Challenges of Data Assimilation in MMF

Milija Zupanski and Dusanka Zupanski

Cooperative Institute for Research in the Atmosphere  
Colorado State University  
Fort Collins, CO 80523-1375  
ZupanskiM@CIRA.colostate.edu
Outline

- Challenges of multi-scale modeling framework (MMF)

- Assimilation-prediction as a single mathematical problem
  - Kolmogorov equation
  - feasible for realistic applications, for the first time

- New (old) look at data assimilation: model errors, uncertainty
  - we knew about it before from the theory, now we can estimate them from observations

- Extended role of data assimilation: How data assimilation can help in MMF development and applications?
  - feasibility of a single assimilation-prediction system opens new avenue of opportunities for data assimilation applications

Milija Zupanski, CIRA/CSU
ZupanskiM@CIRA.colostate.edu
Challenges of multi-scale modeling framework (MMF)

• Cloud-scale interaction
• Climate scale interaction
• Interaction between scales
  - Nonlinearity
  - Uncertainty transfer
• Atmospheric-chemical processes
• Small-scale interaction between atmosphere and land

Question:
How to make a computationally efficient data assimilation, without sacrificing the quality? Never before such a complex system was used to simulate such a wide range of scales!
General Principles

Most general formulation of the assimilation-prediction problem:

Kolmogorov (Fokker-Planck) equation

\[
\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}[p(x,t)f(x,t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2}[p(x,t)g^2(x,t)]
\]

\( p \) – probability density
\( f \) – dynamical model
\( g \) – stochastic forcing (model error)

- **Prediction:** Estimate of the *forecast* probability density
- **Data Assimilation:** Estimate of the *initial* probability density

Implications to weather and climate: **THERE IS ONLY ONE SYSTEM !**
General Principles

Two fundamental theoretical and practical sources for improvement of the assimilation-prediction:

(1) Kalman filtering (includes Kolmogorov equation)
(2) Deterministic chaos (strange attractors)

Most (if not all) known data assimilation methodologies derived, or closely related to the Kalman filtering theory:

- Optimal interpolation
- Variational methods (3D-var, 4D-var)
- Ensemble Kalman filters

The notion of strange attractors implies an existence of a low-dimensional subspace (small number of degrees of freedom)

- Ensemble forecasting
- Ensemble Kalman filters
What do we want from PDF?

(1) Commonly estimated statistical (PDF) parameters
   - Mean (Monte Carlo KF - EnKF, minimum variance – standard KF)
   - Mode (Maximum of PDF: variational, MLEF)

   • Identical for Gaussian PDF and linear models, differ for nonlinear problems
   • Differ for sample-derived mean and mode parameters
   • Differ for non-Gaussian PDFs

(2) An estimate of the PDF width (uncertainty)

   • Covariance, standard deviation
   • Ensemble spread

   • Calculate more than one parameter, if feasible
   • Use all that can improve the knowledge of PDF
What is optimal?
Statistical parameters

PDF

Mean

Mode

PDF

Dynamical state

Dynamical state
Data assimilation

What are the control variable for a prediction model?

- **Initial conditions**
  - best known
- **Model error (including bias)**
  - very little known, yet it may have a dominant impact on the prediction
- **Empirical model parameters**
  - Limited knowledge, often based on a small, inadequate sample of cases
- **Lateral boundary conditions (if needed)**

Can it all be adjusted simultaneously?

- Yes, and it should be (augmented control variable)
- There is an overlap between the model error and empirical model parameters, but it is unknown in practice
Data assimilation using EnKF and related ensemble methods

- At present, the most general, computationally feasible data assimilation approach
  - need further testing in most complex environment

- Single algorithm is used for data assimilation and (ensemble) prediction
  - sample approach to Kolmogorov equation and deterministic chaos

- Can account for model error - new development
  - adjustment of model error and its covariance

- Algorithmically simple, efficient development and maintenance

- Suitable for high-performance computing (HPC)

Milija Zupanski, CIRA/CSU
ZupanskiM@CIRA.colostate.edu
Analysis Error Covariance (MLEF with KdVB model)

