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

[1]

[1] collaboration with Xiaosong Yang (GMU/COLA; now at GFDL)

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
  - Carrera and Neuman, 1986, *Water Resour. Res.*
  - Bennett, 1992, *Inverse Methods in Physical Oceanography*
  - Navon, 1997, *Dyn. Atmos. Oceans*
- The Augmentation Method
  - Jazwinski, 1970, *Stochastic Processes and Filtering Theory*
  - Gelb, 1974, *Applied Optimal Estimation*
  - Anderson, 2001, *Mon. Wea. Rev.*
  - Gershgorin, Harlim, Majda, 2010, *J. Comput. Phys.*

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

$\mathbf{b}_t = \mathbf{b}_{t-1}$	Jazwinski, 1970; Anderson, 2001
$\mathbf{b}_t = \mathbf{b}_{t-1} + \mathbf{w}_t$	Friedland & Grabousky 1982
$d\mathbf{b}/dt = a\mathbf{b}_t + k + d\mathbf{w}_t$	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:

$$\mathbf{H}_z = \begin{pmatrix} \mathbf{H}_x & \mathbf{0} \end{pmatrix} \qquad \text{Interpolation Operator}$$

In this case, the Kalman Filter equations decouple:

$$\boldsymbol{\mu}_x^a = \boldsymbol{\mu}_x^f + \mathbf{K}_x \left(\mathbf{o} - \mathbf{H}_x \boldsymbol{\mu}_x^f \right) \qquad \textit{State}$$

$$\boldsymbol{\mu}_b^a = \boldsymbol{\mu}_b^f + \mathbf{K}_b \left(\mathbf{o} - \mathbf{H}_x \boldsymbol{\mu}_x^f \right) \qquad \textit{Parameter}$$

- 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 $\mathbf{b}_t^f = \beta \mathbf{b}_{t-1}^f + (1 - \beta)\mathbf{b}_{t-1}^a$.
- 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:

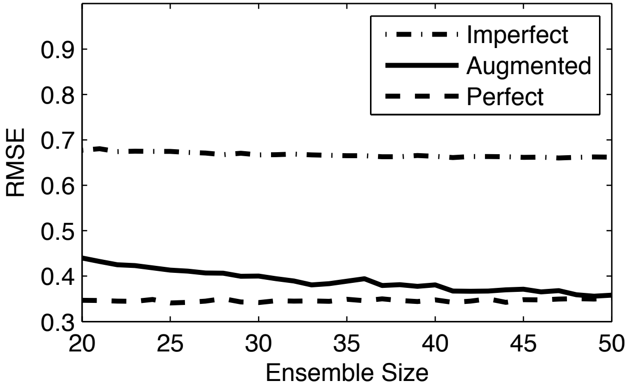
$$\mathbf{x}^* = \begin{pmatrix} x_1 & x_2 & \ldots & x_{40} & f_1 & \ldots & f_{40} & d_1 & \ldots & d_{40} \end{pmatrix}^T$$

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



[2]

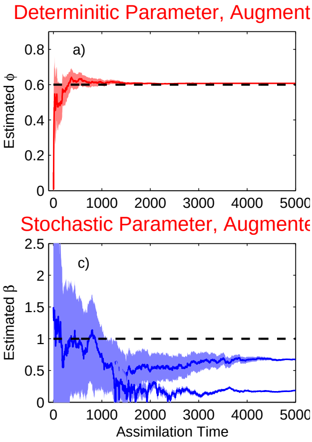
[2] Yang and DelSole 2009

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 β_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 β_t.
- Variations in β_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 β_t are independent (under normal distribution).
- Independence implies KF cannot estimate β_t from x_t

Bayes Theorem

$$\underbrace{p(\beta x|o\Theta)}_{\textit{Posterior}} \propto \underbrace{p(o|x\beta\Theta)}_{\textit{likelihood}} \; \underbrace{p(x|\beta\Theta)}_{\textit{forecast}} \; \underbrace{p(\beta|\Theta)}_{\textit{prior}}$$

where

- o: Observation at time t
- Θ: All observations up to time $t-1$.
- x: State variable
- β: Variance Parameter in stochastic-dynamical model.

Log of the Posterior

$$-2\log p(\boldsymbol{\beta}\mathbf{x}|\mathbf{o}\Theta) =$$

$$\begin{array}{lr}
(\mathbf{o}-\mathbf{H}\mathbf{x})^T\,\mathbf{R}^{-1}\,(\mathbf{o}-\mathbf{H}\mathbf{x}) + \log|\mathbf{R}| + M_o\log 2\pi + & \textit{Likelihood} \\
(\mathbf{x}-\boldsymbol{\mu}_f)^T\,\mathbf{P}^{f\,-1}\,(\mathbf{x}-\boldsymbol{\mu}_f) + \log|\mathbf{P}^f| + M_x\log 2\pi + & \textit{Forecast} \\
(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta)^T\,\Sigma_\beta^{-1}\,(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta) + \log|\Sigma_\beta| + M_\beta\log 2\pi & \textit{Prior}
\end{array}$$

Adjoint Parameter Estimation

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

$$J = (\mathbf{o} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{o} - \mathbf{H}\mathbf{x}) + \\ (\mathbf{x} - \boldsymbol{\mu}_f)^T \mathbf{P}^{f^{-1}} (\mathbf{x} - \boldsymbol{\mu}_f) + (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)^T \Sigma_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta),$$

This is the non-constant part of the posterior provided the forecast covariance $\mathbf{P}^f$ is fixed.

