Joint State-Parameter Estimation by Two-stage Filtering

Naratip Santitissadeekorn*, *Chris Jones***

Department of Mathematics
University of Surrey*
Department of Mathematics
University of North Carolina-Chapel Hill**

DARC Seminar, University of Reading, 19 Nov 2014

Joint state-parameter estimation for EnKF

- Incorrect parameters causes unrealistic state forecasts (leading to ensemble collapse)
- Goal: simultaneously estimate the state vector and calibrating parameters
- Intuitive approach: Augmented method, applying EnKF update to $z = [x, \theta]$.
- Augmented EnKF works well in many cases, but typically not for identifying “stochastic parameters" (e.g. noise and observation covariances).

Previous research

- **Mehra (IEEE, auto. con. 1970)**: Asymptotically unbiased and consistent estimates for Q and R (with known structures) based on innovation sequence, similar to the results in Bélanger (1974).
- **Dee, Cohn, Dalcher, Ghil. (IEEE, auto. con., 1985)**: Used results in Bélanger, but with a better algorithm
- **Dee (MWR, 1995)**: optimize ML for the predictive distribution, designed for 4D-Var
- **Griffith and Nichols (FTC, 2001)**: Augmented 4D-Var, evolving errors with the states
- **Chapnik, Desroziers, Talagrand and Rabier (QJRM Soc. 2004)**: Derive zero gradient conditions for $p(y|r, m)$, which give equations for a fixed-point iteration scheme. Q and R is known up to a scalar r^2 and m^2.
- **Stroud and Bengtsson (MWR 2007)**: Use the Gamma prior and assume linear dynamic. (Hierarchial Bayes)
- **L-Curve with χ^2 diagnosis**: for $Q = m^2 I$ and $R = r^2 I$ in 3D-Var setting

More recent research

- More recent methods are geared toward a combination with EnKF
- **Berry and Sauer (Tellus A, 2013)**: Modifying the updating scheme based on innovation lag-k covariance in Mehra (1970) for EnKF
- **Koyama et al. (MWR, 2010)**: EnKF-EnKF
- **Yang and DelSole (Phys. D, 2010)**:MLE+EnKF, Forecast state $\rightarrow$ update variance by solving a nonlinear equation (required first and second derivatives of $\mathbf{P}^f$ w.r.t variance parameters)$\rightarrow$ Re-forecast state with new variance$\rightarrow$EnKF update for the state
- **Frei and Künsch (MWR 2012)**:Weighted Gaussian mixture, the weight comes from evaluating the predictive density
- All of these works have demonstrated that the augmented EnKF does NOT work well for stochastic parameters.

Augmented ENKF

- Parameter set θ is augmented to the state variable x, say $z_k = [x_k, \theta_k]$.

$$z_{k+1} = \tilde{\Phi}_k^{k+1}(z_k) = \begin{bmatrix} \Phi_k^{k+1}(x_k, \theta_k) \\ \theta_k \end{bmatrix}.$$

$$y_{k+1} = \mathbf{H}x + \varepsilon_k,$$

- The observation operator for the augmented system

$$\tilde{\mathbf{H}} z_k = [\mathbf{H} \quad 0] z_k = \mathbf{H} x_k.$$

- The error covariance matrix takes the form

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_x & \mathbf{P}_{x\theta} \\ \mathbf{P}_{x\theta}^T & \mathbf{P}_\theta \end{bmatrix},$$

- Parameter inference relies on $\mathbf{P}_{x\theta}$

Augmented ENKF

- KF update: $z^a = z^b + \mathbf{K}(y - \tilde{\mathbf{H}} z^b)$
- The Kalman gain

$$\mathbf{K} := \begin{bmatrix} \mathbf{K}_x \\ \mathbf{K}_\theta \end{bmatrix} = \mathbf{P}\tilde{\mathbf{H}}^T(\tilde{\mathbf{H}}^T \mathbf{P} \tilde{\mathbf{H}}^T + \mathbf{R})^{-1}.$$

