Toward improved LETKF assimilation of non-local and dense observation by direct covariance localization in model space

Daisuke Hotta
Meteorological Research Institute,
Japan Meteorological Agency (MRI-JMA)

6th International Symposium on Data Assimilation
9 March 2018
Munich, Germany

Many thanks: Yoichiro Ota (JMA),
Takemasa Miyoshi, Shunji Kotsuki, Guo-Yuan Lien (RIKEN/AICS),
Kosuke Ito (Univ. Ryukyus), Eugenia Kalnay (UMD), Craig Bishop (NRL)

1. Motivation

- The number of observations available for NWP has been steadily increasing:
  - O(10^4) pre-satellite era (~1990s)
  - O(10^5) with Microwave sounders (1990s-2010s)
  - O(10^6) with hyperspectral sounders (AIRS and IASI) (2010s ~)

Most new data are remotely-sensed **non-local** observations.

→ Challenge: to extract as much ***information*** as possible from **dense and non-local observations**

- Important question: How much information can a DA system extract from observations?
- → One way to quantify this: **Degrees of Freedom for Signal (DFS, or *information content*).**

2. What is DFS? (1/2)

- Defined as the trace of the "influence matrix" tr(**S**) = tr(**HK**) = $\sum_i \partial y^a_i / \partial y^o_i$
- Two ways to interpret:
  1. Analysis sensitivity to observations measured in obs space.
  2. The amount of information that the analysis extracts from observations.

Simple illustrative examples:

- **Forecast-Forecast cycle:** analysis is always the same as the background.
  - $\mathbf{y^a} \equiv \mathbf{y^b}$ → **S** is null, DFS= tr(**S**) = 0 (**0% information from obs.**)
- **Direct Insertion:** background is completely replaced by the obs.
  - $\mathbf{y^a} \equiv \mathbf{y^o}$ → **S** is identity, DFS = tr(**S**) = #obs
  - DFS per obs = 1 (**100% information comes from obs.**)

2. What is DFS? (2/2)

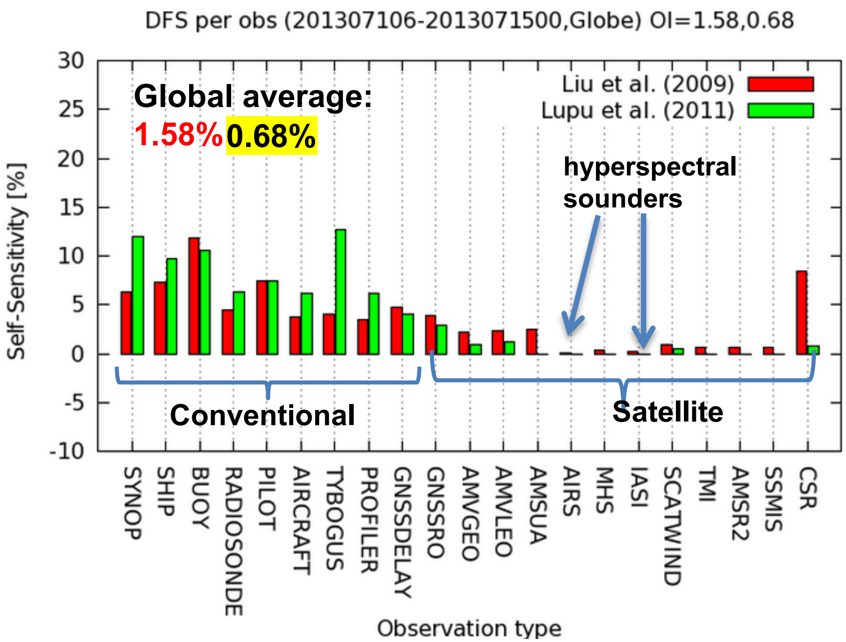
- First introduced to NWP by Fisher (2003) and Cardinali et al. (2004)
- Popular diagnostics for Var, but not many application to EnKFs.
- Liu et al. (2009) derived a simple method to compute DFS in EnKF framework:

  $\mathrm{tr}(\mathbf{S}) = \mathrm{tr}(\mathbf{HK}) = \mathrm{tr}(\mathbf{HAH}^{\mathrm{T}}\mathbf{R}^{-1}) = (\mathbf{R}^{-½}\,\mathbf{Y}^{a})^{\mathrm{T}}(\mathbf{R}^{-½}\,\mathbf{Y}^{a})/(N_{ens}-1)$

3. Ensemble-based DFS diagnostics at JMA

DFS per obs

global LETKF at JMA (50 members)



- Reasonable amount of information (10-15%) coming from **conventional** (***sparse***) observations.

