# Improved Localization of the Ensemble-based Observation Impact Estimate using a Dynamic Group Filter Technique

Nicholas Gasperoni and Xuguang Wang

Center for Analysis and Prediction of Storms
OU School of Meteorology

6th EnKF Workshop – 20 May 2014

# Motivation

- *"How much impact does a subset of assimilated observations have on the forecast?"*

- Data-denial (OSE/OSSEs)
  - Straightforward (compare experiments 'denying' observations with a control experiment)
  - Computational cost is high (need many assimilation/forecast experiments)

- Adjoint Method (e.g. Langland and Baker 2004, Gelaro and Zhu 2008, Cardinali 2009)
  - No need for data denial experiments, just the control run
  - Based on adjoint sensitivity, tangent linear model
  - Limited by linearity, difficulty of developing an adjoint for complex models

# Motivation

- Ensemble-based method (e.g. Liu and Kalnay 2008, Kalnay et al. 2012)
  - Use of ensemble covariances to estimate sensitivities (no data denial, adjoint necessary)
  - Sampling error (ensemble size typically << model degrees of freedom)
  - **Localization can be applied to ameliorate sampling error**
- Localization of ensemble-based impact is not trivial
  - Space/shape
  - Cross-variable
  - ***Time/Location***
- **Goal: Apply dynamic localization method to improve accuracy of ensemble impact metric, learn more about what 'proper' localization should look like**

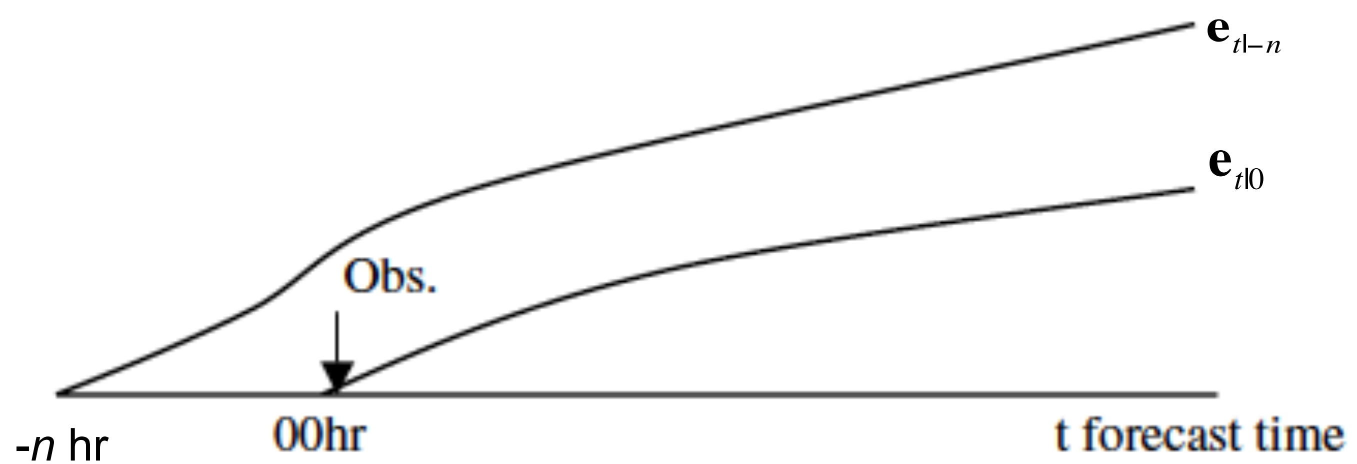
# Ensemble Impact Metric

- Kalnay et al. (2012)
  - Analogous to Langland and Baker (2004) adjoint approach – impact defined as reduction of (squared) error

$$J_{Actual} = \frac{1}{2}\left(\mathbf{e}_{t|0}^{T}\mathbf{e}_{t|0} - \mathbf{e}_{t|-n}^{T}\mathbf{e}_{t|-n}\right)$$

$$\mathbf{e}_{t|0} \equiv \overline{\mathbf{x}}_{t|0}^{f} - \mathbf{x}_{t}^{truth}$$

$$\mathbf{e}_{t|-n} \equiv \overline{\mathbf{x}}_{t|-n}^{f} - \mathbf{x}_{t}^{truth}$$



# Ensemble Impact Metric

$$J_{Actual} = \frac{1}{2}\left(\mathbf{e}_{t|0}^{T}\mathbf{e}_{t|0} - \mathbf{e}_{t|-n}^{T}\mathbf{e}_{t|-n}\right)$$

$$J_{Estimate} = \frac{1}{K-1}(\delta\mathbf{y})^{T}\,\mathbf{R}^{-1}\left[\rho_{I}\circ\left(\mathbf{Y}_{0}^{a}\mathbf{X}_{t|0}^{fT}\right)\right]\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n}\right)$$

- Localization applied to covariances between analysis (in obs space) and model forecast

- Addition of *moving* localization (Kalnay et al. 2012, Ota et al. 2013) led to improved estimates

- *What should proper choice of localization look like?*

# Group Filter Method

- Anderson (2007) – Monte Carlo technique to evaluate sampling errors.
- Uses groups of ensembles ($m$ = 4 groups of $k$ = 16 members, 64 total for this study).
- Each group has a sample regression coefficient at each ob ($l$), grid point ($j$) pair, $\beta_{lj}$

$$\beta_{lj} = \frac{\left(\mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT}\right)_{lj}}{\left(\mathbf{Y}_0^a \mathbf{Y}_0^{aT}\right)_{jj}}$$

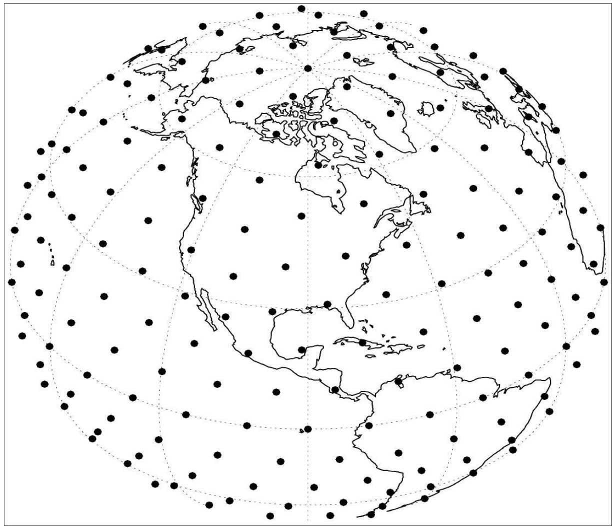
- *Choose weighting factor, $\alpha_{min}$, that minimizes RMS differences between group $\beta$'s, also known as <u>Regression Confidence Factor (RCF)</u>*

$$RCF = \alpha_{\min} = \frac{m - Q^2}{(m-1)Q^2 + m} \qquad Q \equiv \frac{s_\beta}{\left|\overline{\beta}\right|}$$

