A Hybrid System (ETKF-3DVAR) Based Extensive Tests Over a Caribbean Domain

Meral Demirtas

NCAR/RAL/DATC

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

The 8th Adjoint Workshop, 18-22 May 2009

Main headings of this presentation

- Basic ingredients of a hybrid data assimilation system
- Hybrid (ETKF-3DVAR) system implementation in the Data Assimilation Testbed Center (DATC)
- Highlights of preliminary results
- Conclusions and Future Work

Basic ingredients of a hybrid data assimilation system

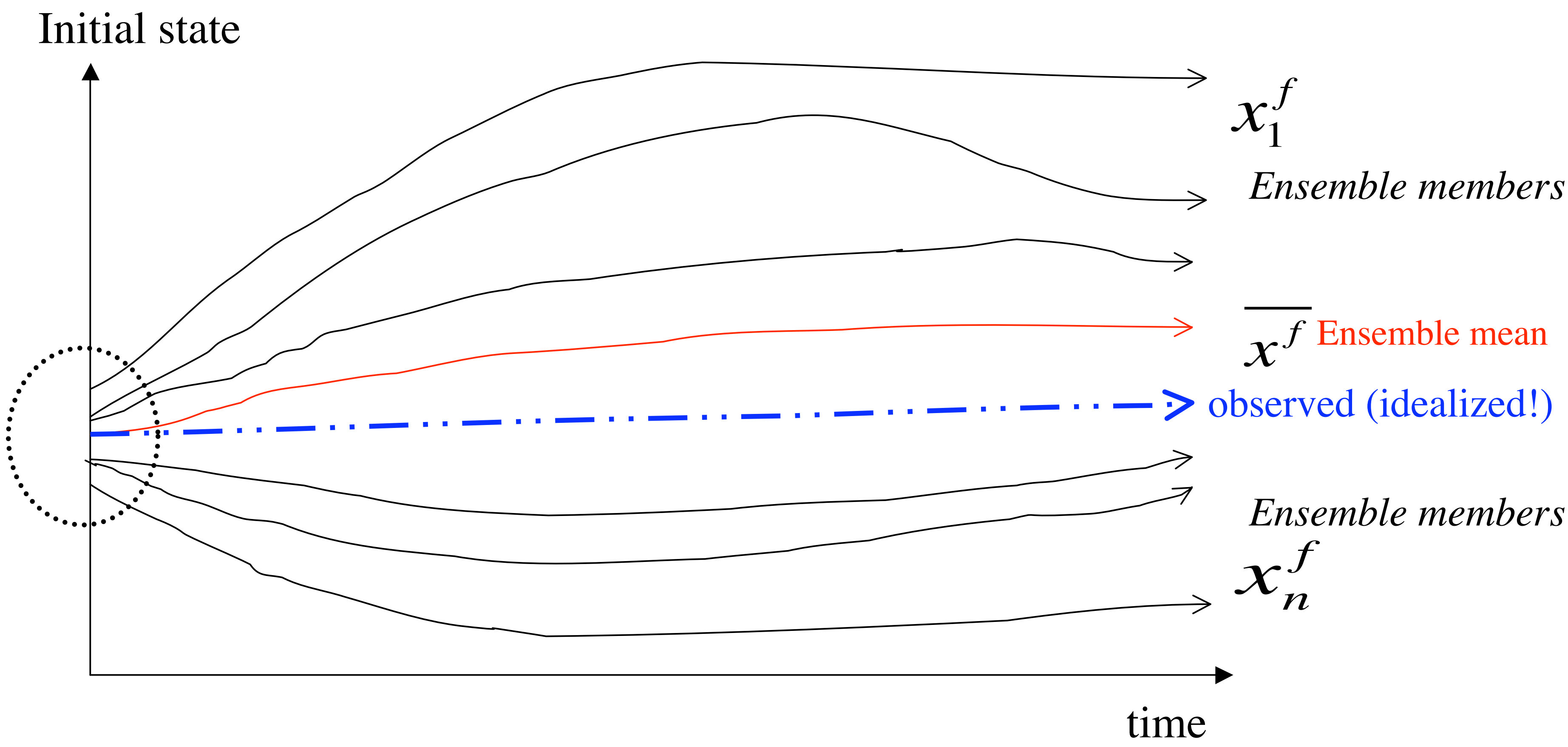
1. Ensemble forecasts: WRF-ensemble forecasts

2. Ensemble updating: Ensemble Transform Kalman Filter (ETKF)

3. Variational data assimilation system: 3D-VAR

4. Hybrid: Combining ensemble based (ETKF) flow-dependent information with climatological background error covariances.

Ensembles to address uncertainties in initial state



Ensemble Basics

Assume the following ensemble forecasts:

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

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:

$$\delta X^f = (\delta x_1^f, \delta x_2^f, \delta x_3^f, \ldots\ldots, \delta x_N^f) \quad n = 1, N$$

