The ensemble-variational method using observation space localization

Le Duc[12], Kazuo Saito[21] , and Sho Yokota[2]
[1] Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[2] Meteorological Research Institute, Tsukuba, Japan

Introduction

The ensemble variational data assimilation method (EnVAR) was originally constructed by using ensemble perturbations to model background covariance. Since the localization matrix in EnVAR is given in the model space, it is taken for granted that observation space localization cannot be employed in EnVAR.

Beside model space localization, observation space localization is another localization method used in ensemble Kalman filter. In this study, we show that observation space localization can also be introduced into EnVAR, thus providing a unified framework for comparison between two localization methods.

EnVAR with observation space localization

To find the analysis increment in EnVAR, we minimize the cost function in the weight space $\boldsymbol{\alpha}$

$$J(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{C}^{-1}\boldsymbol{\alpha} + \frac{1}{2}[\mathbf{d}-\mathbf{HD}\boldsymbol{\alpha}]^T\mathbf{R}^{-1}[\mathbf{d}-\mathbf{HD}\boldsymbol{\alpha}] \quad (1)$$

where **C** is the diagonal block matrix of which each block is the localization matrix **L**, and **D** is the transform matrix from the weight space to the model space

$$\mathbf{C} = \begin{pmatrix} \mathbf{L} & & \\ & \ddots & \\ & & \mathbf{L} \end{pmatrix} \quad (2)$$

$$\mathbf{D} = (Diag(\Delta\boldsymbol{x}_1) \quad \ldots \quad Diag(\Delta\boldsymbol{x}_K)) \quad (3)$$

Here $\Delta\boldsymbol{x}_k$ is the k[th] perturbation and Diag is the operator that converts a vector into a diagonal matrix.

To find for the zero point of gradient of J, we find the zero point of its associated equation in the dual space

$$[\mathbf{HDC}(\mathbf{HD})^T + \mathbf{R}]\boldsymbol{\lambda} = \mathbf{d} \quad (4)$$

We rewrite $\mathbf{C}(\mathbf{HD})^T$ as $(\mathbf{HDC})^T$, then it is easy to verify that

$$\mathbf{HDC} = \mathbf{H}(Diag(\Delta\boldsymbol{x}_1)\mathbf{L} \quad \ldots \quad Diag(\Delta\boldsymbol{x}_K)\mathbf{L}) = \mathbf{H}(\mathbf{L}\odot\Delta\boldsymbol{X}_1 \quad \ldots \quad \mathbf{L}\odot\Delta\boldsymbol{X}_K) \quad (5)$$

where $\Delta\boldsymbol{X}_k$ is the matrix consisting of n identical column vectors $\Delta\boldsymbol{x}_k$. Then observation space localization can be introduced by an approximation

$$\mathbf{H}(\mathbf{L}\odot\Delta\boldsymbol{X}_k) \approx \mathbf{HL}\odot\mathbf{H}\Delta\boldsymbol{X}_k = \mathbf{L_O}\odot\Delta\boldsymbol{Y}_k \quad (6)$$

Here $\mathbf{L_O}$ is the localization matrix between observations and model variables, $\Delta\boldsymbol{Y}_k$ consists of n identical column vectors $\Delta\boldsymbol{y}_k$, which is the k[th] perturbation in the observation space.