- The above augmented system can be rewritten by

$$\begin{aligned} x^a &= x^b + \mathbf{K}_x(y - \tilde{\mathbf{H}} x^b) \\ \theta^a &= \theta^b + \mathbf{K}_\theta(y - \tilde{\mathbf{H}} x^b), \end{aligned}$$

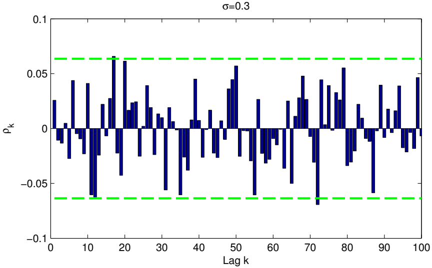
$$\begin{aligned} \mathbf{K}_x &= \mathbf{P}_x \mathbf{H}^T(\mathbf{H}\mathbf{P}_x\mathbf{H}^T + \mathbf{R})^{-1} \\ \mathbf{K}_\theta &= \mathbf{P}_{x\theta}^T \mathbf{H}^T(\mathbf{H}\mathbf{P}_x\mathbf{H}^T + \mathbf{R})^{-1}. \end{aligned}$$

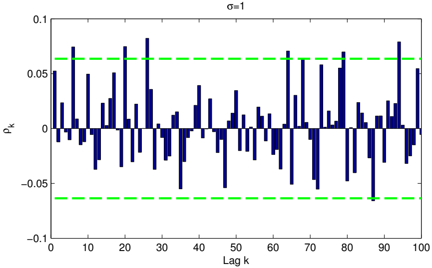
Experiment: AR(1)

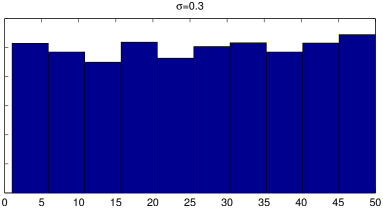
$$x_k = ax_{k-1} + \sigma\eta, \quad \eta \sim \mathcal{N}(0,1)$$

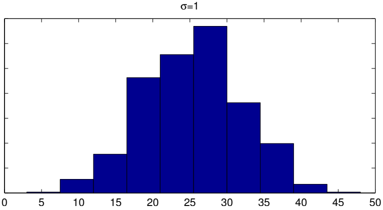
- Prior: $a \sim \mathcal{U}([0.1, 1])$ and $\sigma \sim \mathcal{U}([0.01, 1])$
- $\theta_k = \alpha\theta_{k-1} + (1-\alpha)\bar{\theta} + (h^2 S)\eta_k \quad (\alpha = \sqrt{1-h^2})$.
- Observation noise standard deviation $r^2 = 0.05$
- How do we know if there is inconsistency?
- Many ways: rank histogram, lag-k autocorrelation for the innovation sequence, etc.

Diagnostic: AR(1)



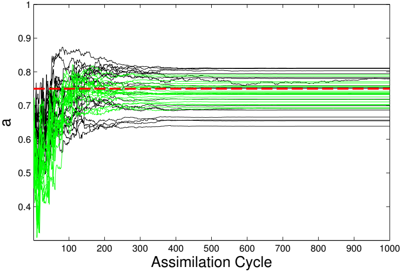


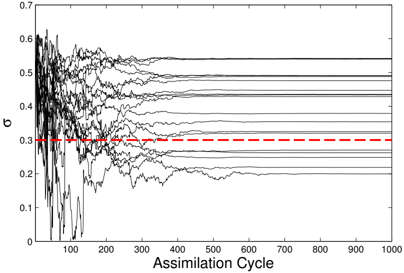




Result: AR(1)

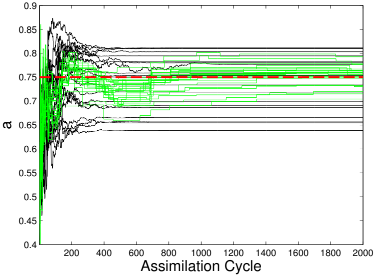
- **Green**: EnKF with unknown AR coefficient a
- **Black**: EnKF with both a and σ unknown

