Validation of a New Algorithm for Empirical Localization of Observations for Ensemble Kalman Filter Data Assimilation in Global and regional Atmospheric Models

Lili Lei and Jeff Anderson
NCAR*

Thanks to Tim Hoar, Nancy Collins, Kevin Rader and Glen Romine

The 6th EnKF Workshop, Buffalo NY, 20 May 2014

*The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Background

- The ensemble Kalman filter (EnKF) uses sample statistics from an ensemble model forecast to estimate flow-dependent background error covariance to determine how an observation modifies the background fields (Evensen 1994).
- Small ensembles lead to spurious correlations between observations and state variables, especially for large separations.
- Localization, a technique to 'localize' the impact of an observation to nearby state variables, reduces spurious error correlations.

Definition of Covariance Localization

Given N ensemble increments for observation y, increments for state variable x are:

$$\Delta x_n = \alpha \hat{b} \Delta y_n, \quad n = 1, \ldots, N$$

where $\hat{b}$ is a sample regression coefficient,

and α is a localization.

Definition of Covariance Localization

Given N ensemble increments for observation y, increments for state variable x are:

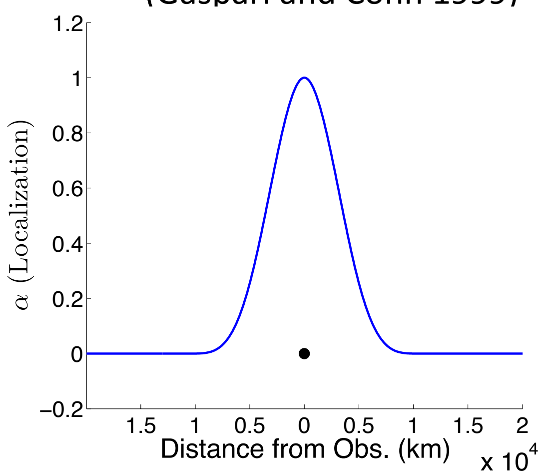
$$\Delta x_n = \alpha \hat{b} \Delta y_n, \quad n = 1, \ldots, N$$

where $\hat{b}$ is a sample regression coefficient,

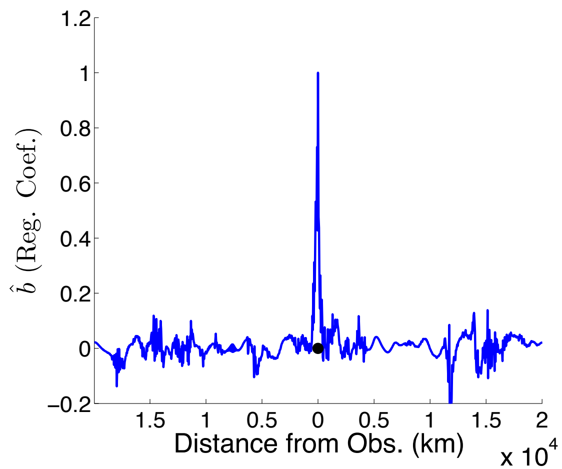
and α is a localization.

GC Localization α

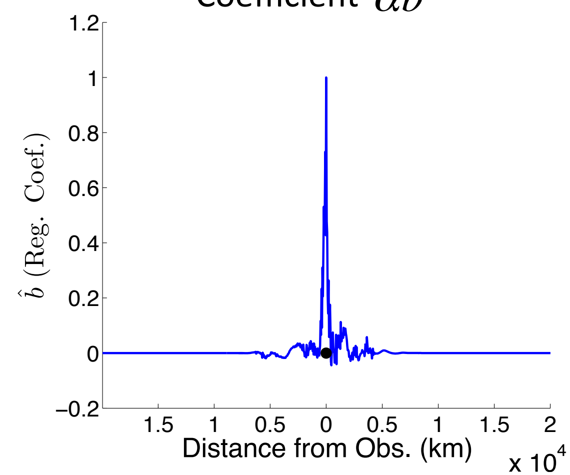
(Gaspari and Cohn 1999)



Regression Coefficient $\hat{b}$

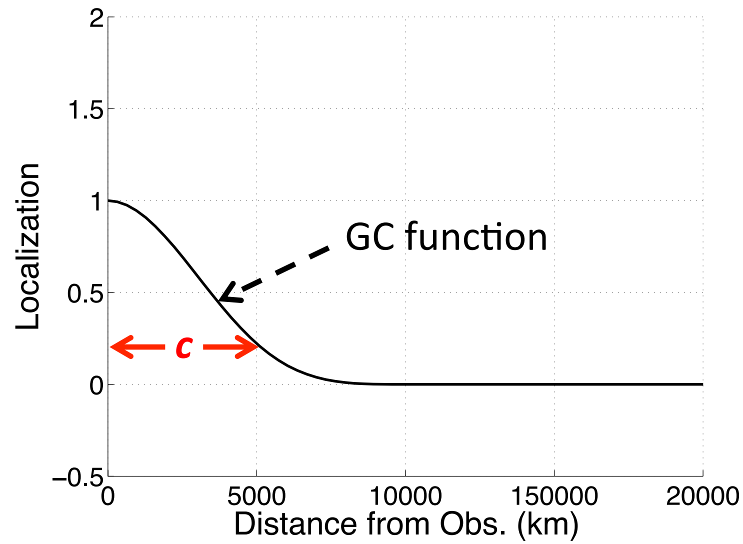


Localized Regression Coefficient $\alpha\hat{b}$



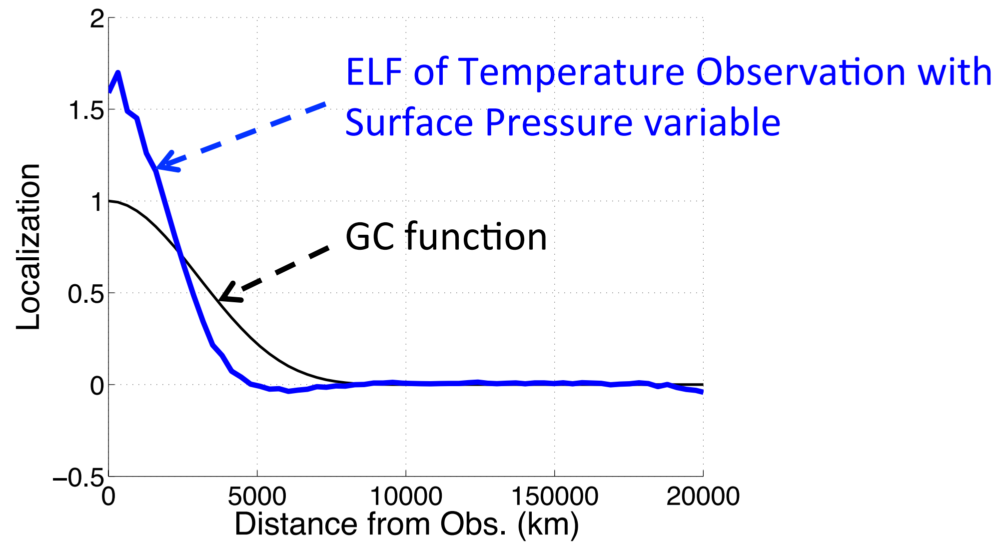
Motivation for a Generic Localization Algorithm

- *The GC function has a single parameter that defines the width of the function.* But tuning even this single parameter can be computationally expensive.
- *The GC function is approximately Gaussian.* But different localization functions are needed for:
  - different observation types (Houtekamer and Mitchell 2005, Anderson and Lei 2013)
  - different state variable kinds (Anderson 2007, 2012)
  - different times (Anderson 2007, Chen and Oliver 2010)
  - different regions (Lei and Anderson 2014).



