AGU100 ADVANCING EARTH AND SPACE SCIENCE

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2018MS001309

Key Point:
- EnKF configuration using linearized observation operator for observation space ensemble perturbations has been developed at NOAA

Correspondence to:
A. Shlyaeva,
anna.v.shlyaeva@noaa.gov

Citation:
Shlyaeva, A., & Whitaker, J. S. (2018). Using the linearized observation operator to calculate observation space ensemble perturbations in ensemble filters. *Journal of Advances in Modeling Earth Systems*, *10*, 1414–1420. https://doi.org/10.1029/2018MS001309

Received 13 MAR 2018
Accepted 26 MAY 2018
Accepted article online 7 JUN 2018
Published online 2 JUL 2018

Using the Linearized Observation Operator to Calculate Observation Space Ensemble Perturbations in Ensemble Filters

Anna Shlyaeva[1] **and Jeffrey S. Whitaker**[2]

[1]Cooperative Institute for Research in Environmental Sciences at the NOAA Earth System Research Laboratory, Physical Sciences Division, University of Colorado Boulder, Boulder, CO, USA, [2]Physical Sciences Division, NOAA Earth System Research Laboratory, Boulder, CO, USA

Abstract Within the National Oceanic and Atmospheric Administration National Weather Service, the hybrid ensemble-variational system (Gridpoint Statistical Interpolation, GSI) is run together with the 80-member ensemble square root filter (EnSRF) operationally for the global forecast data assimilation system. EnSRF uses observation operator from GSI: current operational configuration requires 81 runs of GSI in the observation operator mode to run EnSRF (for each of the 80 ensemble members and for the ensemble mean). To reduce data assimilation cycle computation time, a GSI-EnSRF configuration that requires a single run of the GSI system in the observation operator mode was developed. In this configuration EnSRF uses full observation operator for the ensemble mean and linearized observation operator for the ensemble perturbations. Comparison of the two approaches shows that using linearized observation operator for ensemble perturbations compared to using full observation operator does not change the analysis results significantly and allows to reduce overall data assimilation cycle computation time.

1. Introduction

Data assimilation systems combine observations and short-term model forecast, also called background (or ensemble of those), to generate best estimate of model state at a given time (or ensemble, best representing most probable state and its uncertainty).

Various ensemble Kalman filters (EnKFs) are used in data assimilation for numerical weather prediction (e.g., see review in Houtekamer & Mitchell, 2005) to generate initial conditions for ensemble forecasts and/or ensemble background error covariances for the deterministic ensemble variational assimilation such as Buehner et al. (2015), Clayton et al. (2013), and Kleist and Ide (2015).

Within the National Weather Service (NWS)'s global numerical weather prediction system, an 80-member EnKF is run operationally to initialize Global Ensemble Forecast System and to provide ensemble covariances for the hybrid ensemble variational data assimilation (Wang & Lei, 2014) system (Gridpoint Statistical Interpolation analysis system, GSI) that generates the deterministic analysis for Global Forecast System.

In the current NWS configuration, observation operators — operators mapping from the model space to the observation space — are not implemented in EnKF code. Instead, the variational GSI system is run in the observation operator mode to apply observation operators to the ensemble mean and all the ensemble members. This means that one has to run GSI in the observation operator mode $k + 1$ times (where k is the ensemble size) to be able to run the EnKF. This is computationally inefficient if observation operators for some of the ensemble members have to run sequentially, as is common with the EnKF.

An alternative to this approach is to run the observation operator only for the ensemble mean but to have it save both the full observation operator applied to the ensemble mean and the observation operator Jacobian. In this case the observation operator Jacobian may be used to apply the *linearized* observation operator to the ensemble perturbations. Since observation operators for some observations (such as satellite radiances and Global Positioning System [GPS] radio occultation observations) are nonlinear, using linearized observation operator to calculate observation space ensemble perturbations will give different results than using the full observation operator.

