2014 CAHMDA & HEPEX-DAFOH Workshop
September 8-12, 2014
The University of Texas at Austin

Advancing Data Assimilation Science for Operational Hydrology: Methodology, Computation, and Algorithms

Milija Zupanski
Cooperative Institute for Research in the Atmosphere
Colorado State University
[http://www.cira.colostate.edu/ensemble/]

Acknowledgements

- Sara Zhang (NASA GSFC)
- Prof. I. Michael Navon (Florida State University)
- *NASA GPM*: NNX10AG92G
- *NASA MAP*: NNX13AO10G
- *NOAA OAR*: NA14OAR4830122
- NSF CMG: ATM-0930265
- NCAR Computational and Information System Laboratory, Yellowstone
- NASA Advanced Supercomputing at NASA Ames, Pleiades

Outline

✧Challenges of data assimilation for operational hydrology

✧Maximum Likelihood Ensemble Filter (MLEF) – ensemble-variational method

✧Some MLEF results from atmospheric applications

✧Potential benefits of coupled data assimilation

✧Future development

Challenges of DA for operational hydrology: Methodology

- Multi-component control variable
- Error covariance / uncertainty
- Nonlinearity and non-differentiability
  - processes
  - observations
- High dimensionality
- Computations
- Algorithm efficiency and robustness

Multi-component control variable

- Empirical parameters
- Initial conditions
- Systematic model error
- Forcing (e.g., precipitation)

- **State vector (*x*)**

- A (smallest) subset of variables defining a dynamical/physical system
- Typically it refers to the initial conditions only
- In general, it may include initial conditions, model errors, and empirical parameters

$$x = \begin{pmatrix} p & T & wind & q_{cloud} & O3 & q_{soil} & T_{soil} & param_1 & param_2 & \cdots \end{pmatrix}^T$$

$$p = \begin{pmatrix} p_1 & \ldots & p_N \end{pmatrix}^T \quad T = \begin{pmatrix} T_1 & \ldots & T_N \end{pmatrix}^T \quad \ldots$$

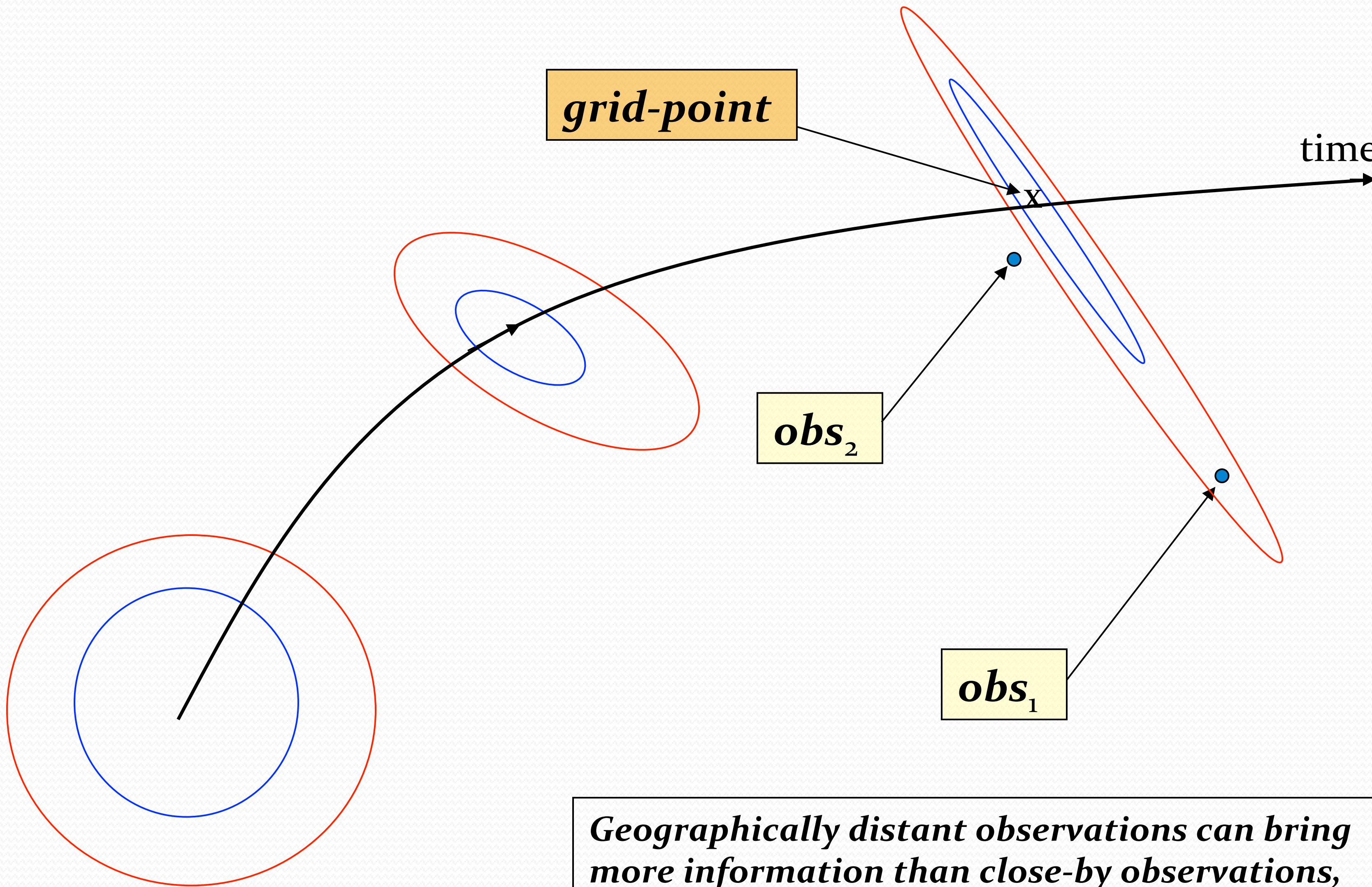
- From mathematical and algorithmic points of view there is nothing different
- However, parameters/model error require a model for uncertainty growth

Practical data assimilation algorithms: Basic methods suitable for operations

Variational DA	Ensemble DA
Forecast uncertainty pre-defined, static	Forecast uncertainty is flow-dependent, ensemble-based
Forecast uncertainty has all required degrees of freedom	Reduced number of degrees of freedom
Maximum a-posteriori estimate	Minimum variance
Iterative minimization	Linear KF solution
Employs adjoint (e.g., transpose) operator	No need for adjoint operator, use difference of nonlinear functions