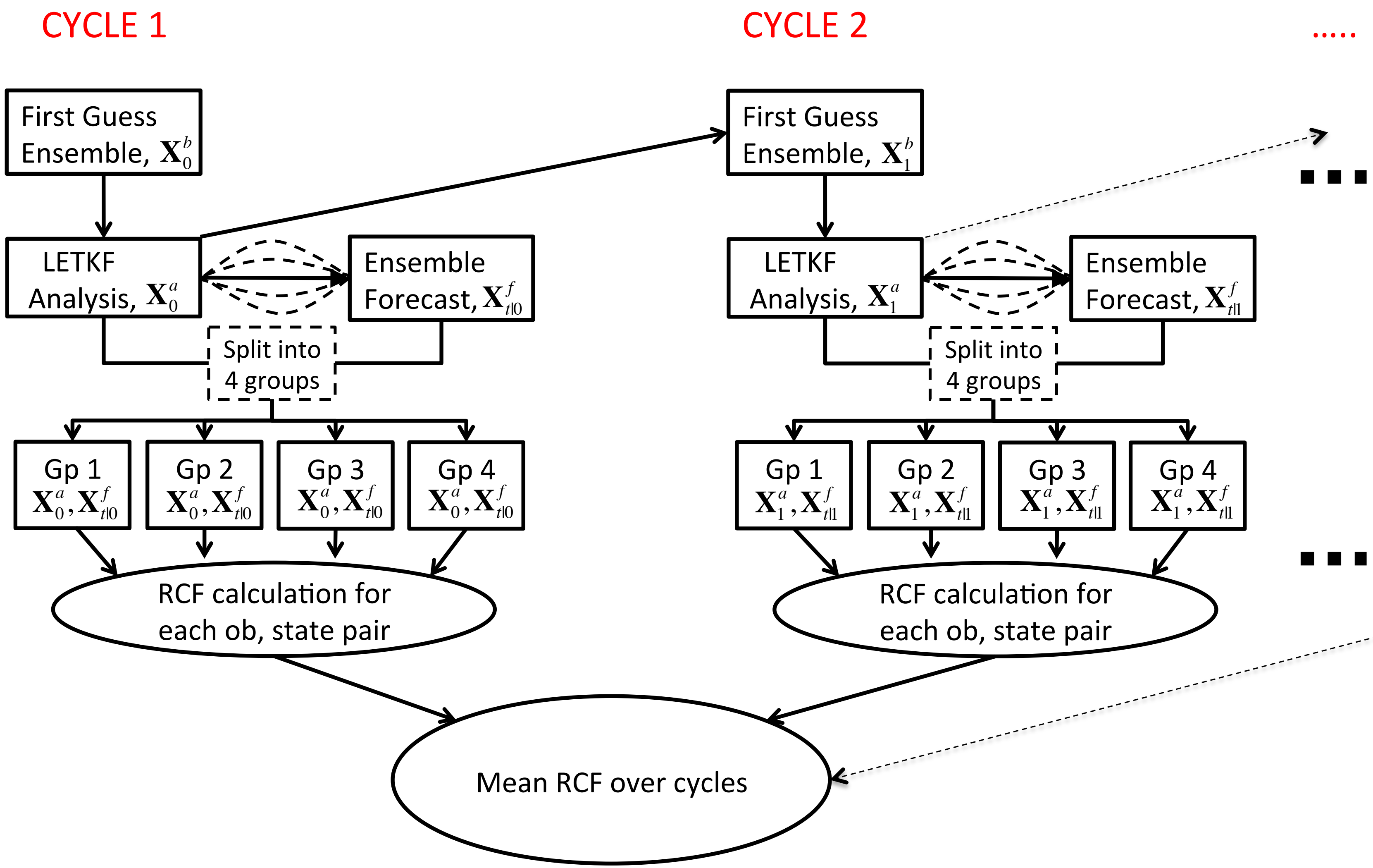
- $m$ is the number of groups
- Every ob, state variable pair has unique RCF, and can be used directly as localization function, valid for a ensemble of size $k$

# Experiment Setup

- LETKF system with Dry, primitive equation **2-layer model** (Zou et al. 1993)
  - Used in several ensemble-based data assimilation experiments because of (e.g., Wang et al., 2007, 2009; Holland and Wang, 2013)
  - **Layer thickness, u, v**
  - Radiative heating and surface drag, with zonal wavenumber-2 terrain, Runge Kutta for forward integration
  - LETKF system setup from Holland and Wang (2013)
    - T31 model resolution
    - 24-hr assimilation cycle length, 1000 cycles total, GC localization (8000 km cutoff)
    - **362 synthetic equally-spaced interface height observations assimilated**
    - Ensembles -> random draw from states of truth run
    - **64 total ensemble members**
- RCF computed for each ob/state pair at each cycle time for 0,1,2,3,4-day forecasts, then averaged over all cycles to reduce noise from sampling error
- Test impact estimate using 16-ensemble members
  - Compare RCF localization to static localizations and no localization



