Introduction to Hybrid Ensemble-Variational Data Assimilation Methods and Recent Research, Development and Application for Global to Storm Scale NWP



Xuguang Wang

School of Meteorology University of Oklahoma, Norman, OK, USA xuguang.wang@ou.edu

Tutorial for Frontiers in Ensemble Data Assimilation for Geoscience Applications, Boulder, CO, 2015





□ Introduction to hybrid DA methods

Examples of recent research, development and application of hybrid DA

- Recent R&D to improve global forecasts
- Recent R&D to improve high resolution hurricane forecasts
- Recent R&D for convective scale weather forecasts over CONUS
- □ Future work and challenges



What is Hybrid?











Hybrid DA system





How to incorporate ensemble in VAR? Method 1

 Direct combination of static and ensemble covariances (Hamill and Snyder 2000; Wang et al. 2007a, 2008a; Kuhl et al. 2013):

$$J(\mathbf{x}') = J_b + J_o$$

$$= \frac{1}{2} \mathbf{x}'^T \left(\frac{1}{\beta_1} \mathbf{B} + \frac{1}{\beta_2} \mathbf{P}^e \circ \mathbf{S} \right)^{-1} \mathbf{x}' + \frac{1}{2} \left(\mathbf{y}^{o'} - \mathbf{H} \mathbf{x}' \right)^T \mathbf{R}^{-1} \left(\mathbf{y}^{o'} - \mathbf{H} \mathbf{x}' \right)$$
Blended static and ensemble covariances
$$\mathbf{P}^e = \sum_{k=1}^{K} \left(\mathbf{x}_k^e \right) \left(\mathbf{x}_k^e \right)^T, \ \frac{1}{\beta_1} + \frac{1}{\beta_2} = 1$$

B 3DVAR static covariance; **R** observation error covariance; *K* ensemble size; **S** correlation matrix for ensemble covariance localization; \mathbf{x}_k^e *k*th ensemble perturbation; \mathbf{x}' total (hybrid) increment; $\mathbf{y}^{o'}$ innovation vector; **H** linearized observation operator; β_1 weighting coefficient for static covariance; β_2 weighting coefficient for ensemble covariance



How to incorporate ensemble in VAR? Method 2

 Extended control variable (ECV) method (Lorenz 2003; Buehner 2005; Wang et al. 2007b, 2008a; Wang 2010):

$$J(\mathbf{x}'_{s}, \boldsymbol{\alpha}) = \beta_{1}J_{s} + \beta_{2}J_{e} + J_{o}$$

= $\beta_{1}\frac{1}{2}\mathbf{x}_{s}^{T}\mathbf{B}^{-1}\mathbf{x}_{s}^{'} + \beta_{2}\frac{1}{2}\boldsymbol{\alpha}^{T}\mathbf{C}^{-1}\boldsymbol{\alpha} + \frac{1}{2}(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}^{'})^{T}\mathbf{R}^{-1}(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}^{'})$
Extra term associated with extended control variable
 $\mathbf{x}' = \mathbf{x}_{s}^{'} + \sum_{k=1}^{K}(\boldsymbol{\alpha}_{k} \circ \mathbf{x}_{k}^{e})$ Extra increment associated with ensemble

B 3DVAR static covariance; **R** observation error covariance; *K* ensemble size; **C** concatenated correlation matrix for ensemble covariance localization; $\mathbf{x}_k^e k$ th ensemble perturbation;

 \mathbf{x}_{s} 3DVAR increment; \mathbf{x}' total (hybrid) increment; $\mathbf{y}^{o'}$ innovation vector;

H linearized observation operator; β_1 weighting coefficient for static covariance;

 β_2 weighting coefficient for ensemble covariance; α extended control variable.



• The solutions from method 1 and method 2 are proved to be equivalent (Wang et al. 2007b).