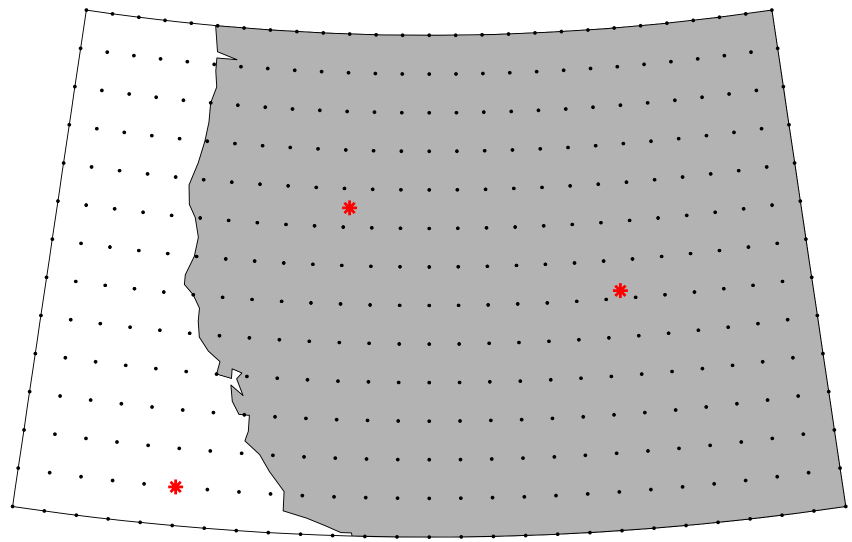
Motivation for a Generic Localization Algorithm

- *The GC function has a single parameter that defines the width of the function.*
- *The GC function is approximately Gaussian.*
- Thus a generic localization algorithm, empirical localization function (ELF), is proposed.
  - ELF provides an estimate for the localization for any possible observation type with a state variable kind (at different times and for different regions).
  - ELF makes few a priori assumptions for the shape of the localization function.
  - ELF has computational cost advantage over tuning the GC halfwidth.
  - ELF can outperform the best GC function.



Empirical Localization

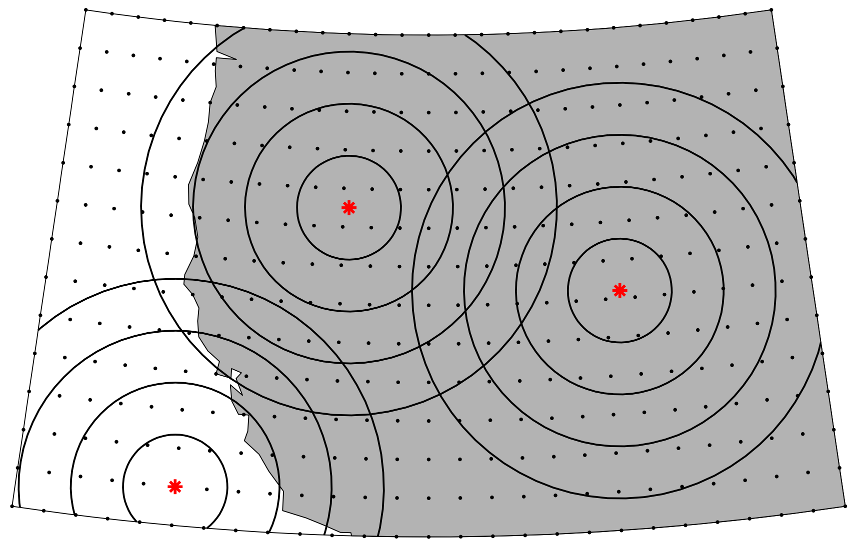
1. Compute separation between each pair of an observation and a state variable;



- Black dots: grid points.
- * Red stars: observations.

Empirical Localization

1. Compute separation between each pair of an observation and a state variable;
2. Divide the set of all pairs into subsets using the separation;

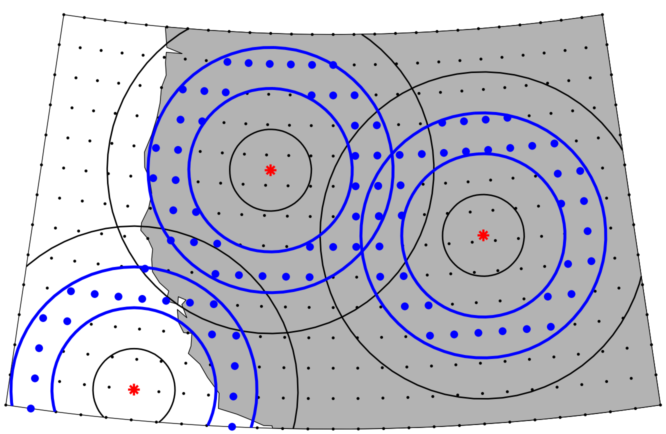


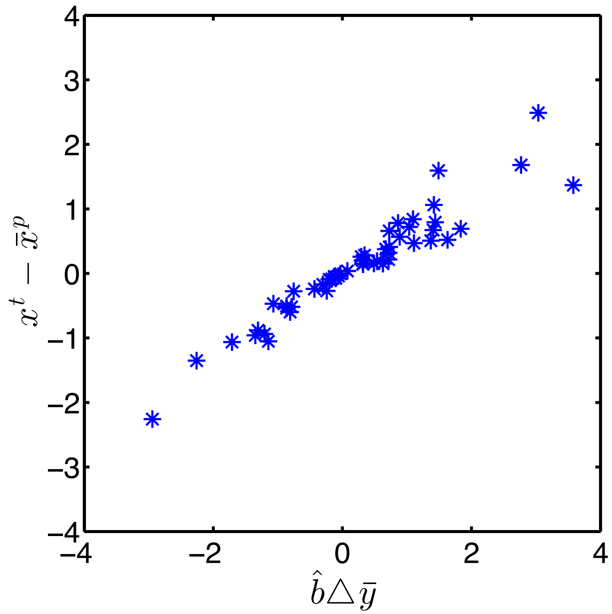
- Black dots: grid points.
- * Red stars: observations.

  Circles: distance ranges from each observation.

Empirical Localization

1. Compute separation between each pair of an observation and a state variable;
2. Divide the set of all pairs into subsets using the separation;
3. Compute the localization for each subset.





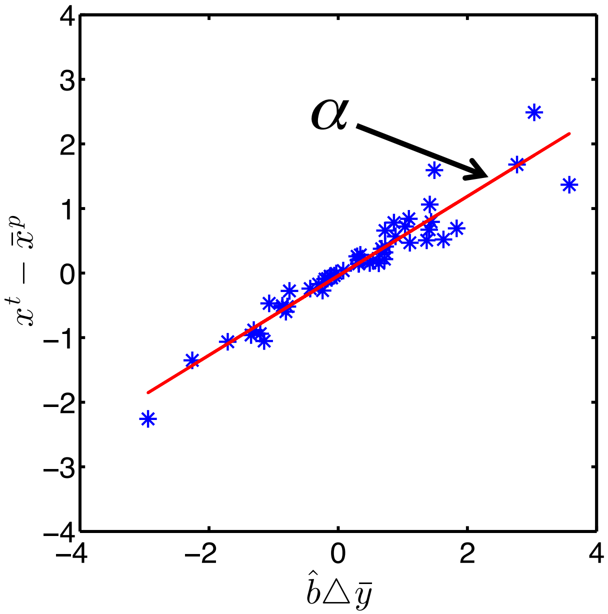
Empirical Localization

1. Compute separation between each pair of an observation and a state variable;
2. Divide the set of all pairs into subsets using the separation;
3. Compute the localization for each subset.

Localization α is computed as the slope of a least square fit.

$$\alpha = \frac{\sum_{k=1}^{K}\left(x_k^t - \bar{x}_k^p\right)\hat{b}_k \Delta\bar{y}_k}{\sum_{k=1}^{K}\left(\hat{b}_k \Delta\bar{y}_k\right)^2}$$

$$J = \sqrt{\sum_{k=1}^{K}(\bar{x}_k^u - x_k^t)^2} = \sqrt{\sum_{k=1}^{K}\left(\bar{x}_k^p + \alpha\hat{b}_k\Delta\bar{y}_k - x_k^t\right)^2}$$



Tests of the Empirical Localization

- The Community Atmospheric Model version 5 (CAM5): Vertical localization and localization for different geographic regions
- The Weather Research and Forecasting Model (WRF): Localization for regions with and without precipitation

ELF in CAM

Experimental Design

Conduct OSSEs in DART/CAM system (Raeder et al. 2012).

