6th WMO Symposium on Data Assimilation , 7 – 11 October 2013, College Park, Maryland

Improving an Ensemble-based Observation Impact Estimate using a Group Filter Technique

Nicholas Gasperoni[a,b] and Xuguang Wang[a,b]

[a]School of Meteorology, University of Oklahoma, Norman, OK, U.S.A.

[b]Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, OK, U.S.A.

Motivation

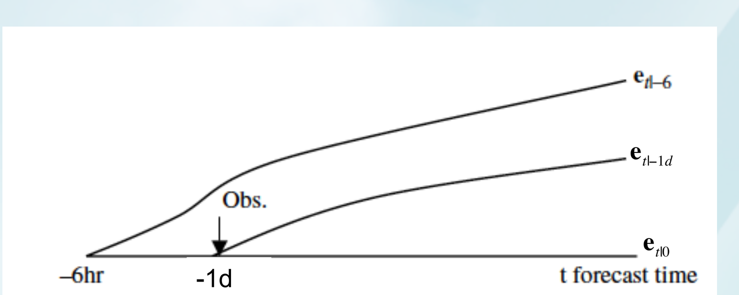
- *"How much impact does a subset of assimilated observations have on the forecast?"*
- Kalnay et al.(2012) developed an ensemble-based metric to estimate the impact from observations on a forecast, using readily-available products from any ensemble filtering system. It is analogous to the adjoint method of Langland and Baker (2004).
- Any attempt to use limited ensembles to estimate model covariances requires localization due to sampling errors.
- We investigate methods to properly specify localization for this estimate.
- Two main objectives:
  1. Use a Monte-Carlo "group filter" technique of Anderson (2007) to learn what a 'proper' localization function may look like
  2. Develop ways to improve the impact estimate over traditional static localization functions

Ensemble-based Observation Impact Metric

$$J_{Actual} = \mathbf{e}_{t|0}^T\mathbf{e}_{t|0} - \mathbf{e}_{t|-1d}^T\mathbf{e}_{t|-1d} \qquad \mathbf{e} = \bar{\mathbf{x}}^f - \bar{\mathbf{x}}^a$$

$$J_{EnsembleEstimate} = \frac{1}{K-1}\left(y - H\left(\bar{\mathbf{x}}^b_{0|-1d}\right)\right)^T \mathbf{R}^{-1}\left[\rho\circ\left(\mathbf{Y}_0^a\mathbf{X}_{t|0}^{fT}\right)\right]\left(\mathbf{e}_{t|0}+\mathbf{e}_{t|-1d}\right)$$

ρ is localization function, which acts on ensemble covariances between the analysis (in obs space) and a forecast of some length t



Schematic plot of time relationship of obs sensitivity impact on forecast at time t (after Langland and Baker, 2004, Fig. 1)

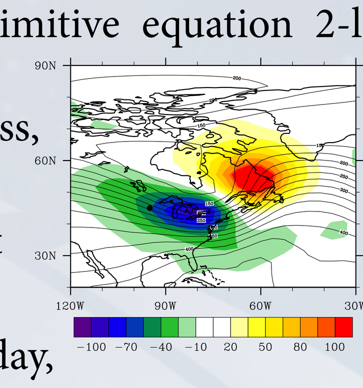
Group Filter Method

- *Anderson (2007)* – Monte Carlo technique to evaluate sampling errors.
- Uses groups of ensembles (m = 4 groups of n = 16 members, 64 total for this study).
- Each group has a sample regression coefficient at each ob (l), grid point (j) pair, $\beta_{lj} = \frac{\left(\mathbf{Y}_0^a\mathbf{X}_{t|0}^{fT}\right)_{lj}}{\left(\mathbf{X}_{t|0}^f\mathbf{X}_{t|0}^{fT}\right)_{jj}}$
- Assume they are samples of 'correct' β.
- **Regression confidence factor (RCF, or α)** – weighting factor minimizes expected RMS differences between the m sample β's $\sqrt{\sum_{j=1}^{m}\sum_{i=1,j\neq i}^{m}\left(\alpha\beta_i - \beta_j\right)^2}$

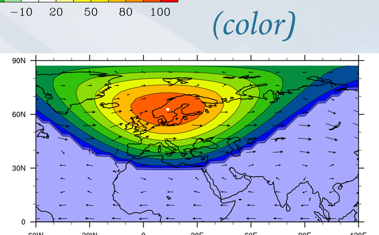
•Each ob, state variable pair has its own RCF value, *envelope of RCFs can be used directly as a localization function*