• Most operational NWP centers use method 2

 For method 2, specific implementation can be different depending on the preconditioning used in variational minimization (e.g., Lorenc 2003 (UK Met); Wang et al. 2008a (WRFVAR hybrid); Wang 2010 (GSI hybrid)).



3DEnVar vs. 4DEnVar Hybrid

• 3DEnVar

e.g. Wang et al. 2008a



$$J(\mathbf{x}'_{s}, \boldsymbol{\alpha}) = \beta_{1}J_{s} + \beta_{2}J_{e} + J_{o}$$

$$= \beta_{1}\frac{1}{2}\mathbf{x}_{s}^{T}\mathbf{B}^{-1}\mathbf{x}'_{s} + \beta_{2}\frac{1}{2}\boldsymbol{\alpha}^{T}\mathbf{C}^{-1}\boldsymbol{\alpha} + \frac{1}{2}(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}')^{T}\mathbf{R}^{-1}(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}')$$
Extra term associated with extended control variable
$$\mathbf{x}' = \mathbf{x}'_{s} + \sum_{k=1}^{K} (\boldsymbol{\alpha}_{k} \circ \mathbf{x}_{k}^{e})$$
Extra increment associated with ensemble

B stat 3DVAR static covariance; **R** observation error covariance; *K* ensemble size; **C** correlation matrix for ensemble covariance localization; \mathbf{x}_k^e *k*th ensemble perturbation; \mathbf{x}_s' 3DVAR increment; \mathbf{x}' total (hybrid) increment; $\mathbf{y}^{o'}$ innovation vector; **H** linearized observation operator; β_1 weighting coefficient for static covariance; β_2 weighting coefficient for ensemble covariance; $\boldsymbol{\alpha}$ extended control variable.



3DEnVar vs. 4DEnVar Hybrid

4DEnVar Wang and Lei 2014 $J(\mathbf{x}'_{s}, \boldsymbol{\alpha}) = \beta_{1}J_{s} + \beta_{2}J_{e} + J_{o}$ $= \beta_{1}\frac{1}{2}\mathbf{x}_{s}^{T}\mathbf{B}^{-1}\mathbf{x}_{s}^{'} + \beta_{2}\frac{1}{2}\boldsymbol{\alpha}^{T}\mathbf{C}^{-1}\boldsymbol{\alpha} + \frac{1}{2}\sum_{t=0}^{T}(y_{t}^{o}-\mathbf{H}_{t}\mathbf{x}_{t}^{'})^{T}\mathbf{R}_{t}^{-1}(y_{t}^{o}-\mathbf{H}_{t}\mathbf{x}_{t}^{'})$ $\mathbf{x}_{t}^{'} = \mathbf{x}_{s}^{'} + \sum_{k=1}^{K}(\alpha_{k} \circ (\mathbf{x}_{k}^{e})_{t})$ Use ensemble temporal covariance to propagate increments over time. No tangent linear adjoint.

B stat 3DVAR static covariance; **R** observation error covariance; *K* ensemble size; **C** Concatenated correlation matrix for ensemble covariance localization; \mathbf{x}_k^e *k*th ensemble perturbation;

 \mathbf{x}_{s} 3DVAR increment; \mathbf{x}' total (hybrid) increment; $\mathbf{y}^{o'}$ innovation vector;

H linearized observation operator; β_1 weighting coefficient for static covariance; β_2 weighting coefficient for ensemble covariance; α extended control variable.

3h



4DEnVar vs. En4DVar Hybrid

• En4DVar

B stat 3DVAR static covariance; **R** observation error covariance; *K* ensemble size; **C** Concatenated correlation matrix for ensemble covariance localization; \mathbf{x}_k^e *k*th ensemble perturbation; **M**, tangent linear model

 \mathbf{x}_{s} 3DVAR increment; \mathbf{x}' total (hybrid) increment; $\mathbf{y}^{o'}$ innovation vector;

H linearized observation operator; β_1 weighting coefficient for static covariance; β_2 weighting coefficient for ensemble covariance; α extended control variable.



