Spring Colloquium on
'Regional Weather Predictability and Modeling'
April 11 - 22, 2005

1) Workshop on Design and Use of Regional Weather Prediction Models, April 11 - 19

2) Conference on Current Efforts Toward Advancing the Skill of Regional Weather Prediction. Challenges and Outlook, April 20 - 22

Probabilistic Assimilation/Prediction

M. Zupanski
Cooperative Institute for Research in the Atmosphere, CSU, Ft. Collins
USA
Spring Colloquium on Regional Weather Predictability and Modeling
The Abdus Salam International Centre for Theoretical Physics
11-22 April, 2005, Trieste, Italy

LECTURE
Probabilistic assimilation/prediction

Milija Zupanski

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

- Why probabilistic?
- Applications
- Future development and research directions
Why probabilistic?

- Uncertainties (errors) of
  - initial conditions
  - empirical parameters
  - model equations
  cause uncertainty (error) of the forecast

- Uncertainties (errors) of observations and of the forecast cause uncertainty (error) of the analysis

- Input errors have probability distribution
  - commonly assumed Gaussian, but it can be generalized

- The posterior (analysis) distribution depends on the prior probability distributions (Bayes theorem)

Probabilistic framework can be used to estimate posterior errors!
Why probabilistic?

- **Probabilistic vs. Deterministic**
  - Suppose the dynamic equations are deterministic, i.e. suppose there is enough information, and huge computers that can calculate all we want.
  - Even then, since the forecast uncertainty is caused by uncertainties in initial conditions, empirical parameters, model equations, lateral boundary conditions, etc., need probabilistic forecasting.

- **Chaotic vs. Deterministic**
  - Weather and Climate models, as well as many other physical systems and models, are chaotic: *extreme sensitivity to initial conditions*.
  - There is a preferred (attractor) subspace.
  - Predicted state vector (forecast) has higher probability to be found in that subspace => *not simply random*.

**Probabilistic framework can capture it all!**
How?

- **Probability Density Function (PDF)**

**Parameters that can be estimated:**

*Mean, Mode, Covariance, higher-order moments, ...*

Log-Normal PDF

- Various estimation methods exist
- Major restriction: highly dimensional PDF

**Balance the need for more parameters with computational resources**
Statistical PDF parameters

Uni-modal PDF

Bi-modal PDF

Mode – dynamical state with highest probability

Mean – mathematical expectation

Mode vs. Mean

- identical for Gaussian errors (linear operators)
- both are important for non-Gaussian errors (nonlinearity, skewness)
**Ensemble Data Assimilation**

**Maximum Likelihood Ensemble Filter (MLEF):**
- maximum likelihood approach (conditional *mode*)
- no perturbed observations
- directly solve the nonlinear problem using iterative minimization

**Ensemble Kalman Filter (EnKF):**
- minimum variance approach (conditional *mean*)
- perturbed, or unperturbed observations
- introduce nonlinearities to the closed-form (linear) KF solution

- Due to nonlinearities of prediction and observation operators, the errors are never strictly Gaussian

- EnKF and MLEF differ, therefore both bring *new information* to the analysis/prediction PDF problem
Possible Applications

• Data assimilation
• Ensemble forecasting
• Parameter estimation
• Model error estimation
• Targeted observations
• Information content of observations
• Advanced parallel computing
• Predictability, nonlinear dynamics

• Any phenomenon with predictive model and observations: Weather, Climate, Ocean, Ecology, Biology, Geology, Chemistry, Cosmology, …

- Wide range of applications
- Unified mathematical framework
Data Assimilation

1) **Analysis**: find optimal mix between the guess and the measurements
2) **Uncertainty** of the analysis estimate

- Irregular geographic location of observations
- Complex prediction model as a guess

*High dimension of unknown parameters:*
Need an efficient mathematical algorithm
Example: 2-D problem with single observation

Forecast uncertainty = $b$

Observation uncertainty = $a$

Optimal analysis

$\mathbf{x}^a = \frac{a^2}{a^2 + b^2} \mathbf{x}^b + \frac{b^2}{a^2 + b^2} \mathbf{y}$

Analysis uncertainty

$c = \frac{ab}{\sqrt{a^2 + b^2}}$

Methods for linear, low-dimensional problems exist (e.g., Kalman Filter)
Can they be extended to high-dimensions?
Ensemble Forecasting

