Workshop on
Mathematical Advancement in Geophysical Data Assimilation
Banff International Research Station
February 3-8, 2008, Banff, Canada

Dynamical Approach to Nonlinear Ensemble Data Assimilation

Milija Zupanski
Cooperative Institute for Research in the Atmosphere
Colorado State University
Fort Collins, Colorado
E-mail: ZupanskiM@CIRA.colostate.edu
URL: http://www.cira.colostate.edu/projects/ensemble/

Acknowledgements:

- *Colorado State Univ.: D. Zupanski, S. Fletcher, C. Kummerow, T. Vonder Haar*
- *Florida State Univ.: I.M. Navon*
- *NASA GMAO: S. Zhang, A. Hou*
- *NCAR Computational and Information Systems Laboratory (Blueice)*
- *NASA Ames Research Center (Columbia supercomputer)*

Overview

- **Dynamics and nonlinearity in data assimilation**
- **A prototype for dynamical ensemble DA algorithm: MLEF**
- **Example 1: MLEF + 1-D soil model**
- **Example 2: MLES + NASA GEOS-5 global model + precipitation**
- **Example 3: MLEF + WRF regional model**
- **Conclusion and Future work**

Dynamics and nonlinearity in data assimilation

(1) Create dynamically consistent analysis

- **DA is a dynamic-stochastic problem**

  - Dynamics reduces the number of degrees of freedom (DOF)

  - Dynamics has a "localization" capability

- **Dynamical requirement:**

  - Analysis and uncertainties are in dynamical balance: $\boldsymbol{x}_a = \boldsymbol{x}_f + \Delta\, \boldsymbol{x}_a$; $\boldsymbol{P}_a^{1/2}$

  - Easier problem: If the first-guess is in balance, need only to make sure that the *analysis correction* is in dynamical balance

- **Possible solutions:**

  1 - *Stage 1*: Get analysis without enforcing dynamical balance; *Stage 2*: filter the noise after analysis to produce the balanced state

  2 - Get balanced analysis by enforcing dynamical balance in DA

Dynamics and nonlinearity in data assimilation

(2) Address nonlinearity in prediction and observations

- **Fundamental theoretical development in DA is based on linear assumptions**

  - Kalman Filtering, Best Linear Unbiased Estimate (BLUE)

- **Real world is nonlinear**

  - Physical processes (i.e. clouds, precipitation)

  - Observation operators (remote sensing, physics related observations)

- **Typical solutions:**

  - Direct

  Use *linear* analysis solution + insert nonlinear operators into linear solution

  - Indirect

  Solution of nonlinear analysis problem by minimizing nonlinear cost function

Forecast error covariance

In DA the (square-root) forecast error covariance defines the uncertainty space in which the analysis is corrected

$$\boldsymbol{x}^a - \boldsymbol{x}^f = \boldsymbol{P}_f^{1/2}\boldsymbol{w} = w_1\boldsymbol{p}_f^1 + w_2\boldsymbol{p}_f^2 + \cdots + w_N\boldsymbol{p}_f^N$$

1) The *analysis increment* is in dynamical balance if the column-vectors of the square-root forecast error covariance are in dynamical balance

2) The *analysis* is in dynamical balance if the first guess and the analysis increment are in dynamical balance

- How well the dynamical balance can be enforced in $\boldsymbol{P}_f$?
- What is the impact of (unbalanced) $\boldsymbol{P}_f$ in cycling of DA?

Means for ensuring the dynamical balance

- **Create balanced first guess**
- **Create balanced forecast error perturbations**

- the dynamical balance in $\boldsymbol{P}_f$ can be enforced by initiating DA with balanced initial conditions and balanced initial ensemble perturbations

- if the initial ensemble perturbations are not in balance, the noise can remain in short-range forecast used to calculate the first guess, eventually contaminating the next cycle analysis

- enforcing the dynamical balance during the analysis is more beneficial then enforcing it after the analysis is completed

- analysis perturbations have to be projected on dynamically consistent directions that will not create a spin-up of the model forecast

- cycling of data assimilation has beneficial impact on dynamical balance

Addressing the nonlinearities and dynamical balance using the Maximum Likelihood Ensemble Filter (MLEF)

- **MLEF developed as a nonlinear extension to Kalman Filter**

  - First-guess obtained using a nonlinear model forecast

  - Square-root forecast error covariance columns defined as dynamically balanced span-vectors of analysis correction subspace