EnVAR system

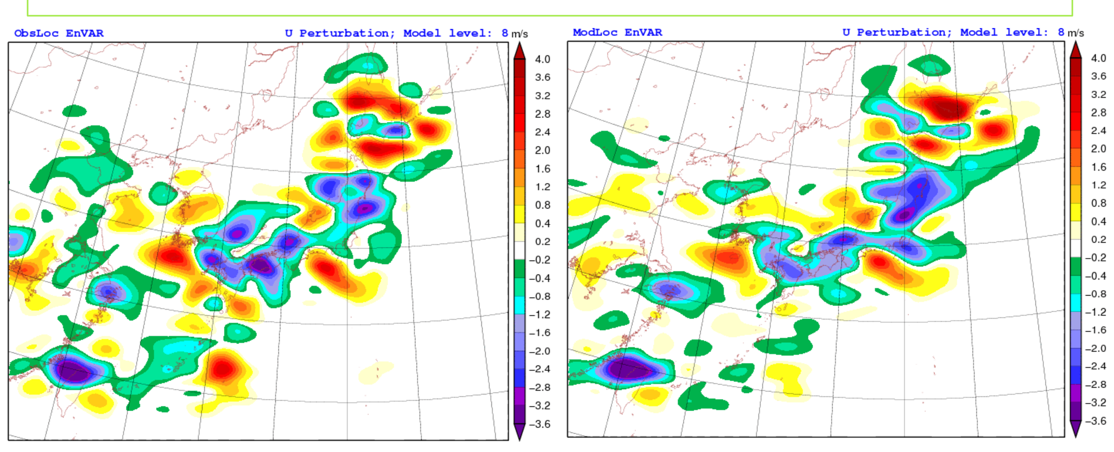
A four-dimensional EnVAR system (NHM-4DEnVAR) was developed using the limited-area operational model NHM of Japan Meteorological Agency. In addition to observation space localization, the system also supports model space localization. Whereas the matrix **L** was derived from the climatological background error covariance, the matrix $\mathbf{L_O}$ used fixed horizontal and vertical localization length scales, which were 250 km and 0.4logp respectively. An local ensemble transform Kalman filter (LETKF) was run in parallel to provide forecast perturbations for NHM-4DEnVAR.

Since the linear equation (4) is no longer symmetric as the original one, we use the GMRES method to solve this linear equation. The adjoint of **H** is not needed in solving this equation, although **H** is still needed. However, we can approximate it by using the full nonlinear operator h.

Real observation experiments

NHM-4DEnVAR with two localization methods (EnVAR-ObsLoc and EnVAR-ModLoc) was run in one month (August 2014) over Japan at the dual resolutions 15 and 5 km. The assimilation cycle was three hours and the number of ensemble members was 50.

Figure 1 shows the analysis increments of u fields at the model level 8 from two experiments for an arbitrary date. For comparison, the analysis increment from LETKF is also plotted. The difference in the three patterns reflects the underlying localization scheme. This of course depends on the specification for the localization matrices. Note the similarity between EnVAR-ObsLoc and LETKF.



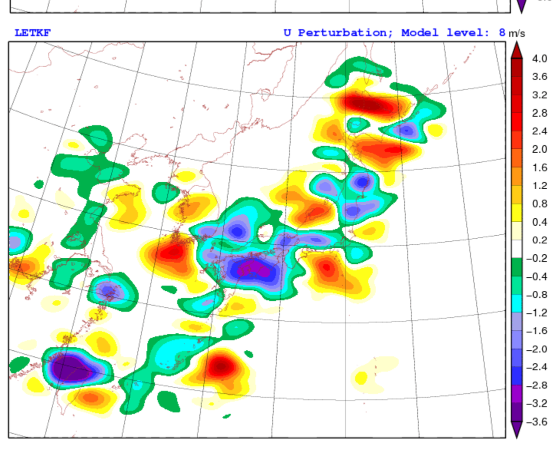


Figure 1. Analysis increments of u fields at the model level 8 with EnVAR-ObsLoc (upper left), EnVAR-ModLoc (upper right), and LETKF (lower left) as reference, respectively.

Verification results

Figure 2 shows averaged RMSEs between radiosonde observations and analyses. Above 700 hPa, EnVAR-ModLoc analyses tend to fit observations closer than EnVAR-ObsLoc and LETKF. If 12-hour forecasts are considered, EnVAR-ModLoc is better than EnVAR-ObsLoc, especially in temperature forecast, above 500 hPa. However, below 500 hPa, EnVAR-ModLoc is slightly worse. We suspect that this resulted from the fact that the localization length scales in EnVAR-ModLoc were smaller than those in EnVAR-ObsLoc at the lower model levels as depicted in Figure 4. In fact, RMSEs were decreased when we increased the localization scales in each experiment, which we do not show here. We will test this conjecture in future by using the same localization scales in both experiments.