Kolmogorov Equation

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

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

Ensemble forecasting can be viewed as an approximate solution method for the Kolmogorov equation
Ensemble Forecasting

1) **Find forecast PDF**
   - PDF parameters: *mean*, *mode*, *covariance*

2) **Initial ensemble perturbations**
   - random search (Monte Carlo) – inefficient for high dimensions
   - dynamically significant subspace – efficient, but need to know where
   - Operational ensemble forecasting
     - adjoint method (ECMWF)
     - breeding method (NCEP)
     - EnKF method (Canadian Met. Service)

- **Important to include all components in a single system**
- **Determine likelihood of predicted weather event**

*Unification of data assimilation and ensemble forecasting improves both!*
Parameter Estimation

- **Estimate empirical parameters:**
  - coefficients (diffusion, …)
  - physical parameterization (cumulus convection, …)

- Generally, empirical parameter are not observed

- Optimal parameter values may be dynamically, or season dependent

- *Strategy:* Utilize indirect observations to improve parameter values

---

Ensemble assimilation/prediction provides most general framework for parameter estimation!
Parameter Estimation

State augmentation approach

\[ x^f_k = M(x^a_{k-1}, \gamma) \]

- forecast (guess)
- analysis from previous time \((k-1)\)
- empirical parameter

Augmented control variable

\[ z = (x, \gamma) \]

Augmented error covariance

\[ P = \begin{pmatrix} P_{x,x} & P_{x,\gamma} \\ P_{\gamma,x} & P_{\gamma,\gamma} \end{pmatrix} \]

\[ \gamma \sim O(10^0) - O(10^2) \quad x \sim O(10^5) - O(10^7) \]

Algorithmically simple to implement, even for most complex models!
Model error estimation

\[ x_{n+1} = M(x_n) + \Phi_n \]

\( \Phi \) – model correction (bias, random error)

Including model error allows realistic fit to observations
Model error estimation

- Scale/phenomenon representation error
  - Errors in model equations
  - Physical parameterization scheme
- Random error, bias
- Estimate of model error relies on indirect observations

**Advantage of using ensemble assimilation/prediction:**
Uncertainty of model error

- Until now, the uncertainty of model error was ALWAYS PRESCRIBED!
- Within ensemble framework:
  - we can estimate uncertainty
  - the uncertainty evolves in time
  - there is a dependence between the model error and model state

Probabilistic tool to learn about models and how to improve them!
Model error and parameter estimation in MLEF

State augmentation approach

\[
\begin{align*}
  z_n &= \begin{bmatrix} x_n \\ b_n \\ \gamma_n \end{bmatrix} = \\
  \begin{bmatrix} M_{n,n-1} & (1 - \alpha^n)G & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} x_{n-1} \\ b_{n-1} \\ \gamma_{n-1} \end{bmatrix} + \\
  \begin{bmatrix} \alpha^n \\ 0 \end{bmatrix} \Phi_0
\end{align*}
\]

\(x\) – state vector (initial conditions); \(b\) – model bias; \(\gamma\) – empirical parameters

Augmented control variable:

\(z = (x_0, b, \gamma)\)

Augmented error covariance:

\[
P = \begin{pmatrix}
  P_{x_0,x_0} & P_{x_0,b} & P_{x_0,\gamma} \\
  P_{b,x_0} & P_{b,b} & P_{b,\gamma} \\
  P_{\gamma,x_0} & P_{\gamma,b} & P_{\gamma,\gamma}
\end{pmatrix}
\]
Model error and parameter estimation with KdVB model

Parameter estimation (diffusion coefficient)

Impact of incorrect diffusion

(10 Ensembles, 101 Observations)

<table>
<thead>
<tr>
<th>Cycle No.</th>
<th>RMS error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.60E-01</td>
</tr>
<tr>
<td>11</td>
<td>1.40E-01</td>
</tr>
<tr>
<td>21</td>
<td>1.20E-01</td>
</tr>
<tr>
<td>31</td>
<td>1.00E-01</td>
</tr>
<tr>
<td>41</td>
<td>8.00E-02</td>
</tr>
<tr>
<td>51</td>
<td>6.00E-02</td>
</tr>
<tr>
<td>61</td>
<td>4.00E-02</td>
</tr>
<tr>
<td>71</td>
<td>2.00E-02</td>
</tr>
<tr>
<td>81</td>
<td>1.00E-02</td>
</tr>
<tr>
<td>91</td>
<td>0.00E+00</td>
</tr>
</tbody>
</table>