  - Nonlinear analysis obtained using iterative minimization of the cost function defined over the analysis correction subspace

  - Full-rank or reduced-rank method

  - Could be extended to non-Gaussian PDFs

  - Not sample based

References: Zupanski 2005; Zupanski and Zupanski 2006; Zupanski et al. 2008; Fletcher and Zupanski 2007a, 2007b)

Mathematical formulation of the MLEF

1) ***Initial state and uncertainty*: Define an initial state and a subspace (span-vectors)**

$$\left[\boldsymbol{x}^0, span\{\boldsymbol{u}_i^0\}\right] \qquad \boldsymbol{x}_i^0 = \boldsymbol{x}^0 + \boldsymbol{u}_i^0; \qquad (i = 1,\ldots,N_E)$$

2) ***Prediction*: Transport the uncertainty span-vectors in time by a prediction model**

$$\left.\begin{aligned} \boldsymbol{x}^t &= M(\boldsymbol{x}^{t-1}) \\ \boldsymbol{x}^t + \boldsymbol{u}_i^t &= M(\boldsymbol{x}^{t-1} + \boldsymbol{u}_i^{t-1}) \end{aligned}\right\} \quad \left[\boldsymbol{x}^t, span\{\boldsymbol{u}_i^t\}\right] \qquad \boldsymbol{u}_i^t = M(\boldsymbol{x}^{t-1} + \boldsymbol{u}_i^{t-1}) - M(\boldsymbol{x}^{t-1})$$

3) ***Analysis*: Maximum a posteriori estimate** $\max P(X \mid Y) = \min\left[-\ln P(X \mid Y)\right]$

- Iterative minimization of cost function over the ensemble subspace $\left[span\{\boldsymbol{u}_i^t\}\right]$

$$\left.\begin{aligned} \boldsymbol{y} &= H(\boldsymbol{x}) \\ \boldsymbol{y} + \boldsymbol{v}_i &= H(\boldsymbol{x} + \boldsymbol{u}_i) \end{aligned}\right\} \quad \left[span\{\boldsymbol{v}_i\}\right] \qquad \boldsymbol{v}_i = H(\boldsymbol{x} + \boldsymbol{u}_i) - H(\boldsymbol{x})$$

Examples with MLEF

- **MLEF with 1-D 1-point soil temperature model**
  - Nonlinear observation operators (fluxes)
  - 1 degree of freedom

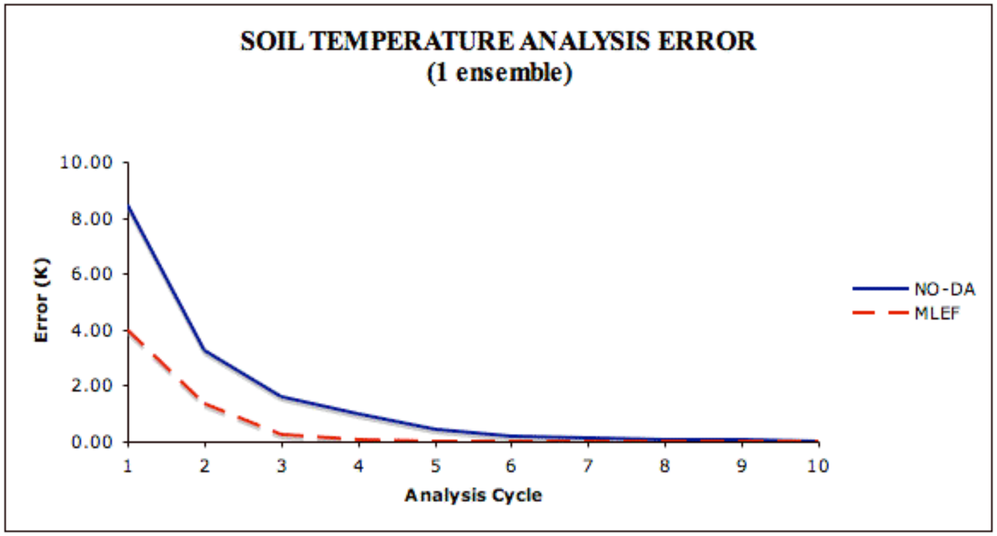
- **MLES with NASA GEOS-5 global model**
  - Assimilation of accumulated precipitation observations (nonlinear)
  - Smoother