Experiment Setup

- LETKF system with dry, primitive equation 2-layer model (Holland and Wang 2012)
- Three variables: Layer thickness, vorticity, and divergence
- 362 simulated interface height observations
- 1000 cycles, cycle length of 1 day, RCF averaged over last 900 cycles
- 64 members total for LETKF analyses (with 8000-km GC loc), split into 4 groups randomly for RCF and impact simulations
- Single-ob and all-ob experiments



LETKF example of background thickness (contour) and analysis increments (color)

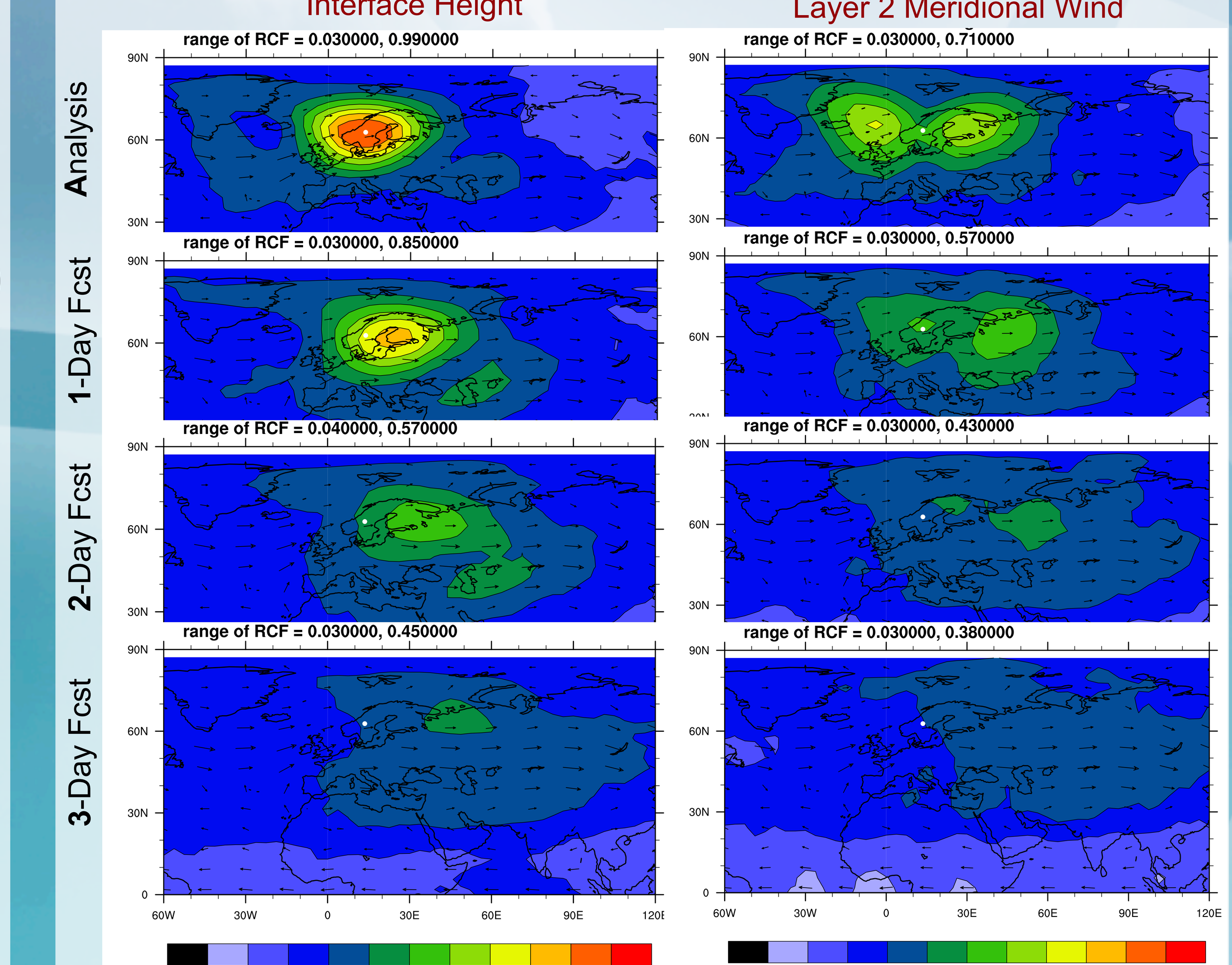


GC Function, 8000-km radius

Results – Group Filter (RCF Functions)

