State and Parameter Estimation in Stochastic Dynamical Models

Timothy DelSole

George Mason University, Fairfax, Va and Center for Ocean-Land-Atmosphere Studies, Calverton, MD

June 21, 2011
Stochastic Parameterizations Involve Tunable Parameters

▶ Berner et al. 2009
- Parameters for spectral AR model for streamfunction.
- In principle, AR parameters should vary with wavenumber.
- Parameters for generating noise with power law behavior.
- Perturbations weighed by dissipation rate implied by numerical dissipation, wave drag, convection.

▶ Shutts 2005
- Cellular automaton stochastic backscatter scheme.
- CA involves numerous parameters (life time, conditions for birth and death, survival rules, spatial smoothing).
- Perturbations weighed by dissipation rate implied by numerical dissipation, wave drag, convection.

▶ Buizza et al. 1999
- Multiplicative noise perturbs parameterized physics tendencies.
- Random numbers drawn from uniform distribution $[-0.5, 0.5]$.
- Random numbers constant over $10^\circ \times 10^\circ$ boxes.
- Random numbers constant for 6 time steps.
How Can Parameters in Stochastic Parameterizations Be Estimated?
How Can Parameters in Stochastic Parameterizations Be Estimated?

- Adjoint parameter estimation
  - Bennett, 1992, *Inverse Methods in Physical Oceanography*

- The Augmentation Method
  - Gelb, 1974, *Applied Optimal Estimation*
Parameter Estimation with the Kalman Filter

\( \mathbf{x} \): State vector
\( \mathbf{b} \): Parameter vector

Usual method: augment state vector with unknown parameters:

\[
\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{b} \end{pmatrix}
\]

Assume parameter update model is

\[
\mathbf{b}_t = \mathbf{b}_{t-1} + \mathbf{w}_t \quad \text{Jazwinski, 1970; Anderson, 2001}
\]
\[
\frac{d\mathbf{b}}{dt} = a \mathbf{b}_t + k + d\mathbf{w}_t \quad \text{Friedland & Grabousky 1982}
\]
\[
\text{Gershgorin, Harlim, Majda, 2010}
\]

How Does This Work?

Variations in \( \mathbf{b} \) cause variations in \( \mathbf{x} \). The covariance between these variables can be used to infer one from the other.
The Update Equations for Augmented State Vectors

Typically, only observations of the state are available:

\[ H_z = \begin{pmatrix} H_x & 0 \end{pmatrix} \quad \text{Interpolation Operator} \]

In this case, the Kalman Filter equations decouple:

\[
\begin{align*}
\mu^a_x &= \mu^f_x + K_x \left( o - H_x \mu^f_x \right) \quad \text{State} \\
\mu^a_b &= \mu^f_b + K_b \left( o - H_x \mu^f_x \right) \quad \text{Parameter}
\end{align*}
\]

- State update is exactly the same as in state-only assimilation.
- State can be updated with existing data assimilation system.
- Parameter update has same structural form as state update.
Illustration with Modified Lorenz96 Model

\[
\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - \frac{x_i}{1 + d_i} + 8 + f_i,
\]

- “true” values of \(d_i\) and \(f_i\) are chosen randomly.
- Note that \(d_i\) is a multiplicative parameter.
- Parameter update model \(b^f_t = \beta b^f_{t-1} + (1 - \beta)b^a_{t-1}\).
- Localization and inflation applied to state and parameters
- \(i = 1, 2, \ldots, 40\).
- 20 observations (every other grid point is observed).
- Augmented state vector has 120 elements:

\[
x^* = (x_1 \ x_2 \ \ldots \ x_{40} \ f_1 \ \ldots \ f_{40} \ d_1 \ \ldots \ d_{40})^T
\]
Additive and Multiplicative Parameter Estimation

Estimate $f_i$ and $d_i$ (additive and multiplicative). Compare with

Imperfect $f_i = 0, d_i = 0$

Perfect $f_i$ and $d_i$ equal to their true values

---

Yang and DelSole 2009
Estimation in Stochastic Models Using Augmented KF

Consider the simplest possible stochastic model

\[ x_t = \phi x_{t-1} + \beta w_t, \]

where \(|\phi| < 1\) and \(w\) is standardized Gaussian white noise.
Why Augmentation Fails for Stochastic Parameters

\[ x_t = \phi x_{t-1} + \beta_t w_t, \]

Distribution of \( x_t \) for fixed \( \beta_t \) and fixed \( x_{t-1} \) is

\[ x_t | \beta_t, x_{t-1} \sim N(\phi x_{t-1}, \beta_t^2 \sigma_w^2). \]

