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:

$$\boldsymbol{x}_k = \tilde{f}(\boldsymbol{x}_{k-1}) + \boldsymbol{\nu}_k$$

$$\mathbf{y}_k = \mathbf{H}\boldsymbol{x}_k + \boldsymbol{\epsilon}_k$$

$$\boldsymbol{\nu}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

$$\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

with a forecast dynamical model

$$\boldsymbol{x}_k^f = f(\boldsymbol{x}_{\mathrm{k}-1}^f, \boldsymbol{\lambda}) + G(\boldsymbol{\theta})$$

$\boldsymbol{\lambda}$: model parameters

(e.g. physical and closure)

$\boldsymbol{\theta}$: stochastic parameters

e.g. amplitude of a stochastic forcing,

spatial decorrelation length

What is data assimilation?

Given a (non-)linear dynamical system:

$$\boldsymbol{x}_k = \hat{f}(\boldsymbol{x}_{k-1}) + \boldsymbol{\nu}_k$$

$$\mathbf{y}_k = \mathbf{H}\boldsymbol{x}_k + \boldsymbol{\epsilon}_k$$

$$\boldsymbol{\nu}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

$$\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

with a forecast dynamical model

$$\boldsymbol{x}_k^f = f(\mathbf{x}_{\mathrm{k-1}}^f, \boldsymbol{\lambda})$$

Find an *optimal* estimate of the evolving state $\boldsymbol{x}^t$ as a combination of the available observations and the forecasted state.

The optimal state at time k is called *analysis:* $\boldsymbol{x}_k^a$

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.

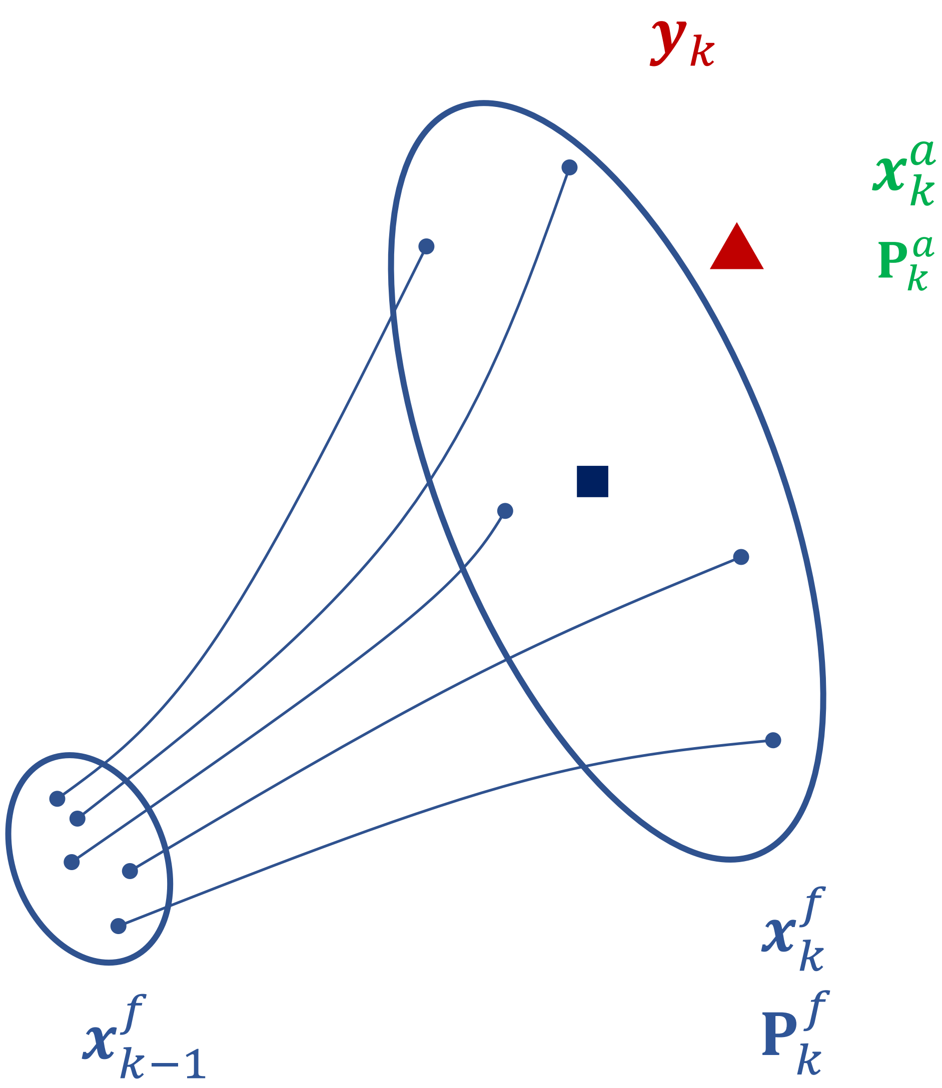
$$\boldsymbol{x}_k^{f(i)} = f\left(\boldsymbol{x}_{k-1}^{a(i)}\right)$$

