Optimization of stochastic parameterizations for model error treatment using nested EnKFs

Guillermo Scheffler
CIMA - CONICET/Universidad de Buenos Aires, Argentina
Universidad Nacional del Nordeste, Corrientes, Argentina

In collaboration with Juan Ruiz and Manuel Pulido

Modèles spatio-temporels en météorologie et océanographie – November 2018 Rennes (France)
Problem overview

Given a (non-)linear dynamical system:

\[
\begin{align*}
x_k &= f(x_{k-1}) + v_k \\
y_k &= Hx_k + \epsilon_k
\end{align*}
\]

\[v_k \sim \mathcal{N}(0, Q)\]
\[\epsilon_k \sim \mathcal{N}(0, R)\]

with a forecast dynamical model

\[
x_k^f = f(x_{k-1}^f, \lambda) + G(\theta)
\]

\(\lambda\): model parameters
   (e.g. physical and closure)

\(\theta\): stochastic parameters
   e.g. amplitude of a stochastic forcing,
   spatial decorrelation length
What is data assimilation?

Given a (non-)linear dynamical system:

\[ x_k = f(x_{k-1}) + v_k \]
\[ y_k = Hx_k + \epsilon_k \]

with a forecast dynamical model

\[ x^f_k = f(x^f_{k-1}, \lambda) \]

Find an *optimal* estimate of the evolving state \( x^e \) as a combination of the available observations and the forecasted state.

The optimal state at time \( k \) is called *analysis*: \( x^a_k \)
Ensemble Kalman Filter (EnKF)

• To be optimal in a statistical sense, the combination of observed and forecasted states must account the observational errors covariance ($R$) and the forecast errors covariance ($P^f$)

• Monte Carlo approach to the filtering equations.

\[
\begin{align*}
    x^f_{k}^{(i)} &= f \left( x^a_{k-1}^{(i)} \right) \\
    \bar{x}_{k}^{a} &= \bar{x}_{k}^{f} + K_{k} (y_{k} - H\bar{x}_{k}^{f})
\end{align*}
\]

\[
    P_{k}^{f} = \frac{1}{N-1} \sum_{i=1}^{N} (x_{k}^{f(i)} - \bar{x}_{k}^{f}) (x_{k}^{f(i)} - \bar{x}_{k}^{f})^T
\]

\[
    K_{k} = P_{k}^{f} H^T (H P_{k}^{f} H^T + R)^{-1}
\]

• $P^f$ is dynamically evolved using statistics of an ensemble of $N$ model integrations
\[ \bar{x}_k^a = \bar{x}_k^f + K_k (y_k - H\bar{x}_k^f) \]

\[ K_k = P_k^f H^T (HP_k^f H^T + R)^{-1} \]
\[
\bar{x}_k^a = \bar{x}_k^f + K_k (y_k - H \bar{x}_k^f)
\]

\[
K_k = P_k^f H^T (H P_k^f H^T + R)^{-1}
\]
State-augmentation: Model parameters can also be inferred, by considering them as state variables.

\[
\begin{align*}
x_{k}^{f(i)} &= \begin{bmatrix} x_{k}^{f(i)} \\ \lambda^{(i)} \end{bmatrix} \\
P_{k}^{f} &= \begin{bmatrix} P_{k}^{xx} & P_{k}^{x\lambda} \\ P_{k}^{\lambda x} & P_{k}^{\lambda\lambda} \end{bmatrix} \\
\bar{x}_{k}^{a} &= \bar{x}_{k}^{f} + K_{k}^{x} (y_{k} - H\bar{x}_{k}^{f}) \\
K_{k}^{x} &= P_{k}^{xx} H^{T} \left( H P_{k}^{xx} H^{T} + R \right)^{-1}
\end{align*}
\]

Initial conditions \( x_{i} \)
Perturbed parameters \( \lambda_{i} \)

Figure from FMI (http://en.ilmatieteenlaitos.fi/)
**State-augmentation:** Model parameters can also be inferred, by considering them as state variables