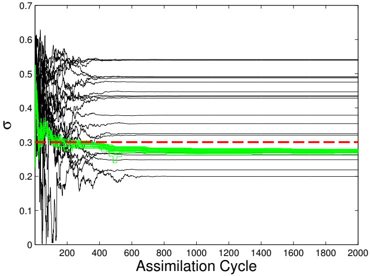




Result: AR(1)

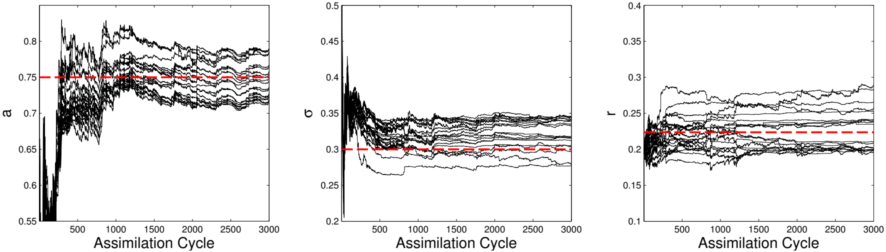
- Again, both parameters *a* and σ are unknown
- EnKF (Black) vs. Two-stage filter (Green)





Result: AR(1)

- a, σ, observation noise (r) unknown
- Using two-stage filtering to estimate them



Result: AR(1)

- Suppose $a < 1$ is known and $\sigma_{k+1} = \sigma_k$.

$$\hat{\sigma}_k^a = \hat{\sigma}_k^f + \frac{cov(x_k^f, \sigma_k^f)}{r + var(x_k^f)}(y_k^o - \hat{x}_k^f)$$

- But $\hat{x}_k^f = a\hat{x}_{k-1}^a + \sigma E[\eta_k] = a\hat{x}_{k-1}^a$.
- $cov(x_k^f, \sigma_k^f)$ and $var(x_k^f)$ depends on σ
- $cov(x_k^f, \sigma_k^f) < a \cdot cov(x_{k-1}^f, \sigma_{k-1}^f)$: covariance is damped after each step.
- Can we rely on $cov(x_k^f, \sigma_k^f)/(r + var(x_k^f))$ to update σ?

- Factorize the joint filtering density as

$$p(x_k, \theta_k | y_{1:k}) = p(x_k | \theta_k, y_{1:k}) p(\theta_k | y_{1:k}).$$

- Applying the (standard) PF to target $p(\theta_k | y_{1:k})$

$$p(x_k, y_k | y_{1:k}) \approx \sum_{i=1}^{N} \omega_k^{(i)} p(x_k | \theta_k^{(i)}, y_{1:k}) \delta(\theta_k - \theta_k^{(i)}),$$

- We need a recursive update formula for $\omega_k^{(i)}$
- We can factor $p(\theta_k | y_{1:k})$ to

$$p(\theta_k | y_{1:k}) \propto p(\theta_k | y_{1:k-1}) p(y_k | y_{1:k-1}, \theta_k)$$

Rao-Blackwellized PF (RBPF)

- The particle weight can be recursively calculated by

$$\omega_k^{(i)} \propto p(y_k | y_{1:k-1}, \theta_k^{(i)}) \omega_{k-1}^{(i)}.$$

- The predictive density conditioned on a particle $\theta_k^{(i)}$ is

$$p(y_k | y_{1:k-1}, \theta_k^{(i)}) \propto \mathcal{N}(y_k; Hx_k^f(\theta_k^{(i)}), HP^f(\theta_k^{(i)})H^T + R),$$

- Assume that $p(x_{1:k} | \theta_k^i, y_{1:k}) \approx$ conditionally Gaussian
- Approximate $p(x_{1:k} | \theta_k^i, y_{1:k})$ by EnKF
- With this setting, one parameter particle is “attached" to one (conditional) EnKF.