$$\overline{\boldsymbol{x}}_k^a = \overline{\boldsymbol{x}}_k^f + \mathbf{K}_k(\boldsymbol{y}_k - \mathbf{H}\overline{\boldsymbol{x}}_k^f)$$

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_k^f \mathbf{H}^\mathrm{T} + \mathbf{R}\right)^{-1}$$

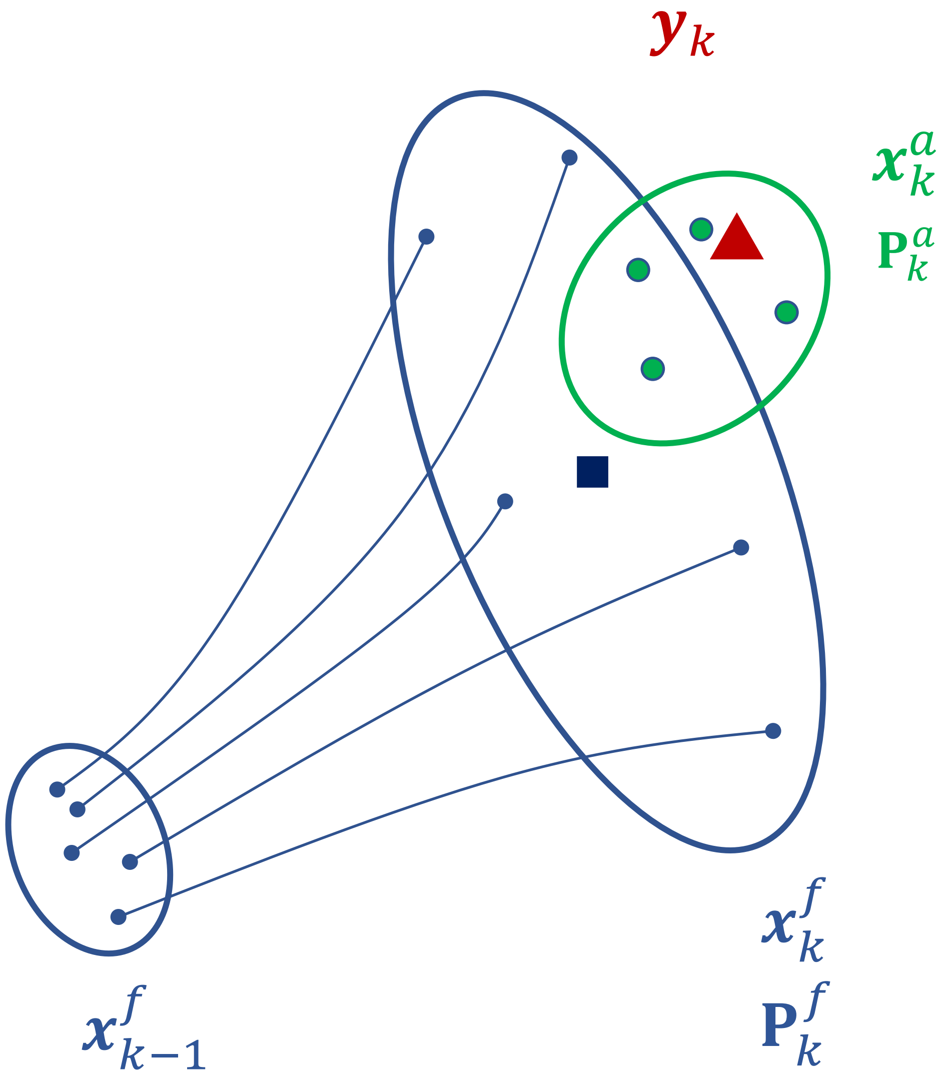
$$\boldsymbol{P}_k^f = \frac{1}{N-1}\sum_{i=1}^{N}\left(\boldsymbol{x}_k^{f(i)} - \overline{\boldsymbol{x}}_k^f\right)\left(\boldsymbol{x}_k^{f(i)} - \overline{\boldsymbol{x}}_k^f\right)^\mathrm{T}$$