Initial conditions $x_i$

Perturbed parameters $\lambda_i$

\[
x^*_k f(i) = \begin{bmatrix} x_k^f(i) \\ \lambda(i) \end{bmatrix}
\]

\[
P_k^f = \begin{bmatrix} P_{xx}^f & P_{x\lambda}^f \\ P_{\lambda x}^f & P_{\lambda\lambda}^f \end{bmatrix}
\]

\[
\bar{x}_k^a = \bar{x}_k^f + K_k^x (y_k - H\bar{x}_k^f)
\]

\[
K_k^x = \begin{bmatrix} P_{xx}^f \end{bmatrix} H^T (HP_{xx}^f H^T + R)^{-1}
\]

\[
\bar{\lambda}_k^a = \bar{\lambda}_k^f + K_k^\lambda (y_k - H\bar{x}_k^f)
\]

\[
K_k^\lambda = \begin{bmatrix} P_{\lambda x}^f \end{bmatrix} H^T (HP_{xx}^f H^T + R)^{-1}
\]

Exploits state-parameter covariances $P_k^{\lambda x}$

Figure from FMI (http://en.ilmatieteenlaitos.fi/)
Estimation of stochastic parameters?

\[ x_k = \hat{f}(x_{k-1}) + \nu_k \]
\[ y_k = Hx_k^t + \epsilon_k \]

Now the dynamical model incorporates a stochastic term that depends on parameter(s) \( \theta \):

\[ x_k^f = f (x_{k-1}^f, \lambda) + G(\theta) \]

\( \theta \): stochastic parameters

e.g. amplitude of a stochastic forcing,
spatial decorrelation length
Two-scales Lorenz-96 system

• True model:

\[
\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b} \sum_{j=J(k-1)+1}^{kJ} Y_j; \quad (1)
\]

\[
\frac{dY_j}{dt} = -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j + \frac{hc}{b} X_{int}\left[\frac{j-1}{J}\right]+1; \quad (2)
\]

\[\rightarrow F=20 \quad \text{chaotic regime} \]
\[n_x=8 \quad \text{slow variables} \]
\[J=256 \quad \text{fast variables} \]
Two-scales Lorenz-96 system

• True model:

\[
\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b} \sum_{j=J(k-1)+1}^{kJ} Y_j; \quad (1)
\]

\[
\frac{dY_j}{dt} = -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j + \frac{hc}{b} X_{\text{int}[j-1]/J+1}; \quad (2)
\]

• Truncated model:

\[
\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + \sum_{d=0}^{D} a_d X_k^d + e_i(\theta, t)
\]

\[
e(t) = \phi e(t - \Delta t) + (1 - \phi^2)\frac{1}{2} z(t); \quad z \sim N[0, \Sigma(\theta)]
\]
**Stochastic parametrization** (AR(1) process with fixed $\phi$)

\[ e(t) = \phi e(t - \Delta t) + (1 - \phi^2) \frac{1}{2} z(t); \quad z \sim N(0, \Sigma(\theta)) \]

Assuming a simple covariance model

\[ \Sigma = \sigma^2 I \]

Even on a twin experiments framework, EnKF with augmented state fails

---

Figure taken from Santitisaadetkorn and Jones (2015). Similar conclusions found by Delsole and Yang (2009)
Parameters associated to the amplitude of (additive) stochastic parameterizations cannot be inferred using an EnKF state-augmentation approach.

\[
P^f_k = \begin{bmatrix} P^{xx}_k & P^{x\lambda}_k \\ P^{\lambda x}_k & P^{\lambda\lambda}_k \end{bmatrix}
\]

On a sequential data assimilation framework, covariance \(P^{\lambda x}_k\) converge to zero, especially if the number of ensemble members increase and the model is sufficiently linear.

Reasons (DelSole & Yang 2009):

i) A state-augmented ensemble integration does not allow to infer \(\text{cov}(x,p)\).

ii) Weak or null correlations between stochastic parameters and observations.

Sophisticated algorithms like Expectation-Maximization are able to estimate these type of parameters (i.e. Dreano et al. 2017).
• Assuming: $\Sigma = \sigma^2 I$

Grid based exploration of the parameter space (a.k.a. brute force searching)

ETKF data assimilation experiment repetitions testing different values of $\sigma$

Average RMSE over different experiments with independent observation error samples

ETKF:
- 2000 assimilation cycles
- 30 member ensemble
- $R = I$
- $H = I$
- NO covariance inflation

A minimum was identified.

