Generating climatological forecast error covariance for 3D-Var with ensemble perturbations: comparison with the “NMC method”

Matthew Wespetal
Advisor: Dr. Eugenia Kalnay
UMD, AOSC Department
For the PSU-UMD DA workshop, June 26, 2017

Goals

- Our interest lies in developing a way to transition from a working ensemble Kalman Filter (EnKF) data assimilation scheme to a hybrid DA scheme

- We also wish to improve on the "NMC method," which is the most common method for constructing the static background covariance for 3D-Var (3D variational data assimilation)

Methods

Our system

- Atmospheric model: SPEEDY (Simplified Parameterizations, primitivE-Equation DYnamics) with T30 resolution and 7 vertical levels

- 3D LETKF with multiplicative inflation

- 3D-Var developed at NCAR for MM5 (Barker et al. 2004)
  - Adapted for SPEEDY by Dr. Miyoshi

Methods

We utilize a simple LETKF/3D-Var hybrid data assimilation scheme (Penny 2014), which keeps both components mostly separate

- Forecast and observations -> LETKF analysis -> 3D-Var analysis of LETKF mean

- $x_{hybrid\_mean} = (1-\alpha)*x_{LETKF\_mean} + \alpha*x_{3DVar}$

- $x_{LETKF\_i\_new} = x_{hybrid\_mean} + x_{LETKF\_i\_old}$

- $0 < \alpha < 1$

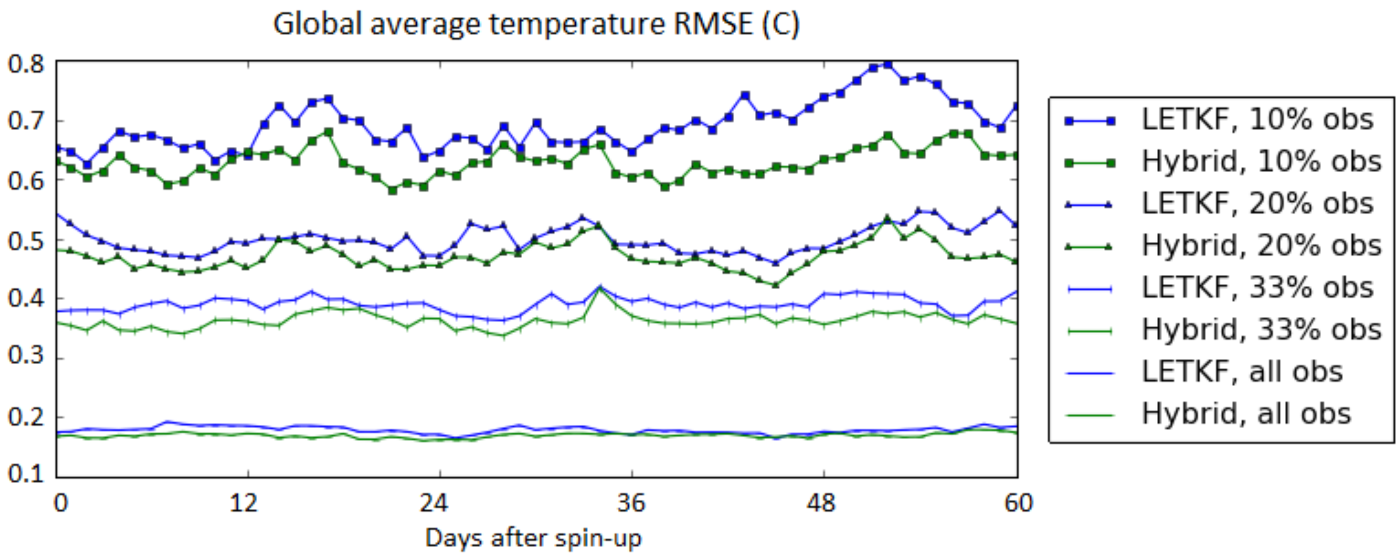
Tested on the Lorenz-96 model and at the ECMWF (Bonavita et al. 2015)