In this paper, we explore this approach and assess how using linearized observation operator for ensemble perturbations affects the EnKF analysis compared to using the full observation operator. A description of the National Oceanic and Atmospheric Administration (NOAA) implementations of two common ensemble filters is given in section 2, the implementation of linearized observation operator in this system is presented in section 3, and the results of the experiments comparing two approaches are discussed in section 4. The conclusions are summarized in section 5.

2. NOAA Ensemble Filter Descriptions

There are two options in the NOAA EnKF code, the serial ensemble square root filter (EnSRF) and the local ensemble transform Kalman filter (LETKF). Currently, the EnSRF is used operationally in the atmospheric global forecast system at NOAA. In both cases, the full observation operator is applied to ensemble mean and ensemble members once, before the EnKF update. The ensemble perturbations in observation space are defined as the difference between the observation operator applied to the mean and the observation operator applied to individual members.

2.1. EnSRF Option of NOAA EnKF

The implementation of NOAA EnSRF is based on the serial EnSRF described in Whitaker and Hamill (2002). For computational efficiency it uses the parallel algorithm as in Anderson and Collins (2007). Within the assimilation, each assimilated observation is used to update both the mean and the ensemble perturbations in state *and* observation space. Updating both the state and observation space ensembles within the observation loop assumes linearization of forward operator around the background ensemble mean.

2.2. LETKF Option of NOAA EnKF

The implementation of NOAA LETKF is based on Hunt et al. (2007). The following parallel algorithm is used: each processor updates a subset of vertical grid columns. The algorithm for distributing grid columns across processors is implemented to minimize load imbalance, using the number of observations in each local volume as an estimate of computational work. The entire observation space prior ensemble is available on each MPI task so that the distribution of vertical grid columns can be arbitrary. MPI version 3 shared memory windows are used to reduce the LETKF memory footprint by allocating a single shared memory segment for the observation space prior ensemble on each compute node.

Relaxation-to-prior spread multiplicative inflation (Whitaker & Hamill, 2012) is used for both the EnSRF and LETKF. Stochastic parameterizations are used to represent model uncertainty within the ensemble forecast step. Observation space covariance localization (Lei & Whitaker, 2015) is used in the EnSRF, while observation error localization (Greybush et al., 2011) is used in the LETKF. The same localization function (based on Gaspari & Cohn, 1999) is used in both cases.

3. Using the Linearized Observation Operator in the EnKF

In the EnKF, the background ensemble mean $\overline{x}$ and background ensemble perturbations $X = \{X_1, X_2, \dots, X_k\}$ are updated with the set of observations. Here $x = \{x_1, x_2, \dots, x_k\}$ is the ensemble of background state vectors, $X_i = x_i - \overline{x}$ and $\overline{x} = 1/k \sum_{i=1}^{k} x_i$.

The EnSRF and LETKF require observation space counterparts of $\overline{x}$ and X: the ensemble mean in observation space $\overline{y}$ and the ensemble perturbations in observation space Y.

If the observation operator $\mathbf{H}$ is linear,

$$\overline{y} = \mathbf{H}\overline{x} \tag{1}$$

$$Y = \mathbf{H}X \tag{2}$$

However, most observation operators H for data used in atmospheric data assimilation systems are nonlinear (e.g., radiative transfer models for satellite radiances). In this case the following is typically used (see, e.g., Houtekamer & Mitchell, 2001):

