9.4

Evaluating the Impact of Satellite Data Density within an Ensemble Data Assimilation Approach

Dusanka Zupanski[1], Louis Grasso[1], Mark DeMaria[2], Manajit Sengupta[1], and Milija Zupanski[1]
[1]Cooperative Institute for Research in the Atmosphere
Colorado State University
Fort Collins, Colorado, U. S. A.
[2] NOAA/NESDIS, Fort Collins, Colorado, U. S. A.

1. INTRODUCTION

Satellite observations from the current and future satellite missions (e.g., GOES-R, NPOESS, CloudSat, and GPM) will have increased spatial, temporal, and spectral resolution. The issues of how to optimally utilize information from these satellite measurements are of fundamental importance. In particular, capabilities of the current data assimilation methods to effectively assimilate dense satellite observations have to be evaluated. Novel methodologies for information content analysis, which are based on information theory and data assimilation, are especially useful for quantifying the impact of different satellite observations.

Ensemble based data assimilation methods can be effectively used for both data assimilation (e.g., Evensen 1994; Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Keppenne 2000; Mitchell and Houtekamer 2000; Anderson 2001; Bishop et al. 2001; van Leeuwen 2001; Reichle et al. 2002a,b; Whitaker and Hamill 2002; Tippett et al. 2003; Zhang et al. 2004; Ott et al. 2005; Szunyogh et al. 2005; Zupanski 2005; and Zupanski and Zupanski 2005) and information content analysis of satellite observations. For example, it was demonstrated in Wei et al. (2005) and Zupanski et al. (2005a,b; 2006) that the ensemble transform matrix of Bishop et al. (2001) can be employed to define information content measures, such as degrees of freedom (DOF) for signal and entropy reduction (e.g., Rodgers 2000). Similarly as in Zupanski et al. (2005a,b; 2006), in this research we employ an ensemble data assimilation approach (Maximum Likelihood Ensemble Filter, MLEF, Zupanski 2005; Zupanski and Zupanski 2005) to evaluate information content measures of simulated atmospheric observations. Unlike in the previous research, we will evaluate the impact of satellite data density on the information content measures.

* *Corresponding author address*: Dusanka Zupanski, Colorado State University/CIRA, Foothills Campus, Fort Collins, CO 80523 (e-mail: Zupanski@cira.colstate.edu)

2. METHODOLOGY

The MLEF seeks a maximum likelihood state solution employing an iterative minimization of a cost function. The solution for an augmented state vector $\boldsymbol{x}$ (including initial conditions, model error, and empirical parameters), of dimension N_{state}, is obtained by minimizing a cost function J defined as

$$J(\boldsymbol{x}) = \frac{1}{2}[\boldsymbol{x}-\boldsymbol{x}_b]^T \boldsymbol{P}_f^{-1}[\boldsymbol{x}-\boldsymbol{x}_b] + \frac{1}{2}[\boldsymbol{y}-H(\boldsymbol{x})]^T \boldsymbol{R}^{-1}[\boldsymbol{y}-H(\boldsymbol{x})] \quad , (1)$$

where $\boldsymbol{y}$ is an observation vector of dimension equal to the number of observations (N_{obs}) and H is a nonlinear observation operator. Subscript b denotes a background (i.e., prior) estimate of $\boldsymbol{x}$, and superscript T denotes a transpose. The $N_{obs} \times N_{obs}$ matrix $\boldsymbol{R}$ is a prescribed observation error covariance. The matrix $\boldsymbol{P}_f$ of dimension $N_{state} \times N_{ens}$ is the forecast error covariance (N_{ens} being the ensemble size).

Uncertainties of the optimal estimate of the state $\boldsymbol{x}$ are also calculated by the MLEF. The uncertainties are defined as square roots of the analysis error covariance ($\boldsymbol{P}_a^{\frac{1}{2}}$) and the forecast error covariance ($\boldsymbol{P}_f^{\frac{1}{2}}$), both defined in terms of ensemble perturbations. The square root of the analysis error covariance is obtained as

$$\boldsymbol{P}_a^{\frac{1}{2}} = \begin{bmatrix} \boldsymbol{p}_a^1 & \boldsymbol{p}_a^2 & \cdots & \boldsymbol{p}_a^{N_{ens}} \end{bmatrix} = \boldsymbol{P}_f^{\frac{1}{2}}(\boldsymbol{I}_{ens}+\boldsymbol{C})^{-\frac{1}{2}} \quad , (2)$$