(Analysis RMSE is quite convex!)
Ensemble 1 – Parameters $\theta_1$

Ensemble 2 – Parameters $\theta_2$

Ensemble NJ – Parameters $\theta_{NJ}$

State estimation

$\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$

$t_1$, $t_2$, $t_3$, Time
In every state assimilation cycle, each ensemble is filtered independently:

\[
\{\bar{x}_{t_1}^a(\theta^1), \bar{x}_{t_2}^a(\theta^1), \bar{x}_{t_3}^a(\theta^1), \ldots \}
\]

\[
\{\bar{x}_{t_1}^a(\theta^2), \bar{x}_{t_2}^a(\theta^2), \bar{x}_{t_3}^a(\theta^2), \ldots \}
\]

\[
\{\bar{x}_{t_1}^a(\theta_{NJ}), \bar{x}_{t_2}^a(\theta_{NJ}), \bar{x}_{t_3}^a(\theta_{NJ}), \ldots \}
\]

We store the mean forecasted state, to compute the parameter update:

\[
\{\bar{x}_{t_1}^f(\theta^1), \bar{x}_{t_2}^f(\theta^1), \bar{x}_{t_3}^f(\theta^1), \ldots \}
\]

\[
\{\bar{x}_{t_1}^f(\theta^2), \bar{x}_{t_2}^f(\theta^2), \bar{x}_{t_3}^f(\theta^2), \ldots \}
\]

\[
\{\bar{x}_{t_1}^f(\theta_{NJ}), \bar{x}_{t_2}^f(\theta_{NJ}), \bar{x}_{t_3}^f(\theta_{NJ}), \ldots \}
\]
The "outer" assimilation cycle exploits covariances between parameters and the mean states.

\[
\theta^a = \theta^f + \sum^K_{k=1} P_{k}^{\theta \bar{x}} H^T [H P_{k}^{\bar{x} \bar{x}} H^T + (H \bar{P}_{k} H^T + R)] (y_k - H \bar{x}_{k}^f)
\]

\[
P_k = \begin{bmatrix}
P_{k}^{\bar{x} \bar{x}} & P_{k}^{\bar{x} \theta} \\
P_{k}^{\theta \bar{x}} & P_{k}^{\theta \theta}
\end{bmatrix}
\]
1. Given $N_J$ parameters $\theta^{1:NJ}$ and $N_J$ independent ensembles of $N_I$ each: $x^{1:N_I,1:NJ}$, and (LxK) observations $y_{1:L,1:K}$

2. State estimation cycle: For each state assimilation cycle $l=1...L$
   
   2.1 For each inner assimilation cycle $k=1...K$
      
      2.1.1 Calculate the ensembles of analysed states $x_{l,k}^{a(j,i)}$ through $N_J$-EnKFs independently
      
      2.1.2 Store the mean predicted observation $H\tilde{x}_{l,k}^{(j)}$ for each ensemble and the average of the forecast error covariance matrix in the observational space $H\tilde{P}_{l,k}H^T = \frac{1}{N_J} \sum_j H\tilde{P}_{l,k}^j H^T$

2.2 Parameter estimation:
   
   2.2.1 Concatenate the K mean predicted observations to construct an $(n_x \times K)$-dimensional ensemble
      
   2.2.2 Construct an aggregated observations vector $y_{l}^* = [y_{l,1}, ..., y_{l,K}]^T$.

   2.2.3. Obtain the updated parameter ensemble mean and perturbations through an ETKF using aggregated matrices from (2.2.1) and (2.2.2)
Isotropic covariance model $\Sigma$: 

$$\Sigma = \sigma^2 I$$

Synthetic truth: $\sigma^2 = 2$

Stochastic parameter estimation. Results for experiments with different observation error samples and initial conditions.

Dots: Analysed state RMSE with nested EnKFs. Exhaustive exploration of parameter space evaluating RMSE (Blue), and $\chi^2$ statistic of rank histograms uniformity (Black)

$H = I$ and $R = I$

$N_f = 15$ (number of filters)

$N_i = 30$ (ens. members)

$\Delta t = 0.005$

$\Delta t_{obs} = 10 \Delta t$
Assuming AR(1) process

\[ e(t) = \phi e(t - \Delta t) + (1 - \phi^2)\frac{1}{2}z(t); \quad z \sim N(0, \Sigma(\theta)) \]

Possible stochastic forcing covariance hypothesis

- Isotropic \( \Sigma \):
  \[ \Sigma = \sigma^2 I \]

- Non-isotropic diagonal \( \Sigma \):
  \[ \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2) \]

- Exponential covariance \( \Sigma \):
  \[ \Sigma \equiv \Sigma_{ij} = \sigma^2 \exp(-\rho|i - j|) \]