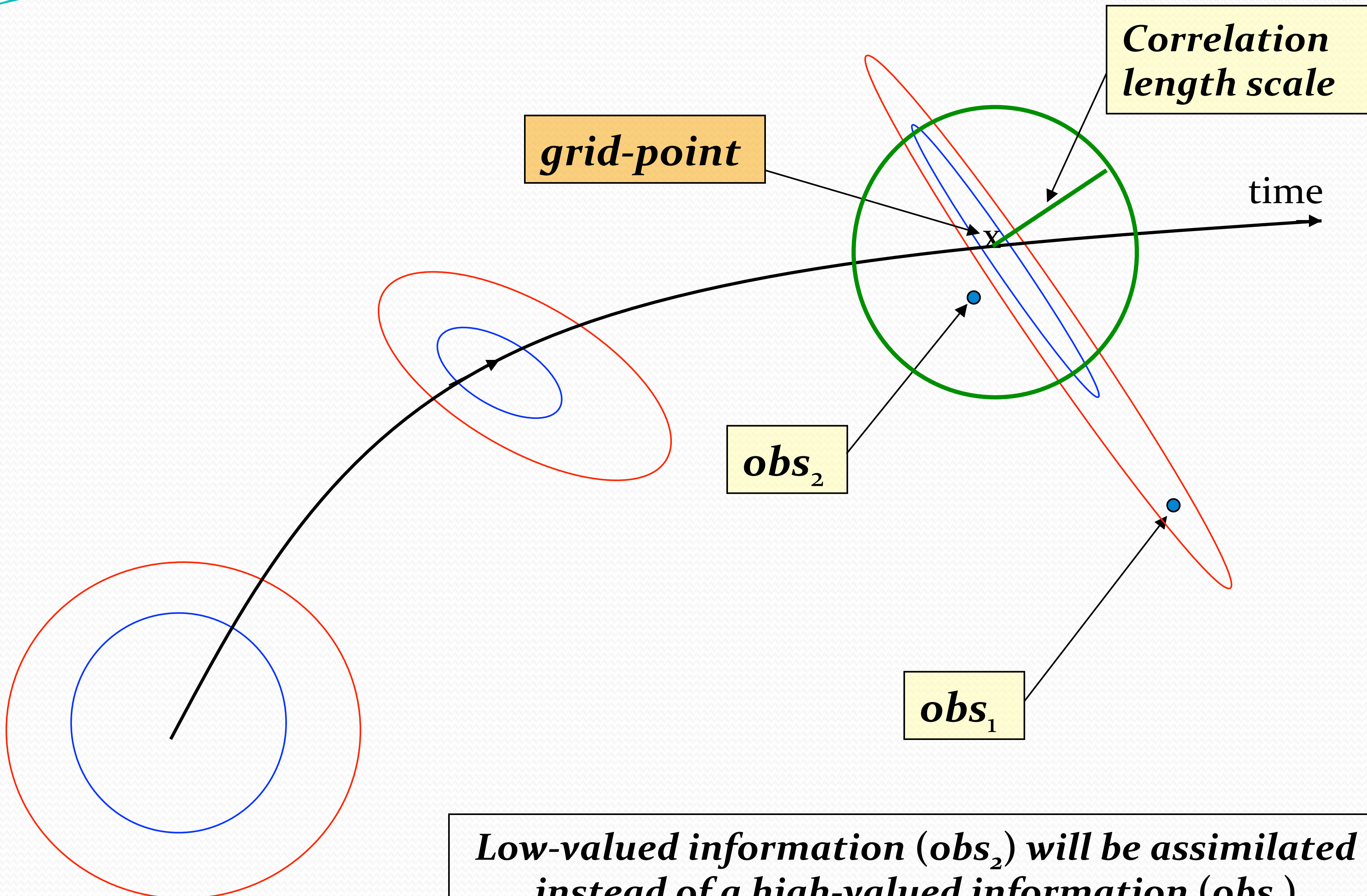
Hybrid variational-ensemble methods are used in weather operations

Flow-dependent forecast error covariance



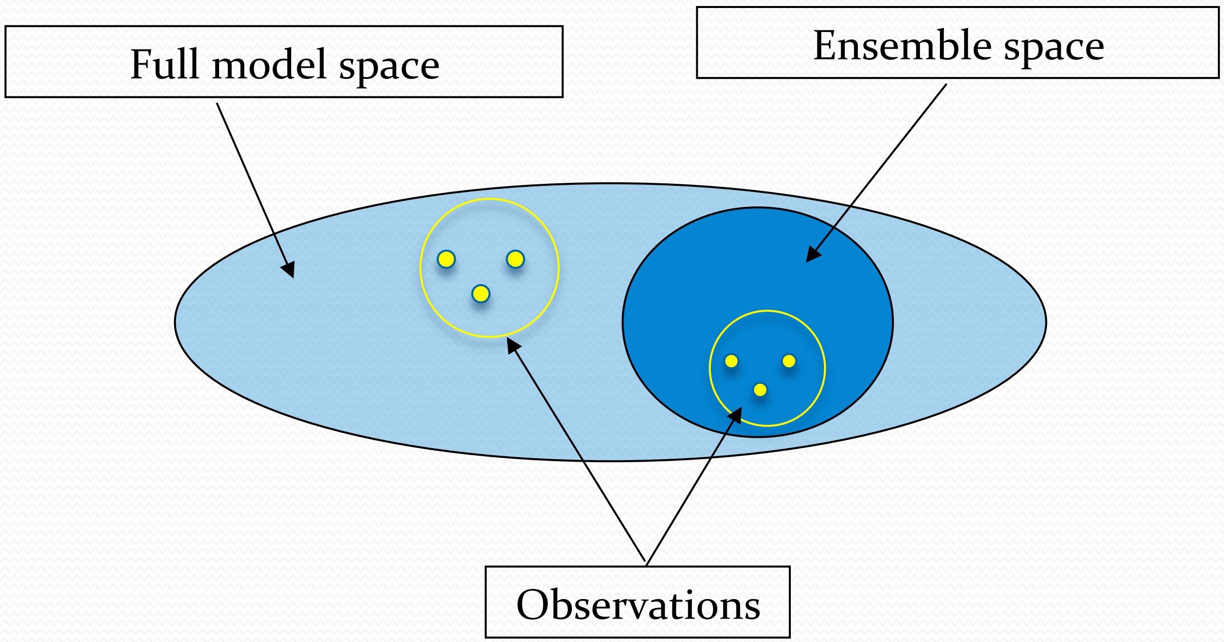
Geographically distant observations can bring more information than close-by observations, if in a dynamically significant region

Impact of static error covariance



Low-valued information (obs_2) will be assimilated instead of a high-valued information (obs_1)

Insufficient rank of forecast error covariance in ensemble methods



Model space dimensions ~ $O(10^7)$

Ensemble space dimensions ~ $O(10^1) - O(10^2)$

Observation outside ensemble space cannot be assimilated!