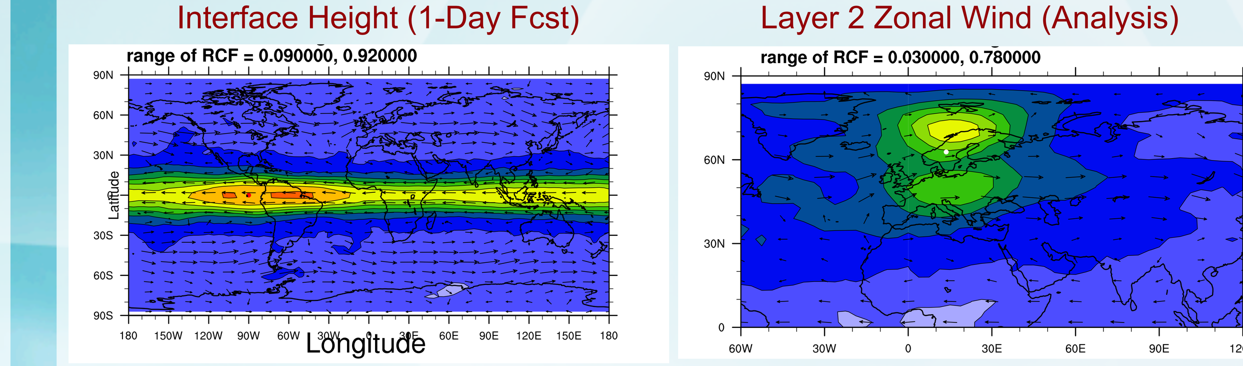
RCF envelope, midlatitude ob: Interface Height

Cross-Variable RCF envelope: Layer 2 Meridional Wind



RCF envelope, *tropical (equatorial)* ob: Interface Height (1-Day Fcst)

Cross-Variable RCF envelope: Layer 2 Zonal Wind (Analysis)



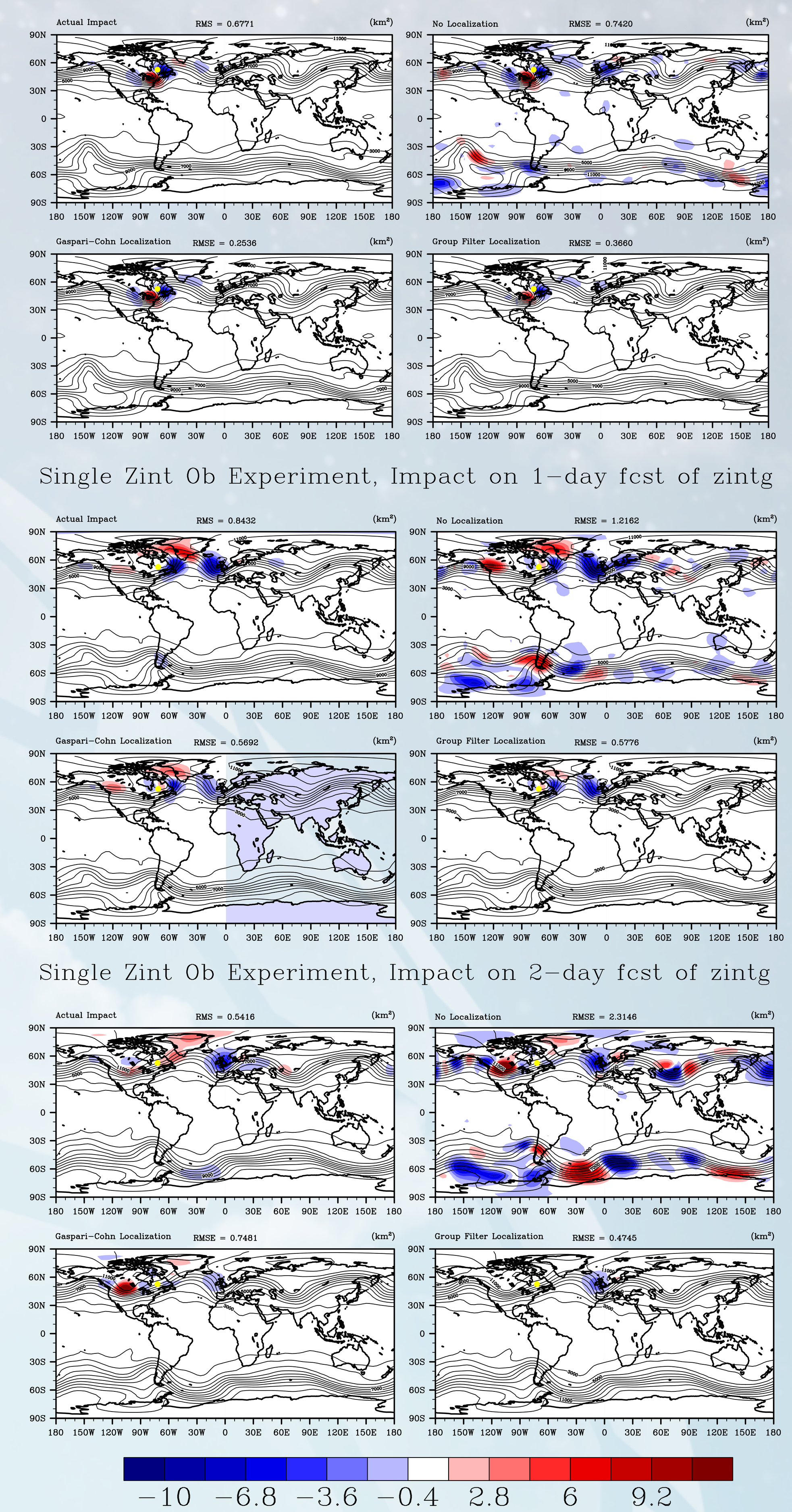
- RCF shifts downstream as forecast time increases
- Signal dampens with time due to nonlinear effects (RCF uses linear regression)
- Averaging over many cycles smooths out noise from sampling error (RCF itself can have sampling error)
- RCF reveals cross-variable model dynamic linking between interface height observations and winds, with bimodal distributions.