CAM5 model:

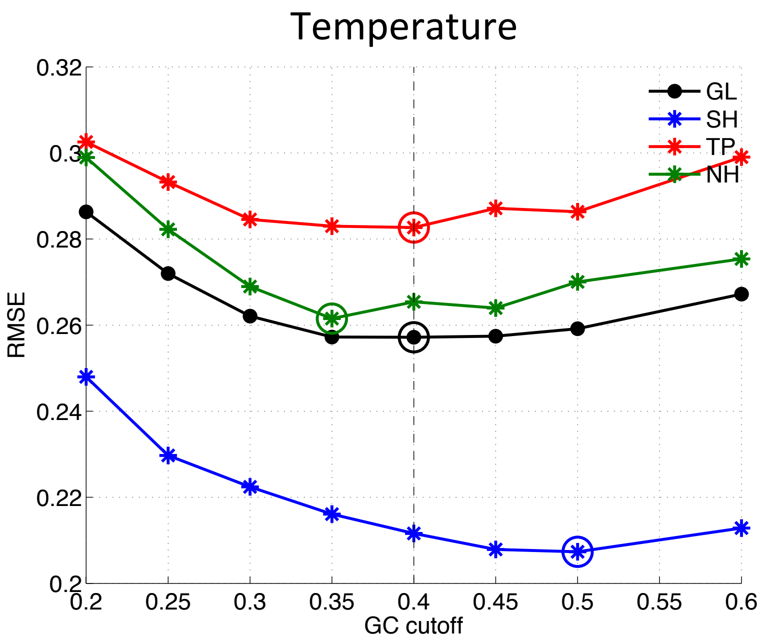
- Atmospheric component of the Community Earth System Model version 1 (CESM1; Gent et al. 2011)
- Finite volume grid with approximately 2° resolution (94x144) and 30 vertical levels
- Default configuration of the Atmospheric Model Intercomparison Project (AMIP; Gates 1992) protocol

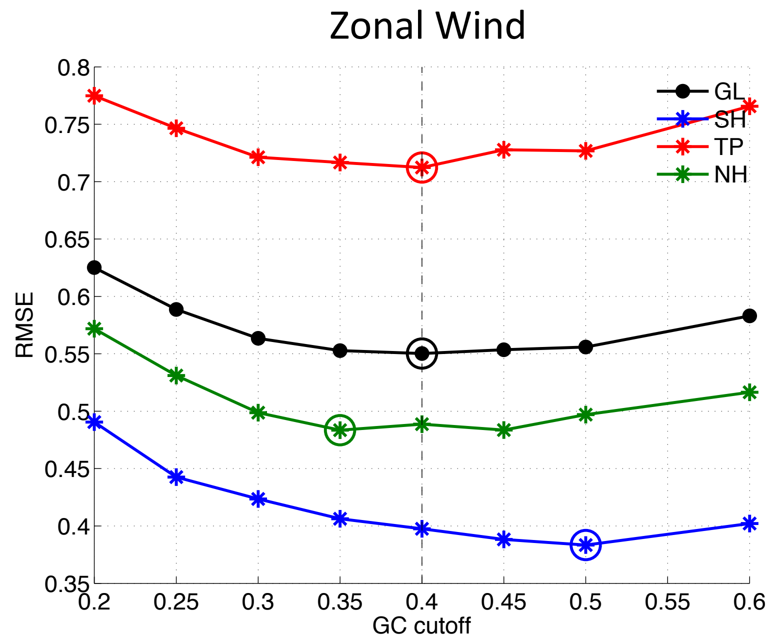
Data assimilation system:

- Ensemble adjustment Kalman filter (EAKF; Anderson 2001) in DART
- Spatially- and temporally-varying state space adaptive inflation (Anderson 2009)
- GC localization as the default

ELF in CAM

RMSE for Different GC Halfwidths





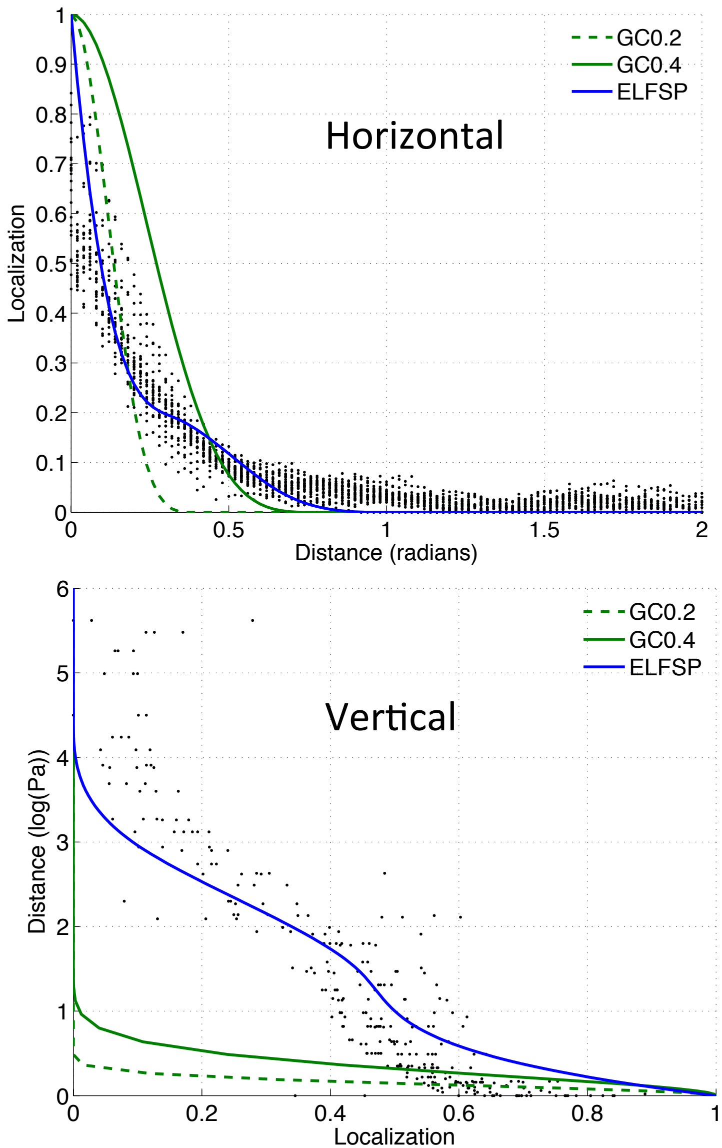
RMSEs for temperature and zonal wind are averaged globally (GL), in the southern hemisphere (SH), tropics (TP) and northern hemisphere (NH).

GC0.4 has smallest globally averaged RMSE, so 0.4 is chosen as the best halfwidth.

Some RMSEs computed for SH, TP and NH separately are smallest for other halfwidths; tuning the GC halfwidth is complex.

ELF in CAM

Horizontal and Vertical Empirical Localization Functions



Empirical localizations (black dots) are computed separately for temperature, zonal and meridional winds at ten levels (30 dots per separation).

A z-test is used to assess the significance of the empirical localization.

A cubic spline (blue line) is applied to the empirical localization to produce the final localization function (ELFSP).

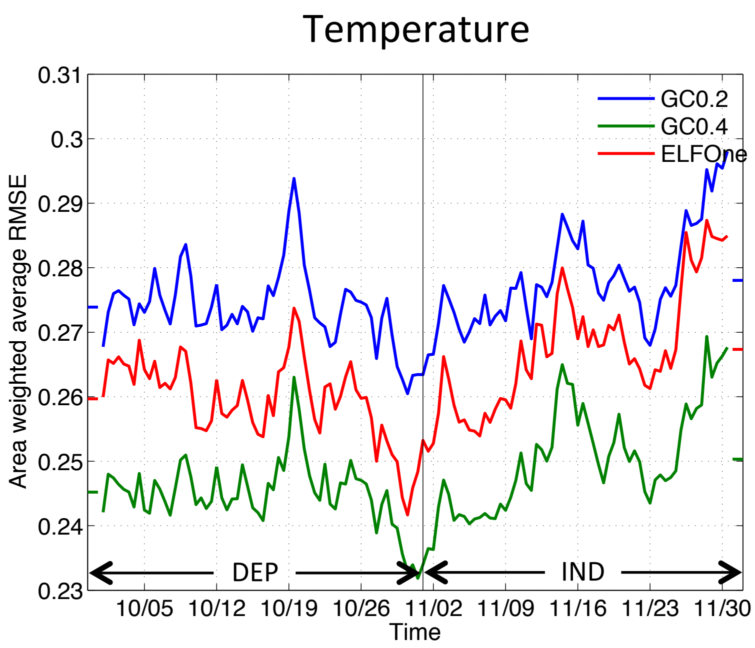
The horizontal ELFSP is smaller than the GC0.2 and GC0.4 at small separations and has a wider tail than GC0.2 and GC0.4.

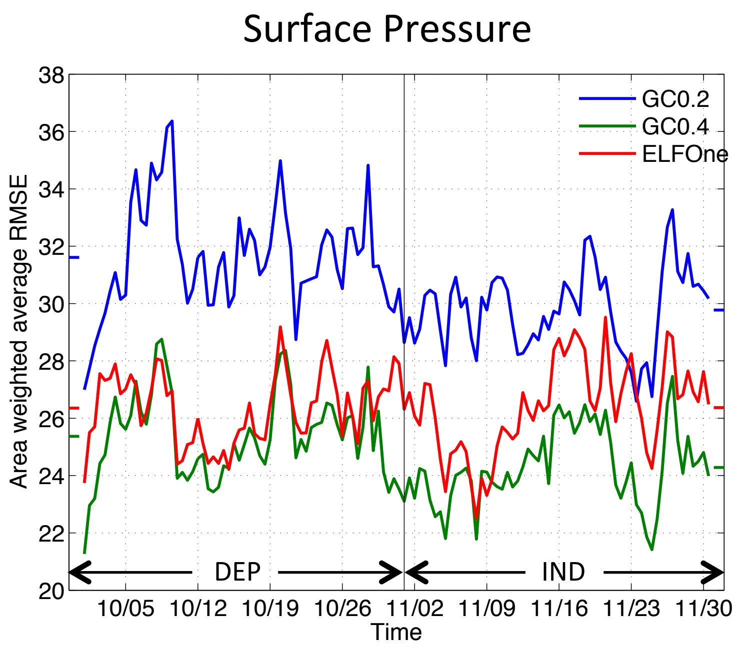
The vertical ELFSP is much broader than the GC0.2 and GC0.4.