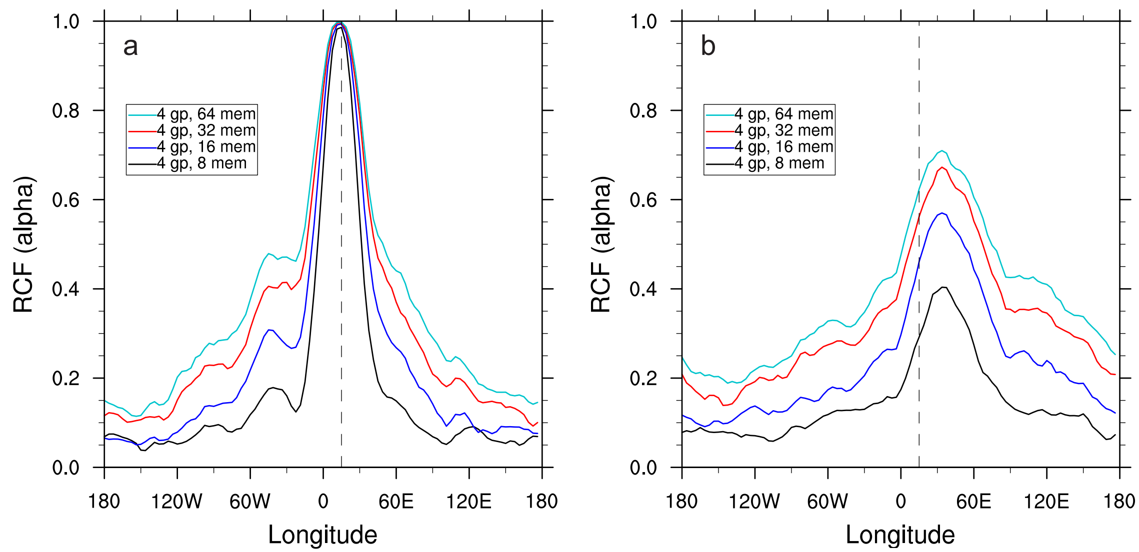
# Experiment Setup



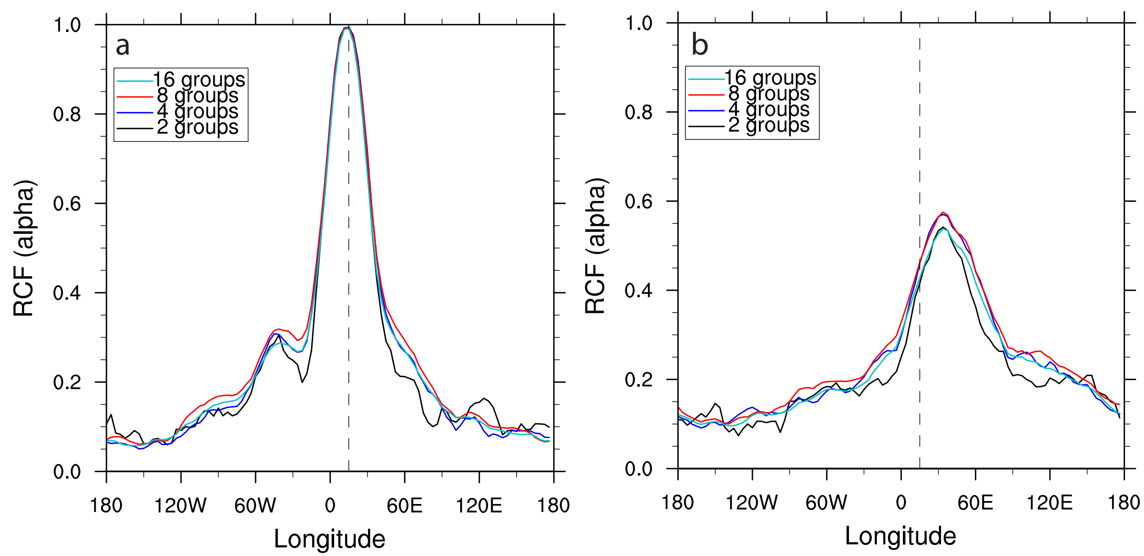
# RCF functions

## Zonal Cross Section, Midlatitude Observation

Sensitivity to number of ensemble members per group

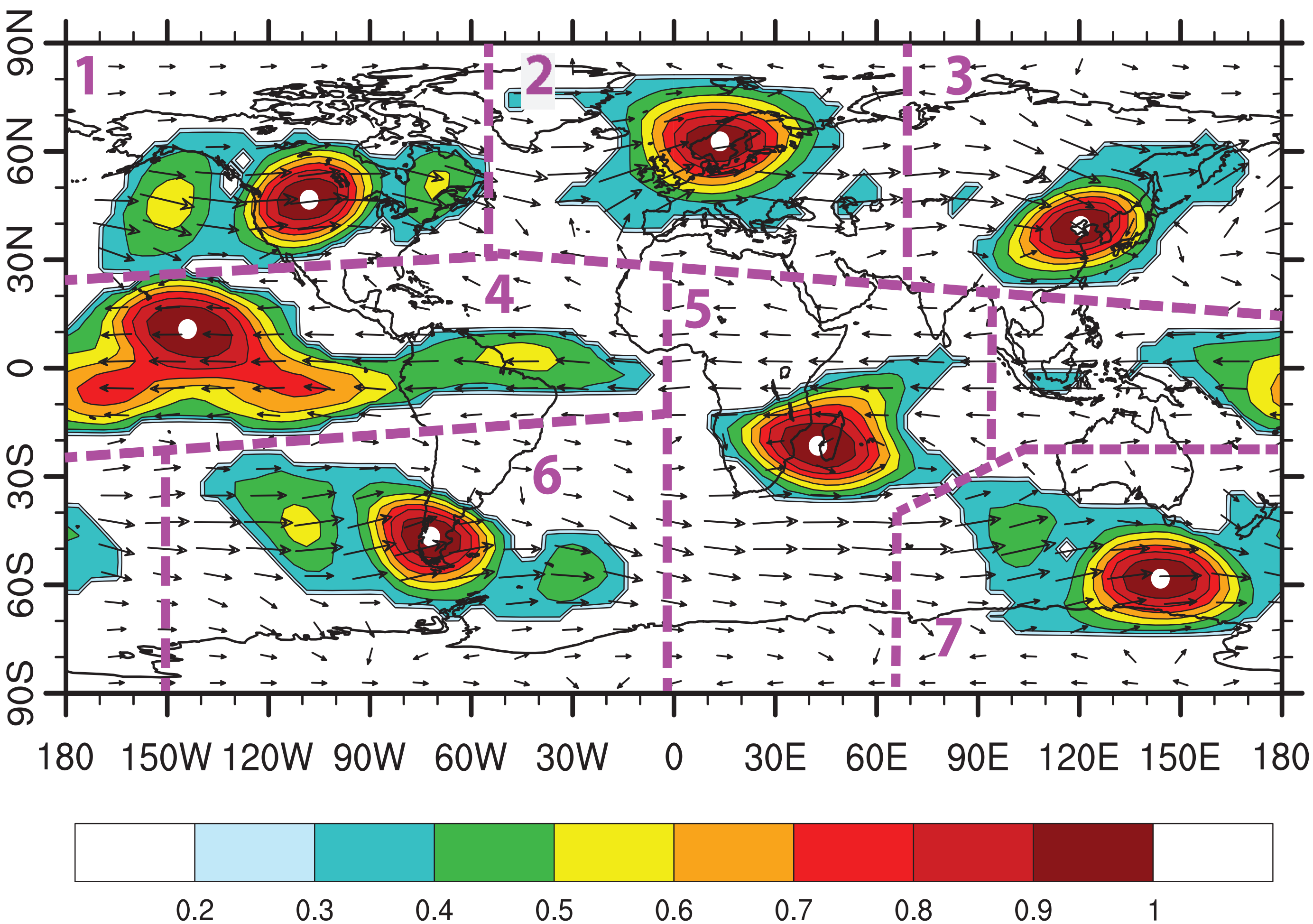



Sensitivity to number of groups (using 16-mem ensemble)



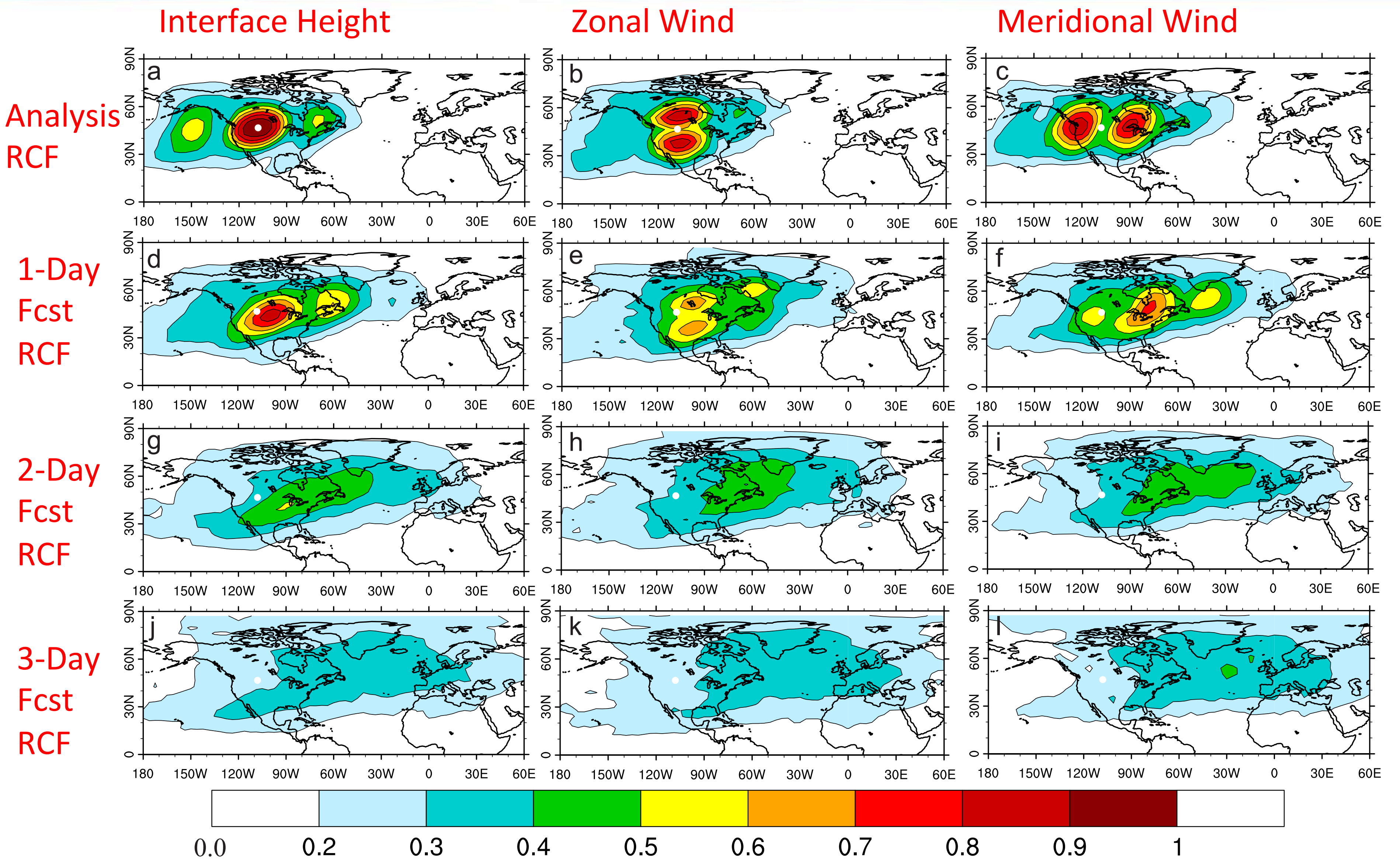
# Examples of RCF Localization functions