- **MLEF with regional WRF MODEL**
  - Linear observation operators, but poor geographical coverage
  - Hurricane simulation

Example 1: ***MLEF + 1-D soil model***

(B. Rajkovic, B. Orescanin - Univ. Belgrade, Serbia)

- **One-point nonlinear soil temperature model** => $\theta_t = M(\theta_{t-1})$
- **Nonlinear observation operators** => $H_1(\theta) \sim \theta^4 \quad H_2(\theta) \sim \dfrac{e^{a_1(\theta-\theta_s)})}{\theta - a_2}$
- **1 DOF => 1 ensemble member**



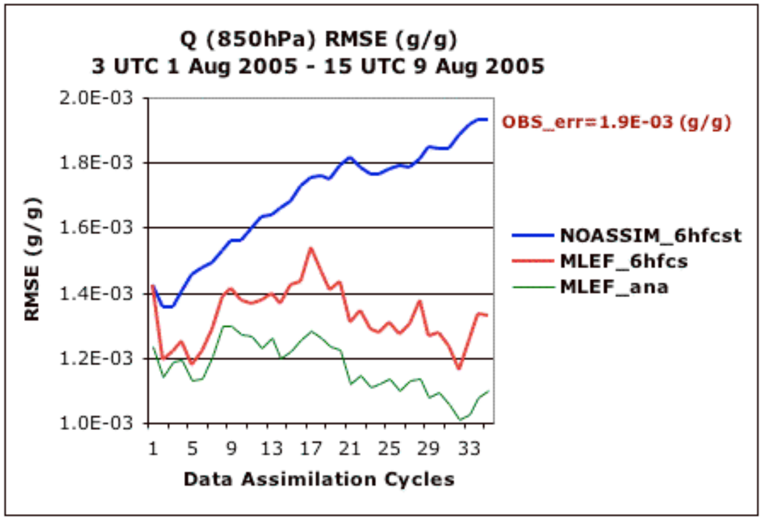
Dynamical method: For a system with 1 DOF, need 1 ensemble!

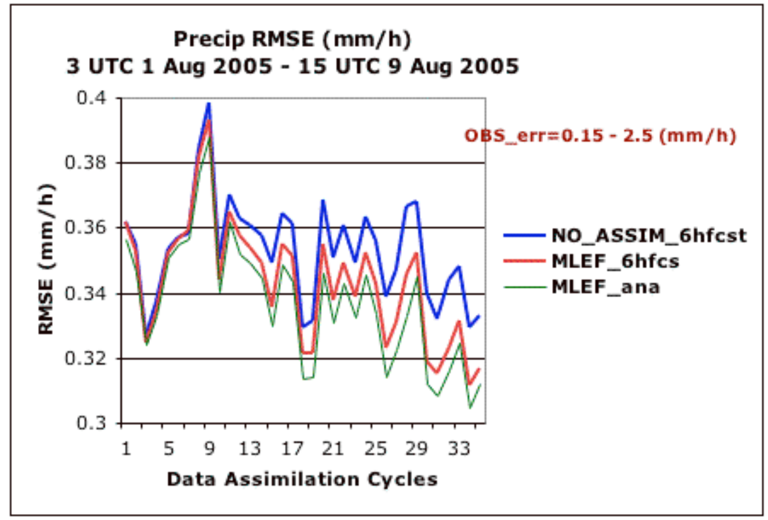
Technically no solution exists with sampling methods.

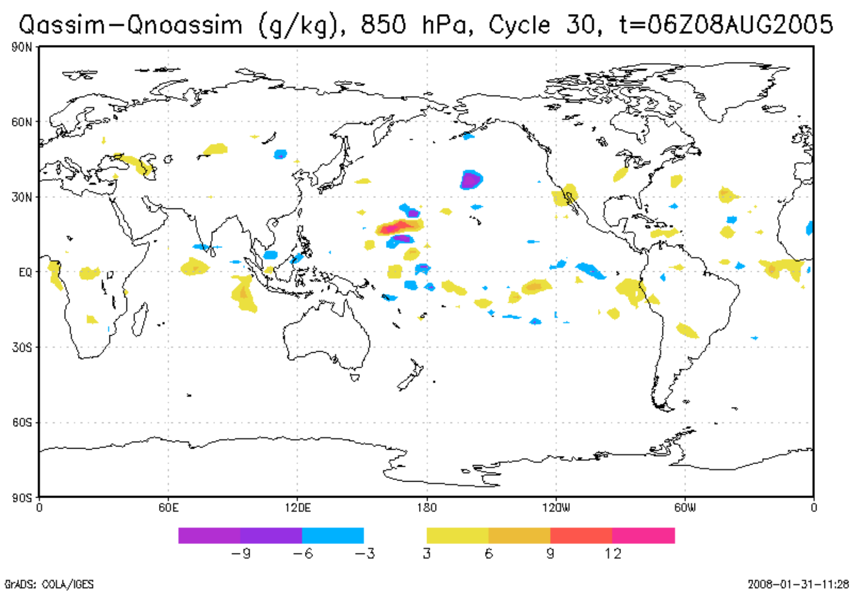
Example 2: **MLES + GEOS-5 + precipitation**
(D. Zupanski, S. Zhang, A. Hou)