The horizontal and vertical ELFSPs are used in a subsequent OSSE (ELFOne).

ELF in CAM

Global Average RMSE for GC0.2, GC0.4 and ELFOne



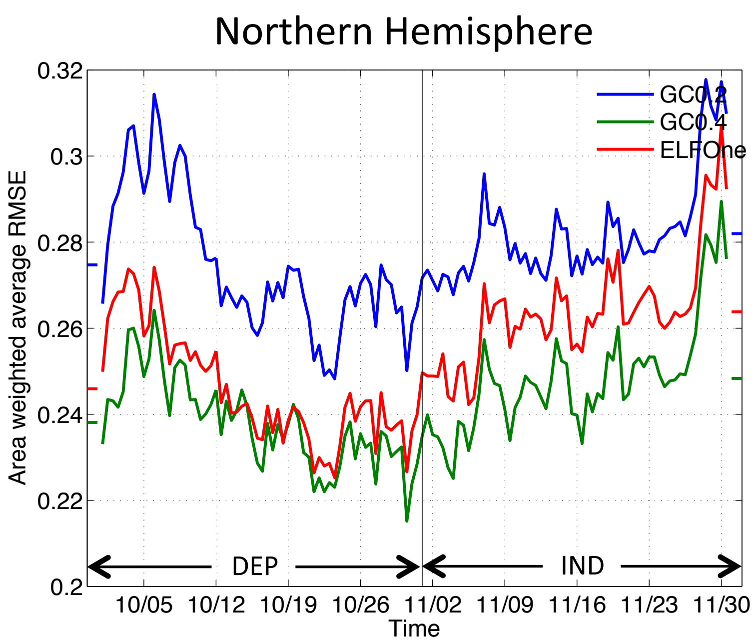


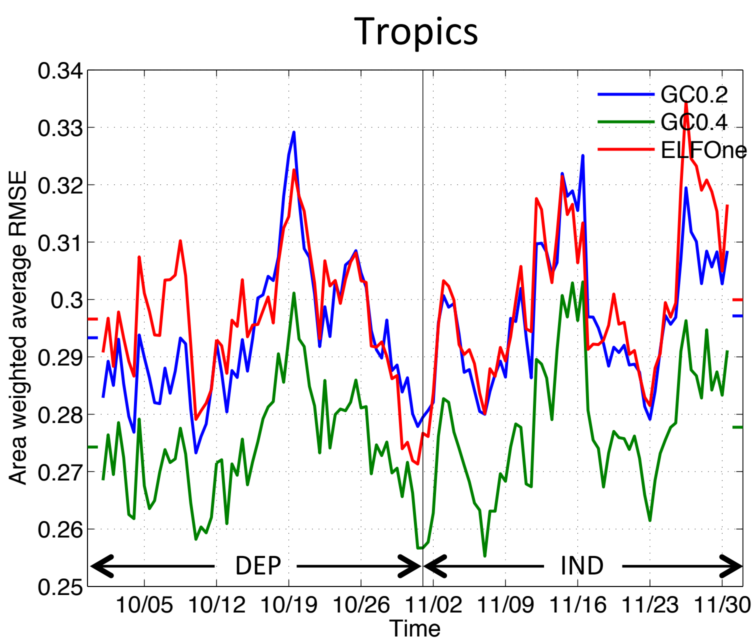
ELFOne has smaller temperature RMSE than GC0.2, but larger RMSE than GC0.4, the best GC.

ELFOne has smaller surface pressure RMSE than GC0.2, and slightly larger RMSE than GC0.4.

ELF in CAM

Temperature RMSE Averaged in NH and TP





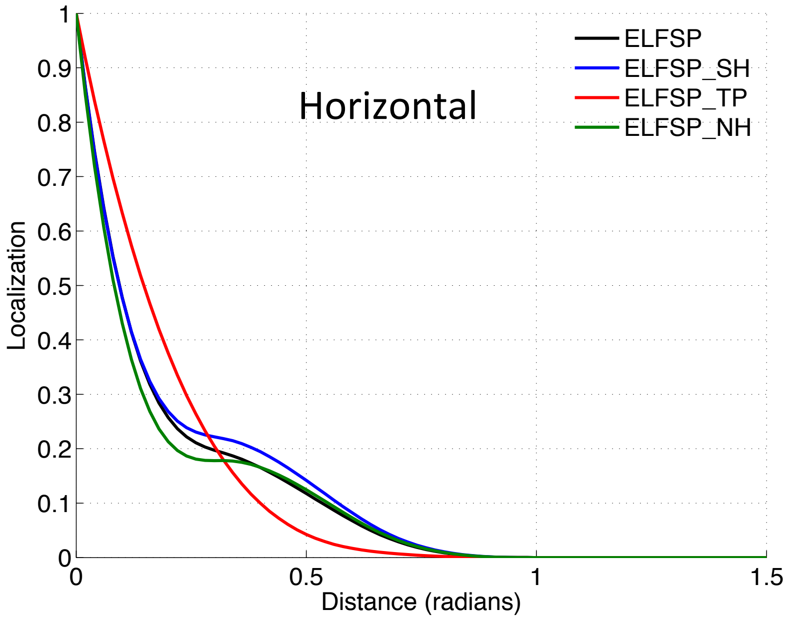
ELFOne has smaller temperature RMSE than GC0.2 in NH and SH.

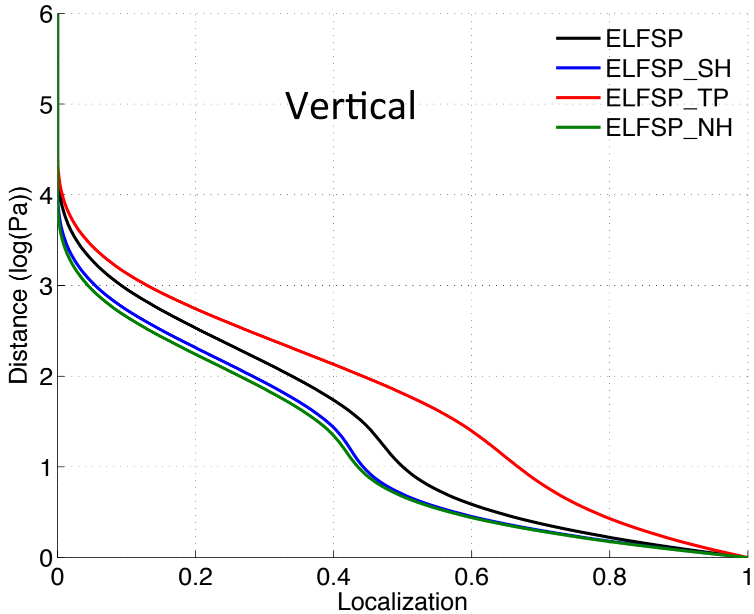
ELFOne has larger temperature RMSE than GC0.2 in TP.

Improvements of ELFOne over GC0.2 are mainly in SH and NH.

ELF in CAM

Horizontal and Vertical ELFs Varying by Region





Horizontal and vertical ELFSPs are computed for the SH, TP and NH separately.

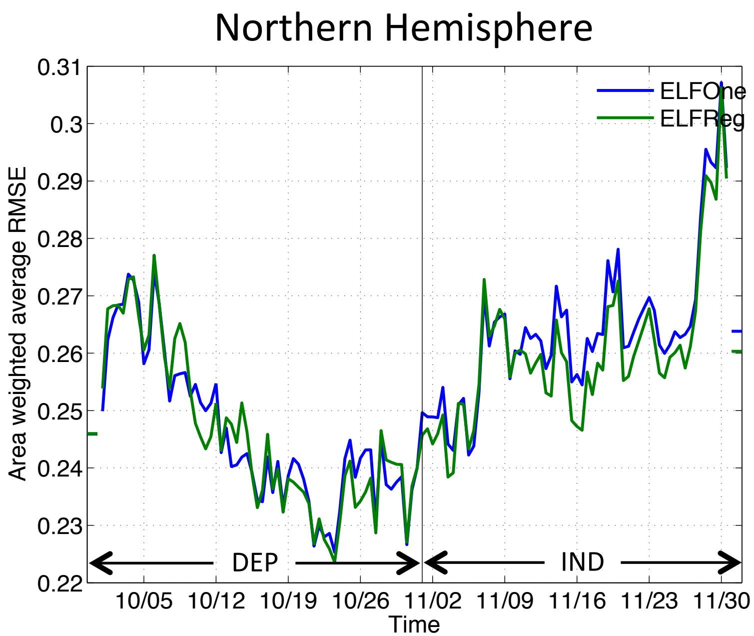
The horizontal ELFSP_SH and ELFSP_NH have similar shape to the global ELFSP. The horizontal ELFSP_TP has a more compact tail than the ELFSP, ELFSP_SH and ELFSP_NH.