RCF for model interface height, at analysis time

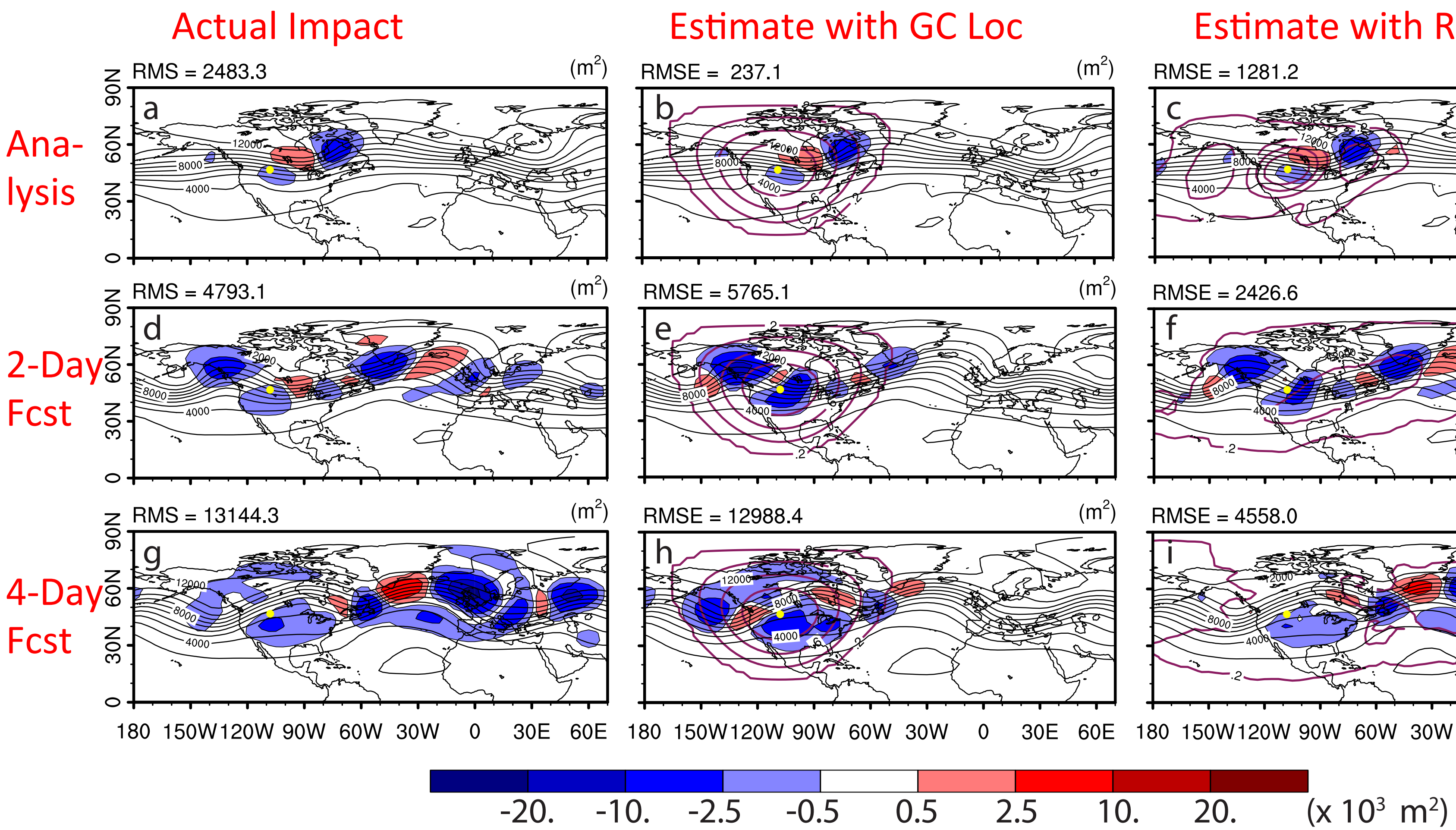



- Triple-peaked structures for some midlatitude obs
- Stretching in zonal direction
- Stretching along equator for tropical obs

# Results – RCF Functions



# Single Observation Impact Experiment



# All-Obs Experiments - Verification

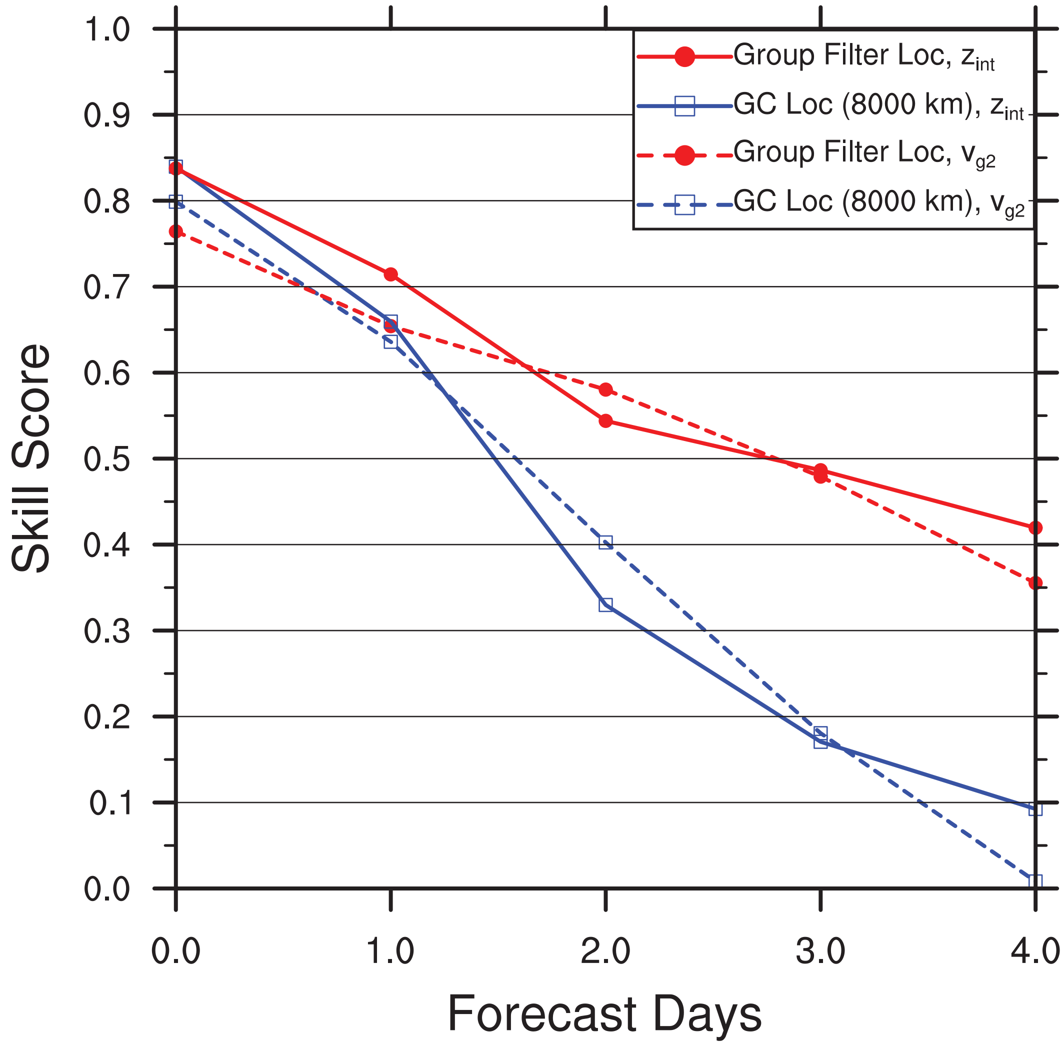
- Impact estimates from each observation summed up at each grid point
- Verified against *actual impact* – actual forecast error reduction in mean ensemble calculated against the truth
- Root mean square error (RMSE) and Skill Score (SS)