Ensemble Transform Kalman Filter (ETKF)

ETKF technique produces ensemble members by re-scaling innovations with a transformation matrix. (Wang and Bishop 2003, Wang et. al. 2004, 2007.)

$$x^a = x^f T$$

T*ransformation matrix*
(solved by Kalman Filter Theory)

An adaptive scalar inflation factor has been introduced by Wang and Bishop (2003) to inflate at time ***i*** by matching spread to innovation vectors, $\prod$:

$$x_i = x_i^f T_i \prod_i$$

Inflation factor
(For the derivation of $\prod$ *see Wang and Bishop 2003.)*

Pros and Cons of ETKF Technique

- Desirable aspects:
  - ETKF is fast (computations are done in model ensemble perturbation subspace).
  - It is suitable for generating ensemble initial conditions.
  - It updates initial condition perturbations.

- Less desirable aspects:
  - ETKF does not localize, therefore it does not represent sampling error efficiently.
  - It may need very high inflation factors.

Why do we need a hybrid system?

- 3D-Var: uses only climatological (static) background error covariances.
- Flow-dependent covariance through ensemble is needed.
- Hybrid combines climatological and flow-dependent background error covariances.
- Hybrid can be more robust for small size ensembles and/or model errors (Wang et al. 2007, 2008a).
- It can be adapted to an existing 3D-VAR system.

The hybrid DA formulation….

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

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

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

(Wang et. al. 2008a)

C: correlation matrix for ensemble covariance localization

$x_1^{'}$ 3D-VAR increment

$x^{'}$ Total increment including hybrid

α Extended control variable

β_1 Weighting coefficient for static 3D-VAR covariance

β_2 Weighting coefficient for ensemble covariance

The hybrid formulation continued….

Conserving total variance requires: $\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1$

<u>Horizontal and Vertical Localization:</u>

- Ensemble covariance horizontal localization is done through recursive filters. Preconditioning designed as:

$$U_1 \approx B^{1/2} \qquad x'_1 = U_1 v_1$$

(Wang et. al. 2008a)

$$U_2 \approx C^{1/2} \qquad \alpha = U_2 v_2$$

Vertical Localization

(taken from Dale Barker's notes…)

- Spurious sampling error not confined to horizontal error correlations.
- Hybrid alpha CVs can be made 3D to damp vertical correlations.
- Two approaches are considered for specifying vertical localization function:
  - a) Empirical function (as in horizontal).
  - b) Use vertical background error covariances to define localization.
- Initial studies use method a).
- Learn from data compression of EOFs to reduce size of alpha CV.

Detailed notes and figures are in attachments…..

The hybrid system implementation and retrospective testing in the DATC

Experiment Set-up

Ensemble size: 10

Test Period: 20070815-20070915

Cycle frequency: 3 hours

Observations: GTS conventional observations

Initial and boundary conditions: GFS (0.5x0.5 degree)