RBPF

- A low-cost simplification for RBPF.
- Approximation 1:

$$p(x_k|\theta_k^{(i)}, y_{1:k}) \approx p(x_k|\hat{\theta}_k, y_{1:k}),$$

- Approximation 2 (actually a consequence of 1):

$$\begin{aligned} p(y_k|y_{1:k-1}, \theta_k^{(i)}) &\propto \mathcal{N}(y_k; Hx_k^f(\theta_k^{(i)}), HP^f(\theta_k^{(i)})H^T + R) \\ &\approx \mathcal{N}(y_k; Hx_k^f(\theta_k^{(i)}), HP^f(\hat{\theta})H^T + R) \end{aligned}$$

- This will split it into two sub-filters: $p(x_k|\hat{\theta}_k, y_{1:k})p(\theta_k|y_{1:k})$
- But the weight update for θ is based only on the point estimate of the state x, but not its uncertainty.
- Temporal smoothing may be used to deal with this issue in practice (Koyama et al., MWR2010)

Two-stage Filter: PF+ENKF

- **Idea**: Run PF for parameter (assuming known states), Run EnKF for states (assuming known parameters)
- **Parameter filter with PF**: use the state estimate $\hat{x}_{t-1}$

$$\theta_t^{(i)} = \alpha\theta_{t-1}^{(i)} + (1-\alpha)\bar{\theta}_{t-1} + \eta_t^{(i)} \qquad 0 < \alpha < 1$$
$$y_t^{(i)} = h(\Phi_{t-1}^{t}(\hat{x}_{t-1}; \theta_t^{(i)}))$$

- **State filter with EnKF**: Use $\hat{\theta}_t$ to find $P(x_t | y_{1:t}, \hat{\theta}_t)$

$$x_t^{(i)} = \Phi_{t-1}^{t}(x_{t-1}^{(i)}; \hat{\theta}_t) + \nu_t^{(i)}$$
$$y_t^{(i)} = h(x_t^{(i)}; \hat{\theta}_t)$$

Shrinkage of Kernel

- **Why not persistence model?**: $\theta_t = \theta_{t-1}$
  - "Sample attrition" issue in re-weighting
  - (Parameter) particles always stay with the same sets of initial guesses
- **Random Walk (Gordon-Salmon-Smith "jittering")**: $\theta_t = \theta_{t-1} + \eta_t$
  - Lead to over-dispersion; posteriors are far too diffuse
  - A way to compensate this over-dispersion is done in the framework of kernel smoothing [Liu&West 99]
- **Liu-West "jittering"**: $\theta_t = \alpha\theta_{t-1} + (1-\alpha)\bar{\theta}_{t-1} + \eta_t$
  - Push samples θ_t toward the ensemble mean before adding a small degree of noise
  - $\eta_t \sim \mathcal{N}(0, h^2 S), \alpha = \sqrt{1-h^2}$, S is sample variance
  - $0.1 \leq h < 0.3$
  - This is helpful to parameter learning only for static parameters!

“Closure" parametrization

- **True model**: L96-fast-slow system

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F - \underbrace{\frac{hc}{b}\sum_{j=1}^{N_y} y_{j,i}}_{f_i}$$

$$\frac{dy_{j,i}}{dt} = cb(y_{j-1,i} - y_{j+2,i})y_{j+1,i} - cy_{j,i} + \frac{hc}{b}x_i y_{N_{y-1},i}$$

- $F = 8$, $N_x = 18$, $N_y = 6$, the coupling strength $h = 1$, the time scale separation $c = 10$ and the magnitude of the fast component $b = 10$
- **Forecast model**: only slow variable, fast-scale process paramatrized by

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F - g(t, x_i)$$

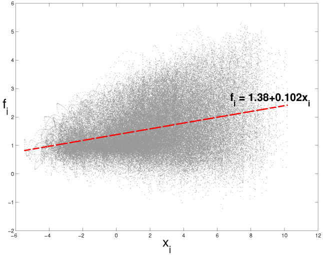
Closure parametrization: L96-fast-slow

- **Warning**: Model reduction is NOT the main goal for this study
- Stochastic: Assume $g(t, x_i) = \theta_1 + \theta_2 x_i + e_i(t)$ [Wilks 2005]