Comparable to DFS in 4D-Var

- Little (<1%) information extracted from **satellite** (***dense***) observations, hyperspectral sounders (AIRS/IASI; ***very dense***) in particular (~0.1%).

An order of magnitude smaller than in 4D-Var

- Why?

4. Why DFS so small for EnKF?

- My Answer: **not enough ensemble size.**
- With simple algebraic argument, we can show, for any EnKF *local* analysis, that DFS = tr(**HK**) < N_{ens} - 1 where N_{ens} is the ensemble size.
- → DFS underestimated (smaller than optimal) whenever #ens << #obs (locally)

5. Proof of $DFS^{ens} < N_{ens} - 1$ (1/2)

Assume you have true **B**. For true (canonical) KF, the following equality holds:

$$(\mathbf{HAH}^T)^{-1} = (\mathbf{HBH}^T)^{-1} + \mathbf{R}^{-1}$$

accuracy of analysis is the sum of accuracy of background and observation

Let $\mathcal{O} \equiv \mathbf{R}^{-1/2}\mathbf{HB}^{1/2}$ (called *observability matrix* in Electrical Engineering/control theory literature) and apply $\mathbf{R}^{-1/2}$ from left and right; we have

$$\mathbf{R}^{-1/2}\mathbf{HAH}^T\mathbf{R}^{-1/2} = \mathbf{R}^{-1/2}((\mathbf{HBH}^T)^{-1} + \mathbf{R}^{-1})^{-1}\mathbf{R}^{-1/2} = ((\mathcal{O}\mathcal{O}^T)^{-1} + \mathbf{I})^{-1}$$

By eigen-decomposing $\mathcal{O}\mathcal{O}^T = \mathbf{U}\mathbf{\Lambda}^b\mathbf{U}^T$, $\mathbf{\Lambda}^b = \mathrm{diag}(\lambda^b_1, \lambda^b_2, \dots, \lambda^b_r, 0, \dots, 0)$

(where $r = \mathrm{rank}(\mathcal{O}\mathcal{O}^T)$), we have

$$\mathbf{R}^{-1/2}\mathbf{HAH}^T\mathbf{R}^{-1/2\,T} = ((\mathcal{O}\mathcal{O}^T)^{-1} + \mathbf{I})^{-1} = \mathbf{U}\mathbf{\Lambda}^a\mathbf{U}^T$$

with $\mathbf{\Lambda}^a = \mathrm{diag}(\lambda^a_1, \lambda^a_2, \dots, \lambda^a_r, 0, \dots, 0)$, $\lambda^a_i = \lambda^b_i/(\lambda^b_i + 1) < 1$

eigenvalues of normalized analysis error covariance in obs space are all less than 1.

→ $DFS^{opt} = \mathrm{tr}(\mathbf{HK}) = \mathrm{tr}(\mathbf{HAH}^T\mathbf{R}^{-1}) = \mathrm{tr}(\mathbf{R}^{-1/2}\mathbf{HAH}^T\mathbf{R}^{-1/2})$

$$= \textstyle\sum_i \lambda^a_i < r = \mathrm{rank}(\mathbf{R}^{-1/2}\mathbf{HBH}^T\mathbf{R}^{-1/2})$$

$$= \min\{\mathrm{rank}(\mathbf{R}), \mathrm{rank}(\mathbf{H}), \mathrm{rank}(\mathbf{B})\} = \#\mathrm{obs} \quad \text{(for most cases)}$$

5. Proof of $\mathrm{DFS}^{ens} < N_{ens} - 1$ (2/2)

Now, consider a local analysis in EnKF *(for now, ignore localization).*

In EnKF, **B** is approximated by $\mathbf{B}^{ens} = \mathbf{X}^b\mathbf{X}^{bT}/(N_{ens} - 1)$.

Since $\mathrm{rank}(\mathbf{R}^{-½}\mathbf{H}\mathbf{B}^{ens}\mathbf{H}^T\mathbf{R}^{-½})$

$= \min\{\mathrm{rank}(\mathbf{R}), \mathrm{rank}(\mathbf{H}), \mathrm{rank}(\mathbf{X}^b)\}$

$= \min\{\ \#obs,\ \ \#obs,\ N_{ens} - 1\ \} = N_{ens} - 1$ (if $N_{ens} < \#obs$)

it follows that

$\mathrm{DFS}^{ens} = \mathrm{tr}(\mathbf{H}\mathbf{K}^{ens}) = \mathrm{tr}(\mathbf{R}^{-½}\mathbf{H}\mathbf{A}^{ens}\mathbf{H}^T\mathbf{R}^{-½})$

$< \mathrm{rank}(\mathbf{R}^{-½}\mathbf{H}\mathbf{B}^{ens}\mathbf{H}^T\mathbf{R}^{-½}) = N_{ens} - 1$

□

- DFS is underestimated in EnKF if $N_{ens} \ll \mathrm{DFS}^{\mathrm{opt}}$

- So what?