$$\overline{y} = H(\overline{x}) \tag{3}$$

$$Y = H(x) - \overline{H(x)} \tag{4}$$

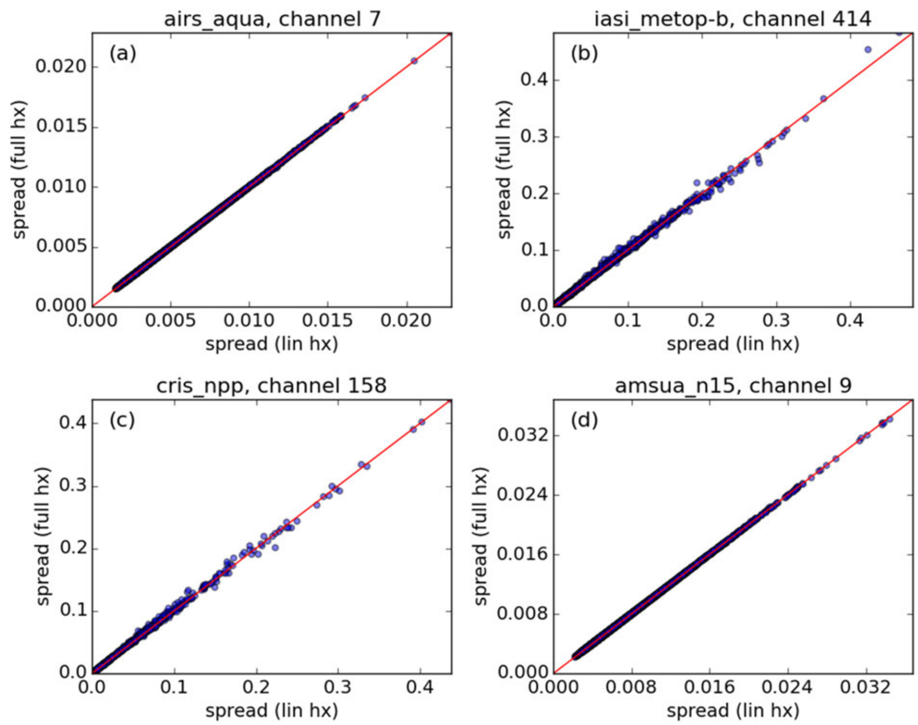

Figure 1. Scatterplots comparing observation space ensemble standard deviations when using linearized observation operator (*x* axis) and full observation operator (*y* axis) for (a) Aqua AIRS channel 7, (b) MetOp-B IASI channel 414, (c) Suomi NPP CrIS channel 158, and (d) NOAA-15 AMSU-A channel 9. AIRS = Atmospheric Infrared Sounder; IASI = Infrared Atmospheric Sounding Interferometer; NOAA = National Oceanic and Atmospheric Administration; AMSU-A = Advanced Microwave Sounding Unit-A.

Here $\overline{H(x)} = 1/k \sum_{i=1}^{k} H(x_i)$ is averaged over full observation operator applied to all ensemble members. This configuration is used in the operational NOAA EnKF, as described in section 2.

Here we propose using observation operator Jacobian $\widetilde{\mathbf{H}} = \frac{\partial H}{\partial x}|_{x=\overline{x}}$ to apply the linearized observation operator to ensemble perturbations; so

$$\overline{y} = H(\overline{x}) \tag{5}$$

$$Y = \widetilde{\mathbf{H}}X \tag{6}$$

In this configuration full observation operator is used to calculate $\overline{y}$, as in equation (3), while the linearized observation operator is used to calculate Y. The application of linearized observation operator is justifiable for small ensemble perturbations.

For the EnKF to use equation (6), the GSI code needs to save $\widetilde{\mathbf{H}}$. This would not be feasible in the operational data assimilation system if the observation operators were not local. We make use of the horizontal and time locality of all observation operators used at NOAA: all of them use simple bilinear interpolation in *xyt* space. This means that for each observation it is sufficient to save only vertical profiles of $\widetilde{\mathbf{H}}$ for all state variables, such as virtual temperature, specific humidity, and ozone, interpolated horizontally and in time to the observation location. They are saved in a sparse-array format to conserve disk space. To obtain Y, the EnKF reads all the state space ensemble perturbations X, interpolates them to the observation locations horizontally and in time, and applies a dot product of these interpolated vertical profiles with the linearized observation operator $\widetilde{\mathbf{H}}$.