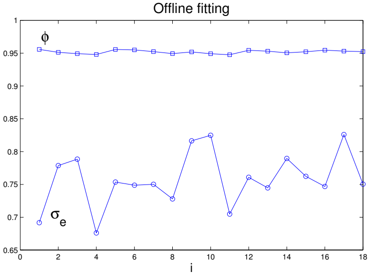
$$e_i(t) = \phi e_i(t - \Delta t) + \sigma_e (1 - \phi^2)^{1/2} \eta_i(t).$$

- Try filter θ_1, θ_2, ϕ, and σ_e

Fitting with the truth run

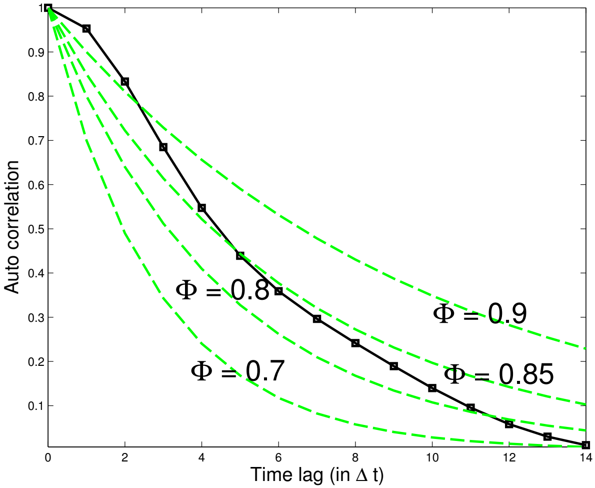


(a) Fitting θ_1 and θ_2



(b) Fitting ϕ and σ_e

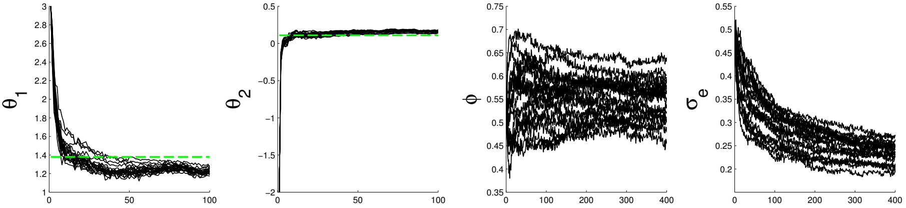
Correlogram



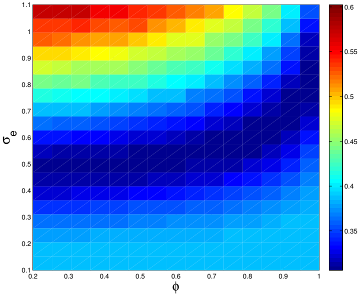
Experimental setup: L96-fast-slow

- The parameters are initially drawn from the following prior distributions; $\theta_1 \sim \mathcal{N}(3, 2)$, $\theta_2 \sim \mathcal{N}(-2, 2)$, $\phi \sim \mathcal{U}(0.1, 0.9)$, and $\sigma_e \sim \mathcal{U}(0.1, 0.9)$.
- Set $\delta t = 50\Delta t$.
- $N = 250$ for Augmented ENKF and $N = 200, M = 50$ for the two-stage filtering
- Augmented ENKF “blows-up"

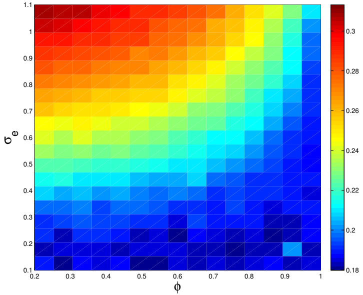
Comparing the posterior



RMSE



(c) no assimilation



(d) state only assimilation