Role of forecast error covariance

Forecast error covariance plays a fundamental role in data assimilation

$$x^a - x^f = P_f H^T \left(H P_f H^T + R \right)^{-1} \left[y - h(x) \right] = P_f z_{obs}$$

Singular Value Decomposition (SVD):

$$P_f^{1/2} = V \Sigma W^T = \sum_i \sigma_i v_i w_i^T$$

$$x^a - x^f = \left(\sum_i \sigma_i^2 v_i v_i^T \right) z_{obs} = \sum_i \mu_i v_i$$

$$\mu_i = \sigma_i^2 v_i^T z_{obs}$$

- Analysis update is defined in the subspace spanned by forecast error covariance SVs

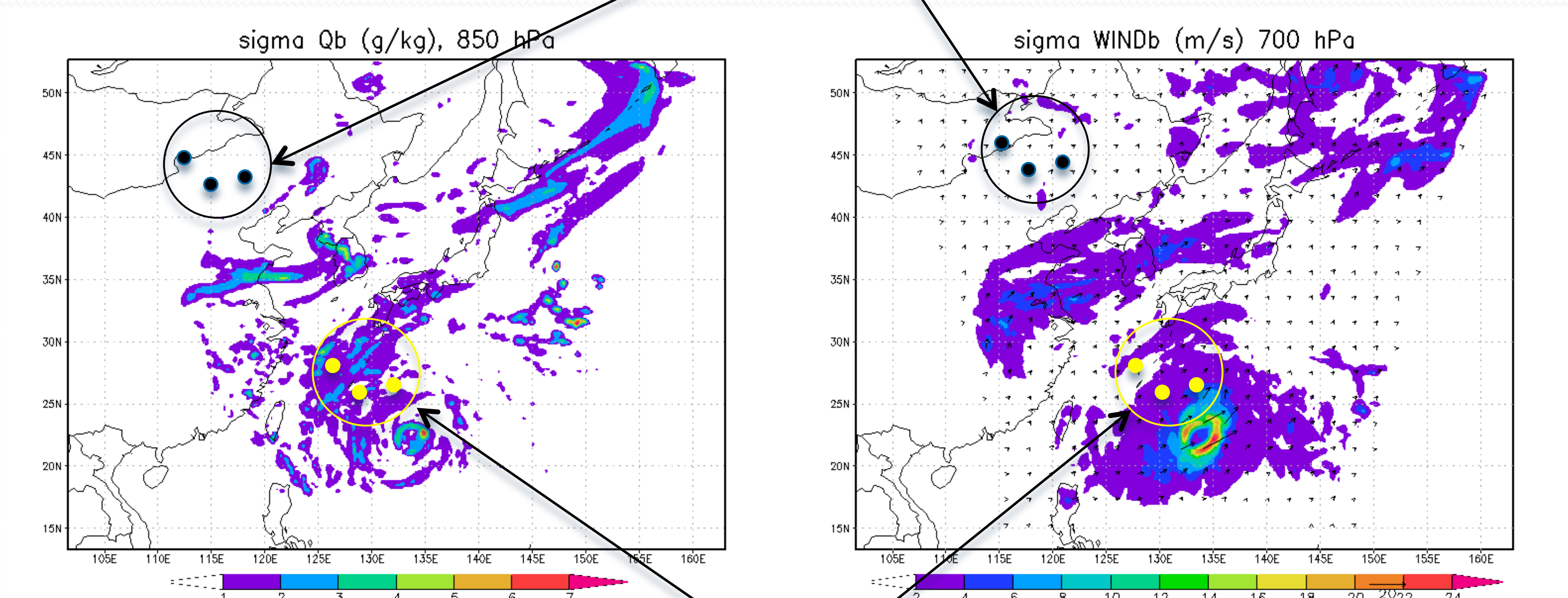
- Transformed observation increments z_{obs} need to have a projection on SVs
- Uncertainty magnitude has to be non-negligible

Forecast uncertainty – 32 ensembles (Typhoon Nobi, valid 03 Sep 2005 0300 UTC)

Insufficient forecast uncertainty prevents successful assimilation

Specific humidity (g/kg)

Wind (m/s)



Sufficient forecast uncertainty makes possible successful assimilation

Nonlinearity (and non-differentiability)

- Physical processes and observation operators are nonlinear
- Closed form solution does not exist for nonlinear DA
- Common approach to nonlinearity is to use iterative minimization

  - *constrained*: Gauss-Newton, Levenberg-Marquardt, …
  - *unconstrained*: Conjugate-gradient, Quasi-Newton, …

