PREPARATIONS FOR GOES-R BRIGHTNESS TEMPERATURE ASSIMILATION: WHY ENSEMBLE DATA ASSIMILATION?

Dusanka Zupanski[1], Louie Grasso[1], Mark DeMaria[2], Arthur Y. Hou[3], and Sara Zhang[3]

[1]*Colorado State University/Cooperative Institute for Research in the Atmosphere, Fort Collins, CO*
[2]*NOAA/National Environmental Satellite, Data, and Information Service, Fort Collins, CO*
[3]*NASA Goddard Space Flight Center/Global Modeling and Assimilation Office, Greenbelt, MD*

Goals

- To reduce risk of the GOES-R mission: Develop and test data assimilation techniques, which can extract maximum information from the future GOES-R observations.
- Examine advantages and disadvantages of different data assimilation approaches (variational, Kalman filter, ensemble-based approaches).
- Focus on information content analysis of satellite observations similar to future GOES-R data (e.g., 10.35 µm brightness temperature).
- Experimental results employing CSU-RAMS atmospheric model and a single column version of the GEOS-5 AGCM are presented.

Why ensemble data assimilation?

Three main reasons :

- ❑ Need for optimal estimate of the atmospheric state + **verifiable uncertainty** of this estimate;
- ❑ Need for **flow-dependent** forecast error covariance matrix; and
- ❑ The above requirements should be applicable to **most complex atmospheric models** (e.g., non-hydrostatic, cloud-resolving, LES).

Are there alternatives?

Two good candidates:

- ❑ **4d-var** method: It employs flow-dependent forecast error covariance, but it does not propagate it in time.
- ❑ **Kalman Filter** (KF): It does propagate flow-dependent forecast error covariance in time, but it is too expensive for applications to complex atmospheric models.

⇓

EnKF is a practical alternative to KF, applicable to most complex atmospheric models.

⇓

A bonus benefit: EnKF does not use adjoint models!

Methodology

Maximum Likelihood Ensemble Filter (MLEF)

(M. Zupanski 2005, MWR; D. Zupanski and M. Zupanski 2006, MWR)

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

$\boldsymbol{x}-\boldsymbol{x}_b=\boldsymbol{P}_f^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-1/2}\boldsymbol{\zeta}$ - Change of variable (preconditioning)

$\boldsymbol{C}=\boldsymbol{Z}^T\boldsymbol{Z}$ - ***C*** is information matrix of dim $Nens\times Nens$

$z^i=R^{-1/2}H[M(x+p_f^i)]-R^{-1/2}H[M(x)]$ - z^i are columns of ***Z***

$p_f^i=M(x+p_a^i)-M(x)$ - $\boldsymbol{p}^i{}_f$ and $\boldsymbol{p}^i{}_a$ are columns of $\boldsymbol{P}_f$ and $\boldsymbol{P}_a$

$\boldsymbol{y}_{obs}$ - Observations vector of dim $Nobs$

$\boldsymbol{x}$ - Model state vector of dim $Nstate>>Nens$

ζ - Control vector in ensemble space of dim $Nens$

$\boldsymbol{x}_n=M_{n,n-1}(\boldsymbol{x}_{n-1})$ - Dynamical forecast model

$y_n=H_n(\boldsymbol{x}_n)$ - Observation operator

Useful properties of the MLEF

- For N_{ens} = N_{state} (full-rank problem), and linear *M* and *H*, **MLEF = classical Kalman filter.**
- For N_{ens} = N_{state}, non-linear *M* and *H*, constant P_f, and assuming ideal Hessian preconditioning in 3d-var, **MLEF = 3d-var.**

⇓

Consistent and easy comparisons of different data assimilation methods (within the same program codes)!

MLEF can be used to calculate **Degrees of freedom** (DOF) **for signal** (d_s), as in Rodgers (2000), in ensemble subspace:

$$d_s=\sum_{i=1}^{N_{ens}}\frac{\lambda_i}{1+\lambda_i}\quad;\quad \lambda_i \text{ -eigenvalues of } \boldsymbol{C}$$

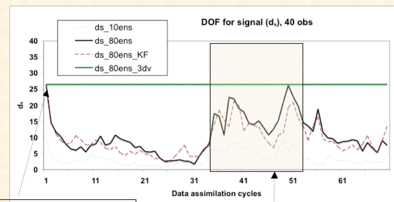
Matrix ***C*** is of relatively small dimensions (N_{ens} x N_{ens}), even for complex atmospheric models and numerous observations.