- Symmetric circulant \( \Sigma \):
  \[ \sigma_{i,i-n} = \sigma_{i,i+n} \]
Exponential covariance model $\Sigma$:

$$\Sigma \equiv \Sigma_{ij} = \sigma^2 \exp(-\rho |i - j|)$$

Synthetic truth: $\sigma^2 = 2$

$\rho = 0.3$
• Symmetric circulant $\Sigma$:

\[
\Sigma = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{12} \\
\sigma_{12} & \sigma_1^2 & \cdots & \sigma_{13} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{12} & \sigma_{13} & \cdots & \sigma_1^2
\end{bmatrix}
\]

Real model

Stochastic parameter estimation Results for experiments with different observation error samples and initial conditions.
• Simultaneous estimation: Multiplicative inflation + $\sigma$

(Isotropic non-correlated case)

$H=I$ and $R=0.15^2I$

$N_J=15$ (number of filters)

$N_I=30$ (ens. members)

$\Sigma = \sigma \Sigma^2 I$

Estimations for different observation error samples and initial conditions.

EnKF analysis RMSE as a function of the inflation factor and $\sigma$ by exhaustive parameter space exploration.
Can other EnKF parameters be inferred using data assimilation schemes?

- Additive inflation parameters?
- P-localization scale?
- Observational error parameters?
- Weighting parameter on hybrid Ensemble-4DVar schemes?

Do not affect model dynamics.

Currently working on a simplification of the scheme for these parameters
Summary

• Stochastic parameters show a poor observability and distinguishability \(\rightarrow\) standard state augmentation methods fail

• We present a hierarchical ensemble DA method to infer (offline) parameters, at a cost comparable to iterative Expectation-Maximization algorithms (Dreano et al. 2017) or SMC\(^2\) (Chopin et al. 2013)

• Estimated parameters minimize analysis RMSE and provide and optimal RMSE/spread ratio

Manuscript available at: https://arxiv.org/abs/1807.10858
\[ p(x_{t+1}, \theta_{t+1} | y_{t,1:K}) = p(x_{t+1} | \theta_{t+1}, y_{t,1:K}) p(\theta_{t+1} | y_{t,1:K}) \]

Sequentially along the DA window

\[ p(\theta_{t+1} | y_{t,1:K}) \propto p(\theta_{t+1} | y_{t-1,1:K}) \prod_{k=1}^{K} p(y_{t,k} | y_{t,1:k-1} | y_{t,1:K}) \]

Accurate marginalization over the full state $x$ is quite expensive

\[ p(y_{t,k} | y_{t,k-1}, \theta_{t+1}) = \int p(y_{t,k} | x_{t,k}, \theta_{t+1}) p(x_{t,k} | y_{t,1:k-1}, \theta_{t+1}) dx_{t,k} \]
\[
p(y_{l,k} | y_{l,k-1}, \theta_{l+1}) = \int p(y_{l,k} | x_{l,k}, \theta_{l+1}) p(x_{l,k} | y_{l,1:k-1}, \theta_{l+1}) \, dx_{l,k} \tag{1}
\]

Instead we assume (Pulido et al. 2018):

\[
p(y_{l,k} | x_{l,k}, \theta_{l+1}) \propto \exp \left[ \left( y_{l,k} - \mathcal{H}(x_{l,k}) \right)^T R^{-1} \left( y_{l,k} - \mathcal{H}_{l,k}(x_{l,k}) \right) \right]
\]

\[
p(x_{l,k} | y_{l,1:k-1}, \theta_{l+1}) \propto \exp \left[ \left( x_{l,k} - \bar{x}_{l,k}^f(\theta_{l+1}) \right)^T P_{l,k}(\theta_{l+1})^{-1} \left( x_{l,k} - \bar{x}_{l,k}^f(\theta_{l+1}) \right) \right]
\]

Replacing in (1):

\[
\prod_{k=1}^{K} p(y_{l,k} | y_{l,k-1}, \theta_{l+1}) \propto \prod_{k=1}^{K} \exp \left[ \left( y_{l,k} - \mathcal{H}_{l,k}(\bar{x}_{l,k}^f) \right)^T \left( H_{l,k} P_{l,k} H_{l,k}^T + R \right)^{-1} \left( y_{l,k} - \mathcal{H}_{l,k}(\bar{x}_{l,k}^f) \right) \right]
\]

Likelihood solved through the nested ensemble Kalman filters