- Ensemble mean of \( x_t \) is independent of \( \beta_t \).
- Variations in \( \beta_t \) affect the ensemble spread, not the mean.
- It can be shown that \( \text{cov}[x_t, \beta_t] = 0 \) if \( \text{cov}[x_0, \beta_0] = 0 \).
- Vanishing covariance implies \( x_t \) and \( \beta_t \) are independent (under normal distribution).
- Independence implies KF cannot estimate \( \beta_t \) from \( x_t \).
Bayes Theorem

\[ p(\beta \mid o, \Theta) \propto p(o \mid \beta, \Theta) \cdot p(x \mid \beta, \Theta) \cdot p(x \mid \beta, \Theta) \cdot p(\beta \mid \Theta) \]

where

- \( o \): Observation at time \( t \)
- \( \Theta \): All observations up to time \( t - 1 \).
- \( x \): State variable
- \( \beta \): Variance Parameter in stochastic-dynamical model.
Log of the Posterior

\[-2 \log p(\beta x | o \Theta) = \]

\[(o - Hx)^T R^{-1} (o - Hx) + \log |R| + M_o \log 2\pi + \]

\[(x - \mu_f)^T P_f^{-1} (x - \mu_f) + \log |P_f| + M_x \log 2\pi + \]

\[(\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta) + \log |\Sigma_\beta| + M_\beta \log 2\pi \]

\text{Likelihood} \quad \text{Forecast} \quad \text{Prior}
Adjoint Parameter Estimation

Adjoint parameter estimation in data assimilation (Navon 1997) is based on minimizing the functional

\[
J = (o - Hx)^T R^{-1} (o - Hx) + (x - \mu_f)^T P_f^{-1} (x - \mu_f) + (\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta),
\]

This is the non-constant part of the posterior provided the forecast covariance \( P_f \) is fixed.

But in stochastic parameter estimation, \( P_f \) is not constant!
Stochastic Parameter Estimation Requires Varying $P^f$

Setting the derivative of the posterior to zero and solving gives the (generalized) maximum likelihood estimate.

\[
\frac{\partial \text{posterior}}{\partial x} = 0 \quad \text{Standard Kalman Filter update (for fixed } \beta) \]

\[
\frac{\partial \text{posterior}}{\partial \beta} = 0 \quad \text{New nonlinear equation to solve for } \beta
\]

The key difference from past studies (e.g., adjoint methods) is that I do not assume that $\partial P^f / \partial \beta$ vanishes.
Estimating Derivatives of Covariance Matrices

Generate ensemble with fixed $\beta + \Delta \beta$, another with fixed $\beta - \Delta \beta$:

$$
\frac{\partial P^f}{\partial \beta} = \frac{P^f(\beta + \Delta \beta) - P^f(\beta - \Delta \beta)}{2\Delta \beta}
$$
Connections

GMLE does not distinguish stochastic and deterministic parameters

- deterministic parameters characterized by \( \partial \mu_f / \partial \beta \neq 0 \).
- stochastic parameters characterized by \( \partial P_f / \partial \beta \neq 0 \)

Augmentation is equivalent to GMLE of deterministic parameters, if \( \Sigma_\beta \) is interpreted as the spread of the parameter ensemble.
Stochastic Parameter Estimation

\[ x_t = \phi x_{t-1} + \beta w_t, \]

Deterministic Parameter, Augmented

Stochastic Parameter, Augmented

Deterministic Parameter, GMLE

Stochastic Parameter, GMLE
Parameter Estimation in Stochastic Lorenz Model

Slightly modified version of Hansen and Penland (2006) model:

\[
\begin{align*}
    dx &= -a(x - y)dt \\
    dy &= (rx - xz - y)dt + r_s x \circ dw \\
    dz &= (xy - bz)dt
\end{align*}
\]

---

**Deterministic Parameter, Augmented**

**Stochastic Parameter, Augmented**

**Deterministic Parameter, GMLE**

**Stochastic Parameter, GMLE**
Two-Stage Ensemble Generation

State-parameter estimation can be achieved in two stages.

1. State-only data assimilation produces ensemble $X^a$.
2. Ensemble is “corrected” to account for parameter estimation:

$$X^{aa} = X^a \left( I + \delta_+ w w^T \right)$$

where

$$w = -X^a T \left( P^{f-1} \frac{\partial P^f}{\partial \beta_k} P^{f-1} (\mu_a - \mu_f) - P^{f-1} \frac{\partial \mu_f}{\partial \beta_k} \right)$$

$\delta_+ = \text{even more complicated!}$

This algorithm takes advantage of an existing ensemble filter.
Summary

- Proposed deriving state and parameter estimates for stochastic dynamical models from generalized maximum likelihood theory.
- Stochastic parameter estimation requires accounting for dependence of forecast covariance on the parameter.
- Solution obtained in two-stages: first a standard Kalman Filter, followed by “correction” to take into account parameter update.
- Proposed solution outperforms augmentation methods for estimating stochastic parameters.
- We show that augmentation method is useless for stochastic parameter estimation (contrary to statements in the literature).
- Method requires generating new ensembles for each stochastic parameter being estimated. More innovative methods?