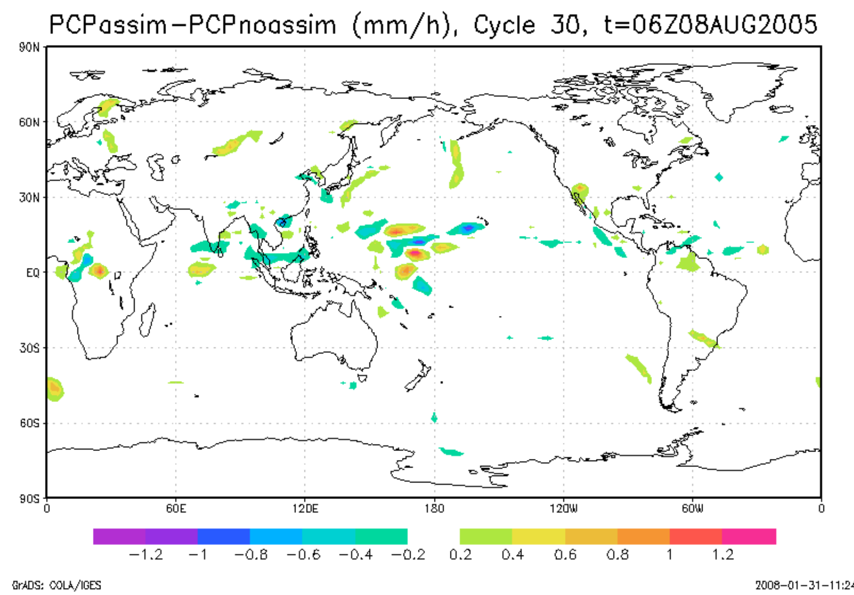
- *Goal*: Assimilate multi-sensor precipitation observations (accumulated rainfall observations from TRMM and SSM/I) to improve precipitation analysis
- Update specific humidity to improve fit to the precipitation obs and pseudo data
- Other variables (U,V,T,Ps) are taken from the NASA G5DAS 3d-Var GSI analyses
- *Model resolution*: 2x2.5 degrees, 72 vertical levels (144x91x72), N_S ~1,000,000
- *Assimilation*: 6-hour assimilation cycle, 8 days of assimilation
- *Method*: Maximum Likelihood Ensemble *Smoother* (MLES)
- 32 ensembles
- Error covariance localization based on a *modified* local-domains approach (10x10 points in each local domain)