6. Implications
What's wrong if $DFS^{ens} \ll DFS^{opt}$?

If $DFS^{ens} \ll DFS^{opt}$, it means…

- Analysis increment is smaller than under optimality
  - analysis increment (in obs space) is $\mathbf{HKd}=\mathbf{HAH^{T}R^{-1}d}$
  - so if DFS (=tr($\mathbf{HAH^{T}R^{-1}}$)) is underestimated, so is analysis increment
- **(more important)** Since DFS = tr($\mathbf{R^{-1/2}\,HAH^{T}\,R^{-1/2}}$) is a measure of analysis spread in obs space, **underestimated DFS implies overconfidence in analysis (overconfident posterior).**

  → ***Requires strong covariance inflation***, but inflating too much is no good since that would lead to inaccurate representation of "the errors of the day" (*i.e., destroy flow-dependence*)

Digression: Interpreting the counterintuitive results from the recent literature

- "Using ***less*** obs is ***better***"
  - ECMWF global LETKF (Hamrud et al. 2015, *MWR*)
  - Convective-scale COSMO-LETKF (Schraff et al. 2016, *QRJMS*)
  - Radar DA at RIKEN, Japan (Poster 2.3 by Guo-Yuan Lien)
  - Meta-analysis of the literature by Tsyrulnikov (2010; *COSMO Newsletter No.* **10**):

    "Optimal localization scale occurs when local analysis domain is small enough so that "ensemble size (is) commensurable with the number of ***observed degrees of freedom*** within [the local patch]"

→ Justification with DFS argument:
Locally assimilating **more obs than #ens** results in **overconfident analysis spread** (requiring **unreasonably large inflation**) and also **smaller-than-optimal analysis increment**.

7. Proposed Solution: **B**-localization through *modulated ensemble*

- DFS underestimation is a quantitative manifestation of the well-known (but vaguely defined) ***rank deficiency issue***.
- → Resolved by covariance localization.
- PO-EnKF or serial enSRF:
  - Replacing $\mathbf{B}^{ens}\mathbf{H}^{T}$ with $\boldsymbol{\rho}_{o} \circ (\mathbf{B}^{ens}\mathbf{H}^{T})$ increases effective rank of $\mathbf{B}^{ens}$ in local analysis, mitigating the DFS underestimation
- By contrast, **R**-localization employed in LEKTF does not resolve the issue ∵ local analysis is still solved in (N_{ens} – 1)-dim space spanned by the perturbations, even with **R**-localization
- Can we somehow increase the rank within ensemble-transform framework?
- We can, by **B-localization through modulated ensemble approach** (C. Bishop, pers. comm. at EnKF workshop 2016)

7. Proposed Solution:
B-localization through *modulated ensemble*

- **B-localization with modulated ensemble** (Bishop and Hodyss, 2009; ECO-RAP paper Part II)
  - $\boldsymbol{\rho}\circ(\mathbf{B}^{\mathrm{ens}}) = \boldsymbol{\rho}\circ(\mathbf{X}\mathbf{X}^T)/(N_{\mathrm{ens}}-1) = (\mathbf{Z}\mathbf{Z}^T)/(N_{mode}N_{\mathrm{ens}}-1)$
  - where
    $\mathbf{Z} = [\mathbf{w}_1\circ\mathbf{x}_1, \mathbf{w}_1\circ\mathbf{x}_2, \ldots, \mathbf{w}_1\circ\mathbf{x}_{N\mathrm{ens}}; \ldots; \mathbf{w}_{Nmode}\circ\mathbf{x}_1, \mathbf{w}_{Nmode}\circ\mathbf{x}_2, \ldots, \mathbf{w}_{Nmode}\circ\mathbf{x}_{N\mathrm{ens}}]$
    with $\boldsymbol{\rho} \approx \mathbf{W}\mathbf{W}^T$, $\mathbf{W}=[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{Nmode}]$

- (My) intuitive interpretation:
  - Localized empirical covariance is identical to the empirical covariance of many *modulated* ensembles, each "raw" member $\mathbf{x}_i$ "localized" with many different localization modes $\mathbf{w}_j$