$$x_{k+1} = x_k + \alpha_k d_k$$

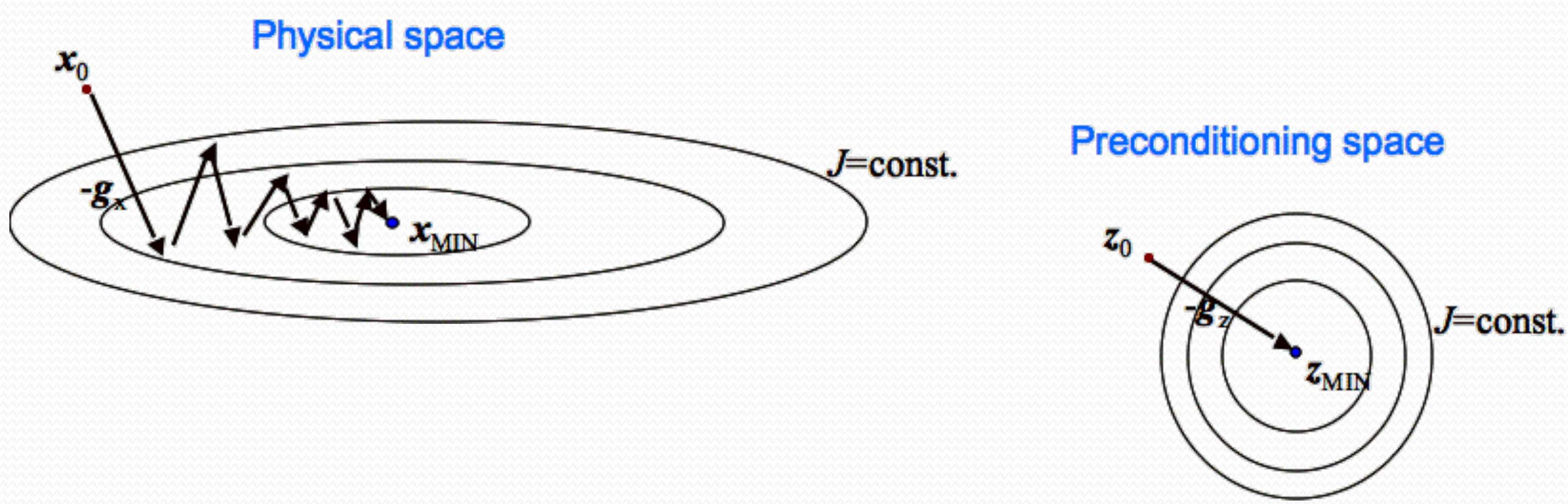
$$G d_k = -g_k$$

- Choose minimization algorithm adequate for the problem
  - use non-smooth algorithms if function/gradient discontinuities (e.g., LMBM)
  - use genetic/simulated annealing algorithms if multi-modal pdf
- Compromise between accuracy and efficiency

Hessian preconditioning

(Hessian matrix = second derivative of the cost function)

- **Optimal Hessian preconditioning:**
  - Improves minimization efficiency
  - Improves the accuracy (e.g., avoids error saturation)
  - Increases the robustness of minimization



Convergence is independent of the first guess in the transformed space

Computation: High dimensionality impacts the calculation of matrix inverse, thus Hessian preconditioning

(1) variational: neglect "difficult" matrix in inversion and apply *nonlinear* iterative solution method

$$\left[P_f^{-1} + H^T R^{-1} H\right]^{-1} \approx P_f \qquad \Rightarrow \qquad x_{k+1} = x_k + \alpha_k P_f H^T R^{-1}(y - h(x_k))$$

(2) ensemble: use reduced rank (RR) matrix inversion and compute *linear* solution

$$\left(P_f^{-1} + H^T R^{-1} H\right)^{-1} \approx \left[\left(P_f^{-1} + H^T R^{-1} H\right)^{-1}\right]_{RR} \qquad \Rightarrow \qquad x = x^f + \left[\left(P_f^{-1} + H^T R^{-1} H\right)^{-1} H^T R^{-1}\right]_{RR} [y - h(x^f)]$$

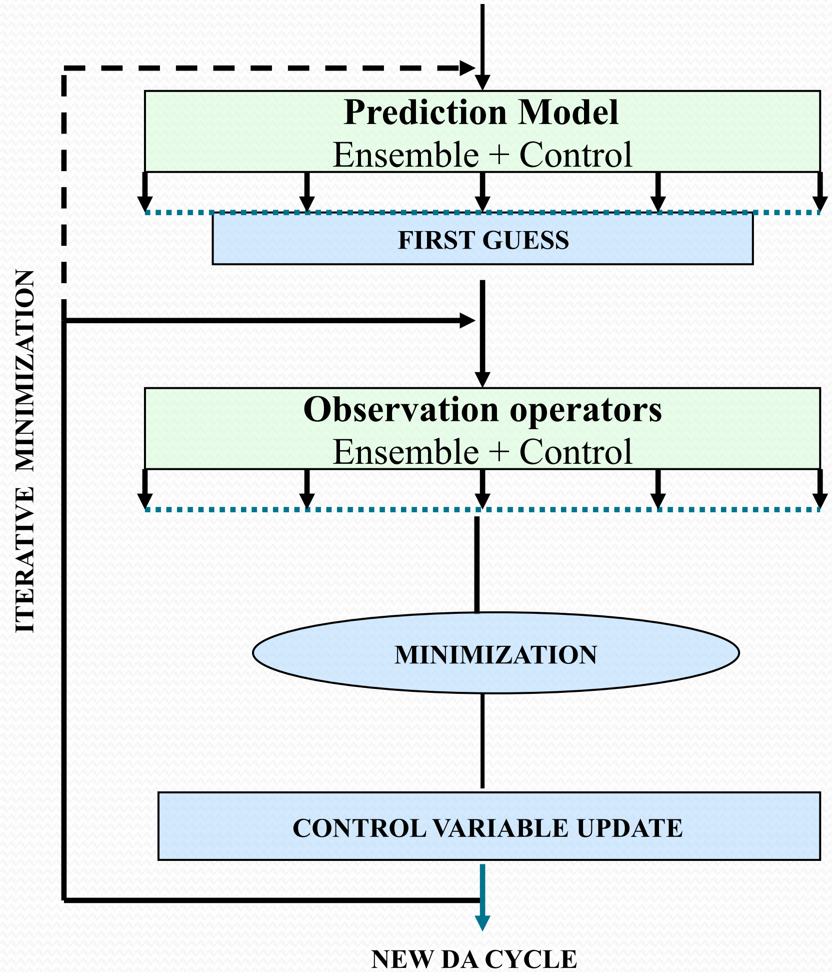
(3) reduced rank hybrid: reduced rank matrix inversion and *nonlinear* iterative solution method

$$\left[P_f^{-1} + H^T R^{-1} H\right]^{-1} \approx \left[\left(P_f^{-1} + H^T R^{-1} H\right)^{-1}\right]_{RR} \qquad \Rightarrow \qquad x_{k+1} = x_k + \alpha_k \left[\left(P_f^{-1} + H^T R^{-1} H\right)^{-1} H^T R^{-1}\right]_{RR} [y - h(x_k)]$$

Computational overhead ultimately impacts the choice of DA methodology

A hybrid data assimilation method: Maximum Likelihood Ensemble Filter (MLEF)

- ❑ Use optimal Hessian preconditioning
- ❑ Employ *most adequate* nonlinear iterative minimization algorithm
- ❑ Modular algorithm structure facilitates using a variety of models and observation operators
- ❑ Applicable to nonlinear and high-dimensional problems



MLEF algorithm

Forecast: Evolve uncertainty in time with ***nonlinear*** dynamical model *m*

$$x^f = m(x^a) \qquad\qquad x_i^f = m(x^a + p_i^a)$$

$$p_i^f = m(x^a + p_i^a) - m(x^a)$$

Analysis: Minimize ***arbitrary nonlinear*** cost function

$$f(x) = \frac{1}{2}\left(x - x^f\right)^T P_f^{-1}\left(x - x^f\right) + \frac{1}{2}\left(y - h(x)\right)^T R^{-1}\left(y - h(x)\right)$$