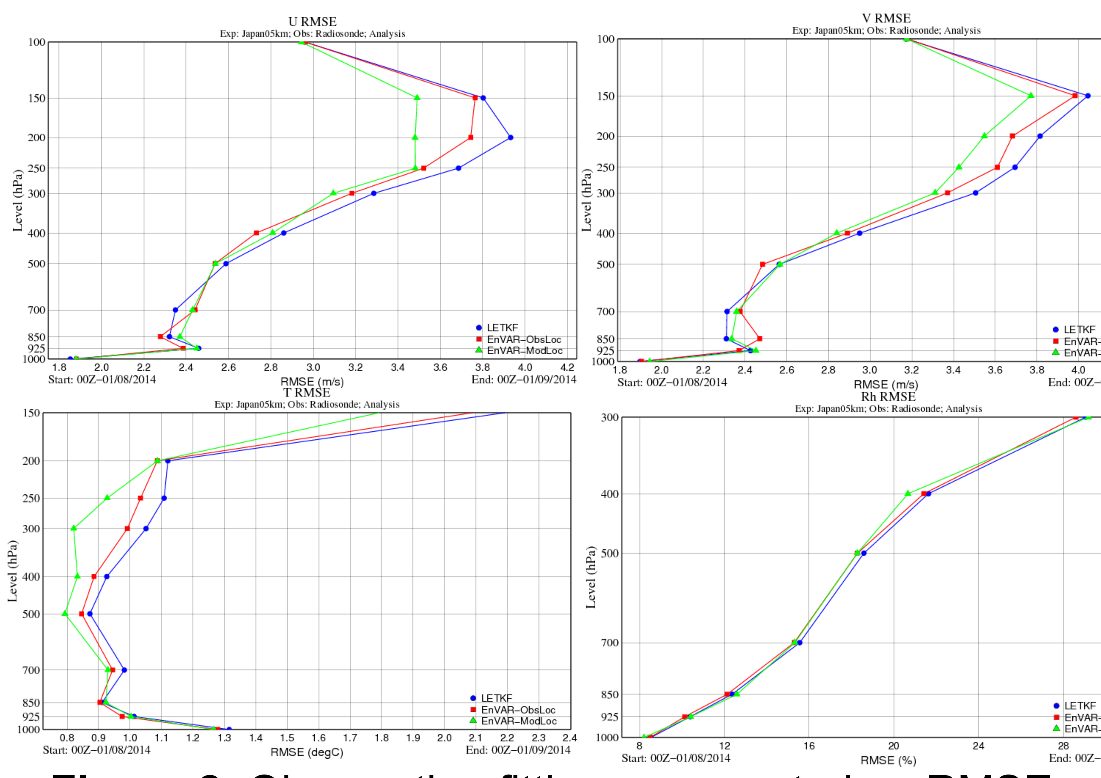


Figure 2. Observation fitting represented as RMSEs of analyses of zonal wind (upper left), meridional wind (upper right), temperature ((lower left), and relative humidity (lower right) by EnVAR-ObsLoc, EnVAR-ModLoc and LETKF.