where $\boldsymbol{I}_{ens}$ is a diagonal identity matrix of dimension $N_{ens} \times N_{ens}$, and $\boldsymbol{p}_a^i$ are column vectors representing analysis perturbations in ensemble subspace. Matrix ***C*** of dimension $N_{ens} \times N_{ens}$ is defined as

$$C = Z^T Z \quad ; \; z^i = R^{-\frac{1}{2}} H(x + p_f^i) - R^{-\frac{1}{2}} H(x) \; , (3)$$

where vectors z^i are columns of the matrix ***Z*** of dimension $N_{obs} \times N_{ens}$. Note that, when calculating z^i, a nonlinear operator *H* is applied to perturbed and unperturbed states ***x***. Vectors p_f^i are the columns of the square root of the forecast error covariance matrix obtained via ensemble forecasting employing a nonlinear dynamical model *M* (e.g., an NWP model)

$$P_f^{\frac{1}{2}} = \begin{bmatrix} p_f^1 & p_f^2 & \ldots & p_f^{N_{ens}} \end{bmatrix} \quad ;$$
$$p_f^i = M(x_a + p_a^i) - M(x_a) \; , \qquad (4)$$

where x_a is the optimal solution for the model state (analysis).

Equations (1)-(3), referred to as analysis equations, are solved iteratively in each data assimilation cycle, while equation (4), referred to as a forecast equation, is used to advance the columns of the forecast error covariance matrix $P_f^{\frac{1}{2}}$ from one cycle to another.

Measures of information content of observations referred to as DOF for signal and entropy reduction are often used in information theory (e.g., Rodgers 2000). In data assimilation applications these measures are commonly defined in terms of analysis and forecast error covariances, P_a and P_f, (e.g., Wahba 1985; Purser and Huang 1993; Wahba et al. 1995; Rodgers 2000; Rabier et al. 2002; Fisher 2003; Johnson 2003; Engelen and Stephens 2004). The information measures can also be calculated employing the eigenvalues λ_i^2 of thr matrix ***C***, defined in (3), that we also refer to as *the information matrix in ensemble subspace*. Thus, the following formulas for DOF for signal d_s and entropy reduction *h* can be used:

$$d_s = \sum_i \frac{\lambda_i^2}{(1+\lambda_i^2)} \; ; \; h = \frac{1}{2} \sum_i \ln(1+\lambda_i^2) \, , \qquad (5)$$

which are essentially the same formulas as in Rodgers (2000). The difference is that the eignevalues of the information matrix defined in ensemble subspace (***C***) are used in our formulation, while in the formulation of Rodgers (2000), the eigenvalues of the information matrix, defined either in the model space or in the observation space, are used. The advantage of the information matrix defined in ensemble subspace is that it is commonly a small matrix (of dimensions $N_{ens} \times N_{ens}$), so it is possible to evaluate the full eigenvalue spectrum of it, even when using complex NWP models and numerous observations. This property is especially appealing for calculating information content of numerous satellite observations. A potential disadvantage is that a small ensemble size might be insufficient to accurately determine the information measures. However, even relatively small ensemble sizes (e.g., 10-50 ensemble members) have been proven useful in determining information content measures (e.g., Wei et al. 2005; and Zupanski et al. 2005a,b, 2006).

We employ the methodology described above to determine the impact of satellite data density on the information content measures. To simulate satellite observations of various densities, we use the Colorado State University Regional Atmospheric Modeling System (RAMS, Pielke et al., 1992; Cotton et al., 2003), coupled with a radiative transfer model. Experimental results for a severe weather case will be presented at the conference. Implications of these results on combining information from multiple sensors of future satellite missions will also be discussed.

Acknowledgements

This research was supported by the GOES-R Risk Reduction Project under NOAA Grant NA17RJ1228. The views, opinions, and findings in this report are those of the authors and should not be construed as an official NOAA and or U. S. Government position, policy, or decision.

References

Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.,* **129**, 2884–2903.

Bishop, C. H., B. J. Etherton, and S. Majumjar, 2001: Adaptive sampling with the ensemble Transform Kalman filter. Part 1: Theoretical aspects. *Mon. Wea. Rev.,* **129**, 420–436.

Cotton, W.R., R.A. Pielke, Sr., R.L. Walko, G.E. Liston, C. J. Tremback, H. Jiang, R. L. McAnelly, J. Y. Harrington, M.E. Nicholls, G. G. Carrió and J. P. Mc Fadden 2003: RAMS 2001: Current Status and future directions. *Meteor. Atmos. Phys.*, **82**, 5-29.