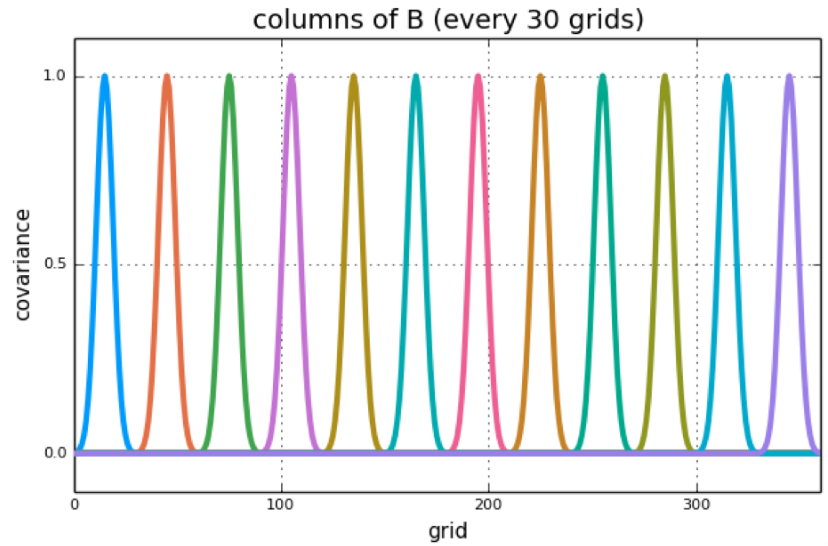
  → LETKF with model-space **B**-localization can be achieved by performing regular ETKF (w/o localization) using the modulated $N_{mode} \times N_{\mathrm{ens}}$-member perturbations.

8. Exposition with a simple covariance model
Experimental setup

Simplest possible scenario following Bishop and Hodyss (2009; ECO-RAP paper Part I):

- 1D periodic domain with **#grid=360**.
- **B** and **R** perfectly known. All errors unbiased and Gaussian. **R** diagonal.
- Perfect generation of $\mathbf{X}^b$ (i.e., $\mathbf{B}^{ens}=\mathbf{X}^b\mathbf{X}^{bT}/(K-1)$ converges to **B** with $K \rightarrow \infty$)
- Equally-spaced obs assimilated, **#obs=120**.
- No cycling.
- All experiments repeated 1,000 times and averaged.

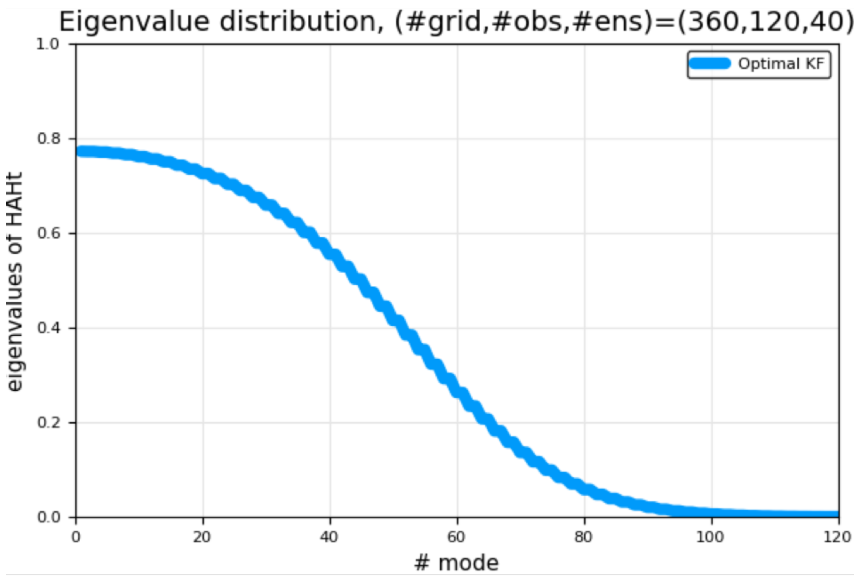
- Specification of **B**:
  - superposition of sinusoids
  - Fourier transform of a Gauss function
  - virtually zero correlation beyond 15-grid interval.
  - Variance is 1 everywhere (B_{ii}=1)



8. Exposition with a simple covariance: Role of localization

We focus on eigen-spectrum of $\mathbf{R}^{-1/2}\mathbf{H}\mathbf{A}^{opt}\mathbf{H}^T\mathbf{R}^{-1/2}$ because DFS is proportional to the area below this curve.