But in stochastic parameter estimation, $\mathbf{P}^f$ is not constant!

Stochastic Parameter Estimation Requires Varying $\mathbf{P}^f$

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

$$\frac{\partial \text{ posterior}}{\partial \mathbf{x}} = 0 \qquad \text{Standard Kalman Filter update (for fixed } \beta\text{)}$$

$$\frac{\partial \text{ posterior}}{\partial \beta} = 0 \qquad \text{New nonlinear equation to solve for } \beta$$

The key difference from past studies (e.g., adjoint methods) is that I do not assume that $\partial \mathbf{P}^f / \partial \beta$ vanishes.

Estimating Derivatives of Covariance Matrices

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

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

Connections

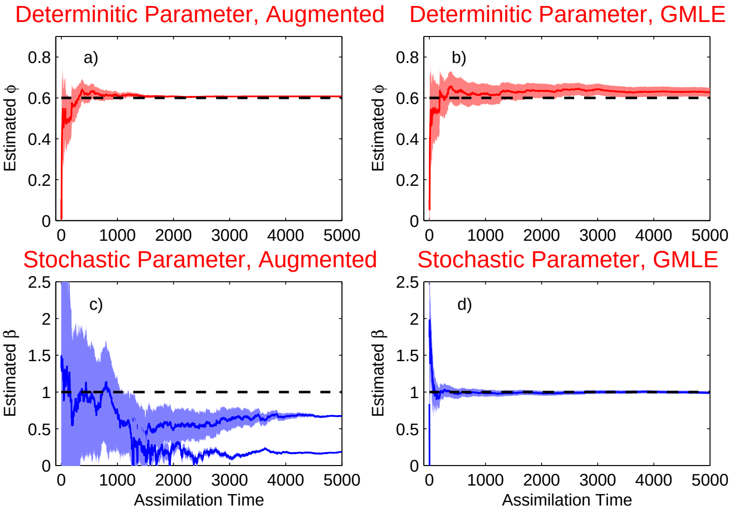
GMLE does not distinguish stochastic and deterministic parameters

- deterministic parameters characterized by $\partial \boldsymbol{\mu}_f / \partial \boldsymbol{\beta} \neq 0$.
- stochastic parameters characterized by $\partial \mathbf{P}_f / \partial \boldsymbol{\beta} \neq 0$

Augmentation is equivalent to GMLE of deterministic parameters, if Σ_β is interpreted as the spread of the parameter ensemble.

Stochastic Parameter Estimation

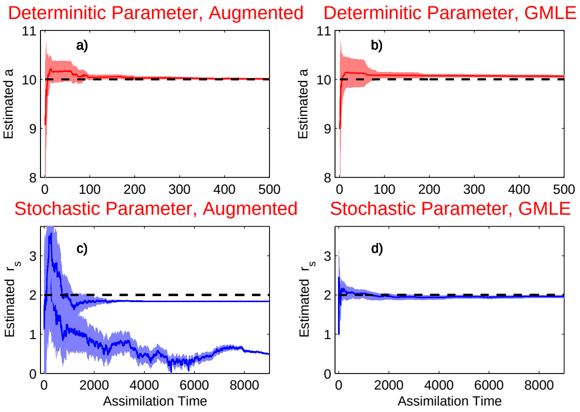
$$x_t = \phi x_{t-1} + \beta w_t,$$



Parameter Estimation in Stochastic Lorenz Model

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

$$dx = -a(x - y)dt$$
$$dy = (rx - xz - y)dt + r_s x \circ dw$$
$$dz = (xy - bz)dt$$



Two-Stage Ensemble Generation

State-parameter estimation can be achieved in two stages.

1. State-only data assimilation produces ensemble $\mathbf{X}^a$.
2. Ensemble is "corrected" to account for parameter estimation:

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

where

$$\mathbf{w} = -\mathbf{X}^{a\,T} \left(\mathbf{P}^{f\,-1} \frac{\partial \mathbf{P}^f}{\partial \beta_k} \mathbf{P}^{f\,-1} \left(\boldsymbol{\mu}_a - \boldsymbol{\mu}_f \right) - \mathbf{P}^{f\,-1} \frac{\partial \boldsymbol{\mu}_f}{\partial \beta_k} \right)$$

$\delta_+ =$ 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?