⇓

It is practical to calculate information content of millions of satellite observations (channels).

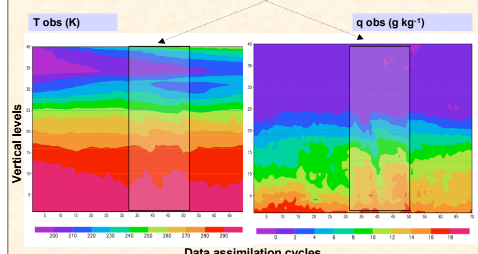
Experiment 1: Evaluate different data assimilation methods (MLEF, KF, and 3d-var) for the purposes of information content analysis of future GOES-R data.

- Single column version of the GEOS-5 AGCM (includes full physics package)
- 40 level model, two control variables: ***T*** and ***q***
- 10 and 80 ensembles, 40 obs of ***T*** and ***q*** per data assimilation cycle (partly observed system)
- 6-h data assimilation interval
- ARM observations used as forcing
- Model simulated "observations" with random noise
- ***H=I*** (identity operator), ***R***=const in all cycles



First cycle always indicates high information measures

More information in later cycles due to dynamics

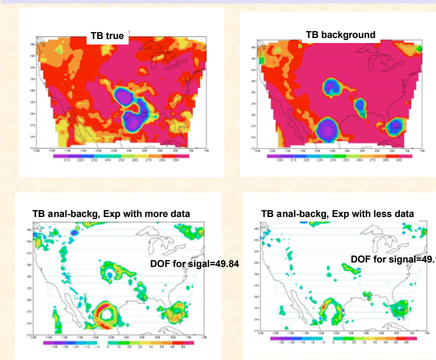


Results summary

- The MLEF and KF diagnose dynamically dependent information measures because of the update of the forecast error covariance. DOF for signal captures changes in the model state (***T*** and ***q***).
- Small ensemble size (10 ens), even though it does not produce perfect analysis, it captures well main data signals.
- DOF for signal obtained employing the 3d-var method is insensitive to the actual changes in the model state, because the forecast error covariance is kept constant in all data assimilation cycles.
- ***Forecast error covariance update is crucially important for information content analysis ⇒ Ensemble-based methods are desirable for evaluation of information measures of GOES-R data.***

Experiment 2: Employ the MLEF to assimilate simulated 10.35 µm brightness temperature data

- **CSU-RAMS** with 2-moment microphysics
- N_{state}=3564000 (includes CCNs), N_{ens}=50
- Δx=50km, 60 vertical levels
- 1-h data assimilation interval
- Hurricane Lili case
- Preliminary results from first data assimilation cycle



Results summary

- The MLEF corrects the background field to eliminate incorrectly positions clouds.
- Creating new clouds is more difficult: it requires a spin-up time (more data assimilation cycles).
- As in GEOS SCM experiment the DOF for signal is approximately equal to the ensemble size in first data assimilation cycle. Nevertheless, using more observations results in an increase of DOF for signal.

Future

Evaluate information measures of brightness temperature observations in consecutive data assimilation cycles, and examine the impact of forecast error covariance update.

Acknowledgements

This research was supported by the GOES-R Risk Reduction Project under NOAA Grant NA17RJ1228 and by NASA grants 621-15-45-78, NAG5-12105 and NNG04GI25G. The views, opinions, and findings in this report are those of the authors and should not be construed as an official NOAA, NASA, and or U. S. Government position, policy, or decision. Computational resources are provided by Columbia supercomputer at Ames Research Center, by halem Linux cluster computer at Goddard Space Flight Center, and by CIRA Linux cluster at Colorado State University.

References

Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific, 238 pp.
Zupanski D. and M. Zupanski, 2006: Model error estimation employing ensemble data assimilation approach. *Mon. Wea. Rev*., **134**, 1337-1354.
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.*
Zupanski, M., 2005: Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon. Wea. Rev*., **133**, 1710–1726.

Zupanski@cira.colostate.edu