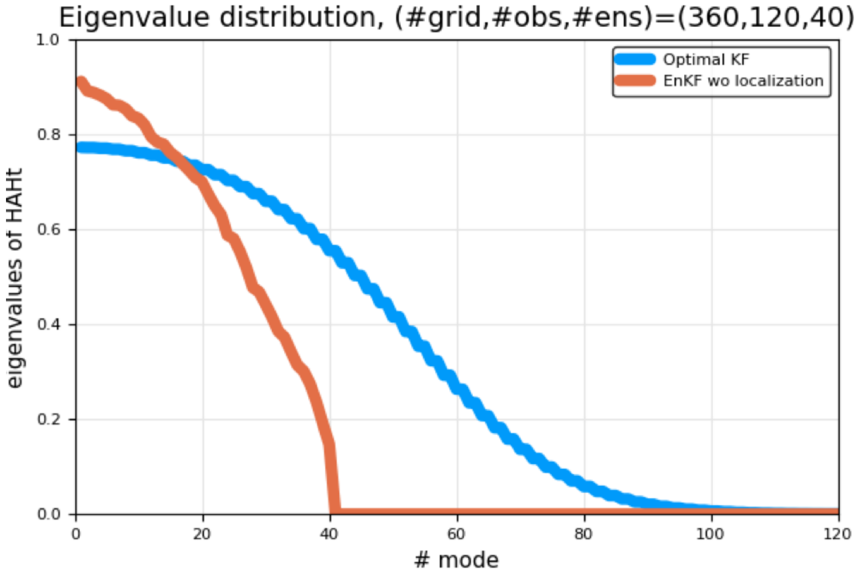
DFS= $tr(\mathbf{R}^{-1/2}\,\mathbf{HAH}^T\,\mathbf{R}^{-1/2}) = \sum_i \lambda^a_i$



- $\mathbf{HA}^{opt}\mathbf{H}^T$ computed as $((\mathbf{HB}^{true}\mathbf{H}^T)^{-1}+\mathbf{R}^{-1})^{-1}$

8. Exposition with a simple covariance: Role of localization

40-member ETKF without localization

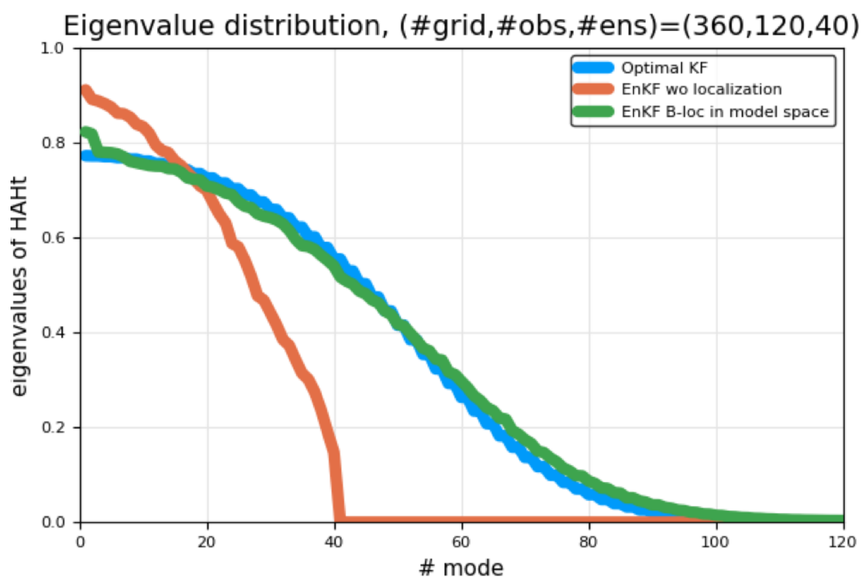


↑#ens=40

- $\mathbf{HA}^{ens}\mathbf{H}^{T}$ computed as $((\mathbf{HB}^{ens}\mathbf{H}^{T})^{-1}+\mathbf{R}^{-1})^{-1}$
- with raw $\mathbf{B}^{ens}=\mathbf{X}^{b}\mathbf{X}^{bT}/(N_{ens}-1)$ (without localization)
- K=40 member ensemble.
- Abrupt truncation at $N_{ens}-1$=39th mode.