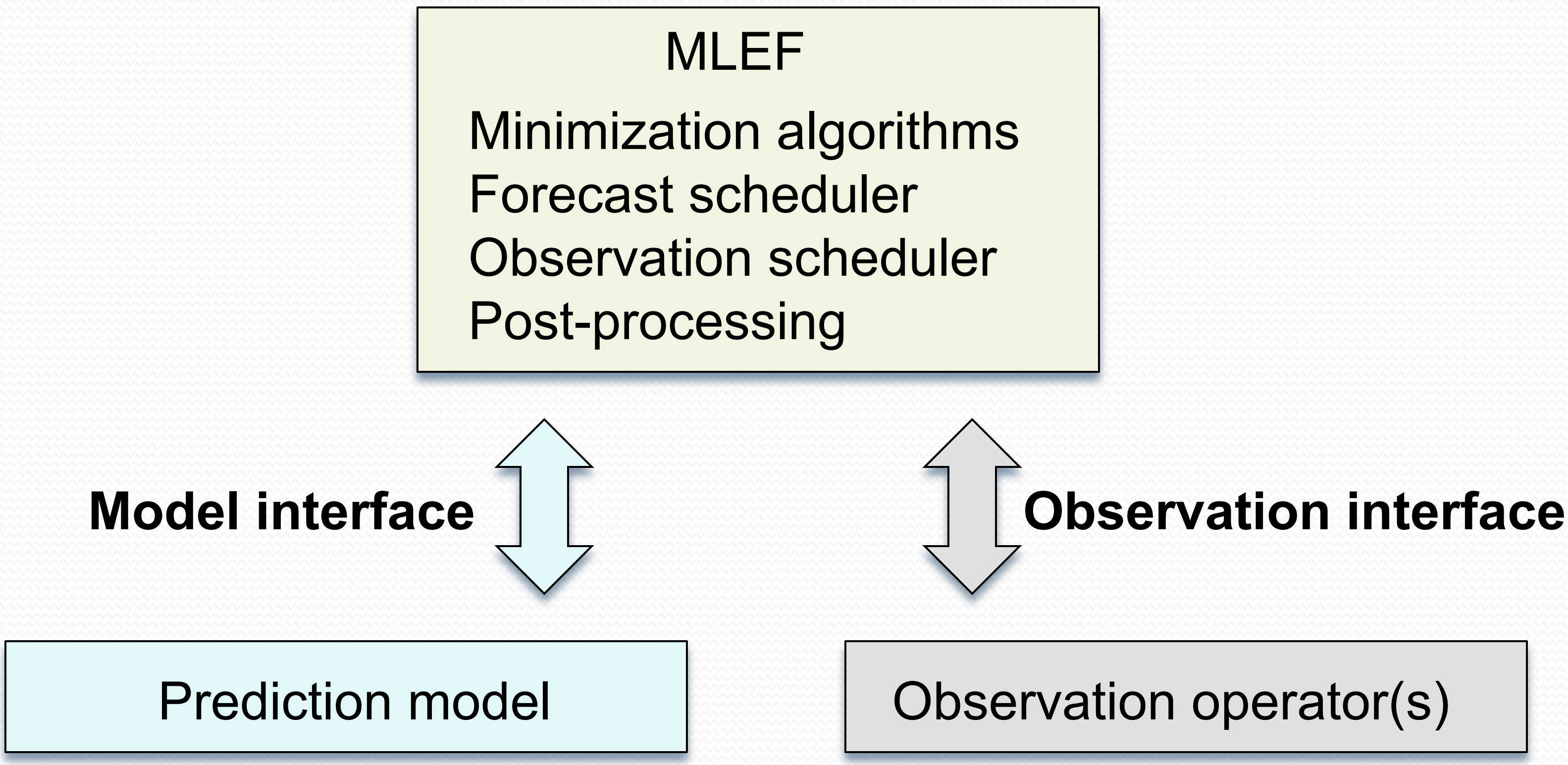
$$x_{k+1} = x_k + \alpha_k d_k$$

Analysis error covariance estimated from the inverse Hessian at the minimum

- ❑ **Reduced rank for high-dimensional state**
- ❑ **Full-rank for low-dimensional state**

Modular algorithm

- User-friendly compilation and experiment specifications
- MPI – optional
- Fortran 90/95 - based

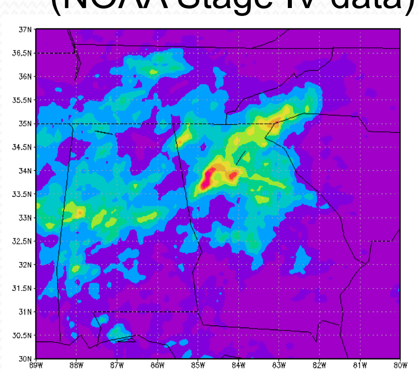


NASA Global Precipitation Mission: Downscaling satellite precipitation information using ensemble data assimilation

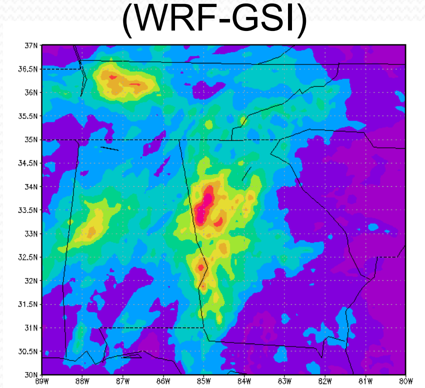
- NASA GPM: Downscaling satellite precipitation information using ensemble data assimilation
- Assimilate precipitation-affected microwave satellite radiances (TMI, AMSU-A/B, AMSR-E, MHS) and NOAA operational observations
- Cloud-scale data assimilation with NASA WRF model (27-9-3 km) and GSI/SDSU observation operator (S. Zhang et al. 2013, *MWR*)

Surface precipitation short-term forecasts verification
(accumulated during 15-22 Sep 2009 in the southeast US flood region)

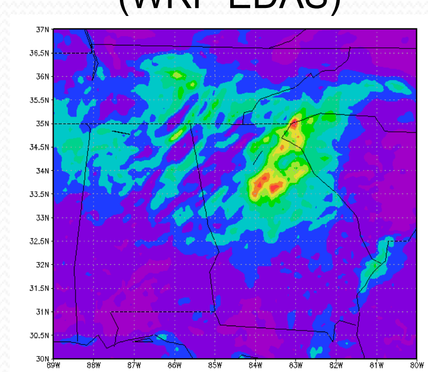
Ground-based Verification
(NOAA Stage IV data)



3DVAR, no AMSR-E,TMI
(WRF-GSI)



MLEF, with AMSR-E, TMI
(WRF-EDAS)

