H \): Observations (forward) operator
\( R \): Observation error covariance (Instrument + representativeness)
\( y'_o \): Observation innovations
\( J_c \): Constraints (physical quantities, balance/noise, etc.)

\( B \) is typically static and estimated a-priori/offline
Motivation from Var

• Current background error covariance (applied operationally) in VAR
  – Isotropic recursive filters
  – Poor handle on cross-variable covariance
  – Minimal flow-dependence added
    • Implicit flow-dependence through linearization in normal mode constraint (Kleist et al. 2009)
    • Flow-dependent variances (only for wind, temperature, and pressure) based on background tendencies (Saha et al., 2010)
  – Tuned NMC-based estimate (lagged forecast pairs)
Current background error for GFS

- Although flow-dependent *variances* are used, confined to be a rescaling of fixed estimate based on time tendencies
  - No multivariate or length scale information used
  - Does not necessarily capture ‘errors of the day’
- Plots valid 00 UTC 12 September 2008
Kalman Filter in Var Setting

Forecast Step

\[
\begin{align*}
\mathbf{x}^b &= M(\mathbf{x}^a) \\
\mathbf{B}_{KF} &= M A_{KF} M^T + Q \\
\mathbf{x}^a &= \mathbf{x}^b + K \left( \mathbf{y} - H \mathbf{x}^b \right) \\
\mathbf{K} &= B_{KF} H^T \left( R + H B_{KF} H^T \right)^{-1} \\
A_{KF} &= (I - K H) B_{KF}
\end{align*}
\]

Analysis

Extended Kalman Filter

• Analysis step in variational framework (cost function)

\[
J_{KF}(\mathbf{x}') = \frac{1}{2} (\mathbf{x}')^T B_{KF}^{-1} (\mathbf{x}') + \frac{1}{2} (\mathbf{y}_o - H \mathbf{x}')^T R^{-1} (\mathbf{y}_o - H \mathbf{x}')
\]

• \(B_{KF}\): Time evolving background error covariance
• \(A_{KF}\): Inverse [Hessian of \(J_{KF}(\mathbf{x}')\)]
Motivation from KF

- **Problem**: dimensions of $A_{KF}$ and $B_{KF}$ are huge, making this practically impossible for large systems (GFS for example).

- **Solution**: sample and update using an ensemble instead of evolving $A_{KF}/B_{KF}$ explicitly

\[
\begin{align*}
\text{Forecast Step: } & \quad X^a \rightarrow X^b \\
\text{Analysis Step: } & \quad X^b \rightarrow X^a \\
B_{KF} & \approx \frac{1}{N-1} X^b (X^b)^T \\
A_{KF} & \approx \frac{1}{N-1} X^a (X^a)^T
\end{align*}
\]

$N$: Ensemble Size
Hybrid Variational-Ensemble

• Incorporate ensemble perturbations directly into variational cost function through extended control variable
  – Lorenc (2003), Buehner (2005), Wang et. al. (2007), etc.

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

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

\[
\frac{1}{\beta_f} + \frac{1}{\beta_e} = 1
\]

\(\beta_f\) & \(\beta_e\): weighting coefficients for fixed and ensemble covariance respectively

\(x_t\): (total increment) sum of increment from fixed/static \(B\) (\(x_f\)) and ensemble \(B\)

\(\alpha^n\): extended control variable; \(x^n_e\): ensemble perturbations

\(L\): correlation matrix [localization on ensemble perturbations]
Hybrid with (global) GSI

• Control variable has been implemented into GSI 3DVAR*
  – Full $\mathbf{B}$ preconditioning
    • Working on extensions to $\mathbf{B}^{1/2}$ preconditioned minimization options
  – Spectral filter for horizontal part of $\mathbf{A}$
    • Eventually replace with (anisotropic) recursive filters
  – Recursive filter used for vertical
  – Dual resolution capability
    • Ensemble can be from different resolution than background/analysis
      (vertical levels are the exception)
  – Various localization options for $\mathbf{L}$
    • Grid units or scale height for vertical
    • Level dependent specification
  – Option to apply TLNMC (Kleist et al. 2009) to analysis increment

$$
\mathbf{x}' = \mathbf{C} \left[ \mathbf{x}_f' + \sum_{n=1}^{N} (\alpha^n \circ \mathbf{x}_c^n) \right]
$$

*Acknowledgement: Dave Parrish for original implementation of extended control variable
Single Observation

Single 850mb Tv observation (1K O-F, 1K error)
Dual-Res Coupled Hybrid

- member 1 forecast
- member 2 forecast
- member 3 forecast
- high res forecast
- EnKF member update
- GSI Hybrid Ens/Var
- high res analysis
- recenter analysis ensemble
- member 1 analysis
- member 2 analysis
- member 3 analysis

Previous Cycle

Current Update Cycle
Hybrid Var-EnKF GFS experiment

- **Model**
  - GFS deterministic (T574L64; post July 2010 version – current operational version)
  - GFS ensemble (T254L64)
    - 80 ensemble members, EnKF update, GSI for observation operators

- **Observations**
  - All operationally available observations (including radiances)
  - Includes early (GFS) and late (GDAS/cycled) cycles as in production

- **Dual-resolution/Coupled**
  - High resolution control/deterministic component
    - Includes TC Relocation on guess
  - Ensemble is recentered every cycle about hybrid analysis
    - Discard ensemble mean analysis

- **Satellite bias corrections**
  - Coefficients come from GSI/VAR

- **Parameter settings**
  - 1/3 static B, 2/3 ensemble
  - Fixed localization: 800km & 1.5 scale heights

- **Test Period**
  - 15 July 2010 – 15 October 2010 (first two weeks ignored for “spin-up”)
Geopotential Height RMSE

Northern Hemisphere

Significant reduction in mean height errors
Stratospheric Fits

Improved fits to stratospheric observations
Forecasts from hybrid analyses fit observation much better.
Tropical Wind Errors (72-h) vs “consensus” analysis

Deep-Layer (850-200 hPa) mean vector wind error (tropics)

DLM == mean of vector wind at 850,700,500,400,300,200 hPa.
“Consensus analysis” = 1/3(ECMWF + UKMET + NCEP)
GSI/EnKF Hybrid vs GSI opnl track errors

Hybrid has significantly lower track errors than operational GSI (using static covariance)
GSI/EnKF Hybrid vs GSI opnl intensity bias

Hybrid has lower intensity bias out to day 4.
HVEDAS (3D) for GDAS/GFS

- Prototype dual-resolution, two-way coupled hybrid Var/EnKF system outperforms standard 3DVAR in GFS experiments
  - 2010 Hurricane Season (August 15 through October 31 2010) run off-site
  - Emphasis on AC, RMSE, TC Tracks

- Plan underway to implement into GDAS/GFS operationally
  - Target: Spring 2012 (subject to many potential issues)
    - Porting of codes/scripts back to IBM P6 (completed)
    - Cost analysis (will everything fit in production suite?)
    - More thorough (pre-implementation) testing and evaluation
      - More test periods (including NH winter)
      - Other/more verification metrics
    - Interaction with other planned DA updates

- Issues
  - More tuning
  - How should EnKF be used within ensemble forecasting paradigm?