$$SS = 1 - \frac{RMSE}{RMSE_{ref}} = 1 - \frac{\left[\sum_{k=1}^{N}\left(\Delta e^2_{actual,k} - \Delta e^2_{ens\ est,k}\right)^2\right]^{1/2}}{\left[\sum_{k=1}^{N}\left(\Delta e^2_{actual,k}\right)^2\right]^{1/2}}$$

- $RMSE_{ref}$ = RMS of the actual impact field, or can be thought of as the RMSE of an "impact estimate" of 0 at all grid points.

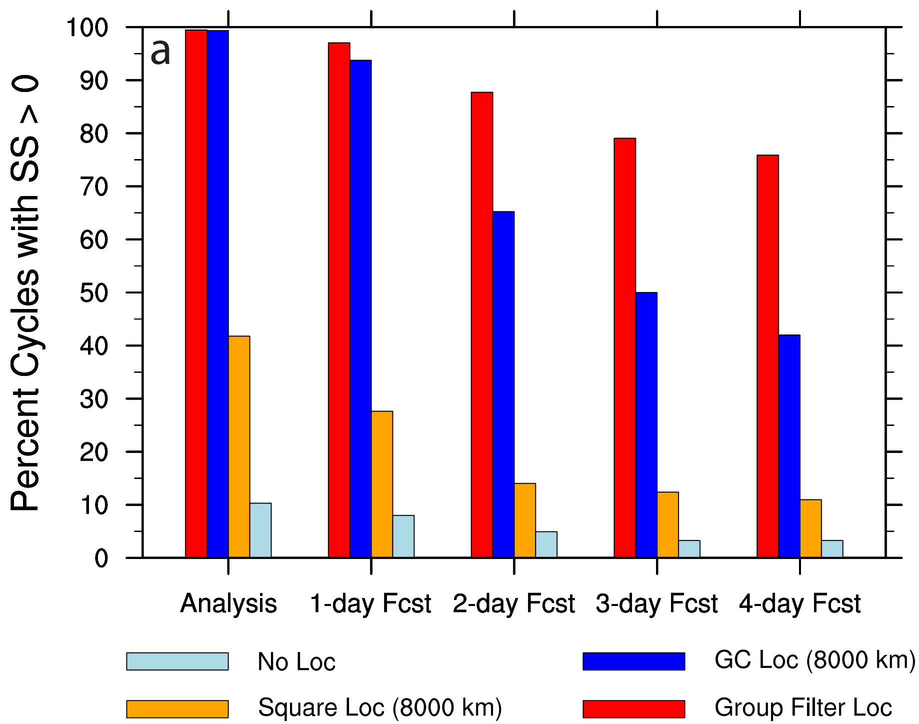
# All-Obs Experiments - Verification

- Time-mean SS, averaged over all grid points
- For impact on forecast model interface height (solid) and meridional wind (dashed)
- At analysis (t = 0), GC localization outperforms RCF localization
- At 1-day forecast and beyond, RCF outperforms GC, differences grow in magnitude