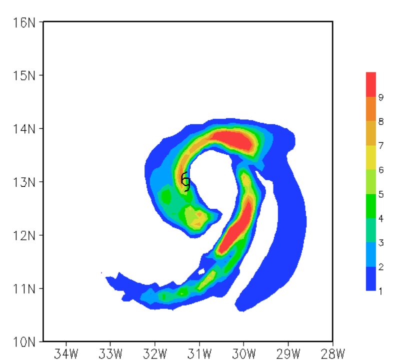


All-sky MSG SEVIRI (infrared) assimilation: Hurricane Fred (2009)

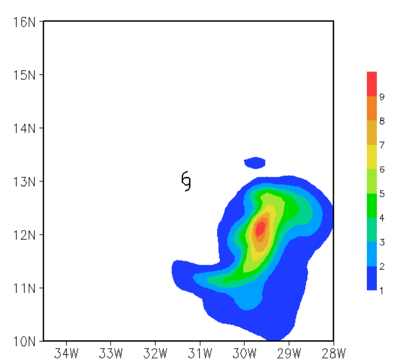
- JCSDA and NOAA GOES-R: Assimilation of all-sky infrared satellite radiances in hurricane core area
- NOAA hurricane WRF (HWRF) model (2011) (inner nest at 9 km) and GSI/CRTM
- 1-hour assimilation interval

Analysis of clouds (e.g., cloud condensate)
(hurricane Fred (2009), M. Zhang et al. 2013)

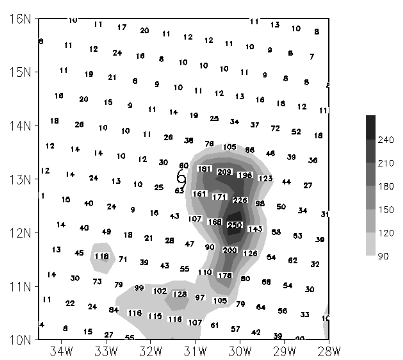
Cloud condensate: Clear-sky



Cloud condensate: All-sky

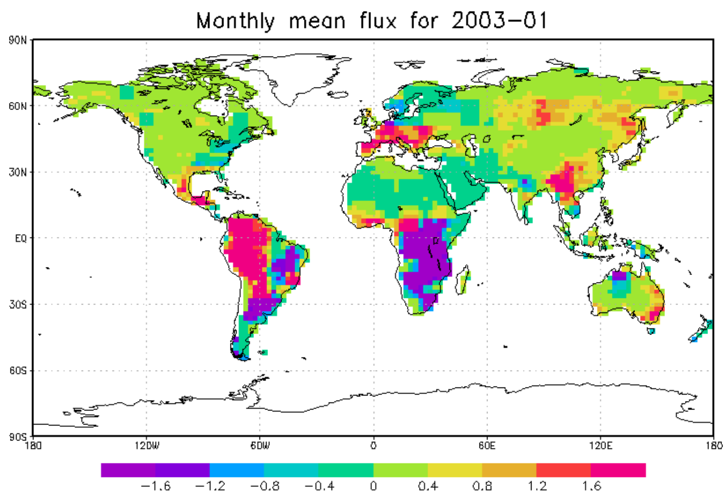


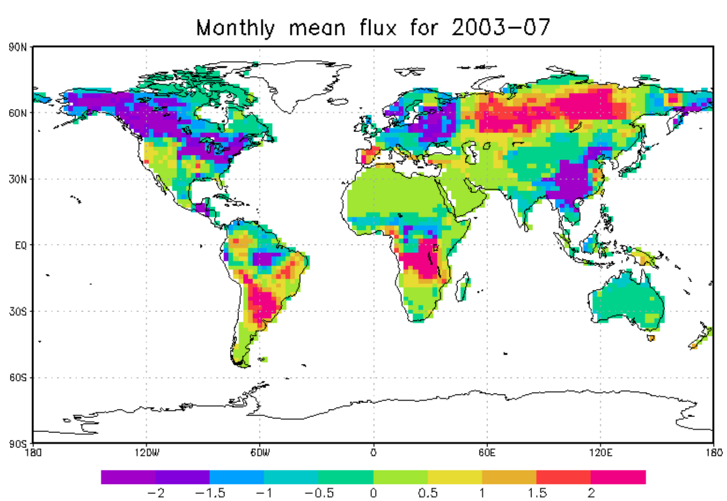
Verification: AMSU-A NOAA-16 retrieved cloud liquid water



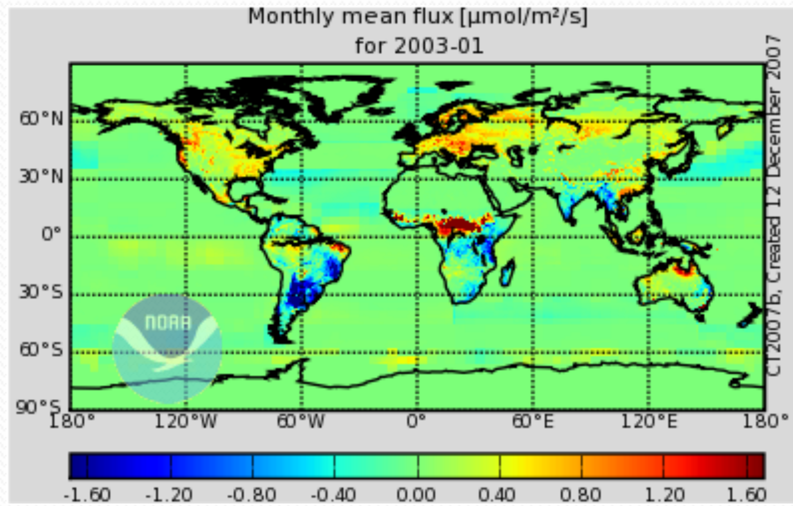
Carbon data assimilation - comparison of monthly mean fluxes (*Lokupitiya et al. 2008, JGR*)