4. Experiments

To assess the effect of using the linearized observation operator instead of the full observation operator for the ensemble perturbations, observation space ensemble perturbations calculated using both approaches are compared in section 4.1. In section 4.2 we compare results of cycled data assimilation experiments with the EnSRF using the two approaches.

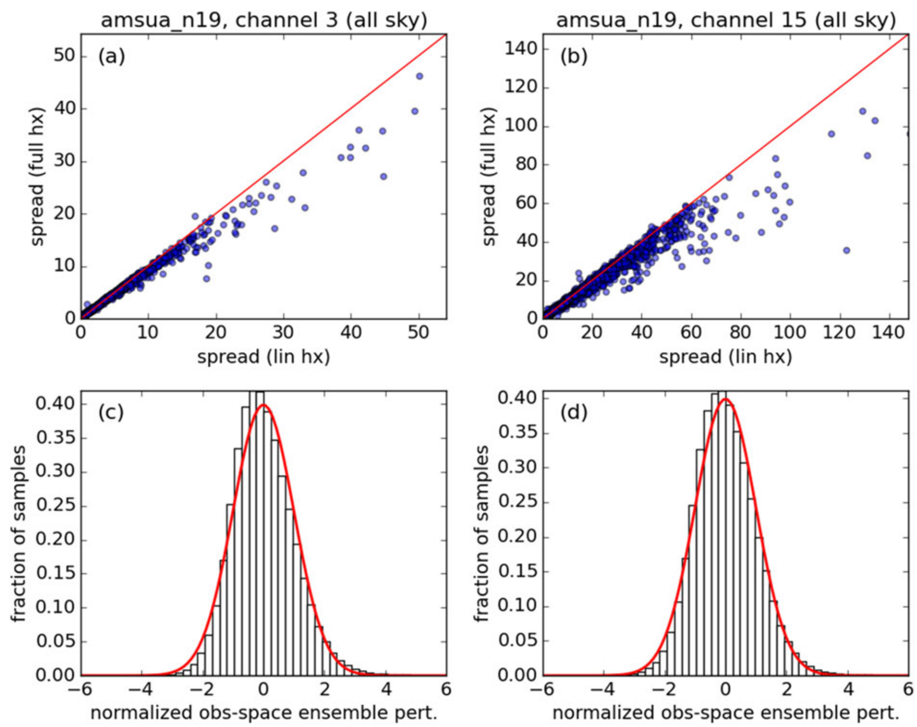

Figure 2. Comparison of all-sky NOAA-19 AMSU-A channels 3 and 15 when using linearized and full observation operator. (a, b) Scatterplots comparing observation space ensemble standard deviations when using linearized observation operator (x axis) and full observation operator (y axis). (c, d) Histograms of normalized ensemble perturbations in the observation space when using full observation operator; the red curve is the standard Gaussian distribution for comparison. The histograms were obtained by normalizing all of the ensemble perturbations $H(x)$ by the corresponding ensemble mean and standard deviation and plotting them together. NOAA = National Oceanic and Atmospheric Administration; AMSU-A = Advanced Microwave Sounding Unit-A.

4.1. Statistics of Ensemble Perturbations in Observation Space

In the proposed modification ensemble perturbations in observation space differ from the operational implementation when the observation operator is nonlinear. Among the observation operators used in the GSI are the GPS bending angle observation operator (Cucurull et al., 2013) and the Community Radiative Transfer Model observation operator (Han, 2006) used for satellite radiances such as AMSU-A (Advanced Microwave Sounding Unit), AIRS (Atmospheric Infrared Sounder), ATMS (Advanced Technology Microwave Sounder), CrIS (Cross-track Infrared Sounder), HIRS/4 (High-resolution Infrared Radiation Sounder), IASI (Infrared Atmospheric Sounding Interferometer), and MHS (Microwave Humidity Sounder).

To assess how much using linearization of different observation operators affects observation space ensemble perturbations, we compare the ensemble standard deviation (spread) in observation space calculated from the same background ensemble using the linearized (equation (6)) and full (equation (4)) observation operator.