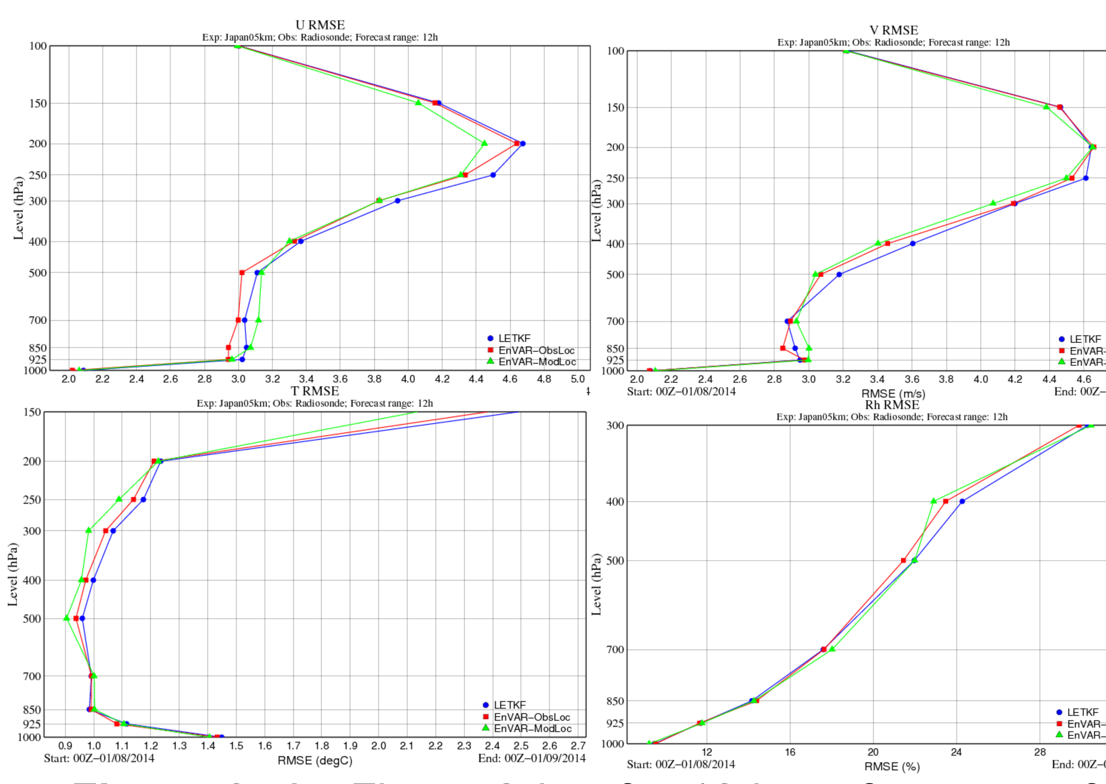


Figure 3. As Figure 2 but for 12-hour forecasts of NHM using EnVAR-ObsLoc, EnVAR-ModLoc and LETKF analyses as initial conditions.

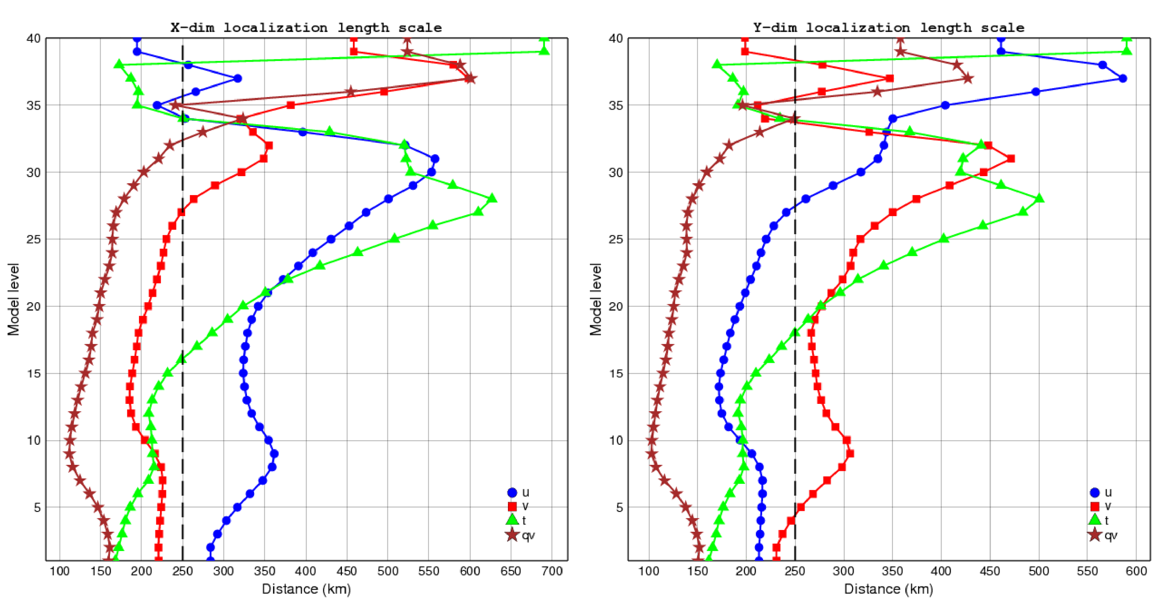


Figure 4. Horizontal localization length scales of different variables used in EnVAR-ModLoc and EnVAR-ObsLoc.