Summary of Hybrid flavors

	Number of time levels of ensemble perturbations incorporated in the DA window during the variational minimization	Tangent linear and adjoint of the forecast model
3DEnVar hybrid	One, usually valid at the center of the DA window	Not needed
4DEnVar hybrid	Multiple	Not needed, same static covariance is used for multiple time levels, equivalent to assuming a numerical model of identity matrix
En4DVar hybrid	One, usually valid at the beginning of the DA window	Needed



Why Hybrid?

	VAR (3D, 4D)	EnKF	hybrid	References (e.g.)
Benefit from use of flow dependent ensemble covariance instead of static B		yes	yes	Hamill and Snyder 2000; Wang et al. 2007a,2008ab, 2009, 2013; Buehner et al. 2010ab; Wang 2011; etc.
Robust for small ensemble			yes	Wang et al. 2007b, 2009;
Model space covariance localization	yes		yes	Campbell et al. 2010
Flexible to add various dynamical/physical constraints	yes		yes	Wang et al. 2013
Built in outer loops for nonlinearity treatment	yes		yes	
Use of various existing capabilities in VAR, e.g. dual resolution hybrid	yes		yes	

Static vs. flow dependent covariance





Cross-variable increment



- •Hurricane IKE 2008
- •WRF ARW: Δx=5km

•Observations: radial velocity from two WSR88D radars (KHGX, KLCH)

•WRFVAR hybrid DA system (Wang et al. 2008ab)

Li et al., 2012



50 member

5 member



Wang et al. 2007b, 2009

Hybrid benefits from variational constraints: example from GSI hybrid for GFS



 Hybrid with full ensemble covariance was better than EnKF due to the use of tangent linear normal mode balance constraint





□ Introduction to hybrid DA methods

- Examples of recent research, development and application of hybrid DA
- Recent R&D to improve global forecasts
- Recent R&D to improve high resolution hurricane forecasts
- Recent R&D for convective scale weather forecasts over CONUS
- □ Future work and challenges

GSI hybrid for GFS: GSI 3DVar vs. 3DEnVar Hybrid vs. EnKF



 3DEnVar Hybrid was better than 3DVar due to use of flow-dependent ensemble covariance

 3DEnVar was better than EnKF due to the use of tangent linear normal mode balance constraint (TLNMC)

Wang et al., MWR, 2013, 141, 4098-4117

GSI hybrid for GFS: NCEP pre-implementation test



30

 $\mathbf{g})$

% RMSE CHANGE

-15

-15

24h

24h

48h

4Åh

72h

7Żh

TR W850

96h

120h

120h

144h

144h

2

-15

24h



96h

7Żh

120h

144h

SH Z1000

b)



Courtesy: Daryl Kleist

Temporal evolution of error covariance by GSI 4DEnVar



Wang and Lei, 2014, MWR



GSI hybrid for GFS: 3DEnVar vs. 4DEnVar

Results from Single Reso. Experiments (Wang and Lei 2014, MWR)

- 4DEnVar improved general global forecasts
- 4DEnVar improved the balance of the analysis
- Performance of 4DEnVar improved if more frequent ensemble perturbations used
- 4DEnVar approximates nonlinear propagation better with more frequent ensemble perturbations
 - TLNMC improved global forecasts





GSI hybrid for GFS: 3DEnVar vs. 4DEnVar



Q Approximation to nonlinear propagation



Wang and Lei, 2014, MWR



Verification of hurricane track forecasts



- 3DEnVar outperforms GSI3DVar.
- 4DEnVar is more accurate than 3DEnVar after the 1-day forecast lead time.
- Negative impact if using less number of time levels of ensemble perturbations.
- Negative impact of TLNMC on TC track forecasts.