- Hybrid showed improvement over LETKF in cases with few observations and small ensemble size

Methods

LETKF/3D-Var gain hybrid experiments on SPEEDY

- OSSE with global simulated rawinsonde and wind/temperature satellite observation profiles



- Hybrid $\alpha = 0.5$

- 16 LETKF ensemble members

Hybrid shows improvement over LEKTF on SPEEDY with few observations

Methods: 3D-Var setup

3D-Var requires a parameterization of a climatological model forecast error covariance: $B = E[ee^T]$, where e is model forecast error

B construction methods that we consider:

1) "NMC":
   - Developed by Parrish and Derber (1992) and named for its use in the NCEP 1996 reanalysis
   - Assumes that forecast error can be estimated by the difference in two forecasts, one shorter (the "truth") and one longer, verifying at the same time
   - Forecasts are initialized from different points in a reanalysis

2) Ensemble perturbations:
   - Begin with an ensemble of perturbed model states and forecast the ensemble; then the perturbation of an ensemble forecast, relative to the forecast mean, estimates forecast error
   - Forecasts can be the ensemble forecasts of an EnKF reanalysis

Methods: 3D-Var setup

The NCAR (Barker et al.) parameterization for 3D-Var splits B into

$$B = U\,U^T = U_p\,U_v\,U_h\,A\,A^T\,U_h^T\,U_v^T\,U_p^T$$

A: error variance

U_p: normalized variable error correlations

U_v: normalized vertical spatial error correlations

U_h: normalized horizontal spatial error correlations

U_p contains regression coefficients that transform correlated wind, temperature and pressure to uncorrelated vorticity and unbalanced mass

The horizontal spatial correlations are fitted to Gaussian distributions; U_h contains parameters for a filter that applies the Gaussian distribution

U_v is not used for the SPEEDY 3D-Var because of the small vertical resolution (7 levels)

Methods: 3D-Var setup

Parameters in U_h and U_p are calculated by averaging over samples of forecast error, *e*

We test two estimates for forecast error *e* (with variations):

1) NMC method: *e* = difference in two forecasts, one shorter (the "truth") and one longer, verifying at the same time

2) Perturbations method: *e* = a perturbation from an ensemble of forecasts, relative to the forecast mean

Variations:

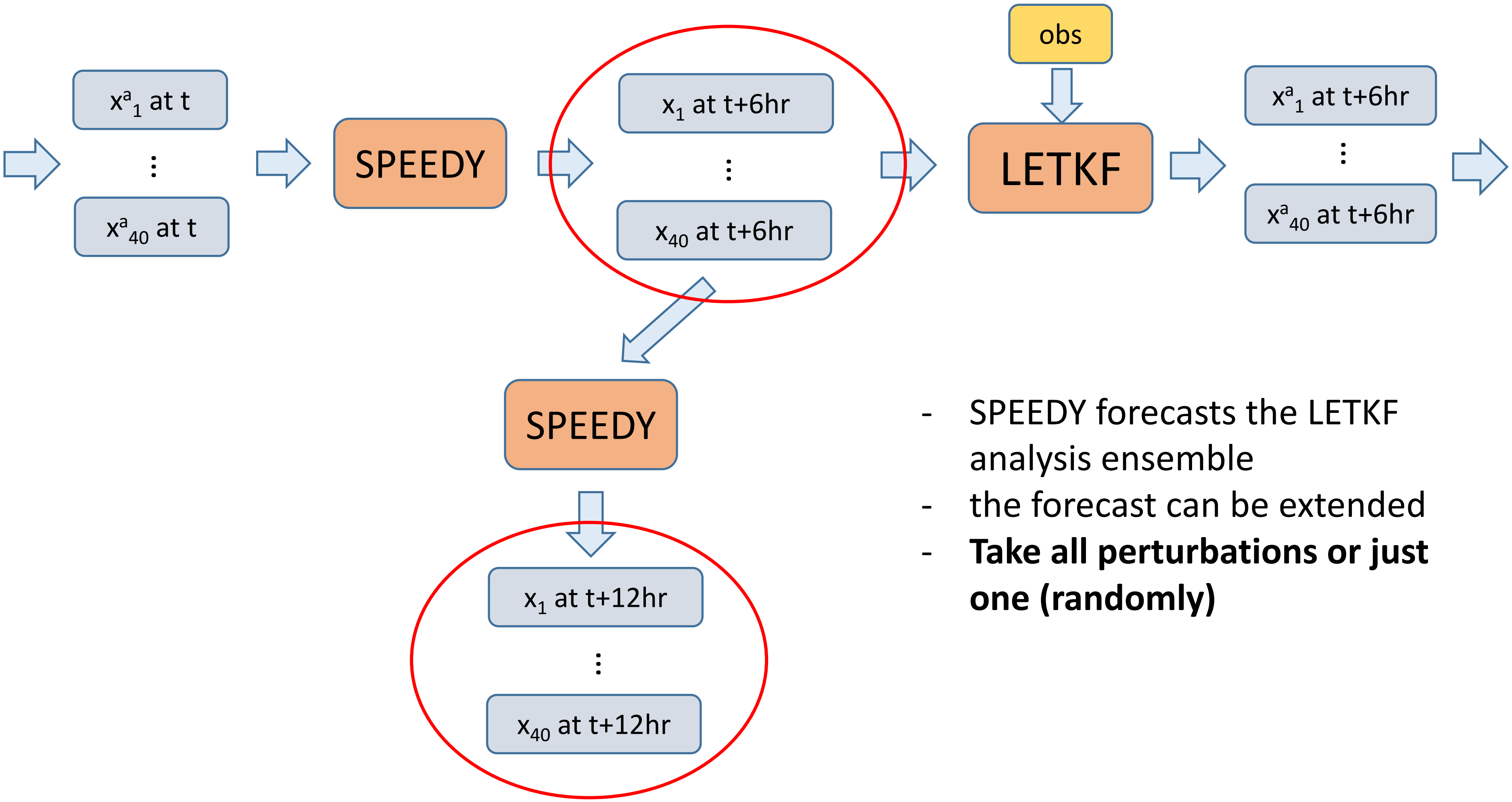
- Considering an ensemble of forecasts, one, some or all of the perturbations can be chosen

- Forecast differences can be obtained from two individual forecasts or from two ensembles of forecasts

- The length of the ensemble forecast, and the length difference between a pair of forecasts compared, can be varied