Horizontal resolution: 45km

Number of vertical levels: 57

Model top: 50 hPa

Tools used

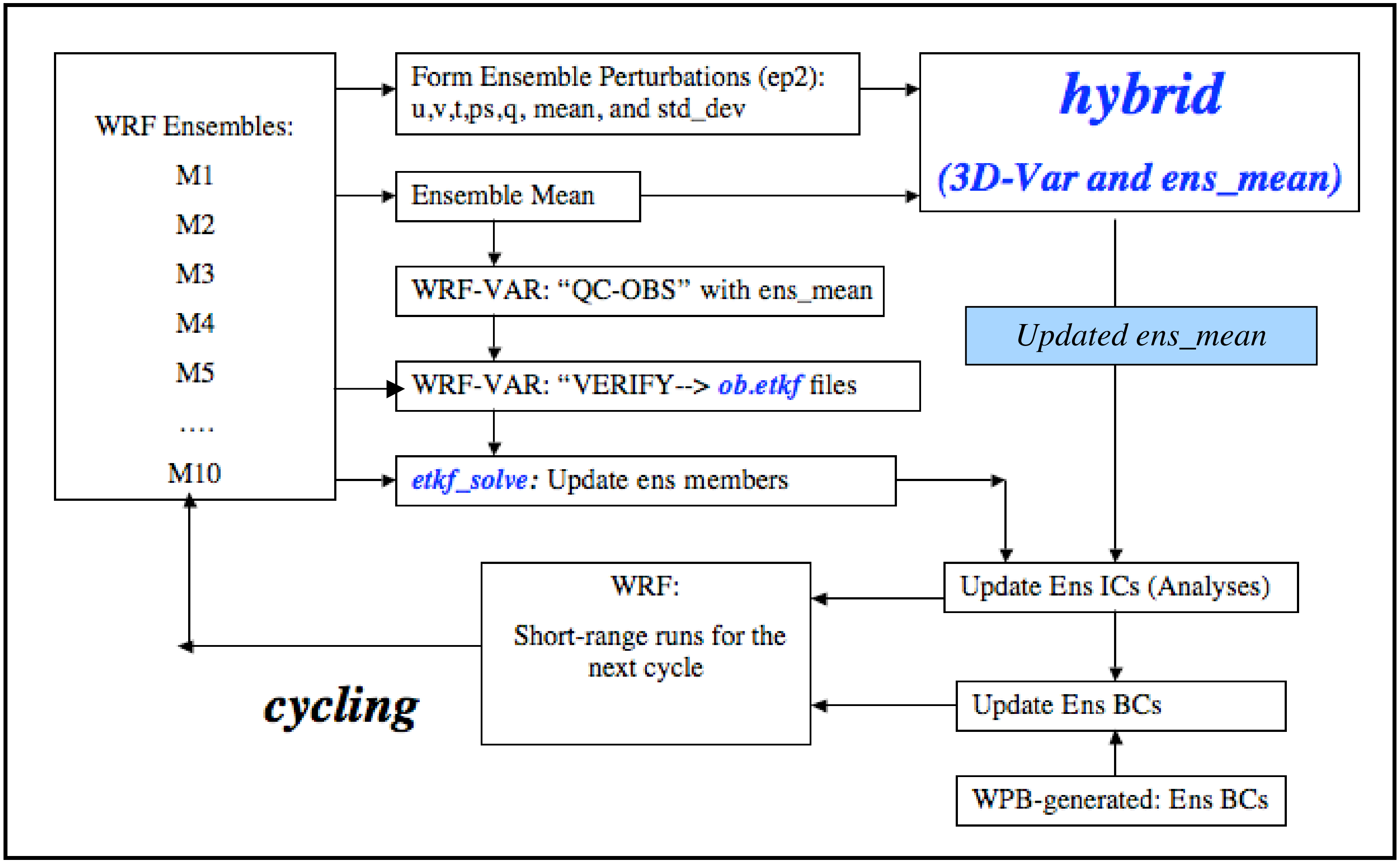
- WRF: Ensemble and deterministic forecasts
- ETKF: Update ensemble perturbations
- Hybrid (WRF-VAR): Update ensemble mean
- Ensemble ICs/LBCs: Produced by adding spatially correlated Gaussian noise to GFS forecasts (Torn et al. 2006). (WRF-VAR and some additional tools used.)

Notes from implementation experience:

- A new step has been implemented to generate filtered (quality controlled) observations by eliminating observations largely deviated from the ensemble mean. This new step has helped to perform much stable runs.

- We have also tested different ETKF inflation factor generation mechanisms:
  - When modest inflation factor generation mechanism is used WRF runs were stable throughout test period, but ensemble spread was small. (On average, square-rooted inflation factor was 4.)
  - High inflation factor generation mechanism (Wang and Bishop 2003) provided better ensemble spread, but presented computational instabilities for few WRF ensemble members. (On average, square-rooted inflation factor was 12.)

WRF-VAR-ETKF Hybrid DA System Implementation at Data Assimilation Testbed Center (NCAR/DATC)



Retrospective Runs Performed

- Base runs: WPS, REAL and WRF
- Background error covariance data generation for the 3D-VAR part.
- Three hourly full cycling with conventional observations:
  - CYC1: Hybrid (ETKF and 3D-VAR)
  - CYC2: Only standard 3DVAR