Figure 1 shows scatterplots comparing observation space ensemble spread for selected channels from AIRS, IASI, CrIS, and AMSU-A. The observation space ensemble spread for those satellite channels is very similar when using the linearized or full observation operator. Similar behavior is exhibited for most instruments and channels.

The most nonlinear observation operators appear to be the ones for AMSU-A channels 1–4 and 15 and MHS channel 4 where all-sky radiances (in both clear and cloudy conditions) are used. Figures 2a and 2b show scatterplots of observation space ensemble standard deviations computed using the linearized and full observation operators for AMSU-A channels 3 and 15. Ensemble perturbations in the observation space for these channels appear to be slightly overestimated when using the linearized observation operator. Figures 2c and 2d show histograms of ensemble perturbations in observation space when using full observation operator for the same observations.The histograms for these AMSU-A channels are somewhat non-Gaussian.

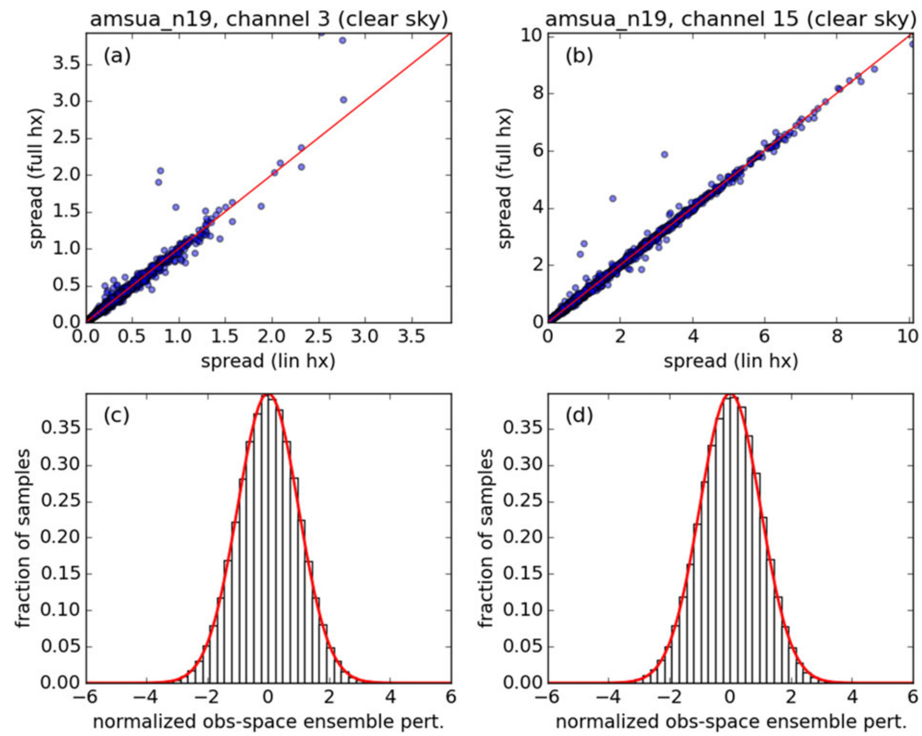

Figure 3. Comparison of clear-sky NOAA-19 AMSU-A channels 3 and 15 when using the linearized and full observation operators. (a, b) Scatterplots comparing observation space ensemble standard deviations when using linearized observation operator (*x* axis) and full observation operator (*y* axis). (c, d) Histograms of normalized ensemble perturbations in the observation space when using full observation operator; the red curve is the standard Gaussian distribution for comparison. The histograms were obtained by normalizing all of the ensemble perturbations $H(x)$ by the corresponding ensemble mean and standard deviation and plotting them together. NOAA = National Oceanic and Atmospheric Administration; AMSU-A = Advanced Microwave Sounding Unit-A.