8. Exposition with a simple covariance: Role of localization

40-member Model-space **B**-localization using modulated ensemble retaining 20 localization modes (localization scale tuned to give best analysis RMSE)



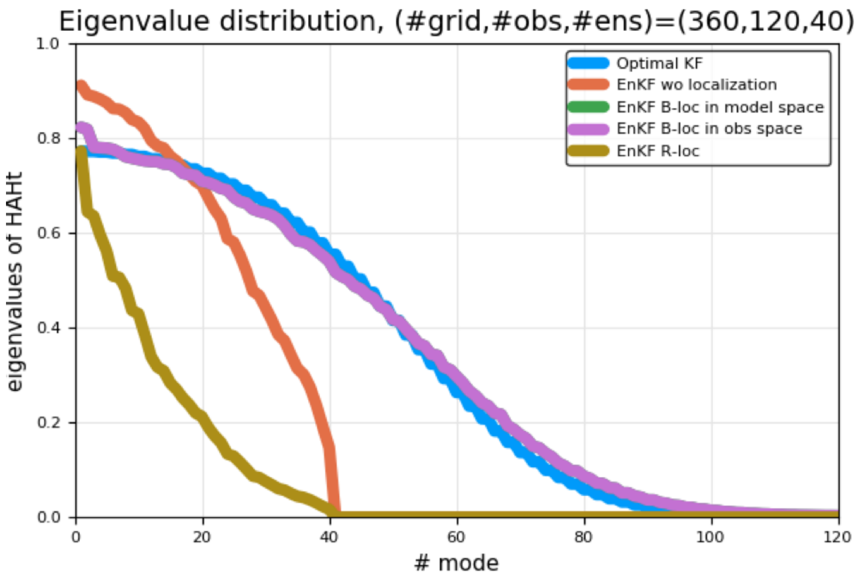
↑#ens=40

- Almost **perfectly recovers** the true (optimal) eigen-spectrum.
- → **B**-localization very effective when assimilating dense obs.

8. Exposition with a simple covariance: Role of localization

40-member LETKF with **R**-localization

(localization scale tuned to give best analysis RMSE)



↑#ens=40

- $\mathbf{H}\mathbf{A}^{\mathbf{R}\text{-loc}}\mathbf{H}^{\mathrm{T}}$ computed *for each grid* as $\mathbf{Y}^{b}\{(K\text{-}1)\mathbf{I}+\boldsymbol{\rho}_{\mathbf{R}}\circ\mathbf{R}^{-1}\}^{-1}\mathbf{Y}^{b\mathrm{T}}$, then synthesized.

Zero eigenvalues beyond N_{ens}–1=39th mode.