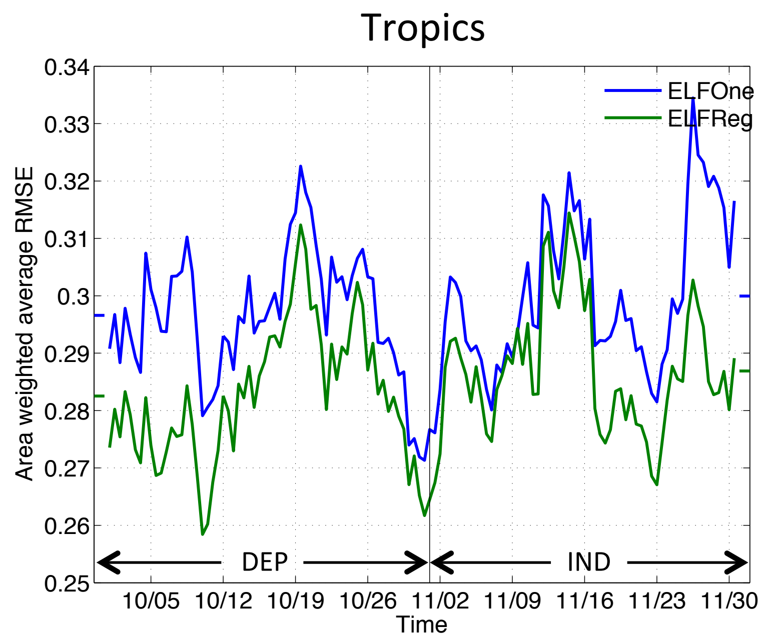
The vertical ELFSP_SH and ELFSP_NH are similar with smaller magnitude than the global ELFSP. The vertical ELFSP_TP is broader than the global ELFSP.

Horizontal and vertical ELFSPs varying by region are used in a subsequent OSSE (ELFReg).

ELF in CAM

Temperature RMSE Averaged in NH and TP





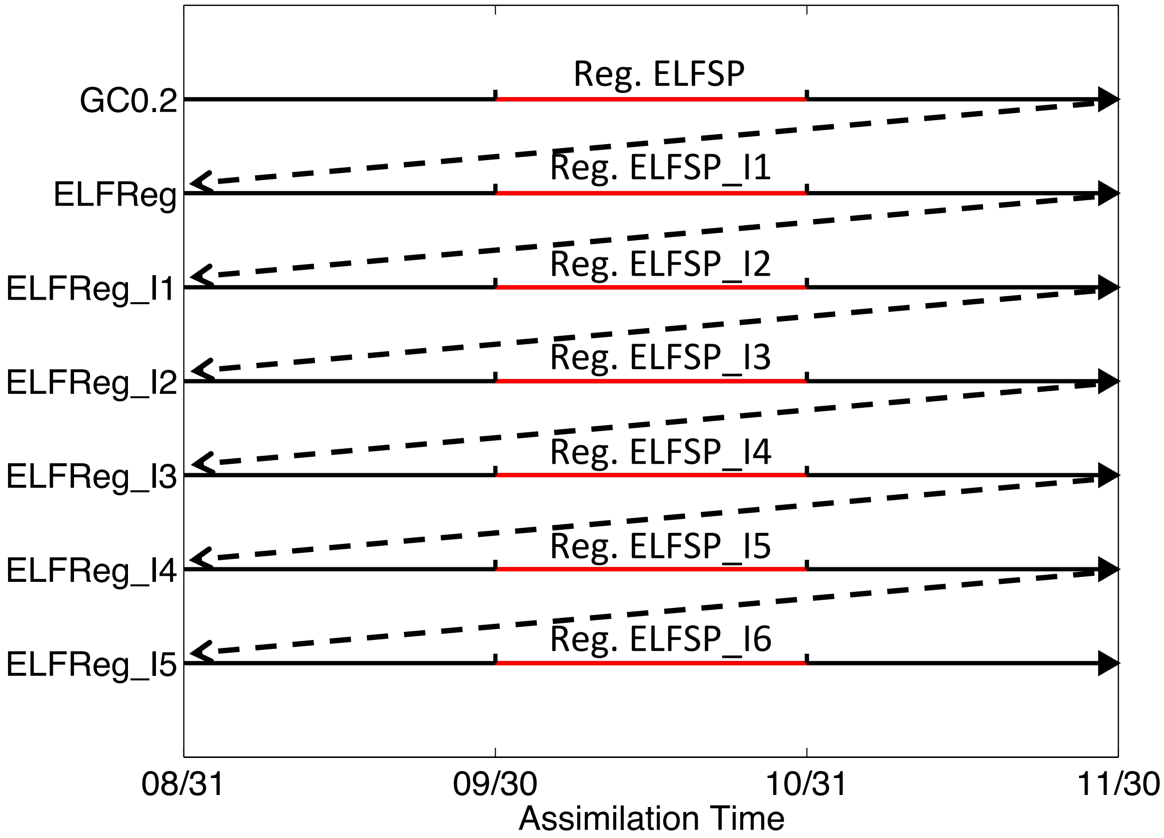
ELFReg has slightly smaller temperature RMSE than ELFOne in NH and SH.

ELFReg has smaller temperature RMSE than ELFOne in TP.

ELFReg has smaller globally averaged RMSE than ELFOne.

ELF in CAM

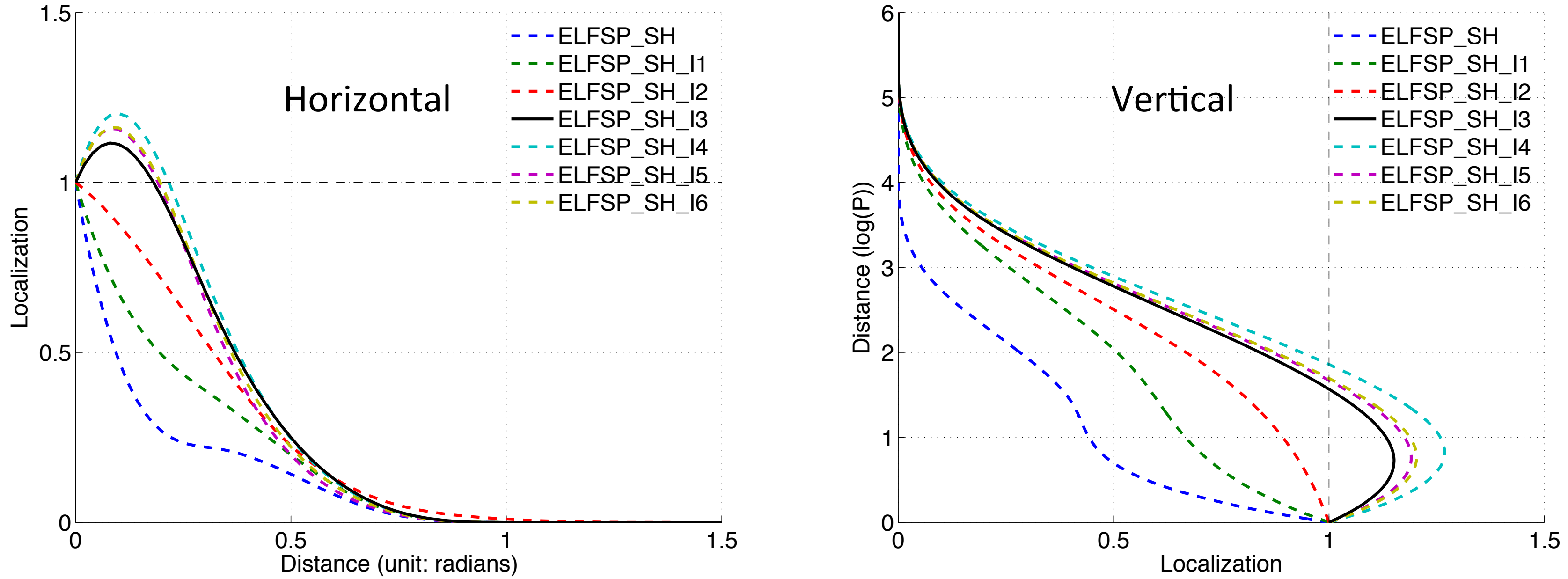
Convergence of the ELF with Cubic Spline Fit (ELFSP)



Five OSSEs (ELFReg_I#, #=1,…,5) are conducted iteratively. Each OSSE uses the regional ELFSPs computed from the output of the previous OSSE.