Methods: 3D-Var setup

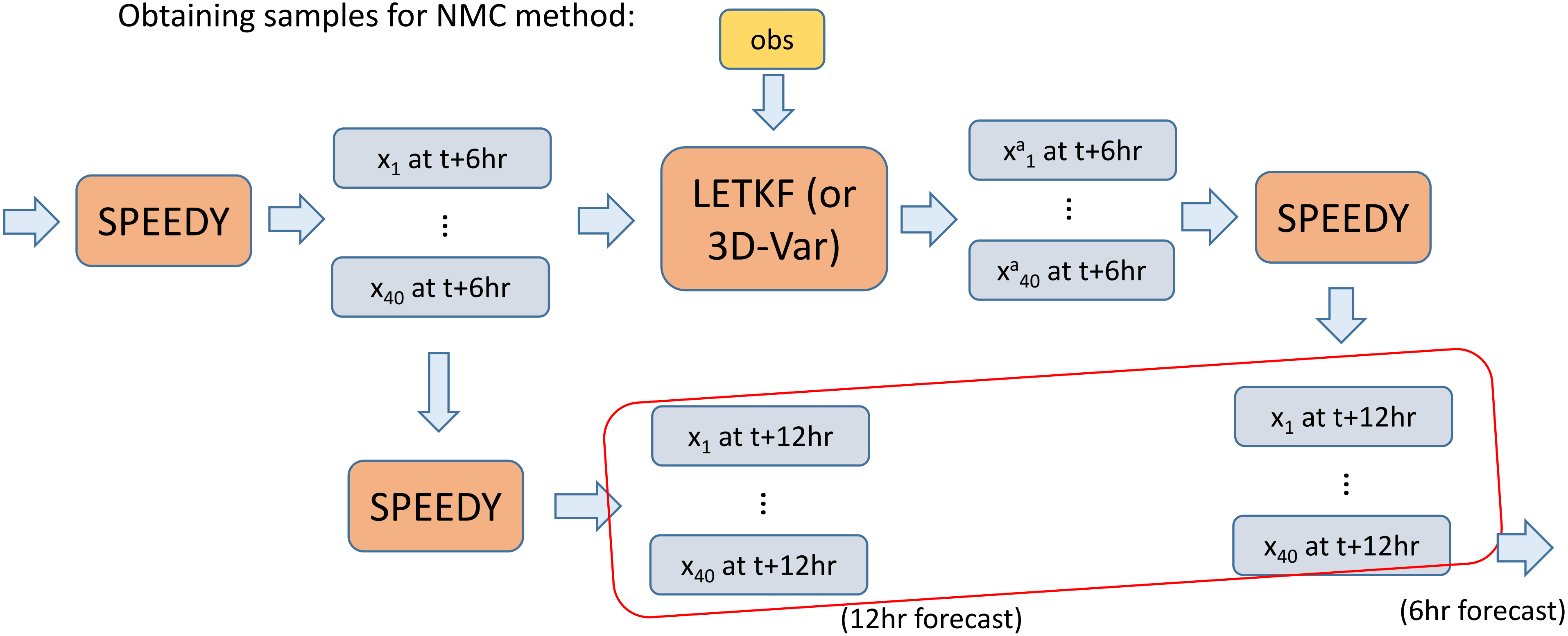
Obtaining samples for the perturbations method: 40-member LETKF OSSE 6hr cycle “reanalysis”



- SPEEDY forecasts the LETKF analysis ensemble
- the forecast can be extended
- **Take all perturbations or just one (randomly)**

Methods: 3D-Var setup

Obtaining samples for NMC method:



- Compare with a forecast from previous forecast or from even earlier analysis step
- **Either take differences between all members, or just one difference (in latter case, use 3D-Var instead of LETKF)**

Results

- We did 3D-Var experiments with statistics from both methods, with variations to find the best setup for each method

- We then compared the two methods with the LETKF/3D-Var hybrid

Results: 3D-Var RMSE (min samples)

Experiments were performed over Jan-Apr

Comparison of RMSE error +/- standard deviation of 3D-Var experiments with NMC and perturbation error statistics, **made taking 1 sample of error each 6hr from the data**

3D-Var statistics method (minimum number of error samples)	Optimum variance scaling	RMSE of U (m/s)	RMSE of V (m/s)	RMSE of T (C)	RMSE of Ps (mb)
Pert: 6h forecast perturbations	2	**0.80±0.07**	0.84±0.06	**0.27±0.01**	14.4±0.7
Pert: 12h forecast perturbations	2.5	0.81±0.05	**0.83±0.05**	0.27±0.01	**14.3±0.8**
NMC: 24h vs 18h forecasts	5	0.92±0.05	0.94±0.05	0.31±0.01	17.6±0.9
NMC: 24h vs 12h forecasts	4	0.87±0.05	0.91±0.05	0.30±0.01	**17.5±0.8**
NMC: 24h vs 6h forecasts	3.5	**0.85±0.05**	**0.90±0.05**	**0.29±0.01**	17.8±0.8
NMC: 18h vs 6h forecasts	4	0.89±0.05	0.92±0.05	0.30±0.01	17.9±1.0
NMC: 12h vs 6h forecasts	6	0.95±0.05	0.97±0.05	0.32±0.01	19.0±1.0