Results – Impact Estimate

Single Zint Ob Experiment, Impact on 0-day fcst of zintg

Single Zint Ob Experiment, Impact on 1-day fcst of zintg

Single Zint Ob Experiment, Impact on 2-day fcst of zintg



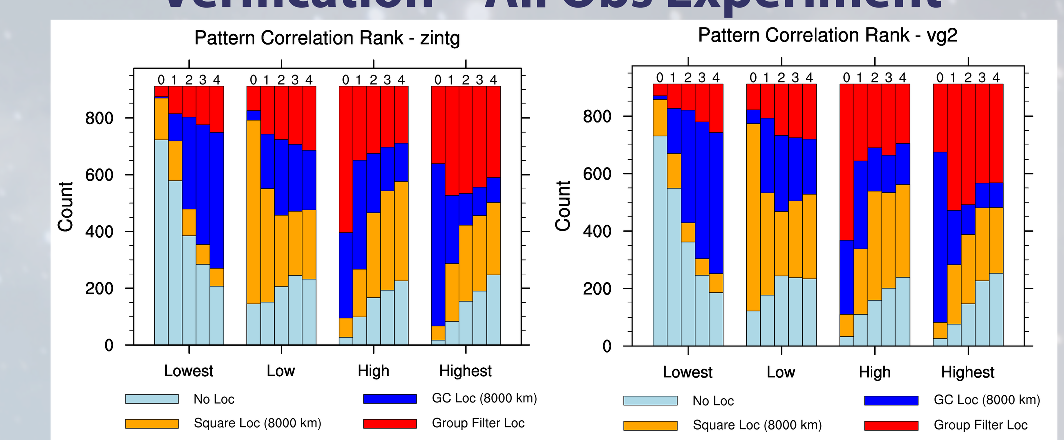
Verification

- Quantify the comparison between *estimated* impact and *actual* impact
- Identify areas where GF method in general works well, and deficiencies that need further investigation
- Use correlation, mean-squared error (MSE), and skill score (SS).