Assimilation results









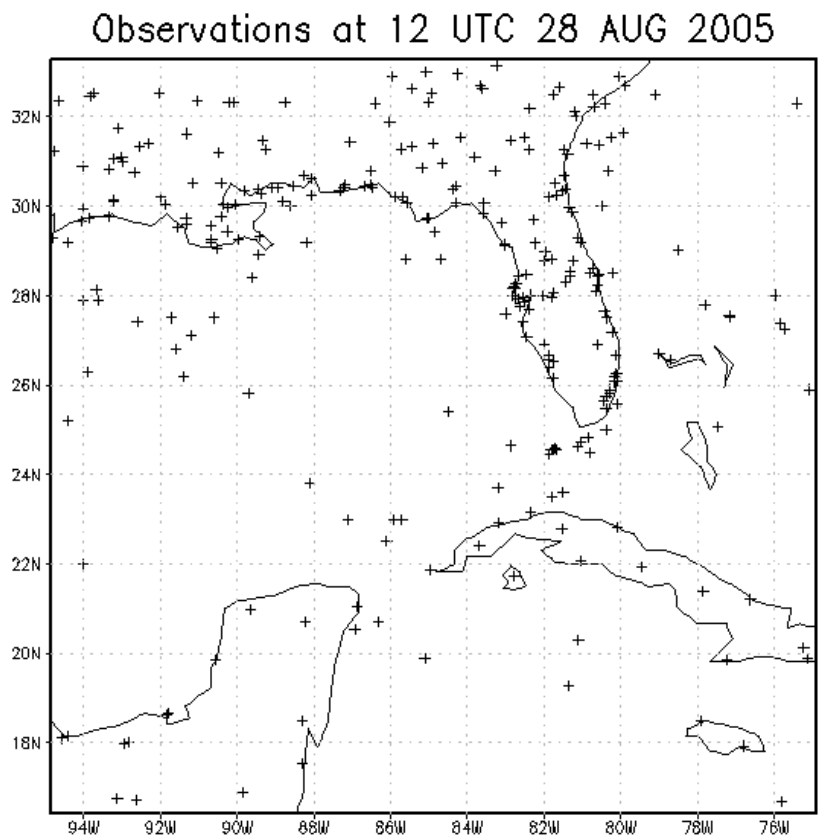
Example 3: **MLEF + WRF regional model**

Hurricane Katrina (landfall 12Z 29 AUG 2005):

- *Model resolution*: 30 km horizontal, 28 vertical levels (75x70x28)
- 6-hour old boundary conditions from the NCEP GFS model
- *Observations*: NCAR upper-air and surface observations (p_s, T, q, u, v)
- *Assimilation*: 6-hour interval, from 26 Aug 00Z - 31 Aug 00Z (5 days)
- *Control variables*: $u, v, \delta\theta, \delta Z, q_v$
- *State vector dimension* ~700,000
- 32 ensembles
- Error covariance localization based on a *modified* local-domains approach (23x25 points in each local domain)

Radiosonde and SYNOP Observations (at 12 UTC)

20 Radiosondes + 409 SYNOPs ~ 1000-3000 observations

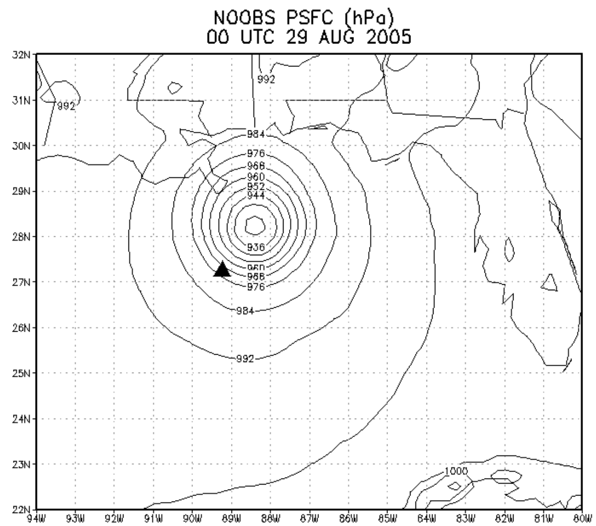


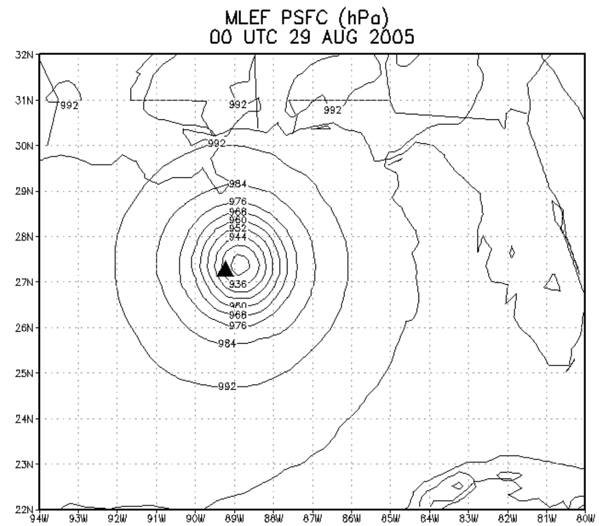
Irregular observation coverage: ***Mostly surface + over the land***

Surface pressure in the MLEF and No-DA experiments

29 AUG 00 UTC

(12-hour before landfall)