NMC-based 3D-Var has higher RMSE than perturbations-based 3D-Var

Results: 3D-Var RMSE (max samples)

Comparison of RMSE error +/- standard deviation of 3D-Var experiments with NMC and perturbation statistics, **made taking 40 samples of error each 6hr from the data**

3D-Var statistics method (maximum number of error samples)	Optimum variance scaling	RMSE of U (m/s)	RMSE of V (m/s)	RMSE of T (C)	RMSE of Ps (mb)
Pert: 6h forecast perturbations	2.5	**0.74±0.06**	**0.77±0.05**	**0.26±0.01**	**13.4±1.1**
Pert: 12h forecast perturbations	2.5	0.80±0.04	0.83±0.04	0.26±0.01	14.1±0.7
NMC: 24h vs 18h forecasts	10	0.87±0.07	0.88±0.07	0.30±0.02	19.0±0.8
NMC: 24h vs 12h forecasts	6.5	0.82±0.06	0.84±0.06	**0.28±0.01**	17.6±1.0
NMC: 24h vs 6h forecasts	6	**0.82±0.05**	0.85±0.05	0.29±0.01	**17.6±0.7**
NMC: 18h vs 6h forecasts	7	0.83±0.04	**0.84±0.05**	0.29±0.01	17.6±0.7

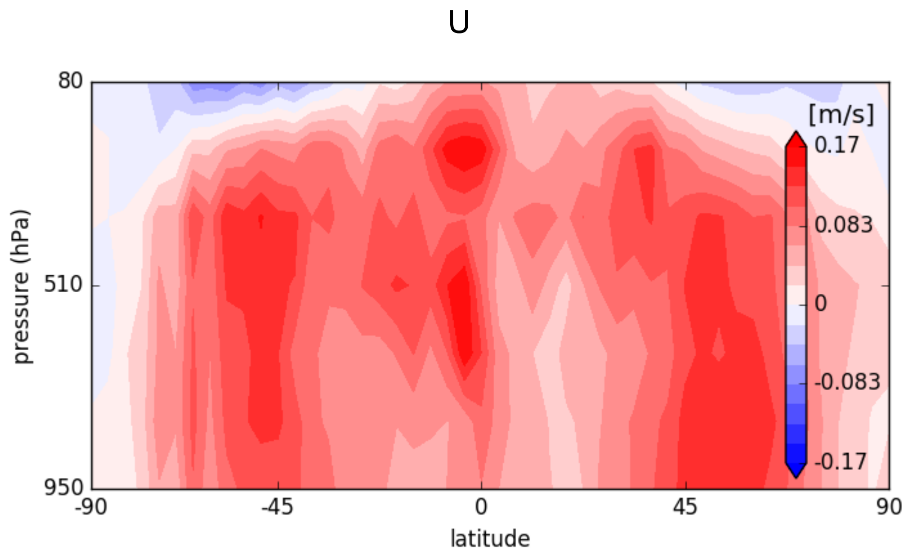
More samples used to construct the 3D-Var climatology leads to lower RMSE
NMC-based 3D-Var still has higher RMSE than perturbations-based 3D-Var