- **P**^{f} is dynamically evolved using statistics of an ensemble of *N* model integrations



$$\overline{\boldsymbol{x}}_k^a = \overline{\boldsymbol{x}}_k^f + \mathbf{K}_k(\boldsymbol{y}_k - \mathbf{H}\overline{\boldsymbol{x}}_k^f)$$

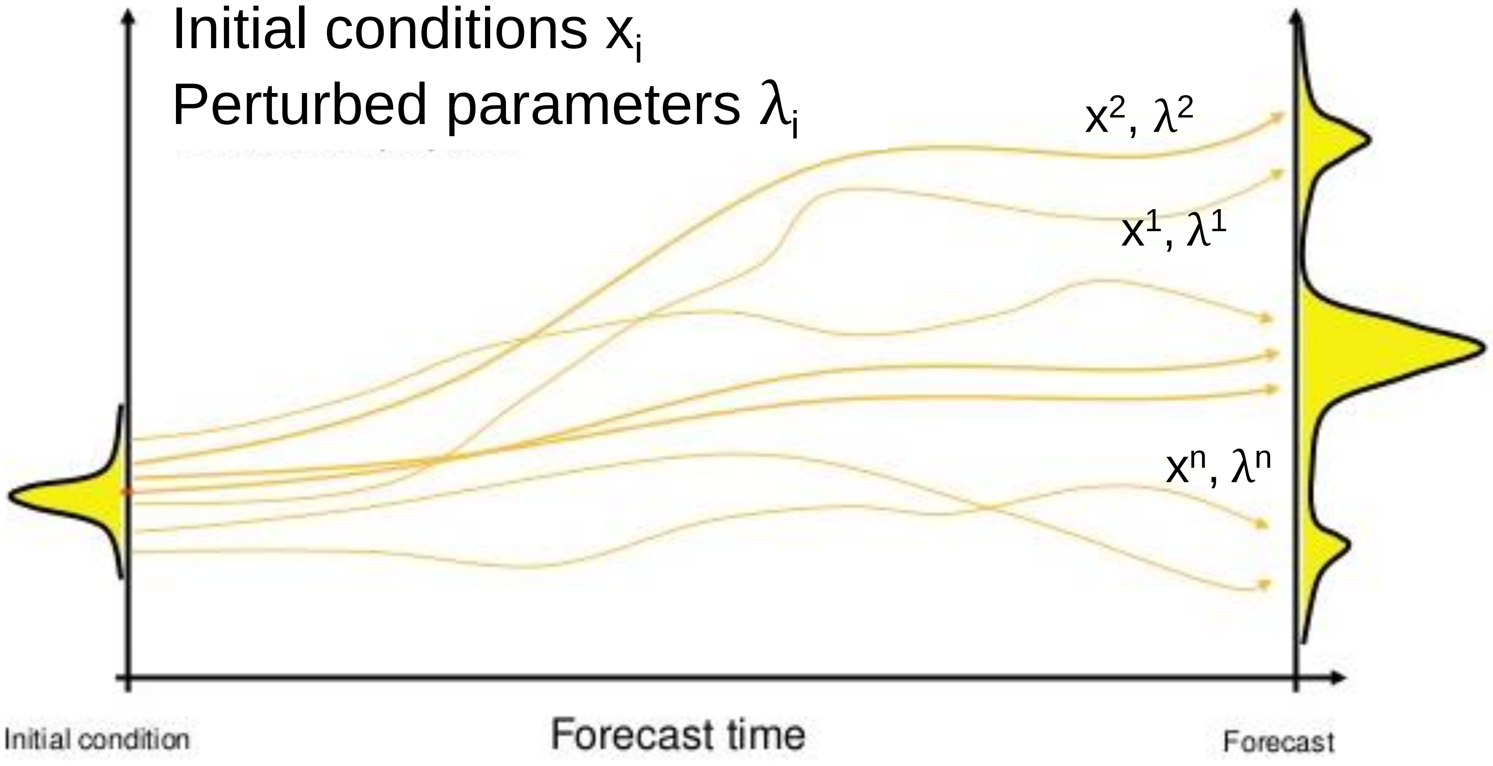
$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_k^f \mathbf{H}^\mathrm{T} + \mathbf{R}\right)^{-1}$$