NO-DA

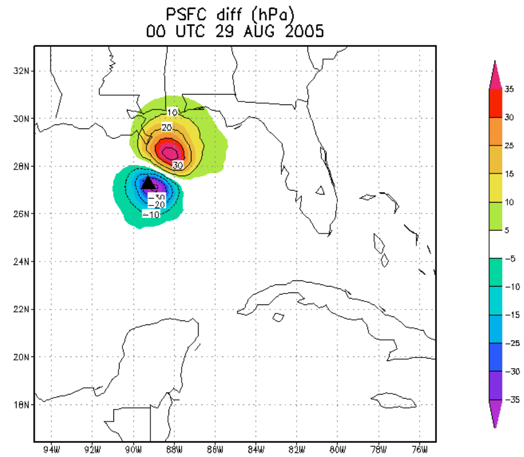
(no assimilation)

MLEF

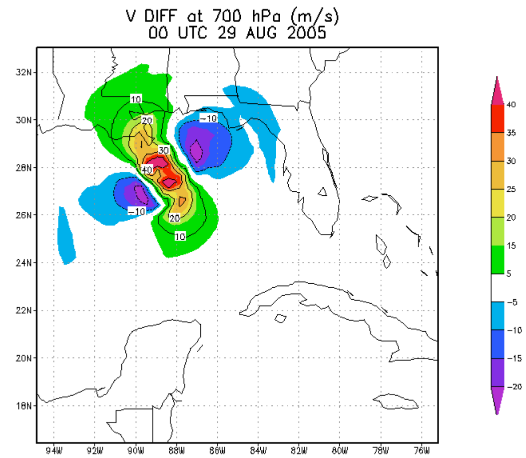
(with assimilation)

Data assimilation with MLEF improves the position and intensity of the hurricane

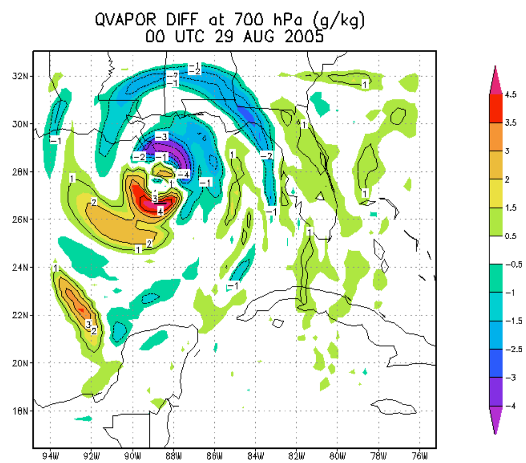
6-hour forecast difference between the No-DA and MLEF experiments (29 AUG 00 UTC)



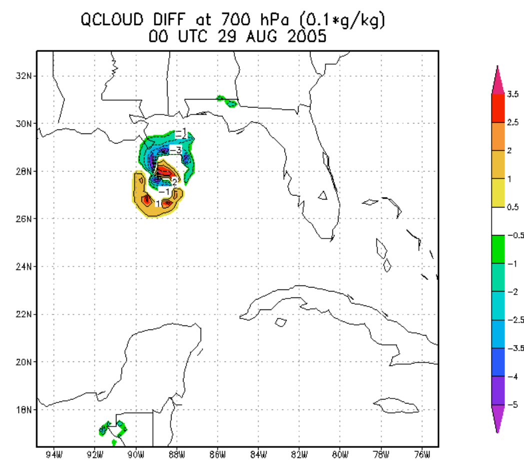
Surface pressure (hPa)



V-wind at 700 hPa (m/s)



Spec humid at 700 hPa (g/kg)



Cloud water at 700 hPa (0.1 g/kg)

- **Dynamically consistent and localized impact of data assimilation**
- **Data assimilation “moves” the hurricane in the correct direction (SW)**

Conclusion and Future work

Conclusion:

- **Dynamical approach to nonlinear data assimilation is advantageous:**
  - algorithmically simpler
  - dynamically consistent analysis does not require further attention
- **Nonlinear operators are efficiently handled by iterative minimization**

Future work:

- **Focus on clouds and precipitation assimilation**
- **High-resolution WRF model (1-5 km)**
- **Further improve error covariance localization by introducing dynamical correlations**

More information about the MLEF and related research at:
http://www.cira.colostate.edu/projects/ensemble