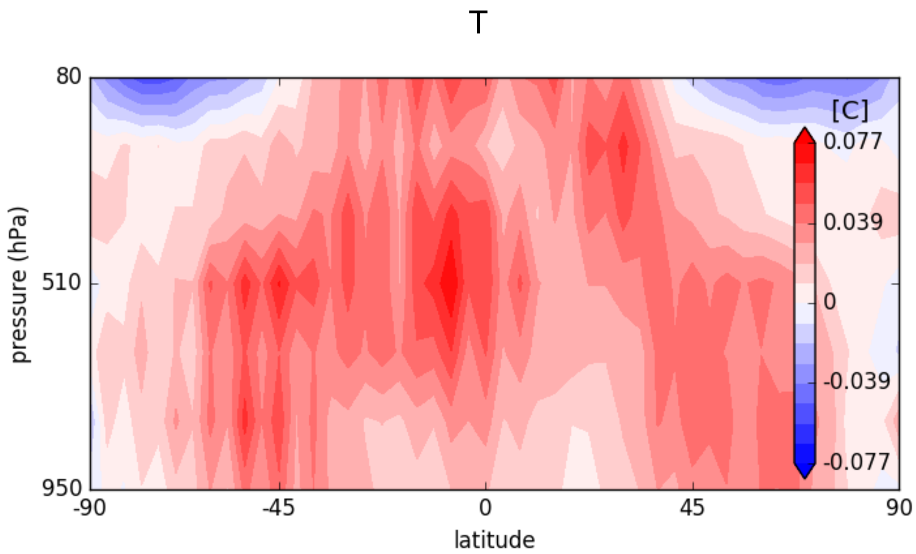
Results: 3D-Var RMSE

Gridded 3D-Var RMSE comparison (3 month average):

NMC (24hr vs 6hr) vs
ensemble perturbations (6hr)

Red: perturbations-based 3D-Var does better than NMC-based 3D-Var





Results: Hybrid RMSE

RMSE error +/- standard deviation of LETKF/3D-Var runs using four different statistics variations, with all observations

DA procedure	Optimum variance scaling	RMSE of U (m/s)	RMSE of V (m/s)	RMSE of T (C)	RMSE of Ps (mb)
LETKF		0.47 ± 0.02	0.47 ± 0.02	0.178 ± 0.06	9.7 ± 0.5
LETKF/3DVAR (40 Pert 6h)	**1.5**	**0.44 ± 0.02**	**0.45 ± 0.02**	**0.166 ± 0.05**	9.4 ± 0.5
LETKF/3DVAR (Single Pert 6h)	1.5	0.45 ± 0.02	0.46 ± 0.03	0.168 ± 0.007	**9.4 ± 0.4**
LETKF/3DVAR (Single NMC 24h vs 6h)	1.5	0.46 ± 0.02	0.46 ± 0.02	0.170 ± 0.004	9.7 ± 0.4
LETKF/3DVAR (40 NMC 24h vs 6h)	3	0.46 ± 0.02	0.46 ± 0.02	0.172 ± 0.005	10.1 ± 0.4