Engelen, R. J., and Stephens G. L., 2004: Information Content of Infrared Satellite Sounding Measurements with Respect to CO_2. *J. Appl. Meteor.* **43**, 373–378.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, (C5), 10143-10162.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter/3D-variational analysis scheme. *Mon. Wea. Rev.,* **128**, 2905–2919.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126**, 796–811.

Johnson, C., 2003: Information content of observations in variational data assimilation. Ph.D. thesis, Department of Meteorology, University of Reading, 218 pp. [Available from University of Reading, Whiteknights, P.O. Box 220, Reading, RG6 2AX, United Kingdom.]

Keppenne, C., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.,* **128**, 1971–1981.

Mitchell, H. L, and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.,* **128**, 416–433.

Ott, Edward, B. R. Hunt, I. Szunyogh, A. Zimin, E. Kostelich, M. Corazza, E. Kalnay, D.J. Patil, and J. A. Yorke, 2005: A local ensemble kalman filter for atmospheric data assimilation. Posted http://arXiv.org/abs/physics/0203058. Submitted to *Tellus*

Pielke, R.A., W.R. Cotton, R.L. Walko, C.J. Tremback, W.A. Lyons, L.D. Grasso, M.E. Nicholls, M.D. Moran, D.A. Wesley, T.J. Lee, and J.H. Copeland, 1992: A comprehensive meteorological modeling system-RAMS. *Meteorol. Atmos. Phys.*, **49**, 69-91.

Purser, R.J. and Huang, H.-L. 1993: Estimating effective data density in a satellite retrieval or an objective analysis. *J. Appl. Meteorol.*, **32**, 1092–1107.

Rabier F., Fourrie N., Djalil C., and Prunet P., 2002: Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Quart. J. Roy. Meteor. Soc.,* **128,** 1011–1027.

Reichle, R. H., D. B. McLaughlin, D. Entekhabi, 2002a: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103–114.

Reichle, R.H., J.P. Walker, R.D. Koster, and P.R. Houser, 2002b: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometorology*, **3**, 728-740.

Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific, 238 pp.

Shannon, C. E., and Weaver W., 1949: *The Mathematical Theory of Communication*. University of Illinois Press, 144 pp.

Szunyogh, I., Kostelich E. J., Gyarmati G., Patil D. J., Hunt B. R., Kalnay E., Ott E., and Yorke J. A., 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the NCEP global model. Submitted to *Tellus*.

Tippett, M., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square-root filters. *Mon. Wea. Rev.*, 131, 1485–1490.

van Leeuwen, P. J., 2001: An ensemble smoother with error estimates. *Mon. Wea. Rev.,* **129**, 709–728.

Wahba, G., Johnson D. R., Gao F., and Gong J., 1995: Adaptive tuning of numerical weather prediction models: Randomized GCV in three- and four-dimensional data assimilation. *Mon. Wea. Rev.*, **123**, 3358–3370.

Wei, M., Toth, Z., Wobus R., Zhu, Y., Bishop C. H., and Wang, X., 2005: ETKF-based ensemble perturbations using real-time observations at NCEP. *Tellus* (in press).

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130**, 1913–1924.

Zupanski D. and M. Zupanski, 2005: Model error estimation employing ensemble data assimilation approach. *Mon. Wea. Rev*. (in press).

Zupanski D., Zupanski M., DeMaria M., and Grasso L., 2005a: Critical issues of ensemble data assimilation in application to GOES-R risk reduction program. Extended abstracts of the *AMS Ninth Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, 9-13 January 2005, San Diego, CA, 2 pp.

Zupanski D., Zupanski, M., DeMaria, M., Grasso L., Hou, A.Y., Zhang, S., and Lindsey, D., 2005b: Ensemble data assimilation and information theory. Extended abstracts of the *AMS 21st Conference on Weather Analysis and Forecasting and AMS 17th Conference on Numerical Weather Prediction*, 1–5 August 2005, Washington, D.C., 4pp.

Zupanski D., M. Zupanski, A.Y. Hou, S. Zhang, C.D. Kummerow, and S.H. Cheung, 2006: Information theory and ensemble data assimilation. Submitted to *J. Atmos. Sci.* [also available at ftp://ftp.cira.colostate.edu/Zupanski/manuscripts/Geos5_Information_JAS.pdf]

Zupanski, M., 2005: Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon. Wea. Rev.*, **133**, 1710–1726.