The skewness in the histograms is due to the nonlinearity in the forward operator for cloud-affected radiances. Cloudy radiance assimilation was implemented in the operational NOAA system in 2016 (Zhu et al., 2016) and resulted in about 10% increase in the number of assimilated observations for AMSU-A channels 1–5 and 12% increase for channel 15. Figure 3 shows the same plots as Figure 2 but for only clear-sky radiances. One can see that if only clear-sky radiances are used, distributions are close to Gaussian and the observation space ensemble spread computed with the linearized observation operator is closer to the observation space ensemble spread when the full observation operator is used.

Most of the observation operators currently used are only slightly nonlinear which justifies using the linearized observation operator to calculate observation space ensemble perturbations in the EnKF.

4.2. Results of Data Assimilation Cycling Experiments

Data assimilation cycling experiments were run with T574 version of the spectral Global Forecast System to compare assimilation results using the two approaches. Unlike in the operational configuration, only EnSRF was used for data assimilation, and ensemble members were not recentered around hybrid ensemble variational analysis. We made this choice of the experiment configuration to intensify the effect of different ensemble perturbations in observation space through using full (nonhybrid) ensemble covariances in the assimilation. All available observations were assimilated: radiosonde, aircraft observation, satellite-derived winds, GPS bending angle, and radiances: AMSU-A, AIRS, ATMS, CrIS, HIRS/4, IASI, and MHS. The data assimilation was cycled from 1 January 2016 to 25 January 2016.

Figure 4 shows root-mean-square differences between the background forecast ensemble mean and the radiosonde wind and temperature observations, averaged over the last 2.5 weeks of the experiments. There is no significant difference between the innovations in the two experiments, which shows that neglecting

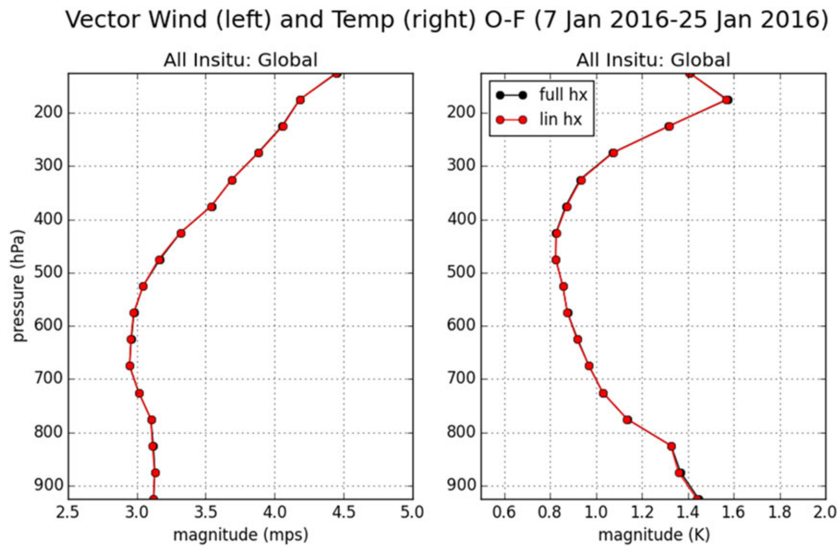

Figure 4. Root-mean-square differences between background ensemble mean forecast vector wind (left) and temperature (right) and radiosonde wind and temperature observations, averaged from 7 January 2016 to 25 January 2016.

the nonlinearity in satellite radiance and GPS bending angle observation operators when computing ensemble perturbations in observation space does not significantly affect the quality of the EnSRF analysis. Since both the EnSRF and LETKF algorithms assume linearization of the observation operator about the background ensemble mean, it is perhaps not surprising that using the full nonlinear observation operator to compute observation space ensemble perturbations does not yield better results than using the linearized operator.

Even though applying linearized observation operator requires extra computations in the EnKF, savings on running the observation operator only once (as opposed to $k+1$ times) allowed to decrease the overall assimilation cycle time on average by 15%. This number, however, depends on the specifics of the configuration: the choice of the number of processors to run observation operators and number of observation operators run in parallel.