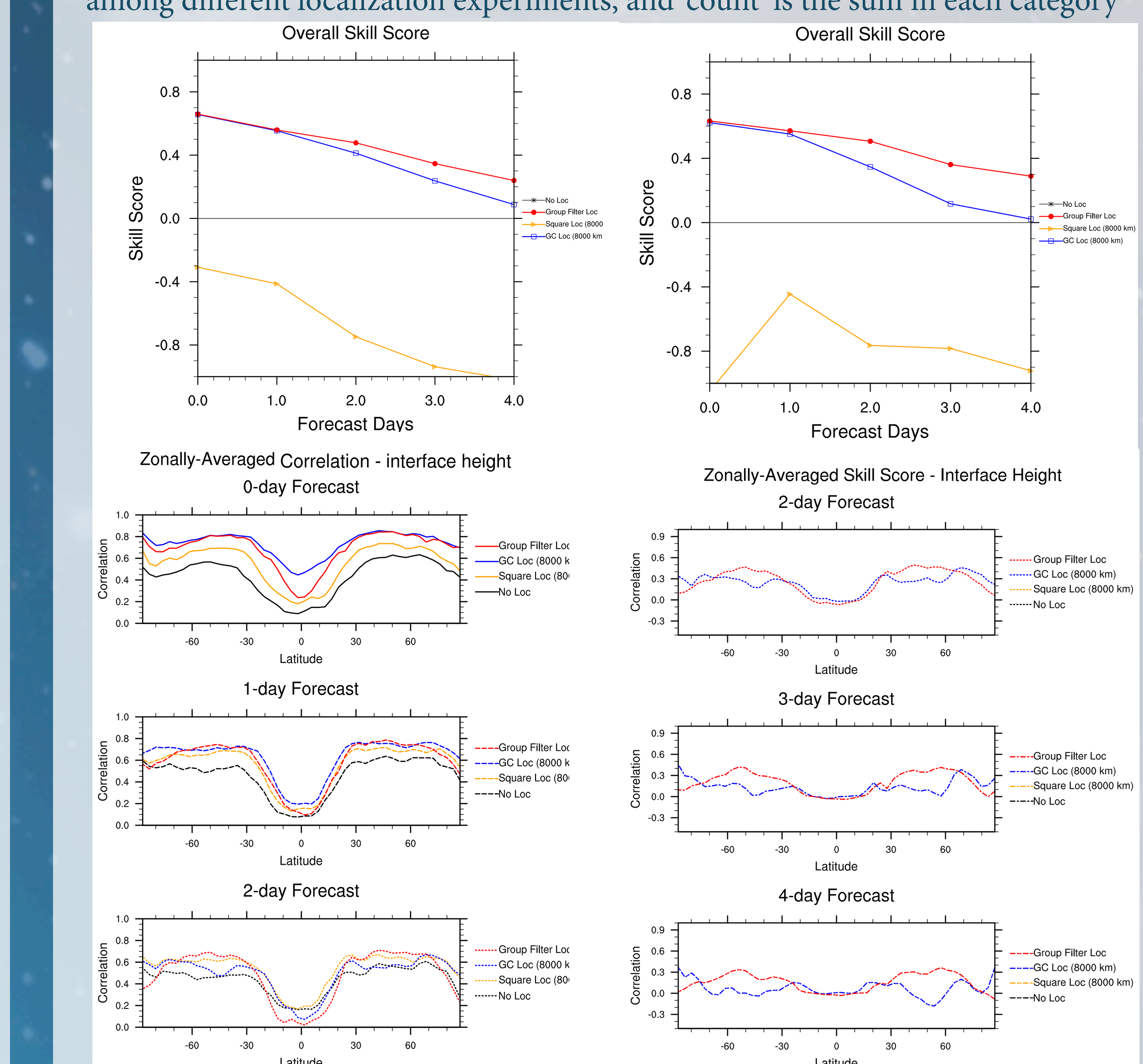
$$SS = 1 - \frac{MSE}{MSE_{ref}} \qquad MSE = \frac{1}{n}\sum_{k=1}^{n}\left(J^k_{EnsEst} - J^k_{Actual}\right)^2 \qquad MSE_{ref} = \frac{1}{n}\sum_{k=1}^{n}\left(J^k_{Actual}\right)^2$$

MSE_{ref} can be thought of as the mean-squared error if your "estimated impact" was zero at all grid points

Verification – All Obs Experiment



Correlation calculated over matched grid point pairs for one cycle, total of 900 map correlation values. For each cycle, correlations were ranked from lowest to highest among different localization experiments, and 'count' is the sum in each category



Conclusions and Future Work

- RCF function shows dynamical flow and time-shift dependencies of covariances between analysis and forecast at some time.
- Application of RCF function as localization to the observation impact estimate shows about as much skill as tuned static Gaspari-Cohn localization, at 0 and 1-day impact. GC or RCF functions are needed to show any positive skill at all.
- On average, group filter localization shows more skill at longer lead-time forecasts and for cross-variable impact.
- This added skill is confined to midlatitudes.
- Localization, and particularly the RCF method, does not work as well at the equator. More investigation is needed here.
- Need to investigate impacts of increasing number of groups and number of ensembles per group
- Investigate use of single-cycle or short-term-averaged RCF, to get dynamic linking more appropriate for cycle-to-cycle dynamical forecast variations, though sampling error in RCF will be an issue (need to 'localize' the RCF localization!)

References

Anderson, J., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*,**230**: 99-111.

Holland, B. and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter . *Q. J. R. Meteo. Soc.*, **139**, 758-770.

Kalnay, E., Ota, Y., Miyoshi, T., and Liu, J., 2012: A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A*, **64**.

Langland, R.H. and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201.

http://caps.ou.edu

First-author email: ngaspero@ou.edu

ARPS Simulated Tornado