ELF in CAM

Convergence of the ELFSP: SH Example



The ELFSP_SHs becomes larger with iterations.

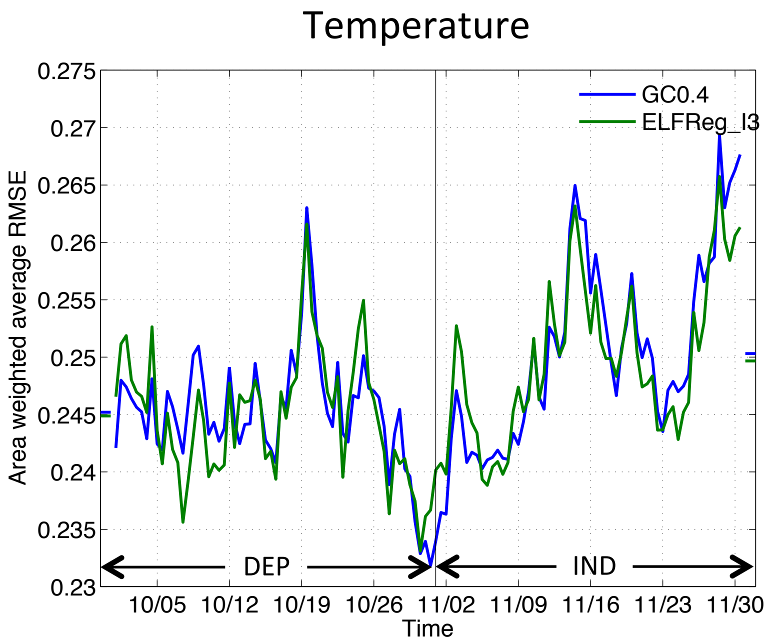
The ELFSP_SHs appear to have mostly converged after 3 iterations.

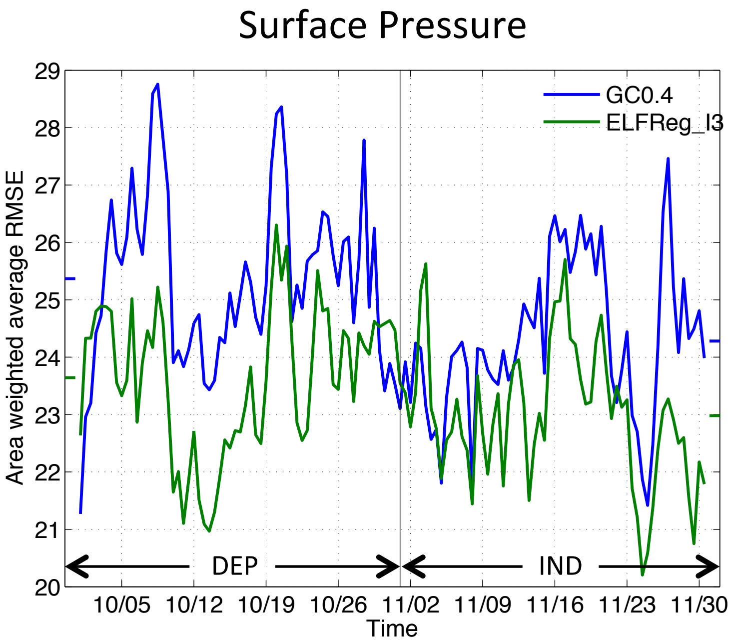
The ELFSP_SHs from iterations 3 to 6 are larger than 1.0 at small separations. This indicates insufficient spread and the empirical localization acts as an inflation.

Empirical localization values larger than 1.0 are set to 1.0 when used in an OSSE.

ELF in CAM

Global Average RMSE for GC0.4 and ELFReg_I3



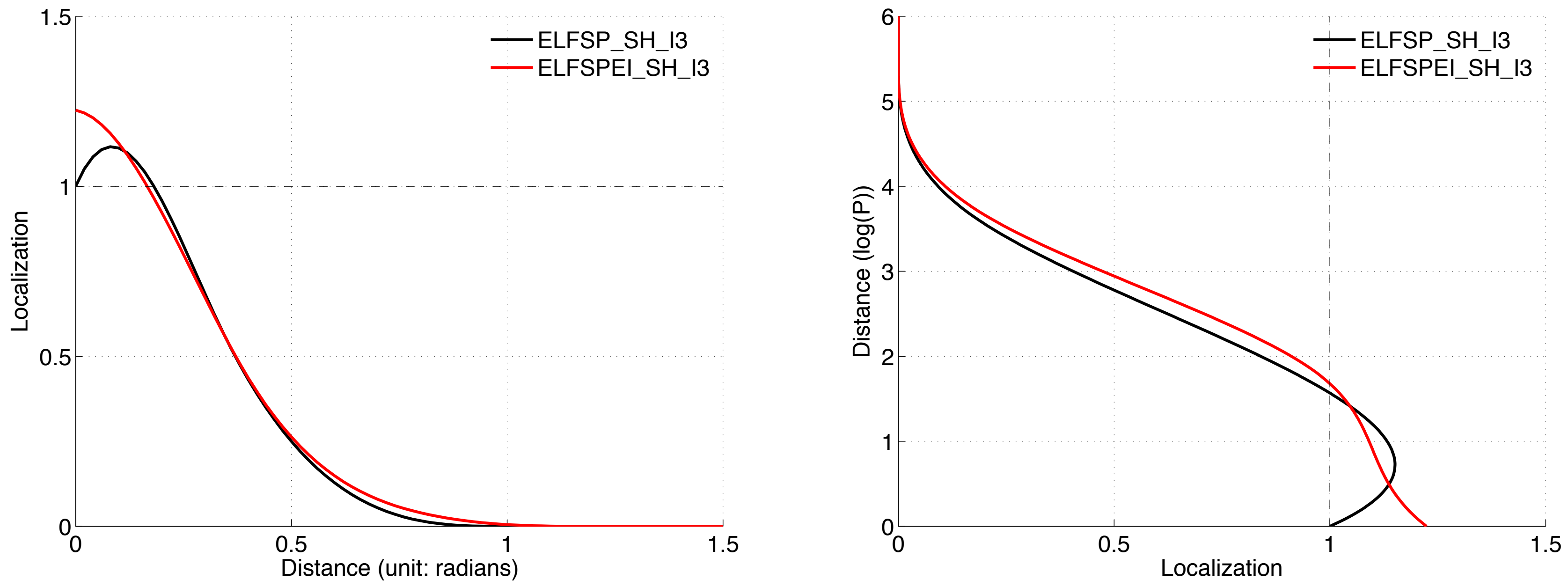


ELFReg_I3 produces slightly smaller temperature RMSE than GC0.4.