□ Introduction to hybrid DA methods

Examples of recent research, development and application of hybrid DA

- Recent R&D to improve global forecasts
- Recent R&D to improve high resolution hurricane forecasts
- Recent R&D for convective scale weather forecasts over CONUS
- □ Future work and challenges



GSI hybrid for HWRF Hurricane Sandy, Oct. 2012







- Complicated evolution
- Tremendous size
- 147 direct deaths across Atlantic Basin
 - US damage \$50 billion

- New York State before and after
- nhc.noaa.gov



Experiment Design



• Model: HWRF

•Observations: radial velocity from Tail Doppler Radar (TDR) onboard NOAA P3 aircraft

• Initial and LBC ensemble: GFS global hybrid DA system

• Ensemble size: 40

Lu et al. 2015a

Experiment Design



• Model: HWRF

•Observations: radial velocity from Tail Doppler Radar (TDR) onboard NOAA P3 aircraft

• Initial and LBC ensemble: GFS global hybrid DA system

• Ensemble size: 40

TDR data distribution (mission 1)



P3 Mission 1

Vr vertical distribution

Q Verification against SFMR wind speed





Comparison with HRD radar wind analysis



Track forecast (RMSE for 7 missions)





Correlation between HRD radar wind analysis and analyses from various DA methods







Experiments for 2012-2013 season





HWRF Dual Resolution Hybrid DA



Lu et al. 2015b



Observations assimilated:

3km domain:

Conventional in-situ data in prepbufr, satellite wind, TDR and tcvital

9km domain:

Conventional in-situ data in prepbufr, satellite wind, TDR and tcvital Satellite radiances

Hybrid vs Hybrid-279

analyzed Edouard structure @2014091518

Pa

50

Hybrid @1km 18Z15

Pressure

29°N

28°N



10 15 20 25 30 35 40 45 50 55 60

27°N 26°N 55°W 58°W 57°W 56°W 54°W 10 15 20 25 30 35 40 45 50 55 60 Hybrid @3km 18Z15 50 29⁴N 28°N ٠ 27°N 26°N 57°W 56°W 55°W 54°W 58°W







56°W

54°W

57°W

58°W



km

Hybrid vs Hybrid-279 analyzed Edouard structure @2014091518

HRD radar along lat 18Z15



HRD radar along lon 18Z15





Hybrid 1800Z15



Hybrid-279 1800Z15



Hybrid-279 1800Z15



40



Hybrid vs Hybrid-279 vs operational HWRF RMSE for all cycles



- Hybrid improved especially MSLP and Vmax forecasts compared to operational HWRF
- Dual resolution hybrid improved MSLP and Vmax forecast for the first 12/18 hours than single 9km resolution hybrid





□ Introduction to hybrid DA methods

Examples of recent research, development and application of hybrid DA

- Recent R&D to improve global forecasts
- Recent R&D to improve high resolution hurricane forecasts
- Recent R&D for convective scale weather forecasts over CONUS
- □ Future work and challenges



Motivation for Hybrid DA Development and Research for Convective Scales

- Convective scale analysis and forecasting is a multi-scale problem, requiring an accurate estimate of both the synoptic/mesoscale environment and the convective scale details.
- Convective scale observations (i.e., radar, satellite radiances) require unique observation operators and inclusion of additional state variables (e.g., hydrometeors).
- Accurate cross-variable covariance is especially important.
- Comparison study among Var, EnKF or 3DEnVar or 4DEnVar hybrid for convective scales including both complex cases with multiple storm modes and interactions, and a largely heterogeneous environment, and tornadic storms is still limited.



May 8th 2003 OKC Tornadic Supercell



- An isolated supercell case that produced F-4 intensity tornadoes in Moore and Oklahoma City (OKC) during about 2210—2240 UTC.
- Supercell maintained well beyond 2300 until about 2400 UTC.