Similar results as with pure 3D-Var, but reduced difference between the methods

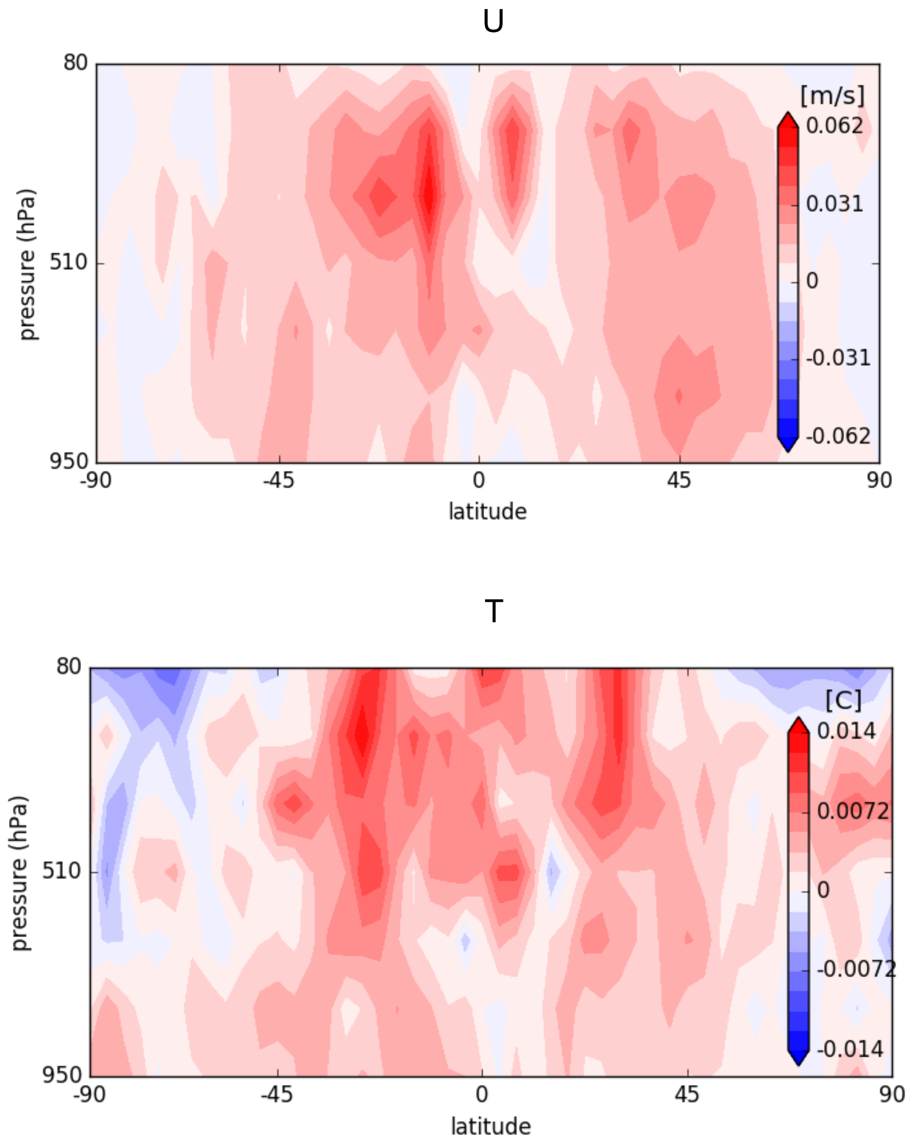
Results: Hybrid RMSE

Gridded hybrid RMSE comparison (3 month average):

NMC (24hr vs 6hr) vs ensemble perturbations

Red: NMC-based 3D-Var does worse than perturbations-based 3D-Var

Difference between methods is now more concentrated in the mid-latitudes



Conclusions

- LETKF background ensemble perturbations can be used to generate error covariance statistics for 3D-Var
- The perturbation error estimation method results in smaller 3D-Var errors than the “NMC method”
- The impact of the new method in the hybrid is smaller than in the 3D-Var

Future plans

- I am porting this hybrid with ensemble statistics to the MGCM (Mars GCM), for use in the ongoing EMARS reanalysis, a joint project by Penn State, UMD, and NASA

- We are preparing a paper for publication

References

- Barker, D. M., W. Huang, Y. Guo, and A, 2003. Bourgeois: A three-dimensional variational (3DVAR) data assimilation system for use with MM5. NCAR Tech. Note 68, 73 pp.
- Barker, D. M., Huang, W., Guo, Y. R., Bourgeois, A. J., and Xiao, Q. N, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.,* **132**, 897-914.
- Penny, S. G, 2014: The hybrid local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **142**, 2139-2149.
- Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. thesis, Dept. of Atmospheric and Oceanic Science, University of Maryland, 226 pp.
- Bonavita, M., M. Hamrud, and L. Isaksen, 2015: EnKF and hybrid gain ensemble data assimilation. Part II: EnKF and hybrid gain results. *Mon. Wea. Rev.,* **143**, 4865-4882.

- Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. thesis, Dept. of Atmospheric and Oceanic Science, University of Maryland, 226 pp.