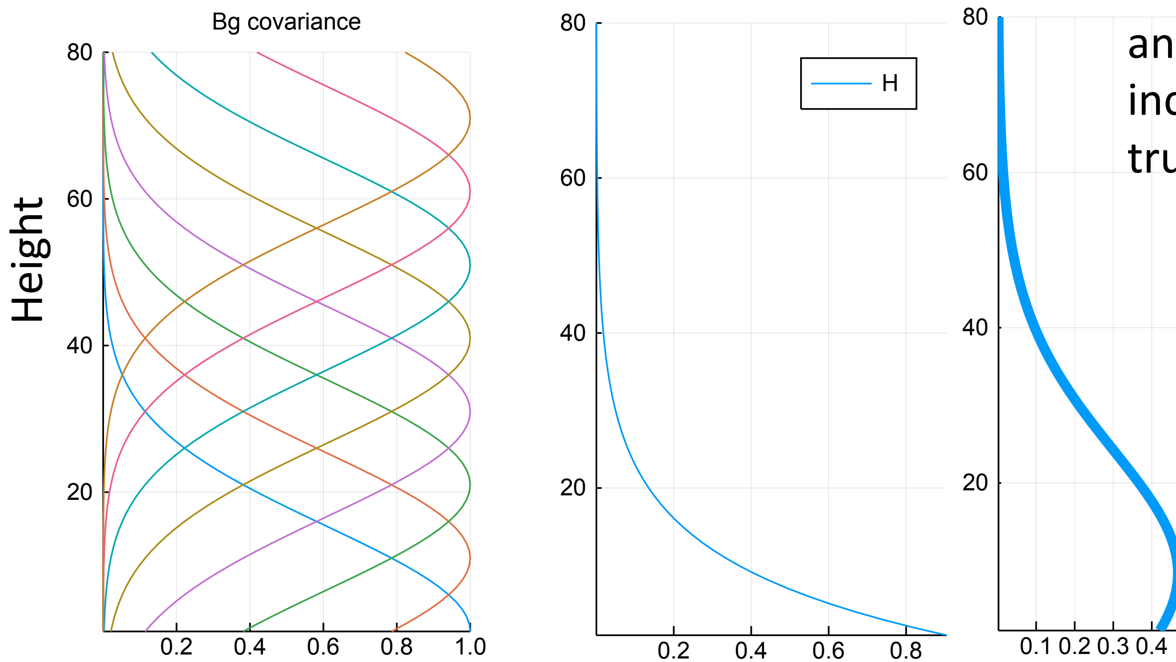
9. Non-local obs and localization

- Another disadvantage of **R**-localization:
  Not clear how to localize impact of obs whose position in physical space is not clearly defined, .e.g.,
  - Satellite radiances
  - Ground-based GNSS obs (e.g., Poster 1.1 by Michael Bender)
  - Surface pressure (!)
- A problem common with obs-space localization on $\mathbf{BH}^T$

→Solution already proposed:
model-space **B** localization (Poster 4.5 by Craig Bishop)

9. Non-local obs and localization
1D toy system mimicking GNSS PWV

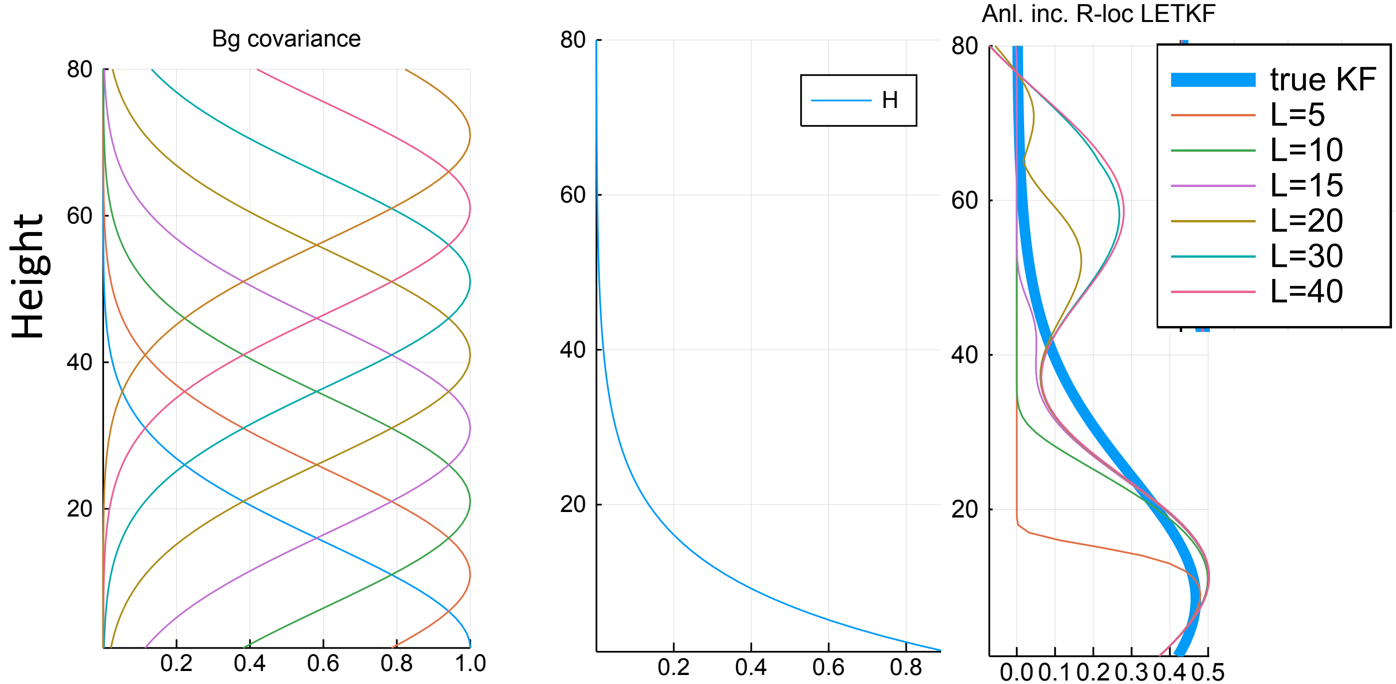
- Bg covariance: Gaussian-shape
- Obs operator: exponentially decays with height
- Single-obs assimilation with true (canonical) KF: $x^a - x^b = \mathbf{K}d \propto \mathbf{B}\mathbf{H}^T = \Sigma_k h_k \mathbf{b}_k$