In-situ

Targeted

Targeted (intelligently placed) observations improve ensemble DA performance

From Zupanski 2004, MWR
[Available at ftp://ftp/cira.colostate.edu/milija/MLEF_mwr.pdf]

Milija Zupanski, CIRA/CSU
ZupanskiM@CIRA.colostate.edu
Possible issues with MMF applications

• How many ensembles can we realistically do with MMF?
  - simultaneous integration of a climate and cloud-system resolving models (CSRM) is quite demanding
  - what is the ultimate goal for resolution of CSRM (impact on climate)

• Exploit the statistical aspect of CSRM
  - each ‘grid-box’ of a climate model includes a sample of CSRM realizations

• Explore new avenues
  - adjust only statistical PDF parameters for the cloud-scales
  - no need for CSRM ensembles, PDF information is already given
  - control forecast using most complex model, ensembles used only for the climate component

Milija Zupanski, CIRA/CSU
ZupanskiM@CIRA.colostate.edu
Why is model error so important in data assimilation?

- Data assimilation system that employs a model is more sensitive to the model performance (as it should be)

- There is an overlap between the model error and the choice of empirical model parameters, but it is unknown in practice
  - best to have both the model error and parameters as components of control variable

- Data assimilation is the most efficient way to learn about model errors, from the comparison with observations
  - model bias
  - empirical parameters

Why not use data assimilation to learn about model biases and parameter values, and eventually correct the model itself?

Milija Zupanski, CIRA/CSU
ZupanskiM@CIRA.colostate.edu
Ensemble data assimilation with KdVB model
Augmented analysis error covariance matrix

From Zupanski and Zupanski 2004, MWR
[Available at ftp://ftp/cira.colostate.edu/milija/MLEF_model_err.pdf]
How to use data assimilation?

• Traditional role

  - model evaluation and validation against observations
  - first develop the model, than worry about data assimilation

Model development issues:

  - initial model testing with adequately defined empirical parameters and constants will be beneficial
  - debugging: model error adjustment in data assimilation will point to programming errors, as well as to the real biases
  - facilitated testing in large sample of cases
Additional role for data assimilation in model development

• Data assimilation and prediction are a single system, why not test them together?
  - speed-up development of a robust model (and data assimilation)
  - it will be used together later anyway

• Empirical model parameters
  - from the very beginning of model development and testing, the parameters used will be adequately estimated from observations, no need to wait
  - even if assimilated observations cannot directly relate to the scales and processes represented by parameters, other observations will improve parameter estimation through the implicit use of model equations

• Model error (bias)
  - may point to the incorrectly specified equations, facilitate debugging
  - actual model biases and errors will be known early in model development, therefore it may be possible to correct the model equations

Milija Zupanski, CIRA/CSU
ZupanskiM@CIRA.colostate.edu
New component to model development effort

- In order to be used during the initial model development, data assimilation ought to be:
  - easy to upgrade, accommodate for evolving model and observations
  - does not require considerable changes

EnKF and related ensemble methods

- Require only minor addition to the model
  - read control variable (initial conditions, model error, parameters)
  - no change required when adding new subroutines and processes to the model

- Simultaneous testing of the assimilation-prediction system
  - robustness of the system, and the model greatly enhanced
  - saves considerable time
  - probabilistic (PDF) evaluation of the prediction system
Implications

- The assimilation-prediction system is a unified system (e.g., Kolmogorov equation), and it can only be beneficial if treated as such from the very beginning in development, to obtain optimal results.

- Most general way to optimally introduce observations in model development is through data assimilation.

- Learn about model errors and biases early, possibly correct them.

- Find about appropriate values for empirical parameters and constants, even before all scales and types of observations are included.

- Data assimilation component of the system can be viewed as a new tool for model development and testing:
  - evaluate interaction between the scales
  - uncertainty transfer, especially between cloud-scales and climate
  - can be used in probabilistic (e.g. ensemble prediction), or deterministic sense (deterministic, control prediction).