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

$$K_k = P_k^f H^T \left(HP_k^f H^T + R\right)^{-1}$$

State-augmentation: Model parameters can also be inferred, by considering them as state variables



$$\mathbf{x^*}_{\mathrm{k}}^{f(i)} = \begin{bmatrix} \mathbf{x}_{\mathrm{k}}^{\mathrm{f(i)}} \\ \boldsymbol{\lambda}^{(i)} \end{bmatrix}$$

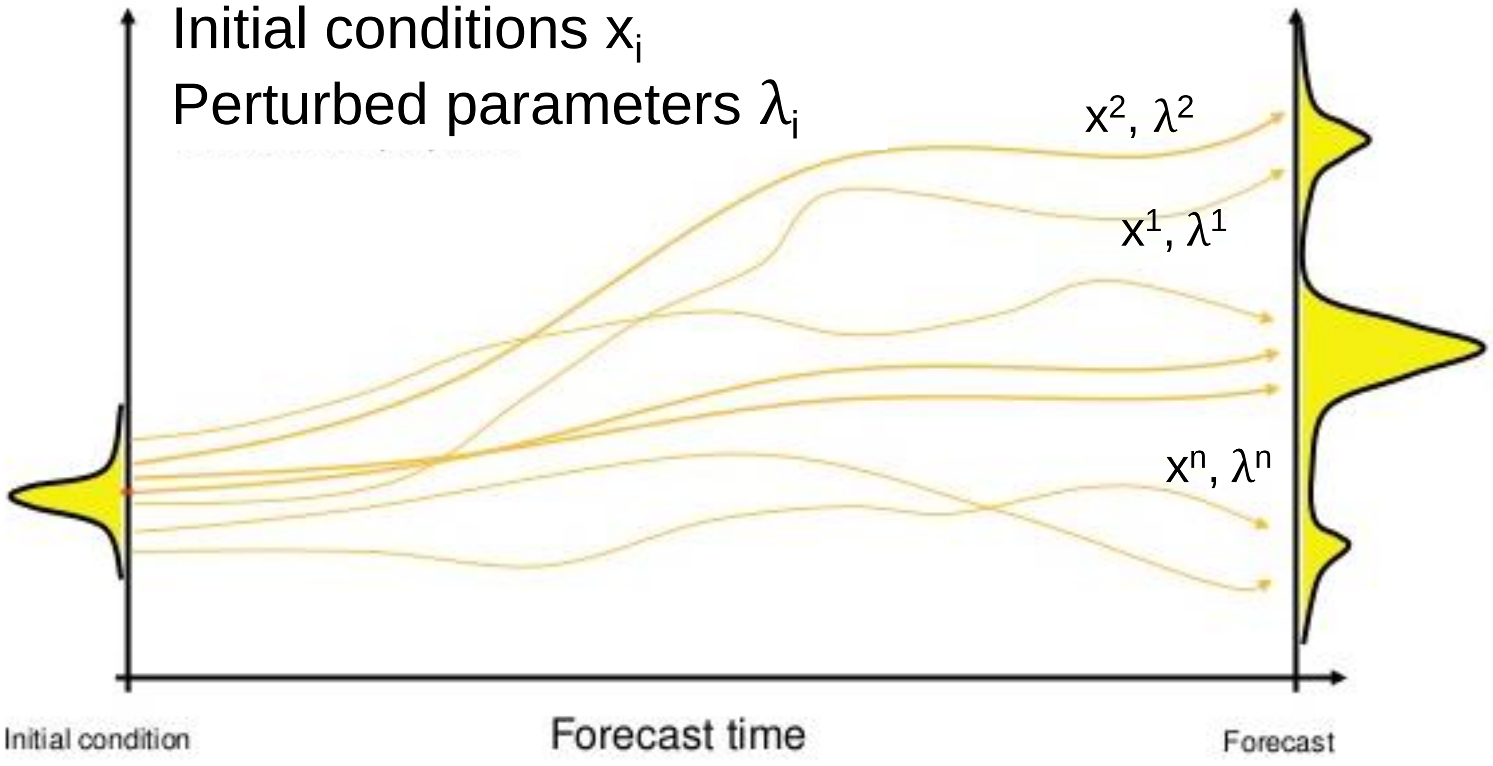
$$\mathbf{P}_k^f = \begin{bmatrix} \mathbf{P}_k^{xx} & \mathbf{P}_k^{x\lambda} \\ \mathbf{P}_k^{\lambda x} & \mathbf{P}_k^{\lambda\lambda} \end{bmatrix}$$