analysis increment with true KF

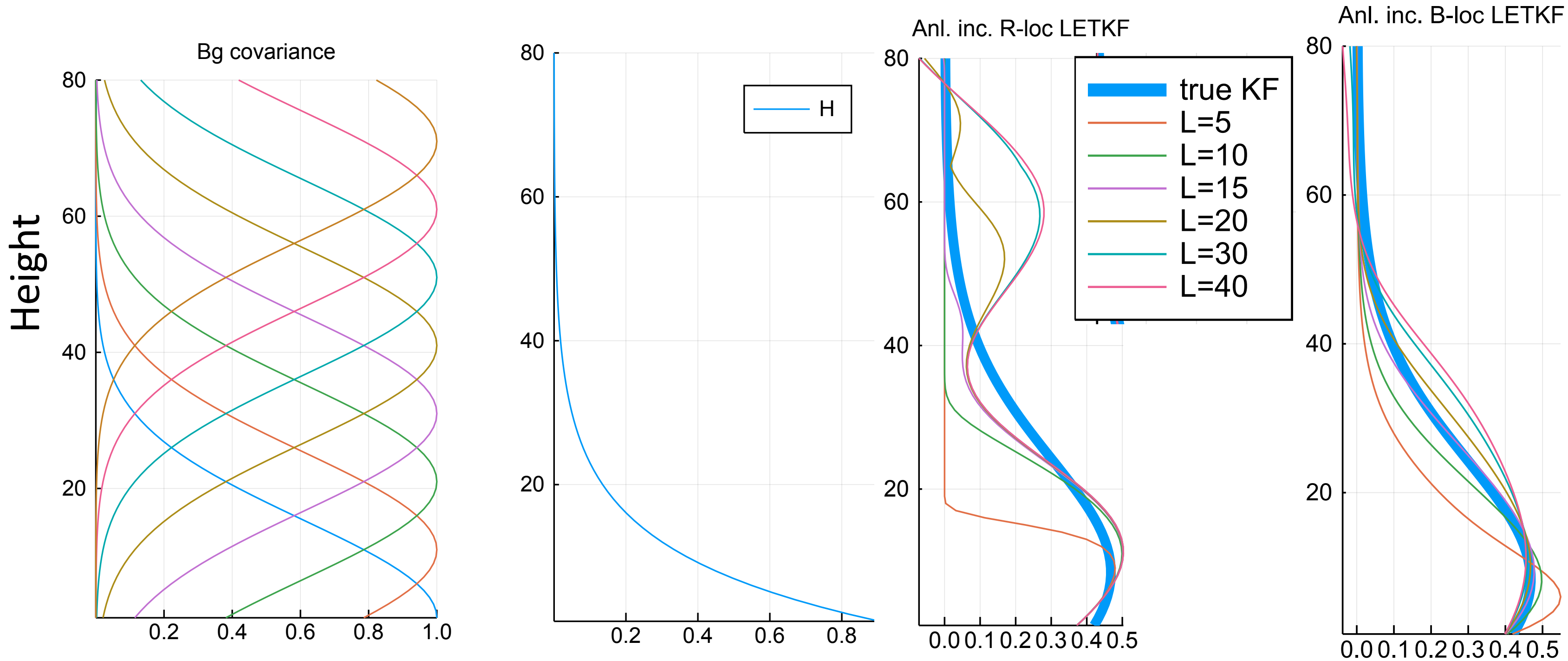
9. Non-local obs and localization
1D toy system mimicking GNSS PWV

- Analysis increment from 20-member LETKF with **R**-localization
- assuming that the obs is ”located” at the surface (k=1)
- Localizations with Gaspari-Cohn, localization scale *L* ranging from 5 to 40
- Small *L* → increment too much localized near the surface
- Large *L* → spurious increments inevitable in the upper atmosphere



9. Non-local obs and localization
1D toy system mimicking GNSS PWV

- Analysis increment from 20-member LETKF with **B**-localization retaining 10 modes
- Localizations with Gaspari-Cohn, localization scale *L* ranging from 5 to 40
- Moderate *L* (15~30) → increment very close to true KF,
  with no spurious increment in the upper layers
- Quite insensitive to the exact choice of *L* → *L* can be tuned rather easily



Summary:

- LETKF with **R**-loc is efficient and works well for relatively sparse and local obs.
- May not be optimal otherwise.
- Model-space **B**-localization with modulated ensemble solves the problems with dense and non-local obs.

My plan over the next few years:

- Investigate whether **B**-loc really improves real NWP
  - Target obs: ground-based GNSS (ZTD or PWV)
  - Target model: convective-scale LAM (JMA-NHM)
  - DA method: **B**-loc in the vertical, **R**-loc in the horizontal

Danke schön!