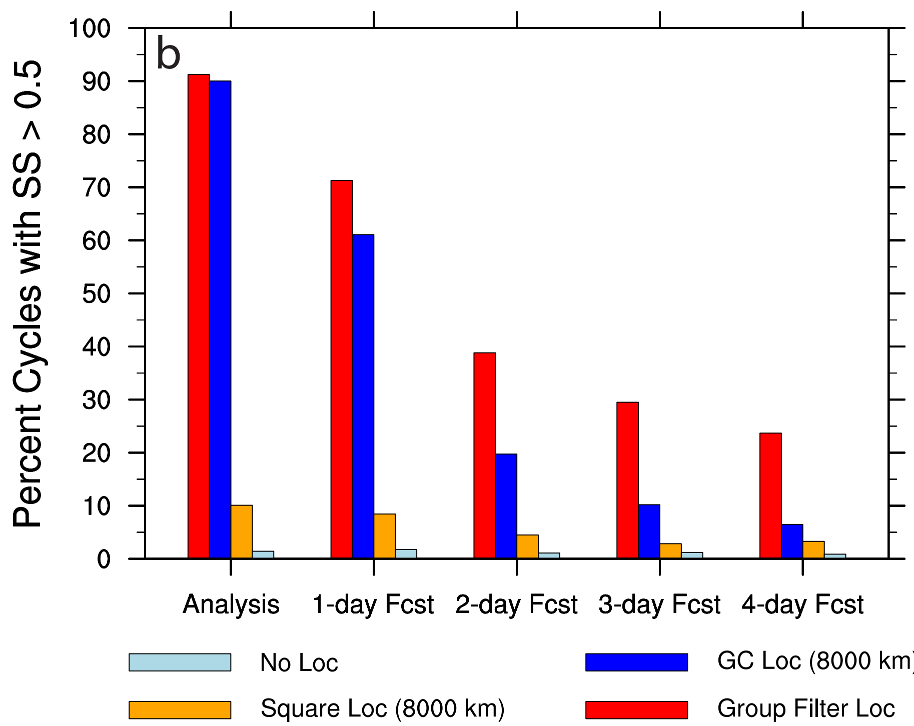


# All-Obs Experiments - Verification

- Cycle-mean SS, Percentage of the 900 analysis cycles which show positive skill

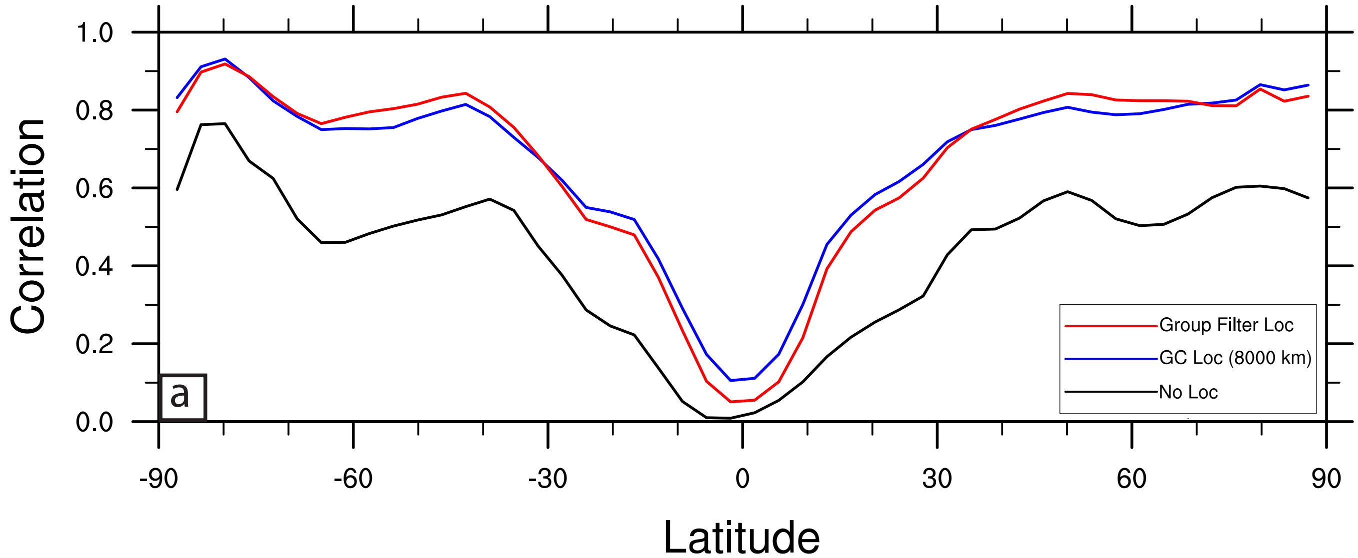



- Cycle-mean SS, Percentage of the 900 analysis cycles which show SS greater than 0.5