Hybrid settings

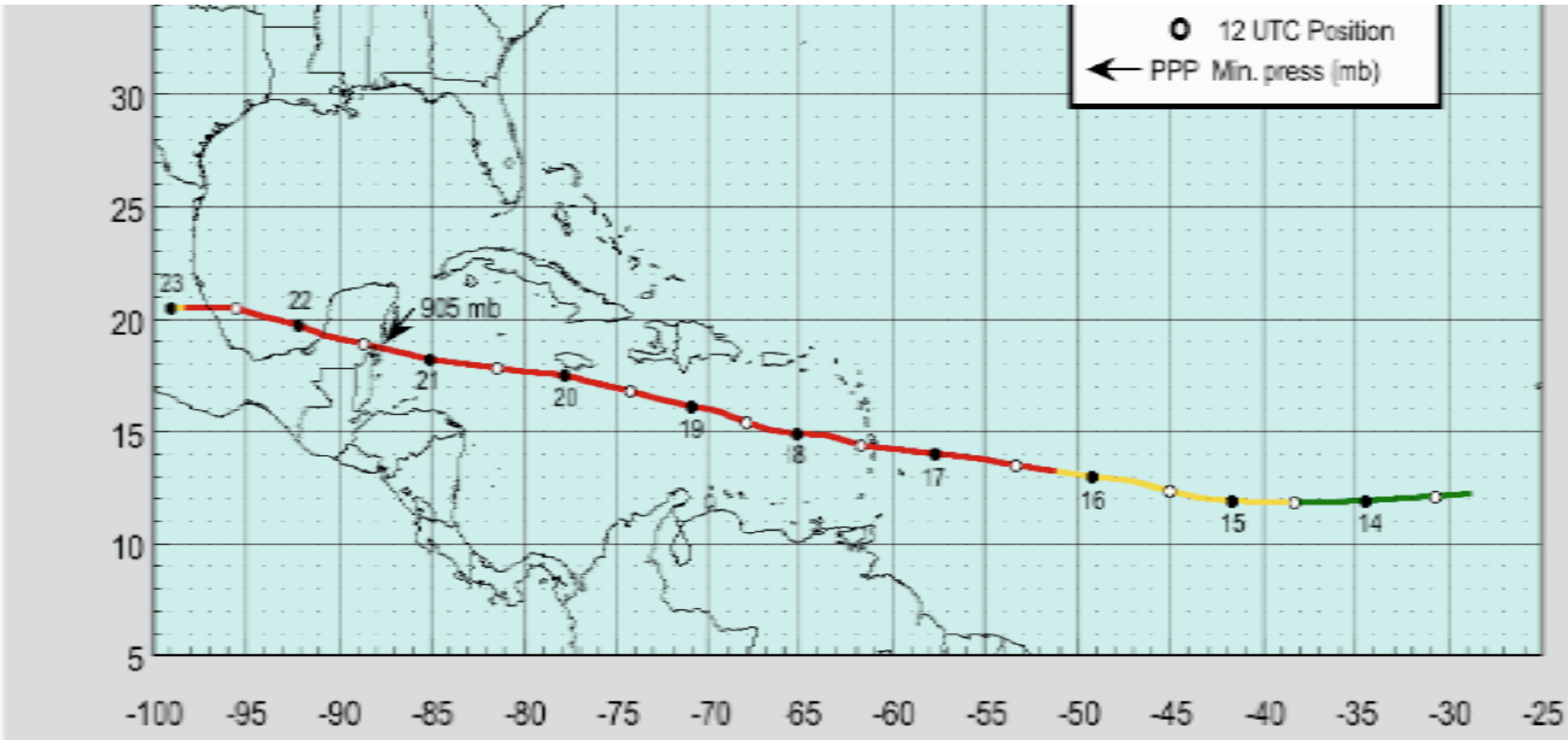
- alpha_corr_scale=1500km (Default)
- je_factor (β_1)=2.0
- jb_factor (β_2)=je_factor/(je_factor -1)=2.0
- alphacv_method=2 (ensemble perturbations on model space)
- ensdim_alpha=10 (ensemble size)