The results of LETKF assimilation using linearized versus full observation operator for ensemble perturbations are similar to those of EnSRF presented in this section.

5. Conclusions

A version of the operational NOAA EnKF that uses linearized observation operators for calculating ensemble perturbations in observation space has been developed to increase computational efficiency of the data assimilation cycle. The computational savings are achieved by running the nonlinear observation operator only once for the ensemble mean and utilizing the observation operator Jacobian to apply the linearized observation operator to ensemble perturbations.

Comparison of ensemble spread in observation space calculated with the full versus linearized observation operators identified the most nonlinear satellite radiance observations: AMSU-A channels 1–4 and 15 when cloud-affected radiances are used. For most other satellites and channels used in data assimilation the nonlinearities are small.

Cycled data assimilation experiments showed no significant difference between using the linearized and full observation operator to calculate observation space ensemble perturbations in the EnKF.

Using the linearized observation operator makes it possible to apply the observation operator after the ensemble covariance localization, and so it will potentially allow for some EnSRF algorithm advancements, such as model space and scale-dependent localization.

Acknowledgments
This research is supported by NOAA/NWS Next Generation Global Prediction System (NGGPS) project. The observation diagnostics files and software used to generate Figures 1–3, as well as averaged diagnostics of cycled data assimilation experiments that are presented in section 4.2, are archived at ftp://ftp.cdc.noaa.gov/Datasets.other/ShlyaevaWhitaker_linhx_2018/.

References

Anderson, J. L., & Collins, N. (2007). Scalable implementations of ensemble filter algorithms for data assimilation. *Journal of Atmospheric and Oceanic Technology*, *24*(8), 1452–1463. https://doi.org/10.1175/JTECH2049.1

Buehner, M., McTaggart-Cowan, R., Beaulne, A., Charette, C., Garand, L., Heilliette, S., et al. (2015). Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. *Monthly Weather Review*, *143*(7), 2532–2559. https://doi.org/10.1175/MWR-D-14-00354.1

Clayton, A. M., Lorenc, A. C., & Barker, D. M. (2013). Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, *139*(675), 1445–1461. https://doi.org/10.1002/qj.2054

Cucurull, L., Derber, J. C., & Purser, R. J. (2013). A bending angle forward operator for global positioning system radio occultation measurements. *Journal of Geophysical Research: Atmospheres*, *118*, 14–28. https://doi.org/10.1029/2012JD017782

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., & Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, *139*(2), 511–522. https://doi.org/10.1175/2010MWR3328.1

Han, Y. (2006). JCSDA Community Radiative Transfer Model (CRTM): Version 1 (NOAA Tech. Rep. NESDIS). U.S. Department of Commerce National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service.

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 123–137. https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Houtekamer, P. L., & Mitchell, H. L. (2005). Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3269–3289. https://doi.org/10.1256/qj.05.135

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1), 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Kleist, D. T., & Ide, K. (2015). An OSSE-based evaluation of hybrid variational ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Monthly Weather Review*, *143*(2), 433–451. https://doi.org/10.1175/MWR-D-13-00351.1

Lei, L., & Whitaker, J. S. (2015). Model space localization is not always better than observation space localization for assimilation of satellite radiances. *Monthly Weather Review*, *143*(10), 3948–3955. https://doi.org/10. 1175/MWR-D-14-00413.1

Wang, X., & Lei, T. (2014). GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. *Monthly Weather Review*, *142*(9), 3303–3325. https://doi.org/10.1175/MWR-D-13-00303.1

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*(9), 3078–3089. https://doi.org/10.1175/MWR-D-11-00276.1

Zhu, Y., Liu, E., Mahajan, R., Thomas, C., Groff, D., Delst, P. V., et al. (2016). All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Monthly Weather Review*, *144*(12), 4709–4735. https://doi.org/10.1175/MWR-D-15-0445.1