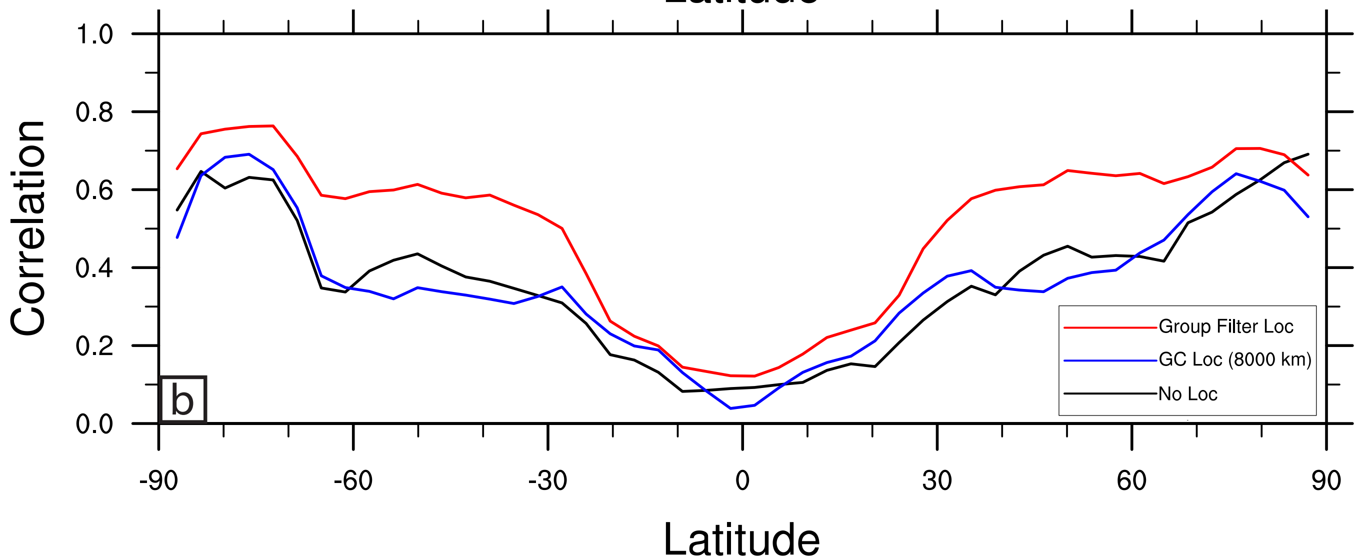


# Zonal Average Pattern Correlation

- Analysis impact



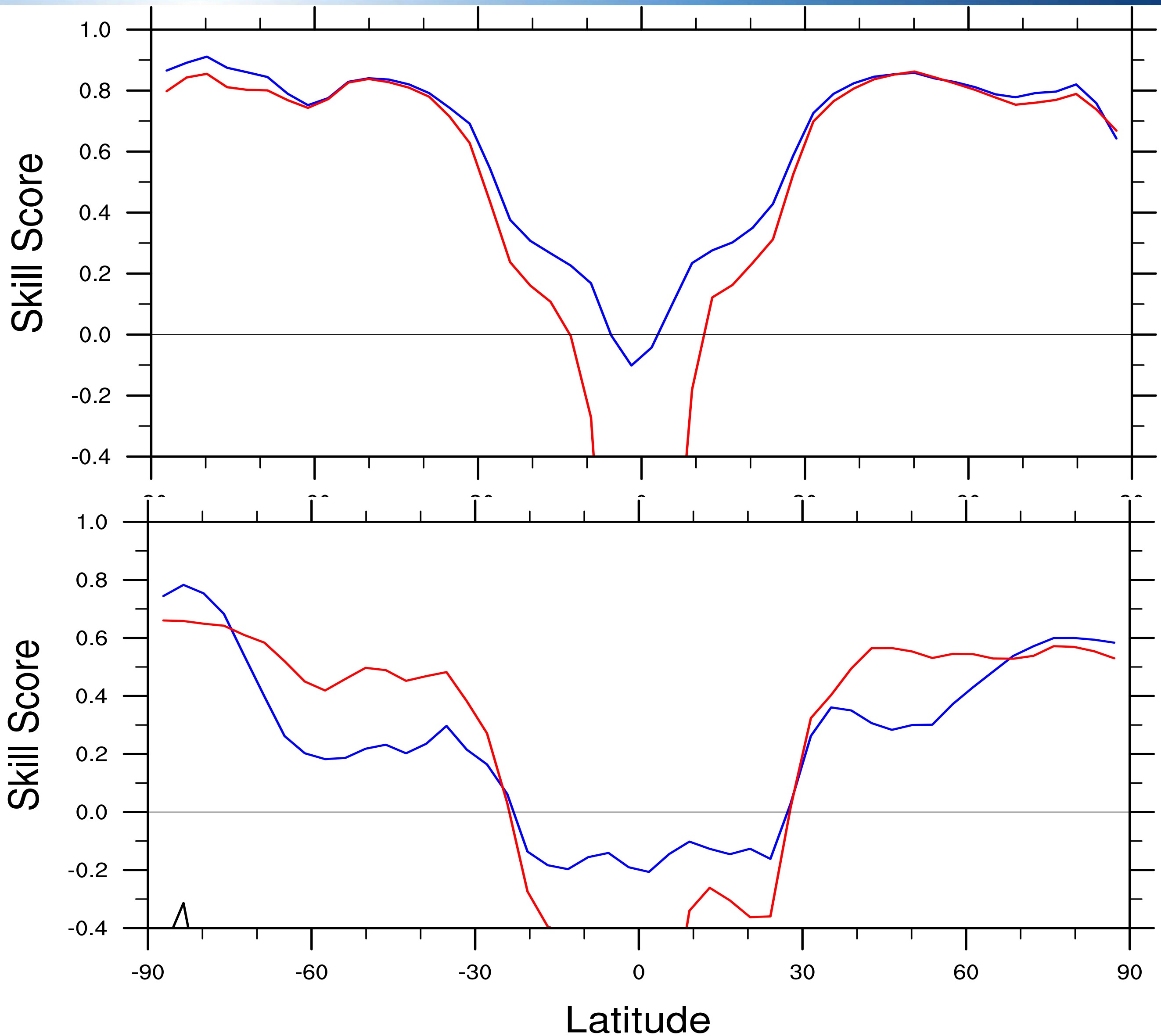

- 2-day impact



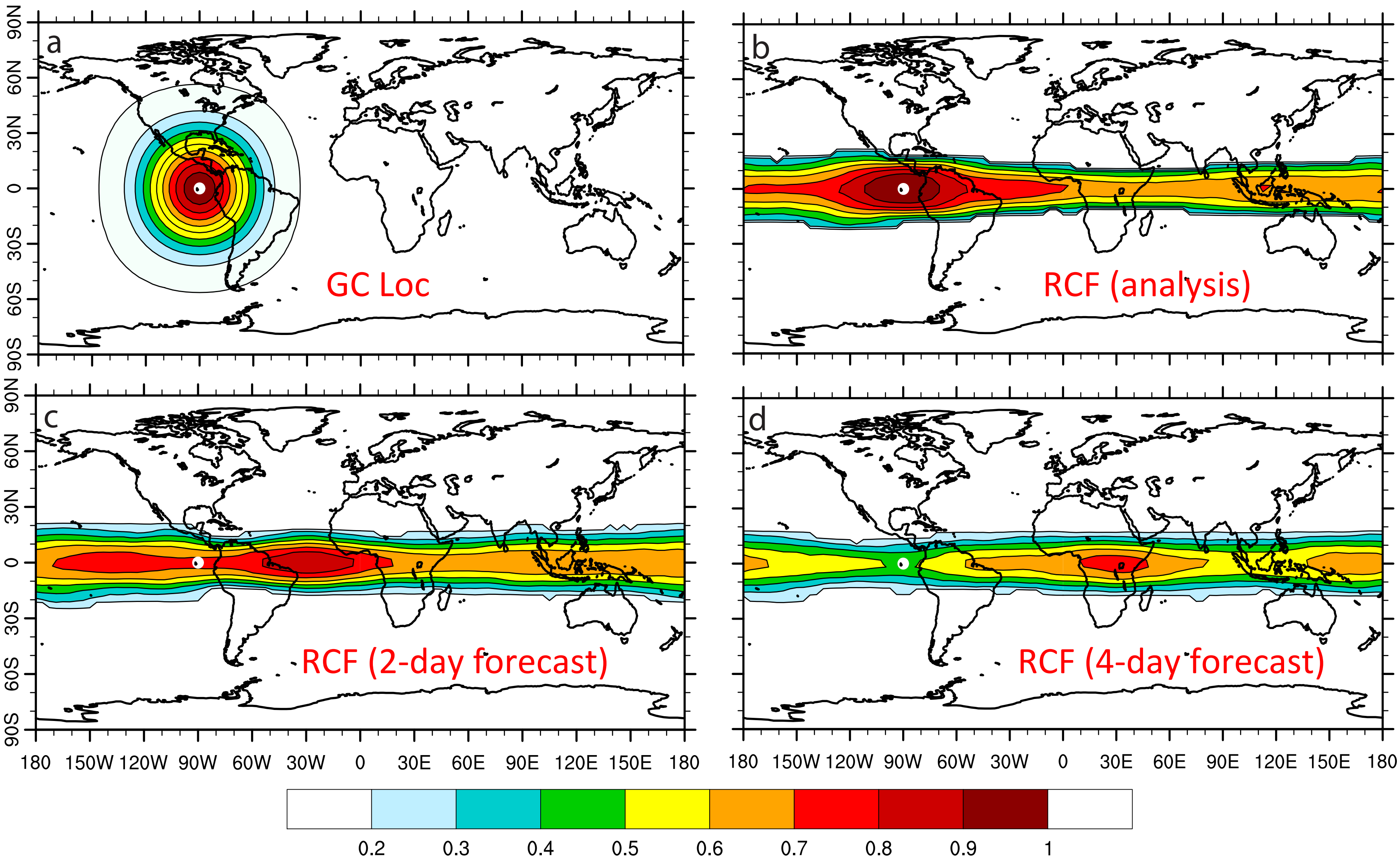
# Zonal Average SS

- Analysis impact

- 2-day impact



# Tropics Issue – RCF functions



# Tropics – Single Ob Experiment

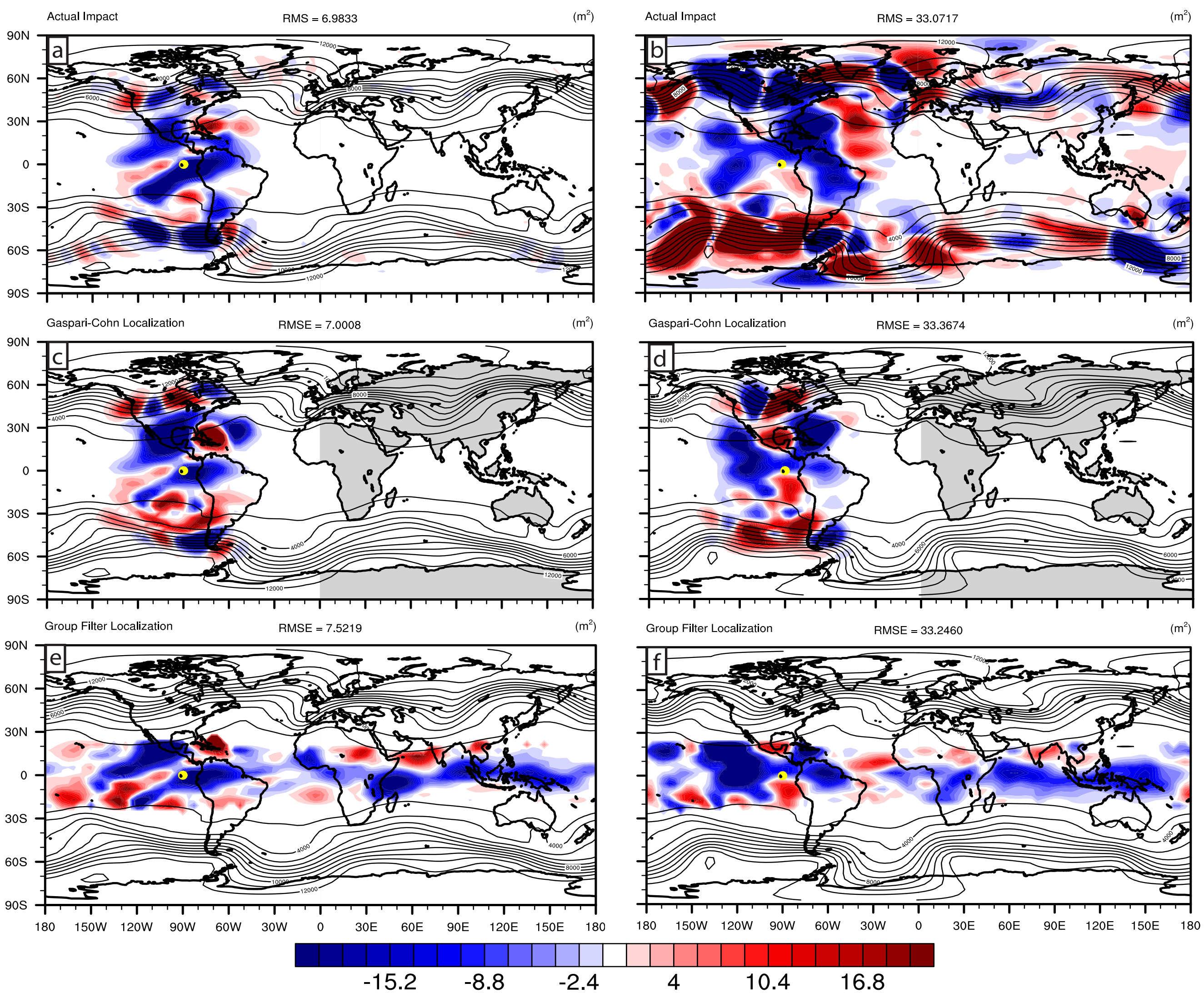
Analysis impact (t = 0)

