A Hybrid Data Assimilation System
(WRF-VAR and Ensemble Transform Kalman Filter)
Based Retrospective Tests

Meral Demirtas
National Center for Atmospheric Research

Dale Barker, Yongsheng Chen, Josh Hacker,
Xiang-Yu Huang, Chris Snyder and Xuguang Wang

10th Annual WRF User’s Workshop
23-27 June 2009, Boulder, Colorado
An outline of my hybrid story

- Why do we need a hybrid data assimilation system?
- What are the basic ingredients of a hybrid system?
- How have we implemented the hybrid system at the Data Assimilation Testbed Center (DATC)?
- What we have found: Highlights of preliminary results
- Summary and conclusions
- Future work
**Why do we need a hybrid system?**

- The WRF 3D-VAR system uses only *climatological* (static) background error covariances.

- *Flow-dependent* covariance through ensemble is needed.

- Hybrid combines *climatological* and *flow-dependent* background error covariances.

- It can be adapted to an existing 3D-VAR system.

- Hybrid can be robust for small size ensembles.
What are the basic ingredients of a hybrid system?

1. Ensemble forecasts: WRF-ensemble forecasts
2. A mechanism to update ensemble perturbations: Ensemble Transform Kalman Filter (ETKF)
3. A data assimilation system: WRF 3D-VAR

It sounds simple…. :-)

Ensembles to address uncertainties in the initial state

- Ensemble members
- Ensemble mean
- Observed (idealized!)

Initial state

Time
Ensemble Basics

Assume the following ensemble forecasts:

\[ X^f = (x_1^f, x_2^f, x_3^f, \ldots, x_N^f) \]

Ensemble mean: \[ \overline{x}^f = \frac{1}{N} \sum_{i=1}^{N} x_n^f \]

Ensemble perturbations: \[ \delta x_n^f = x_n^f - \overline{x}^f \]

Ensemble perturbations in vector form:

\[ \delta X^f = (\delta x_1^f, \delta x_2^f, \delta x_3^f, \ldots, \delta x_N^f) \quad n = 1, N \]
**How to update ensemble perturbations?**

ETKF technique updates ensemble perturbations by rescaling innovations with a transformation matrix (Wang and Bishop 2003).

\[
\chi^a = \chi^f T
\]

Transformation matrix
(solved by Kalman Filter Theory)

An adaptive scalar inflation factor has been introduced to inflate at time \(i\) by matching spread to innovation vectors, \(\Pi\):

\[
\chi_i = \chi_i^f T_i \Pi_i
\]

Inflation factor
(For the derivation of \(\Pi\) see Wang and Bishop 2003.)
The hybrid DA formulation....

Ensemble covariance is implemented into the 3D-VAR cost function via *extended control variables*:

\[
J(x_1', \alpha) = \beta_1 \frac{1}{2} x_1'^T B^{-1} x_1' + \beta_2 \frac{1}{2} \alpha^T C^{-1} \alpha + \frac{1}{2} (y^o' - Hx')^T R^{-1} (y^o' - Hx')
\]

\[
x' = x_1' + \sum_{k=1}^{K} (\alpha_k \circ x_k^e)
\]

*\(C\): correlation matrix for ensemble covariance localization

*\(x_1'\): 3D-VAR increment

*\(x'\): Total increment including hybrid

*\(\alpha\): Extended control variable

*\(\beta_1\): Weighting coefficient for static 3D-VAR covariance

*\(\beta_2\): Weighting coefficient for ensemble covariance

(Wang et. al. 2008)
**How** have we implemented hybrid (3DVAR -ETKF) system at the (DATC)?

**WRF Ensemble:**
- M1
- M2
- M3
- M4
- M5
- Mn

**Compute:**
- Ensemble perturbations (ep2): 
  - u,v,t,ps,q, mean, and std_dev

**WRF-VAR (hybrid)**
- Update: ensemble mean

**Compute:**
- Ensemble mean

**WRF-VAR (QC-OBS):**
- filtered_ob.ascii

**WRF-VAR (VERIFY):**
- ob.etkf ensemble

**ETKF:**
- Update ens perturbations

**WRF ensemble run for the next cycle**

**cycling**

**Update:**
- ens. initial conditions
- ens. boundary conditions

**WRF deterministic forecast run**
A few notes on hybrid settings

- $\text{alpha\_corr\_scale}=1500\text{km} \ (\text{Default})$
- $\text{je\_factor} (\beta_1)=2.0$
- $\text{jb\_factor} (\beta_2)=\frac{\text{je\_factor}}{\text{je\_factor} - 1}=2.0$
- $\text{alphacv\_method}=2 \ (\text{ensemble perturbations on model space})$
- $\text{ensdim\_alpha}=10 \ (\text{ensemble size})$

$\text{N.B. Conservation of total variance requires: } \frac{1}{\beta_1} + \frac{1}{\beta_2} = 1$
Retrospective Test Runs

• Base runs: WPS, REAL and WRF

• Generate background error covariance for 3D-VAR

• Three hourly full cycling with conventional observations:
  ➢ Hybrid (3D-VAR and ETKF)
  ➢ Only standard 3D-VAR
Experiment Set-up

- 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 (*Torn et al. 2006*). (WRF-VAR and some additional tools.)
- Horizontal resolution: 45km
- Number of vertical levels: 57
- Model top: 50 hPa

*For details see: Demirtas et al. 2009*
What we have found:
Highlights of preliminary results
Ensemble Spread: 500 hPa height (m) std. dev.

WRF t+3 valid at 2007081900

ETKF: Adaptive modest inflation factors used

ETKF: Adaptive high inflation factors used
Hybrid gives better RMSE scores for wind compared to 3D-VAR.

ECMWF analysis data (T106) used.
Hybrid gives better RMSE scores for wind compared to 3D-VAR.

ECMWF analysis data (T106) used.
Hybrid gives better RMSE scores for wind compared to 3D-VAR.
Hybrid gives better RMSE scores for wind compared to 3D-VAR.
Summary and Conclusions

• A WRF-VAR-ETKF based hybrid system has been constructed with some enhancements at the DATC.

• The hybrid system has been tested for the 30-day retrospective runs which coincided with the hurricane Dean’s active period. A few computational instabilities noted during WRF runs, otherwise it was stable.

• Ensemble spread is not “the bee’s knees”, but we noted better spread with high inflation factors.

• Verification (RMSE vertical profiles) results of hybrid test are encouraging particularly for the lower troposphere. They are better than those of standard 3D-VAR.
Future Work

- ETKF part: Update inflation generation mechanism in the light of recent applications.

- Hybrid part:
  - Vertical localization: It is currently being tested.
  - Additional isolated runs are needed to evaluate various tunable hybrid parameters:
    - impact of increased weighted contribution from ensembles
    - the impact of smaller/larger horizontal length scale for covariance localization
    - investigating the benefit of tuning background error covariance matrix with ensemble mean based forecasts
    - using higher horizontal resolution
Thanks for listening
my hybrid saga… :-)

References