MLEF

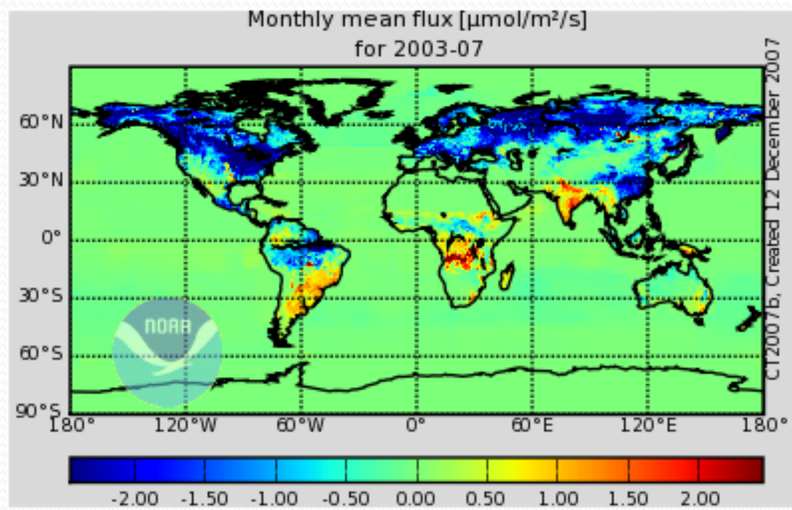




Carbon Tracker



JAN



JUL

Coupled DA: Uncertainty and Information

- Two-component coupled system with variables X_1 and X_2

Mutual information

$$I(X_1, X_2) = H(X_1) + H(X_2) - H(X_1, X_2)$$

Shannon Entropy

$$H\{X\} = -\int p(x)\log[p(x)]dx$$

$$I(X_1, X_2) \leq I(X_1, X_1) + I(X_2, X_2)$$

- ❑ Interpretation: *There are fewer degrees of freedom in a coupled system than in the sum of separate systems*
- ❑ This improves the capability of ensemble coupled DA – since fewer ensembles are needed

2-point DA coupled atmosphere-land system with single-point atmospheric observation

Forecast error covariance

$$P_f = \begin{pmatrix} (\sigma_f^2)_{atm} & \rho_{atm,land} \\ \rho_{atm,land} & (\sigma_f^2)_{land} \end{pmatrix}$$

Atmosphere-Land correlation $\rho_{atm,land}$

De-coupled analysis solution $\rho_{atm,land} = 0$

$$x_{atm}^a = \frac{1}{1+\varepsilon_{atm}^2} x_{atm}^f + \frac{\varepsilon_{atm}^2}{1+\varepsilon_{atm}^2} y_{atm}$$

$$\varepsilon_{atm}^2 = \frac{(\sigma_f^2)_{atm}}{(\sigma_R^2)_{atm}}$$

$$x_{land}^a = x_{land}^f$$

- **Weak coupling**: Coupled forecast, de-coupled DA
- **Atmospheric observation cannot improve land analysis (IC)**

2-point DA coupled atmosphere-land system with single-point atmospheric observation

Forecast error covariance

$$P_f = \begin{pmatrix} (\sigma_f^2)_{atm} & \rho_{atm,land} \\ \rho_{atm,land} & (\sigma_f^2)_{land} \end{pmatrix}$$

Atmosphere-Land correlation $\rho_{atm,land}$

Coupled analysis solution $\rho_{atm,land} \neq 0$

$$x_{atm}^a = \frac{1}{1+\varepsilon_{atm}^2} x_{atm}^f + \frac{\varepsilon_{atm}^2}{1+\varepsilon_{atm}^2} y_{atm}$$

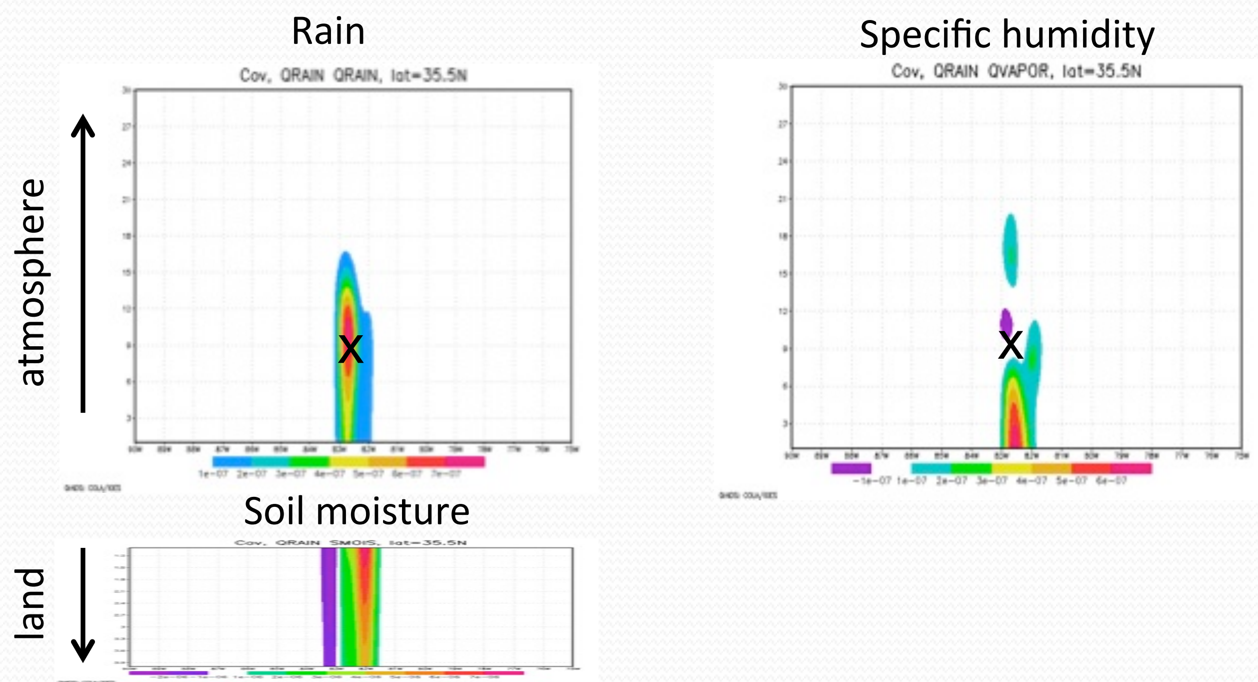
$$\varepsilon_{atm}^2 = \frac{(\sigma_f^2)_{atm}}{(\sigma_R^2)_{atm}}$$

$$x_{land}^a = x_{land}^f + \frac{1}{1+\varepsilon_{atm}^2} \frac{\rho_{atm,land}}{(\sigma_R^2)_{atm}} (y_{atm} - x_{atm}^f)$$

- **Strong coupling**: Coupled forecast, coupled DA
- **Atmospheric observation can improve land analysis (IC)**

Atmosphere-land coupled data assimilation: WRF-NOAH model

- ❑ NASA MAP: Improve impact of cloud and precipitation estimation on land surface
- ❑ NASA Atmosphere-land-chemistry coupled model (**NASA-Unified WRF** – 9km)
- ❑ Evaluate ensemble cross-variable error covariance
- ❑ Analysis response to single pseudo-observation of cloud rain water at 700 hPa