Note that weighting coefficients of ensemble and 3DVAR are equal.

Preliminary results from DATC applications (snapshots)

TC Dean challenge for non localized ETKF!!

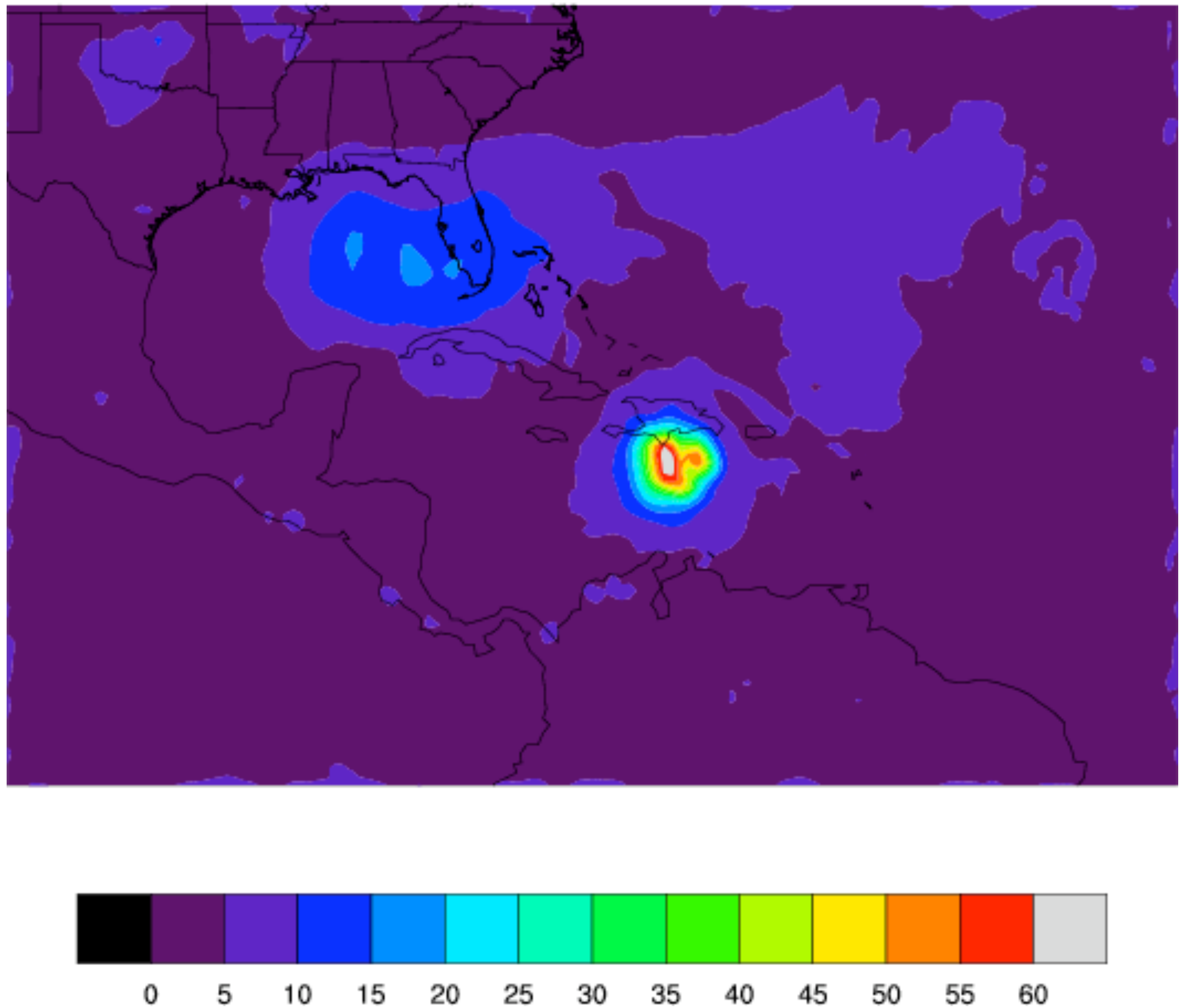
Track Positions for Hurricane Dean: 13-23 Aug 2007