Analysis at 2200 UTC: GSI-3DVar

Wang Y. et al. 2015

22:00:00



 \frown

34

Analysis at 2200 UTC: GSI-Hybrid Wang Y. et al. 2015



Ref and vorticity at 1 km



at 4 km max/min W37.4298 / -7.64254 (m s-1) at 4 km

Prob. fcst. starting 2200 UTC: GSI-hybrid



GSI hybrid extended with different microphysics schemes for reflectivity assimilation







□ Introduction to hybrid DA methods

- Examples of recent research, development and application of hybrid DA
- Recent R&D to improve global forecasts
- Recent R&D to improve high resolution hurricane forecasts
- Recent R&D for convective scale weather forecasts over CONUS

□ Future work and challenges



□Assimilating observations in a system that resolves multiple scales.

Effective methods to sample model error in ensemble background.

□Assimilating advanced/new observations: e.g., cloudy radiance observations.

□Correct location and field alignment errors for storm scale DA.

□Variational constraint for different scales.

□Improving static covariance for storm scales.



References cited

- Hamill, T. M., and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.
- Lorenc, A. C. 2003: The potential of the ensemble Kalman filter for NWP a comparison with 4D-VAR. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183-3203.
- Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013-1043.
- Wang, X., T. M. Hamill, J. S. Whitaker and C. H. Bishop, 2007a: A comparison of hybrid ensemble transform Kalman filter-OI and ensemble square-root filter analysis schemes. *Mon. Wea. Rev.*, **135**, 1055-1076.
- Wang, X., C. Snyder, and T. M. Hamill, 2007b: On the theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222-227.
- Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev.*, 136, 5116-5131.
- Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008b: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. *Mon. Wea. Rev.*, **136**, 5132-5147.
- Wang, X., T. M. Hamill, J. S. Whitaker, C. H. Bishop, 2009: A comparison of the hybrid and EnSRF analysis schemes in the presence of model error due to unresolved scales. *Mon. Wea. Rev.*, **137**,3219-3232.
- Wang, X., 2010: Incorporating ensemble covariance in the Gridpoint Statistical Interpolation (GSI) variational minimization: a mathematical framework. *Mon. Wea. Rev.*, **138**, 2990-2995.



References cited

- Wang, X. 2011: Application of the WRF hybrid ETKF-3DVAR data assimilation system for hurricane track forecasts. *Wea. Forecasting*, **26**, 868-884.
- Wang, X., D. Parrish, D. Kleist and J. S. Whitaker, 2013: GSI 3DVar-based Ensemble-Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single Resolution Experiments. *Mon. Wea. Rev.*, **141**, 4098-4117.
- Wang, X. and T. Lei, 2014: GSI-based four dimensional ensemble-variational (4DEnsVar) data assimilation: formulation and single resolution experiments with real data for NCEP Global Forecast System. Mon. Wea. Rev., 142, 3303-3325.
- Li, Y., X. Wang and M. Xue, 2012: Assimilation of radar radial velocity data with the WRF ensemble-3DVAR hybrid system for the prediction of hurricane Ike (2008). *Mon. Wea. Rev.*, 140, 3507-3524.
- Lu, X., X. Wang, Y. Li, M. Tong, X. Ma and H. Winterbottom, 2015a: GSI-based ensemblevariational hybrid data assimilation for HWRF using airborne radar observations for hurricane initialization and prediction. To be submitted.
- Lu, X. and X. Wang 2015b: GSI-based, continuously cycled, dual resolution, ensemblevariational hybrid data assimilation for HWRF: system description and experiments with Edouard (2014). To be submitted.
- Wang, Y., X. Wang and T. Lei, 2015: Assimilation of Reflectivity Data in GSI-based Hybrid Data Assimilation System Using Three Options of Control Variables for the analysis and prediction of 8 May 2003 Oklahoma City Tornadic Supercell Storm . To be submitted.
- David D. Kuhl, Thomas E. Rosmond, Craig H. Bishop, Justin McLay, and Nancy L. Baker, 2013: Comparison of Hybrid Ensemble/4DVar and 4DVar within the NAVDAS-AR Data Assimilation Framework. *Mon. Wea. Rev.*, 141, 2740–2758.