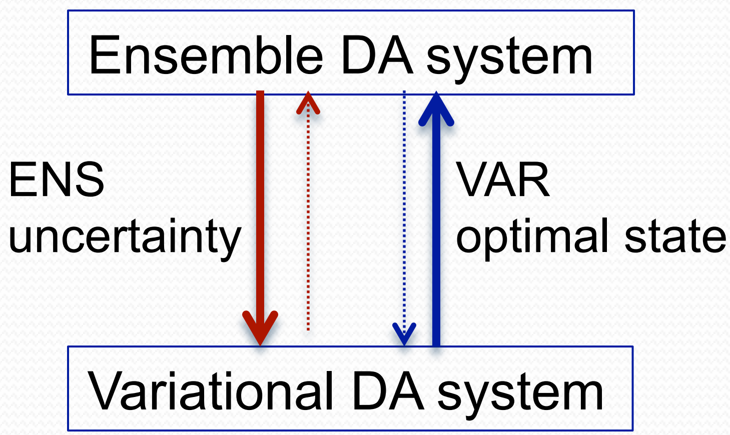


Coupled model history contained in forecast error covariance
➔ instant benefit for DA

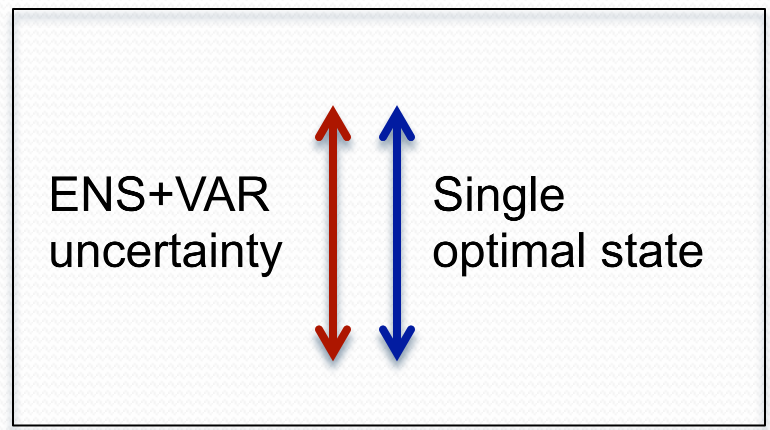
New development: Addressing insufficient rank of forecast error covariance

✧ A typical remedy is hybrid variational-ensemble data assimilation: *combine ensemble and variational error covariances* $P_{HYB} = f(P_{ENS}, P_{VAR})$

Hybrid VAR-ENS DA



New MLEF



One-way interaction due to:
- *Separate* VAR and ENS DA systems
- *Sub-optimal* Hessian preconditioning

Two-way interaction:
- *Single* DA system
- *Optimal* Hessian preconditioning

General spatiotemporal approach: n-dimensional MLEF algorithm

- **n - dimensional control variable and uncertainty**
  - Allow simultaneous adjustment in time and space
  - Increased dimension of state vector
  - Error covariance can include temporal component
  - Error covariance localization is n-dimensional

- **Formal extension of multivariate pdf to all spatial and temporal components**
  - For Gaussian assumption define 4-dimensional cost function

$$f(u) = \frac{1}{2}\left(u - u^f\right)^T P_f^{-1}\left(u - u^f\right) + \frac{1}{2}\left(y - h(u)\right)^T R^{-1}\left(y - h(u)\right)$$

$$u = u(x,y,z,t) \qquad\qquad P_f = P_f(x,y,z,t)$$

Summary

- Operational DA implementation requires simple and efficient codes
- Development of variational and ensemble methods is combined in hybrid variational-ensemble methodology
- Potential value of coupled DA
- New DA methodologies are already available for pre-operational testing
- Important to maintain generality of DA algorithm: potential for collaboration with other groups working with different models and observations
- Modular code provides adaptive framework

  - adding new model and observations only requires new DA interfaces with model and observations

Thank you !

Further information at http://www.cira.colostate.edu/ensemble/

Related publications

Chambon, P., S. Q. Zhang, A. Y. Hou, M. Zupanski, and S. Cheung, 2013: Assessing the impact of pre-GPM constellation microwave precipitation radiance data in the Goddard WRF ensemble data assimilation system. *Quart. J. Roy. Meteorol. Soc.*, DOI:10.1002/qj.2215.

Fletcher, S. J., and M. Zupanski, 2006: A Data Assimilation Method for Log-normally Distributed Observational Errors. *Q. J. Roy. Meteorol. Soc.*, **132**, 2505-2519.

Lokupitiya, R.S., D. Zupanski, A.S. Denning, S.R. Kawa, K.R. Gurney, M. Zupanski, W. Peters, 2008: Estimation of global CO_2 fluxes at regional scale using the maximum likelihood ensemble filter. *J. Geophys. Res.*, **113**, D20110, doi:10.1029/2007JD009679.

Orescanin, B., B. Rajkovic, M. Zupanski, and D. Zupanski, 2009: Soil model parameter estimation with ensemble data assimilation. *Atmos. Sci. Letters*, **10**, 127-131.

Steward, J. L., I. M. Navon, M. Zupanski, and N. Karmitsa, 2012: Impact of Non-Smooth Observation Operators on Variational and Sequential Data Assimilation for a Limited-Area Shallow-Water Equation Model. *Quart. J. Roy. Meteorol. Soc.*, **138**, 323-339.

Tran, A. P., M. Vanclooster, M. Zupanski, and S. Lambot, 2014: Joint estimation of soil moisture profile and hydraulic parameters by ground-penetrating radar data assimilation with maximum likelihood ensemble filter. *Water Resour. Res.*, **50**, doi:10.1002/2013WR014583.

Zhang, S. Q., M. Zupanski, A. Y. Hou, X. Lin, and S. H. Cheung, 2013: Assimilation of precipitation-affected radiances in a cloud-resolving WRF ensemble data assimilation system. *Mon. Wea. Rev.*, **141**, 754-772.

Zupanski, D., S. Q. Zhang, M. Zupanski, A. Y. Hou, and S. H. Cheung, 2011: A prototype WRF-based ensemble data assimilation system for downscaling satellite precipitation observations. *J. Hydromet.*, **12**, 118-134.

Zupanski, D., A.Y. Hou, S.Q. Zhang, M. Zupanski, C.D. Kummerow, and S. H. Cheung, 2007: Application of information theory in ensemble data assimilation. *Q. J. R. Meteorol. Soc.*, **133**, 1533-1545.

Zupanski, M., I. M. Navon, and D. Zupanski, 2008: The Maximum Likelihood Ensemble Filter as a non-differentiable minimization algorithm. *Q. J. R. Meteorol. Soc.*, **134**, 1039-1050.

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