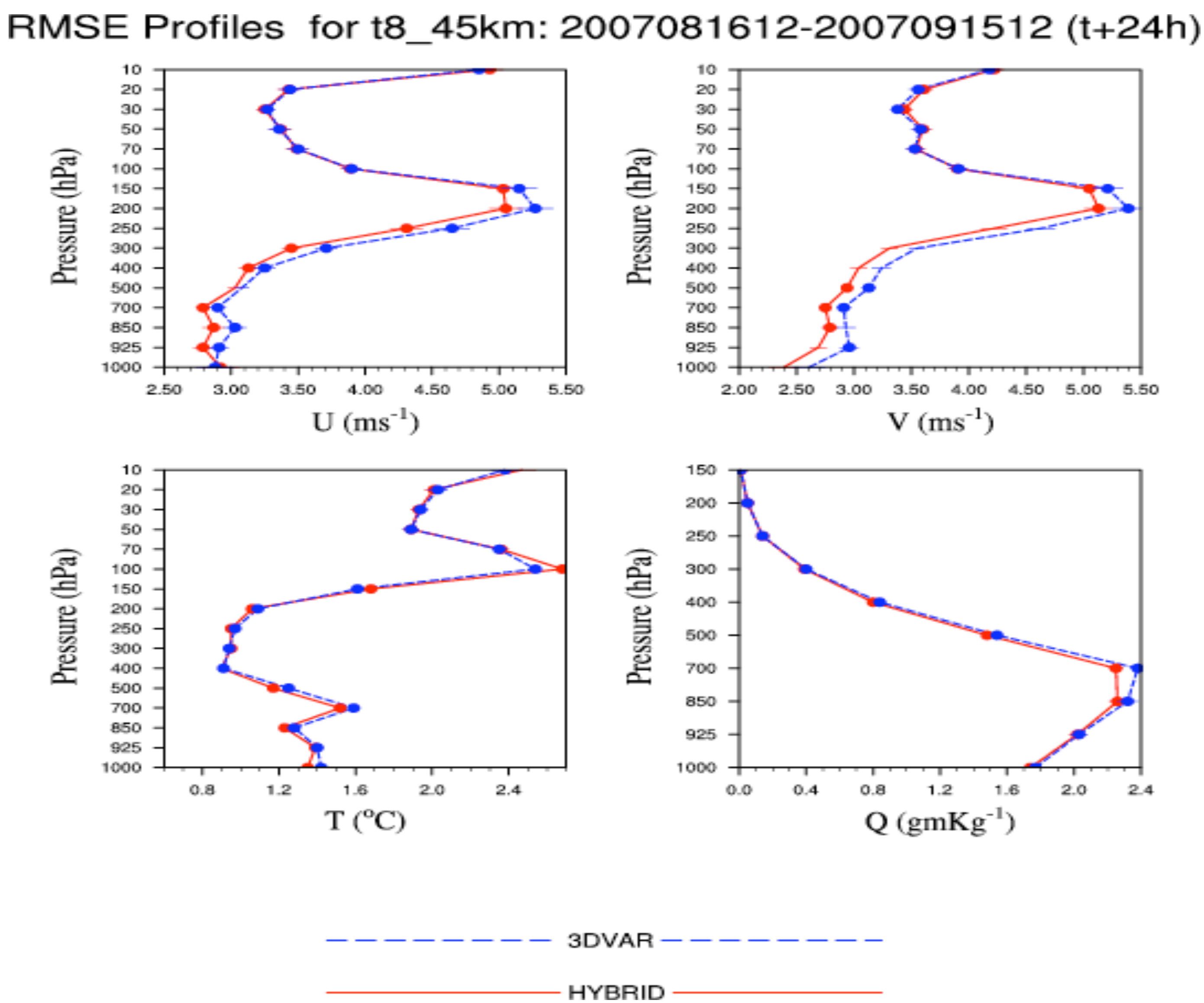


Taken from Tropical Cyclone Report by James L. Franklin, NHC, 2008

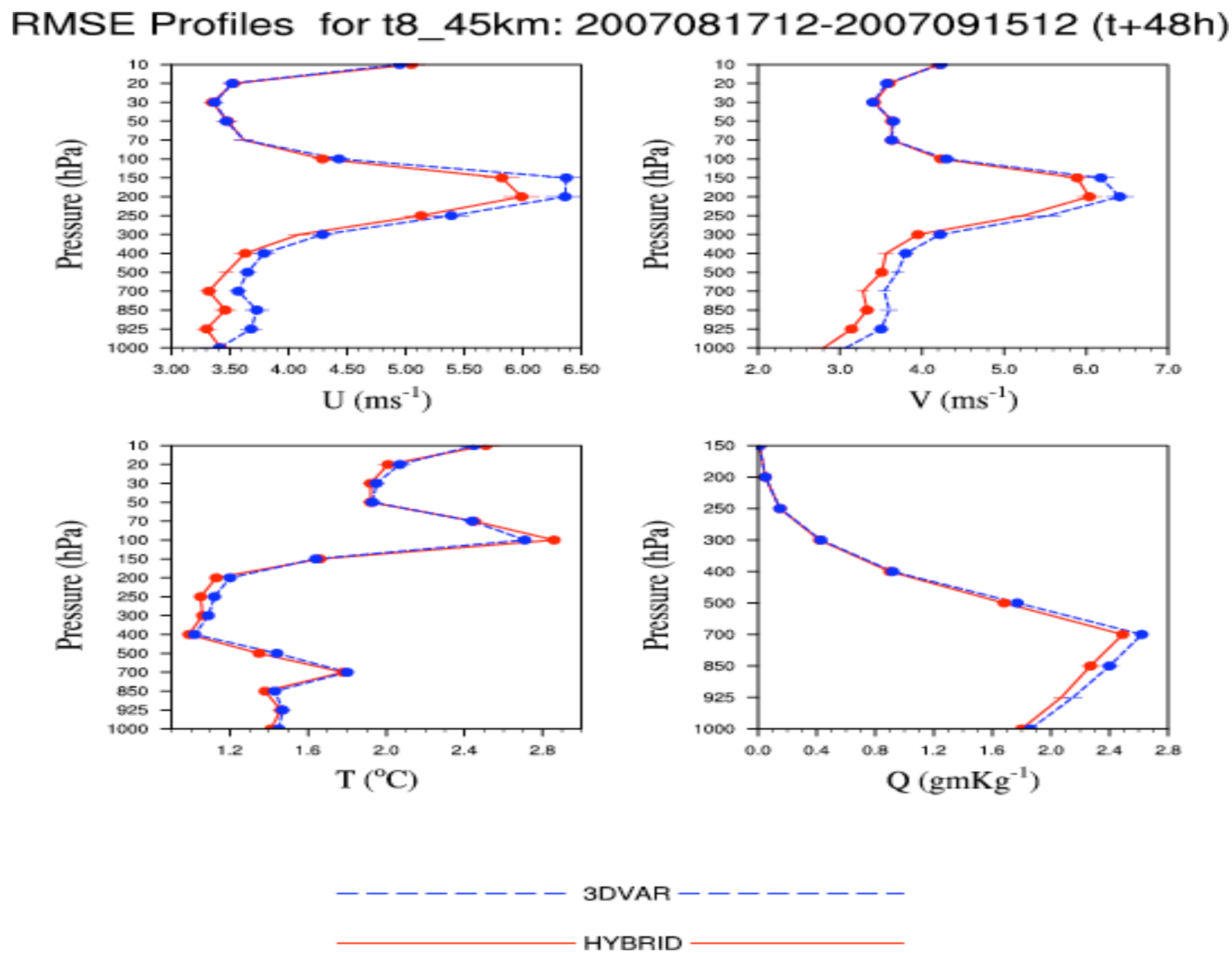
500 hPa height (m) std. dev.