ELFReg_I3 has significantly smaller surface pressure RMSE than GC0.4

ELF in CAM

ELFSPs with Empirical Inflation

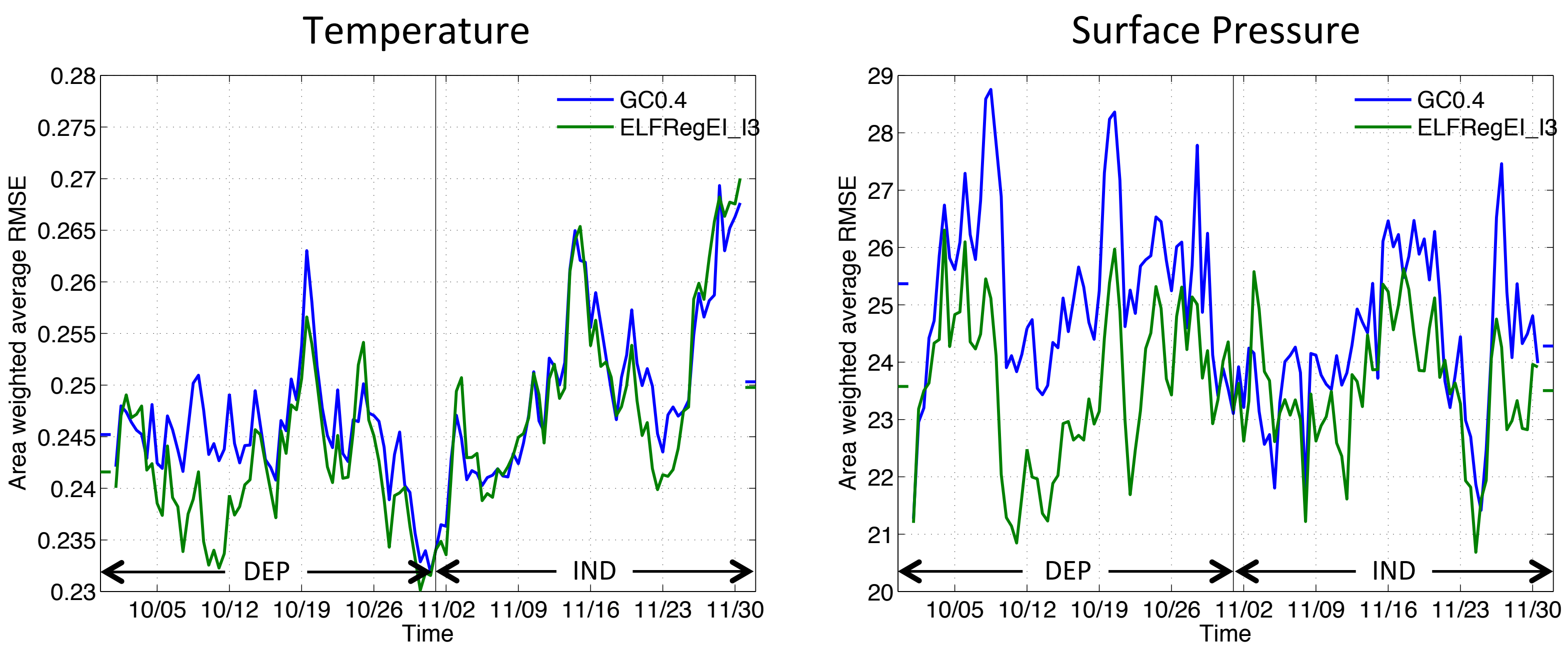


The cubic spline fit of ELF with empirical inflation (ELFSPEI) has values larger than 1.0 at small separations.

Horizontal and vertical ELFSPEIs are used in a subsequent OSSE (ELFRegEI).

ELF in CAM

Global average RMSE for GC0.4 and ELFRegEI_I3



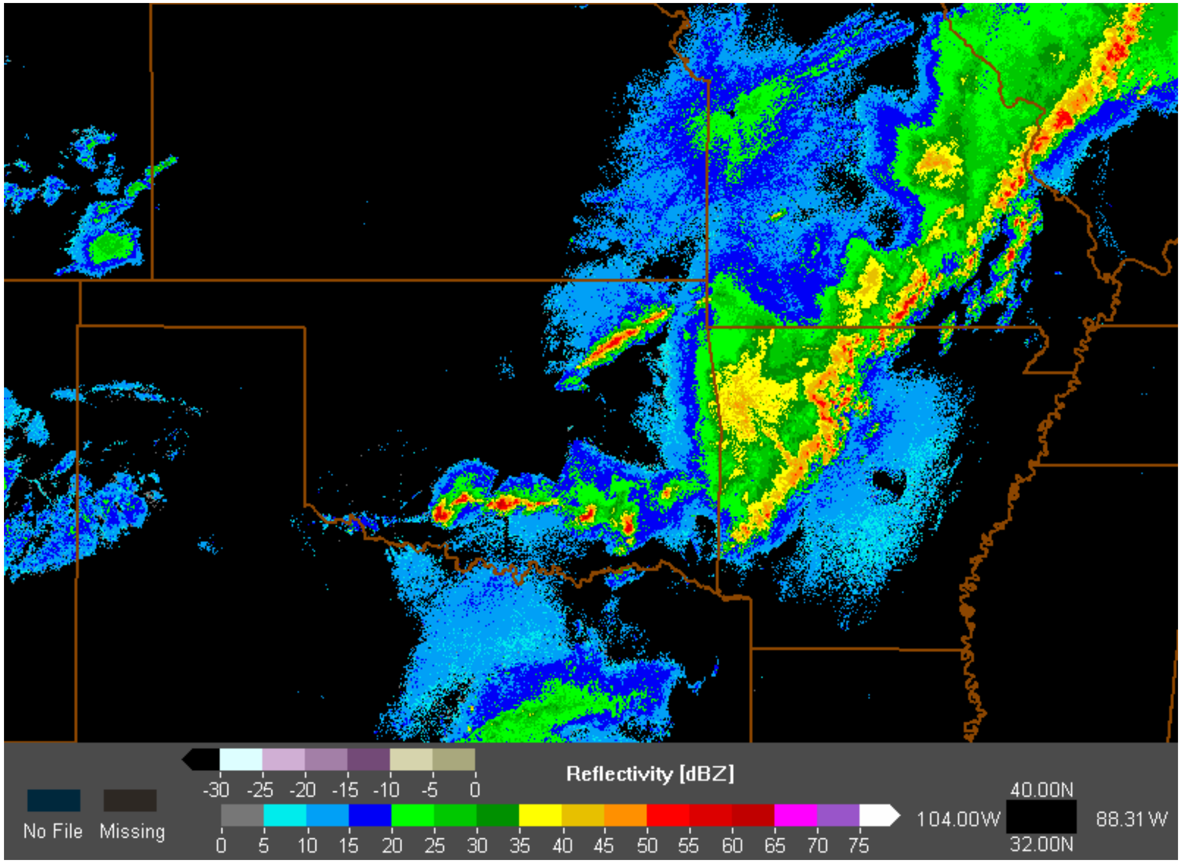
ELFRegEI_I3 produces smaller temperature RMSE than GC0.4.

ELFRegEI_I3 has significantly smaller surface pressure RMSE than GC0.4

Tests of the Empirical Localization

- The Community Atmospheric Model version 5 (CAM5): Vertical localization and localization for different geographic regions
- The Weather Research and Forecasting Model (WRF): Localization for regions with and without precipitation

Is different localization needed for different weather?



(http://nmq.ou.edu/applications/qvs_2d_maps.html)

Experimental Design

Conduct OSSEs in DART/WRF system.

WRF model V3.3.1 :

- CONUS domain with horizontal grid spacing 15 km, 40 vertical layers and model top at 50 hPa
- Model physics: RRTMG long wave and short wave radiation schemes, Thompson 2-moment microphysics scheme, Noah land surface model, MYJ PBL scheme, and Tiedtke cumulus scheme

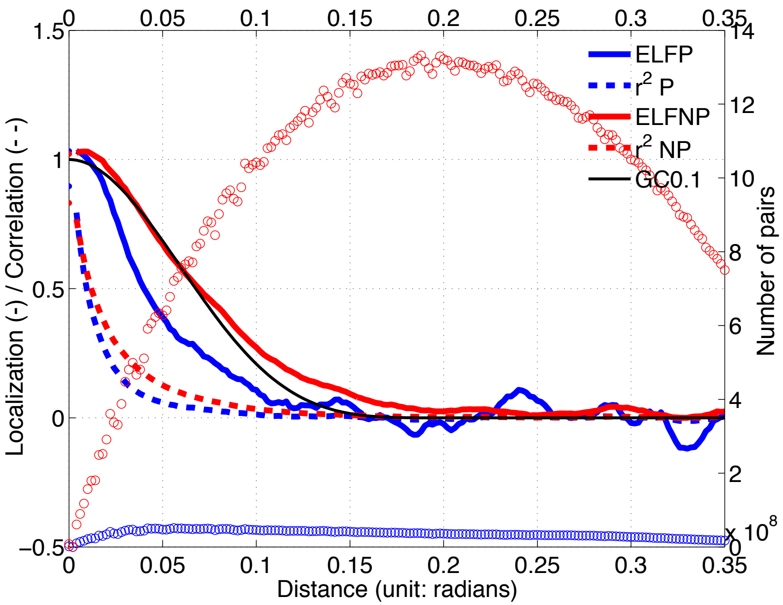
Data assimilation system:

- EAKF in DART
- Spatially- and temporally-varying state space adaptive inflation
- GC localization of halfwidth 0.1 radians as the default