Correct diffusion
Incorrect diffusion
Param. estimation

Estimation of diffusion coefficient

(102 Ensembles, 101 Observations)

<table>
<thead>
<tr>
<th>Cycle No.</th>
<th>Diffusion coefficient value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.00E-02</td>
</tr>
<tr>
<td>11</td>
<td>4.00E-02</td>
</tr>
<tr>
<td>21</td>
<td>6.00E-02</td>
</tr>
<tr>
<td>31</td>
<td>8.00E-02</td>
</tr>
<tr>
<td>41</td>
<td>1.00E-01</td>
</tr>
<tr>
<td>51</td>
<td>1.20E-01</td>
</tr>
<tr>
<td>61</td>
<td>1.40E-01</td>
</tr>
<tr>
<td>71</td>
<td>1.60E-01</td>
</tr>
<tr>
<td>81</td>
<td>1.80E-01</td>
</tr>
<tr>
<td>91</td>
<td>2.00E-01</td>
</tr>
</tbody>
</table>

Correct diffusion
Incorrect diffusion
Param. estimation (0.07)
Param. estimation (0.20)
True value (0.07)
True value (0.20)

Parameter estimation (diffusion coefficient)

Error covariance block matrices

IC-IC

ME-ME

IC-ME

Significant cross-correlation between initial conditions and model error
Targeted observations

**Observing system which adapts to dynamically significant changes**

*Why targeted observations?*
- It is impossible to observe everything (all possible variables), everywhere (finest spatial and temporal resolution)
- Even if it would be possible, sky-rocketing costs

*Strategy:*
- Observe where it matters the most
- Determine the *most likely* error growth (area and variable)
- Remote sensing, aircraft, …

*Self-controlled system:*
- Know *when, where, what* to observe
- Improve where is needed the most
Targeted observations

Map summarizing the extra observations taken during the FASTEX experiment for IOP 17. The black track identifies the location of the cyclone minimum pressure, the colored tracks the aircraft missions, and the red symbols additional radio-soundings (from Montani et al. 1999).

Great interest of operational weather centers to optimize observation network + minimize cost
Information content

- Information content of observations is at the heart of the problem
- Assimilation/prediction system benefits from new information, not necessarily from more observations

Design a system that makes decisions based on the information content of observations!
Information content

• True degrees of freedom of an assimilation/prediction system are determined by the information content of observations and of model state

• The minimal basis of the analysis correction subspace is determined from the singular vectors of the information content matrix:
  \[ R^{-1/2} HP_f^{1/2} = UAV^T \]

Less efficient approaches, traditionally used:
  - model space
  - observation space

New approach, using ensembles:
  - Extract independent information from ensembles
  - Minimize the ensemble size, or optimize the utilization of existing ensembles
Parallel computing

In realistic weather and climate applications:
• dimension of the state variable is large (grid points $\times$ number of variables)
• prediction model is computationally expensive

Computationally most demanding task is ensemble forecasting
- can be efficiently reduced by running ensemble members on separate nodes
- ultimate efficiency: $\text{cost of ensemble forecasting} = \text{cost of one forecast}$

Explore options to parallelize the system, based on:
- model state dimensions
- ensemble size (number)

Take advantage of parallel computing development in general

**Ensemble assimilation/prediction enables superior parallel performance**
Lorenz equation illustrates the complexity of attractor subspace
Predictability

- Highly nonlinear problem
- Probabilistic problem
- Multi-modal PDFs
- Coupled models

In order to learn about model predictability, need:
- Reliable system, close to the truth
  - observations
  - model error (bias) and its uncertainty

Attractor subspace
- Find the basis
- Estimate dimensions

Capability to learn about the predictability and attractor subspace for most complex models!
Future development and research directions

• Non-Gaussian and multi-modal PDFs
  - current assimilation/prediction generally assumes Gaussian errors
  - practical system with or without explicit knowledge of PDF

• Probabilistic tool for model development:
  - address the important issues in the context of PDFs
  - probabilistic view of deterministic models

• Understand predictability of complex weather/climate systems
  - find most natural invariant subspace and conservative variables

• Control theory
  - algorithms for non-smooth (derivative-free) optimization

Theoretical advancement using realistic systems!
Literature