WRF t+3 valid at 2007081900





Hybrid gives better RMSE scores for wind compared to 3D-Var.


RMSE Profiles for t8_45km: 2007081712-2007091512 (t+48h)
Pressure (hPa)
U (ms-1)
V (ms-1)
T (oC)
Q (gmKg-1)
3DVAR
HYBRID

Summary and Conclusions

- A WRF-VAR-ETKF based hybrid system has been constructed with some enhancements in 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 some spread in 500hPa height std. deviation.
- 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

- In our next extensive testbed (Antarctica); we plan to use 20 ensemble members and add AIRS retrievals into obs.
- 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

References

Barker, D. M., 1999: Var scientific development paper 25: the use of synoptic-dependent error structure in 3DVAR. *UK MET Met. Office Technical Reports*, available from the UK Met. Office, Fitzroy Road, Exeter, Devon, EX1, 3PB, UK

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited area ensemble Kalman filters. *Mon. Wea. Rev.,* **134**, 2490-2502.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140-1158.

Wang, X., D. M. 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. M. 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**, 5132–514

Thanks…..

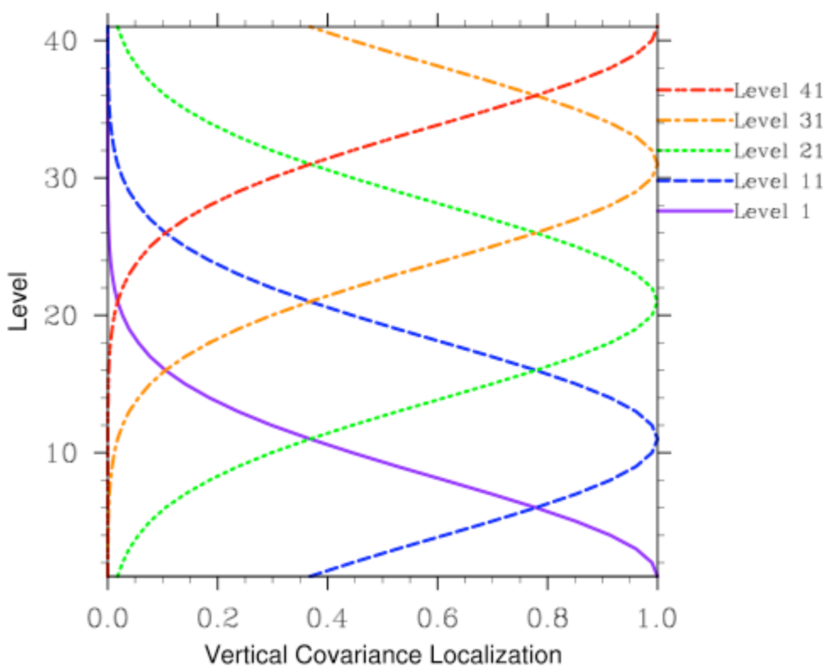
Additional slides for the hbrid vertical localization

Empirical Vertical Covariance Localization

Apply Gaussian Vertical Covariance Localization:

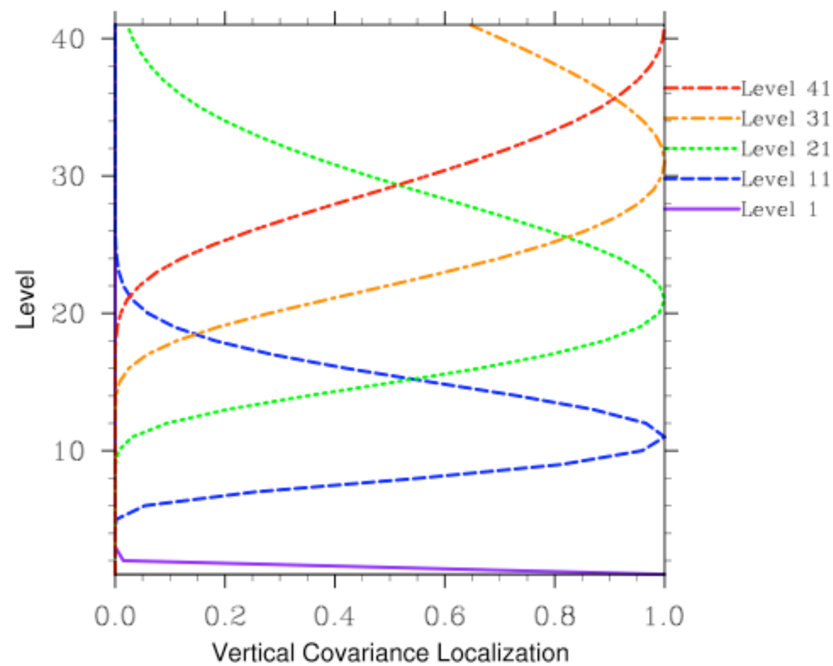
$$\rho(k - k_c) = \exp\left[-\left(k - k_c\right)^2 / L_c^2 \right]$$