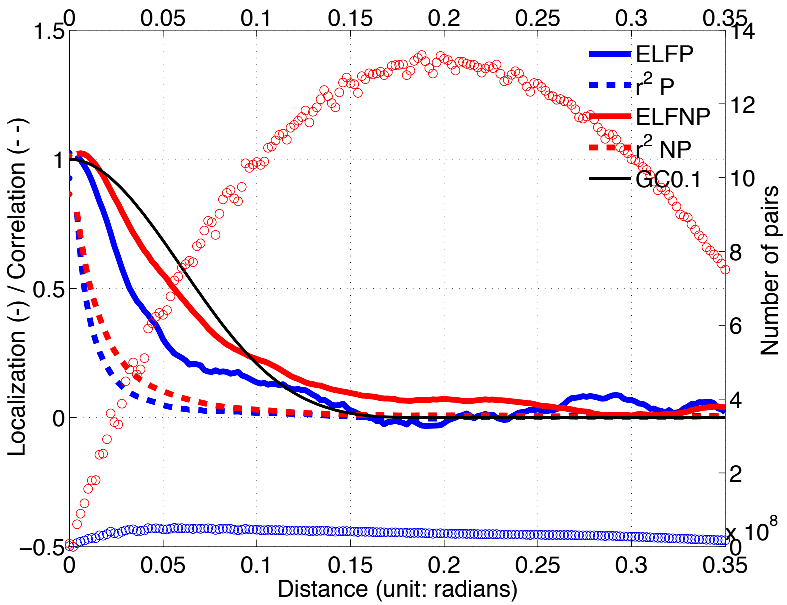
ELF in WRF

Horizontal Empirical Localization Functions

Temperature



U-wind



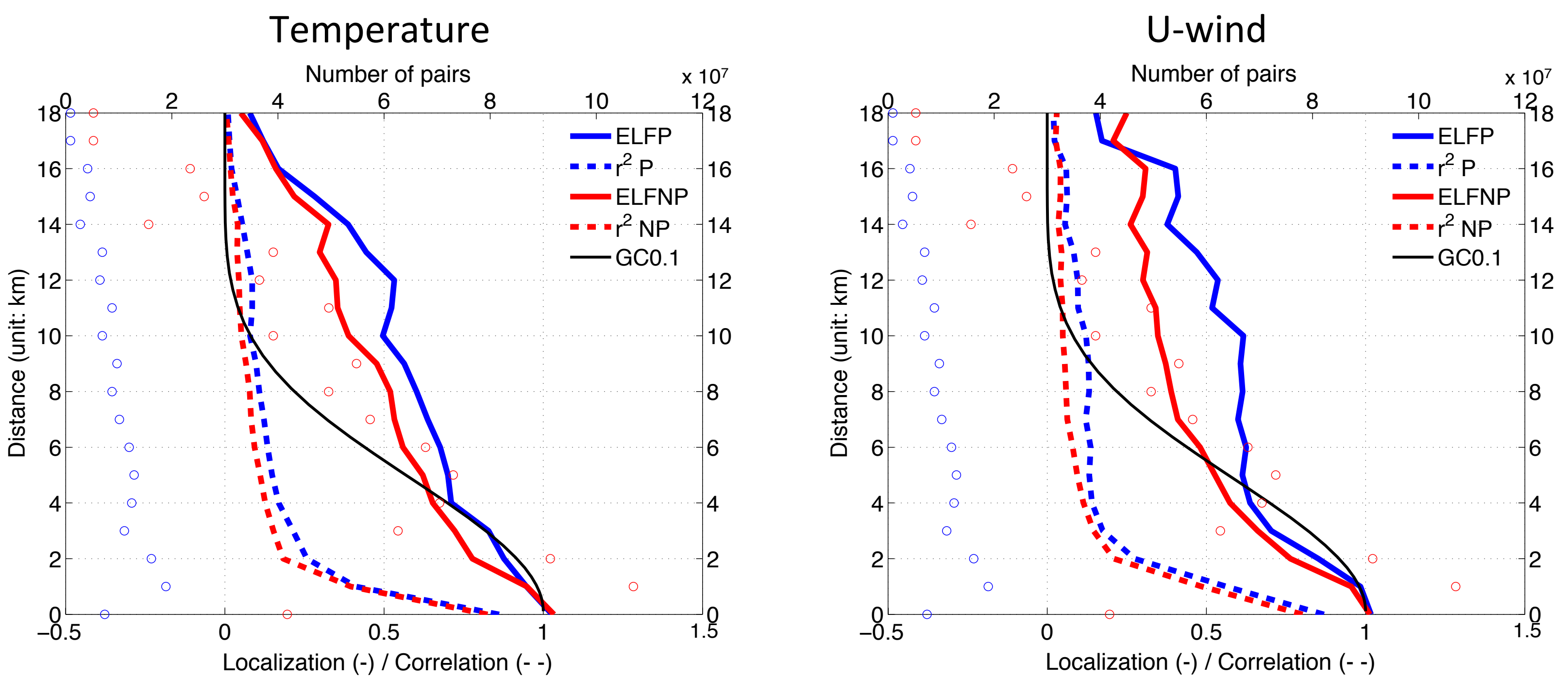
The ELFs for non-precipitating regions (ELFNP) have similar shape to GC0.1, but ELFNP of u-wind is smaller than GC0.1 for small separations.

The ELFs for precipitating regions (ELFP) are narrower than GC0.1 and ELFNP.

The correlation coefficient of ELF for precipitating regions decreases faster than for non-precipitating regions.

ELF in WRF

Vertical Empirical Localization Functions



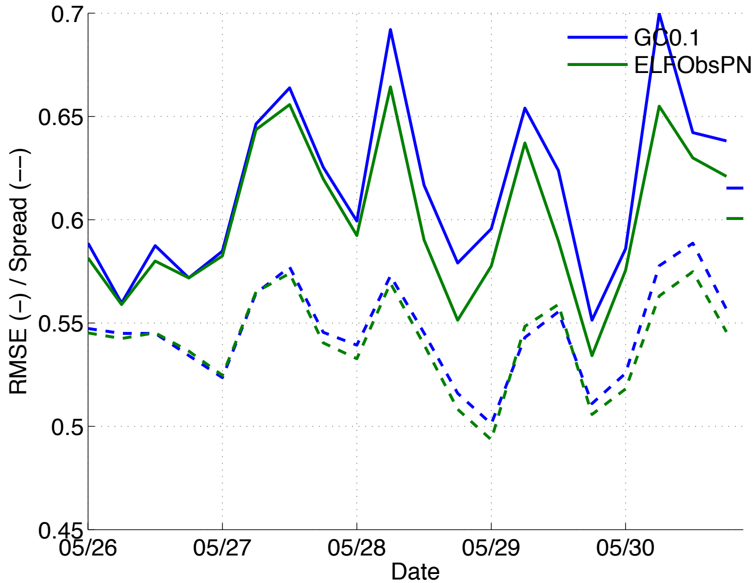
The vertical ELFs for precipitating regions generally have larger localizations.

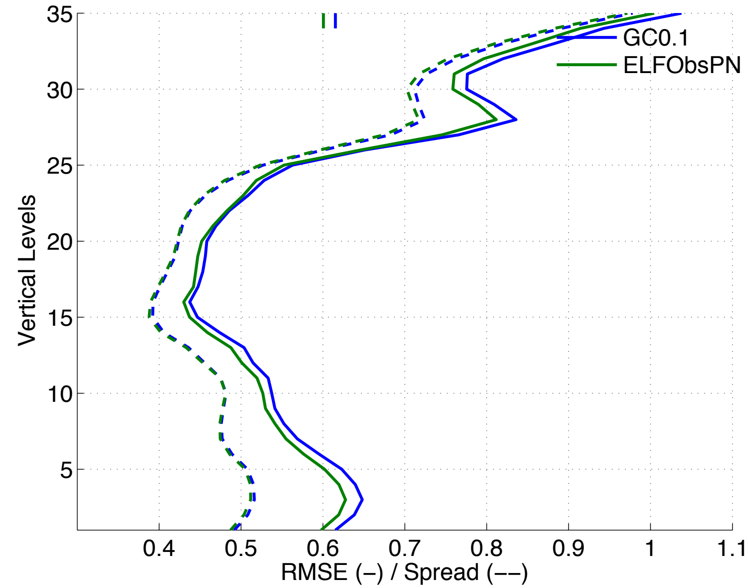
The vertical ELFP of temperature decreases more quickly with height than for u- and v-winds between 4 and 10 km.

The correlation coefficient of ELF for precipitating regions is larger.

ELF in WRF

Average Temperature RMSE for GC0.1 and ELF





ELF has smaller RMSE of temperature, u- and v-winds than GC0.1, the approximately optimal GC case.

The improvements produced by the ELFs occur in nearly every model level .

Conclusions

- The empirical localization algorithm can automatically provide an estimate of the localization function and does not require empirical tuning of the localization scale.
- It can compute an appropriate localization function for any potential observation type and kind of state variable, for different geographic regions and weather.
- It plays the role of empirical inflation when needed.
- The empirical localization function generally outperforms the best GC localization in CAM and WRF.