1-day Forecast impact (t = 1)

Actual Impact

Estimate with GC localization (8000 km)

Estimate with Group Filter localization



# Conclusions

- Group Filter technique revealed underlying model dynamics
- Single-observation experiments displayed the evolving structures of the actual forecast error reduction
  - Increased magnitude and coverage in actual impact with longer lead time
- Applying RCF functions for impact estimates showed overall improvement when verified against the actual forecast error reduction, especially at longer forecast lead times.
  - Biggest contribution from dynamic localization was ability to capture time-evolving component (spatial coverage, magnitude diminishing, shift)
- Noted deficiencies at impact estimates near tropics
  - There is an inherentl relationship between the localization applied at assimilation time and the localization used for the impact estimate
  - Utility of adaptive localization method for ensemble impact metric depends on what localization was used during assimilation

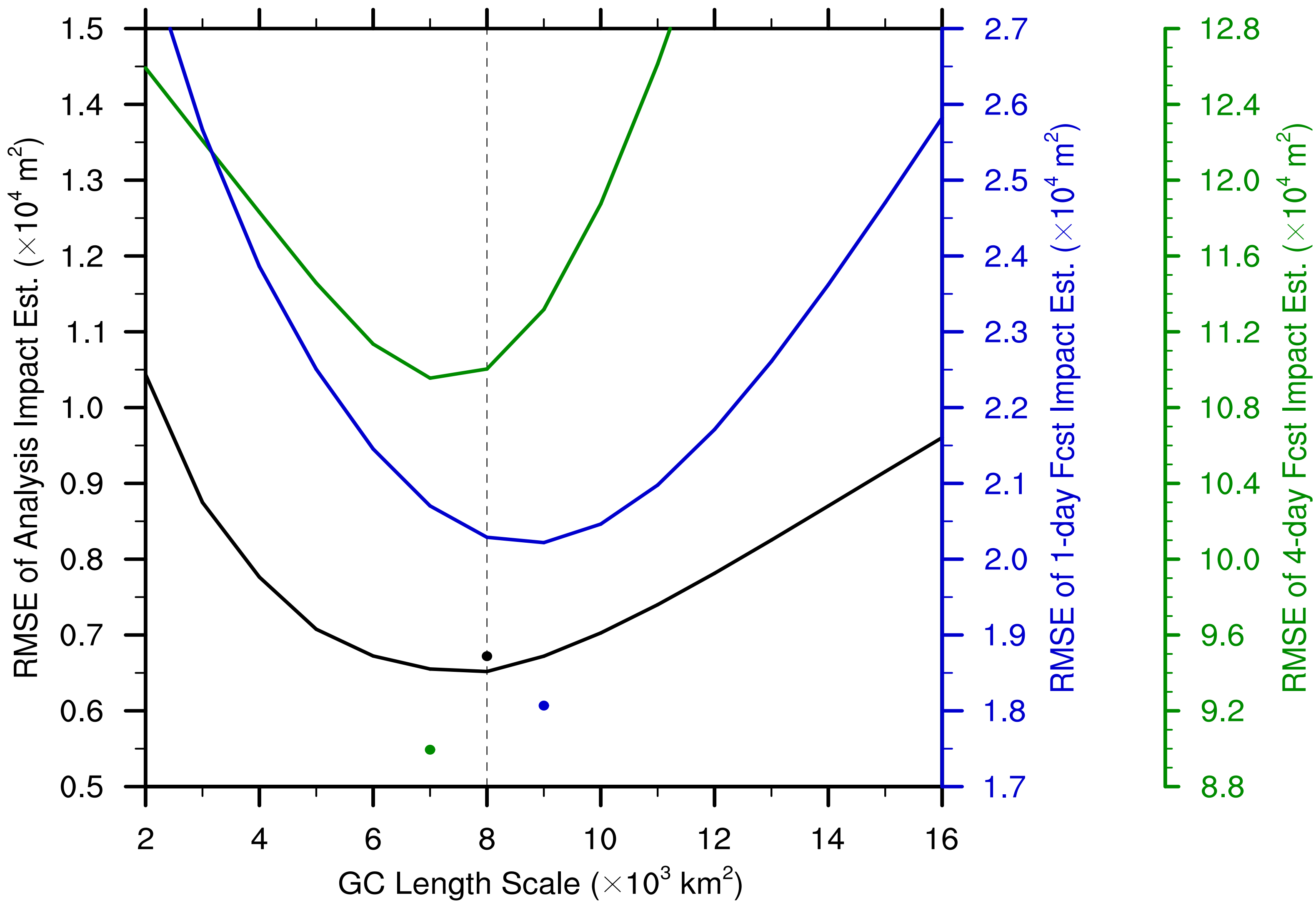
# Future Work

- Impact experiments at mesoscale
  - Dallas testbed, “Network of Networks”
  - Compare ensemble estimated impact to OSEs
  - Apply dynamic localization technique

- Explore application of other adaptive methods besides group filter

**Thank You!**
**Questions?**

# All-Obs Experiments – GC tuning



# Discussion

- Inherent relationship between proper localization choice for impact metric and the localization used during assimilation
- Derivation relies on post-analysis Kalman gain formulation

$$\mathbf{K} = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1} = (k-1)^{-1}\mathbf{X}^a\mathbf{X}^{aT}\mathbf{H}^T\mathbf{R}^{-1}$$

- However, the above is not explicitly valid when localization is applied during assimilation.

$$\mathbf{K} = \mathbf{P}^a_{loc}\mathbf{H}^T\mathbf{R}^{-1} \qquad \mathbf{P}^a_{loc} = (\mathbf{I} - \mathbf{KH})\left(\rho_A \circ \mathbf{P}^b\right)$$

- So, the original formulation is valid only under the *approximation* that $\mathbf{P}^a_{loc} \approx \mathbf{P}^a = (k-1)^{-1}\mathbf{X}^a\mathbf{X}^{aT}$
- This approximation becomes poor as localization is more severe.
- So the proper choice for localization of the impact estimate is tied to the localization used during assimilation