Example 1:
Constant Localization Scale



$$L_c = 10$$

Example 2:
Vertically-Dependent Localization Scale



$$L_c = 20 k_c / 41$$

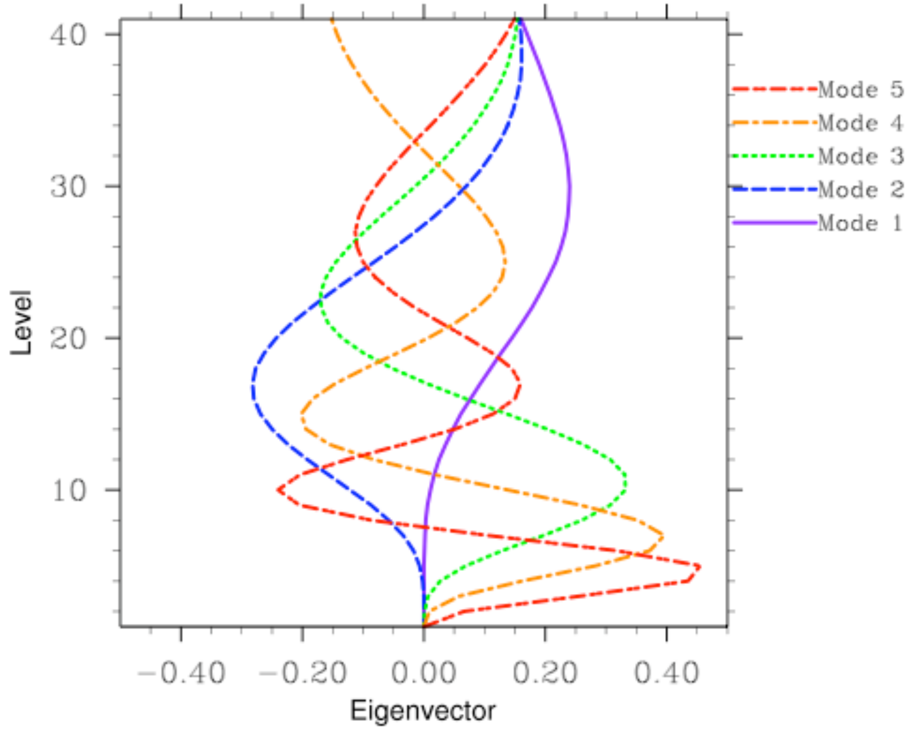
Covariance Localization Decomposition

Example: Gaussian Localization with variable localization scale:

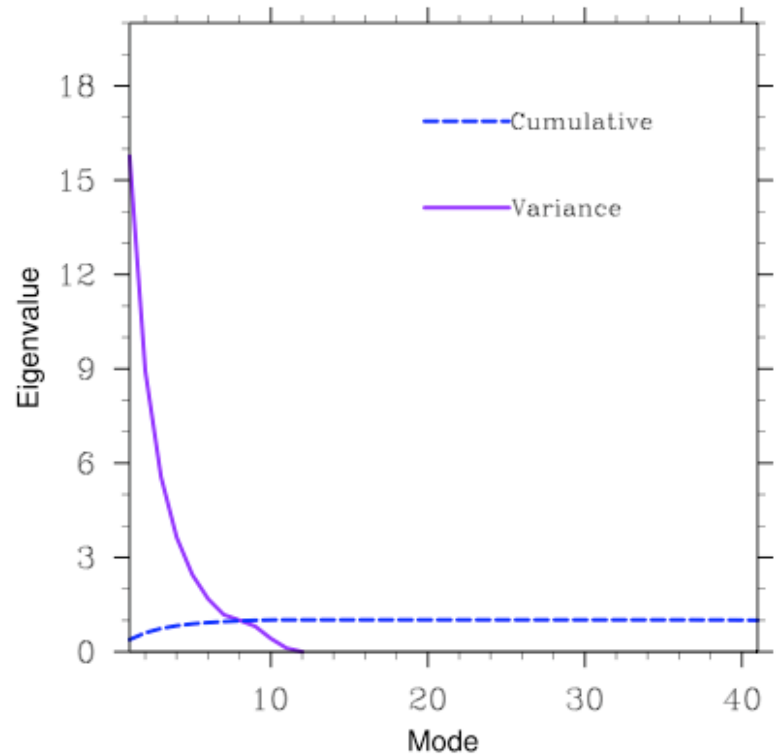
$$\rho(k-k_c) = \exp\left[-\left(k-k_c\right)^2 / L_c^2\right]$$

$$L_c = 20k_c / 41$$

Eigenvectors



Eigenvalues



75% data compression via use of EOFs for covariance localization