$$\overline{\boldsymbol{x}}_k^a = \overline{\boldsymbol{x}}_k^f + \mathbf{K}_k^{\boldsymbol{x}}(\boldsymbol{y}_k - \mathbf{H}\overline{\boldsymbol{x}}_k^f)$$

$$\mathbf{K}_k^x = \mathbf{P}_k^{xx}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_k^{xx}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

Figure from FMI (http://en.ilmatieteenlaitos.fi/)

State-augmentation: Model parameters can also be inferred, by considering them as state variables



$$\mathbf{x}^{*f(i)}_{\mathrm{k}} = \begin{bmatrix} \mathbf{x}_{\mathrm{k}}^{\mathrm{f(i)}} \\ \boldsymbol{\lambda}^{(i)} \end{bmatrix}$$

$$\mathbf{P}_k^f = \begin{bmatrix} \mathbf{P}_k^{xx} & \mathbf{P}_k^{x\lambda} \\ \mathbf{P}_k^{\lambda x} & \mathbf{P}_k^{\lambda\lambda} \end{bmatrix}$$

$$\overline{\boldsymbol{x}}_k^a = \overline{\boldsymbol{x}}_k^f + \mathbf{K}_k^{\boldsymbol{x}}(\boldsymbol{y}_k - \mathbf{H}\overline{\boldsymbol{x}}_k^f)$$

$$\mathbf{K}_k^x = \mathbf{P}_k^{xx}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_k^{xx}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

$$\overline{\boldsymbol{\lambda}}_k^a = \overline{\boldsymbol{\lambda}}_k^f + \mathbf{K}_k^{\lambda}(\boldsymbol{y}_k - \mathbf{H}\overline{\boldsymbol{x}}_k^f)$$

$$\mathbf{K}_k^\lambda = \mathbf{P}_k^{\lambda x}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_k^{xx}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

Exploits state-parameter covariances $\mathbf{P}_k^{\lambda x}$

Figure from FMI (http://en.ilmatieteenlaitos.fi/)

Estimation of stochastic parameters?

$$\boldsymbol{x}_k = \hat{f}(\mathbf{x}_{k-1}) + \boldsymbol{\nu}_k$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k^t + \boldsymbol{\epsilon}_k$$

Now the dynamical model incorporates a stochastic term that depends on parameter(s) $\boldsymbol{\theta}$

$$\boldsymbol{x}_k^f = f(\boldsymbol{x}_{\mathrm{k}-1}^f, \boldsymbol{\lambda}) + G(\boldsymbol{\theta})$$

$\boldsymbol{\theta}$: stochastic parameters

e.g. amplitude of a stochastic forcing,
spatial decorrelation length

Two-scales Lorenz-96 system

- True model:

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=J(k-1)+1}^{kJ} Yj; \qquad (1)$$

k=1…n_x

j=1…J

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

→ F=20 chaotic regime

n_x=8 slow variables

J=256 fast variables

Two-scales Lorenz-96 system

- True model:

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

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

Deterministic parameterization

- Truncated model:

Stochastic parameterization

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

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

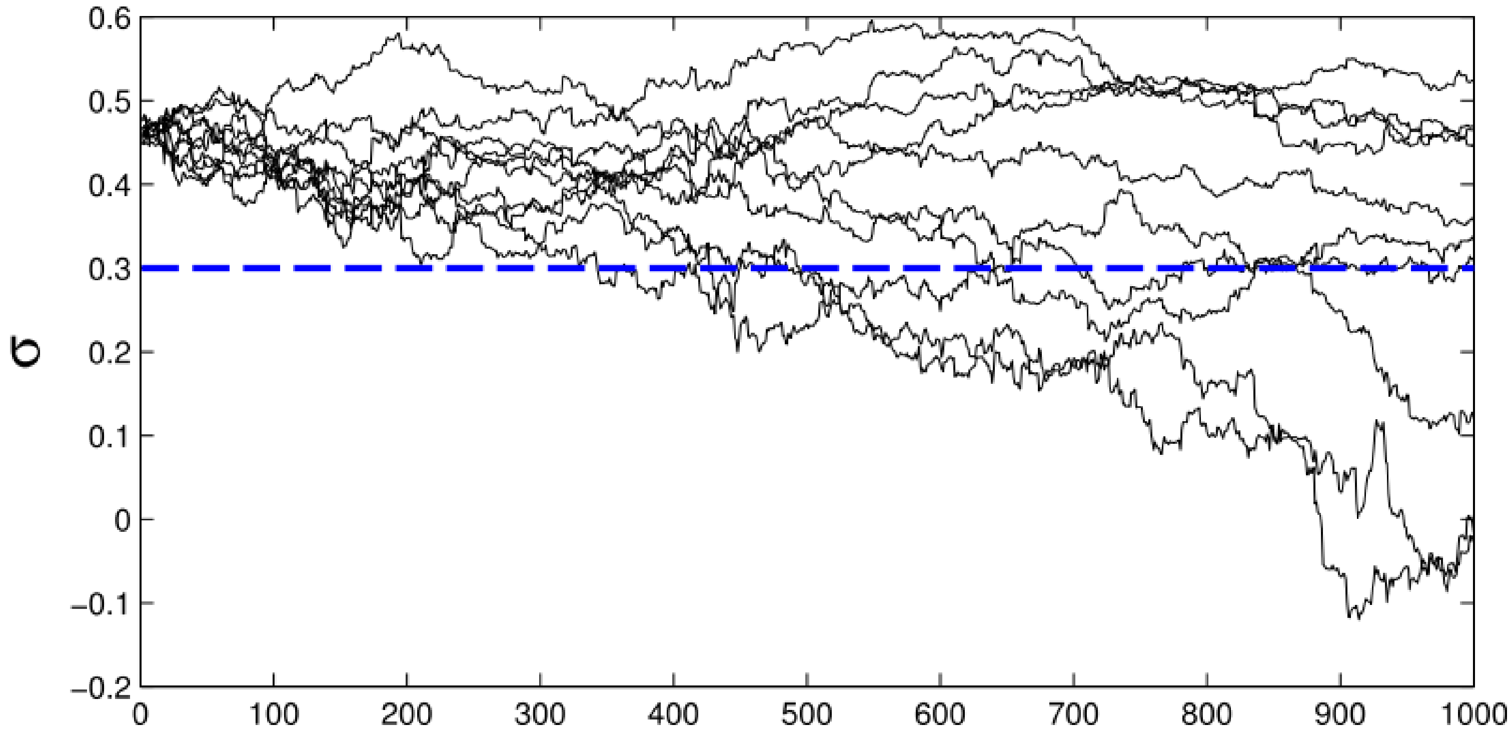
Wilks 2005
Pulido et al. 2016

Stochastic parametrization (AR(1) process with fixed ϕ)

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

Assuming a simple covariance model

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$$



Even on a twin experiments framework,

EnKF with augmented state fails

Figure taken from Santitisaadekorn 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

$$\mathbf{P}_k^f = \begin{bmatrix} \mathbf{P}_k^{xx} & \mathbf{P}_k^{x\lambda} \\ \mathbf{P}_k^{\lambda x} & \mathbf{P}_k^{\lambda\lambda} \end{bmatrix}$$

Reasons (DelSole & Yang 2009):

i) A state-augmented ensemble integration does not allow to infer cov(**x**,**p**)

ii) Weak or null correlations between stochastic parameters and observations

On a sequential data assimilation framework

covariance $\mathbf{P}_k^{\lambda x}$ converge to zero, especially if

the number of ensemble members increase

the model is sufficiently linear

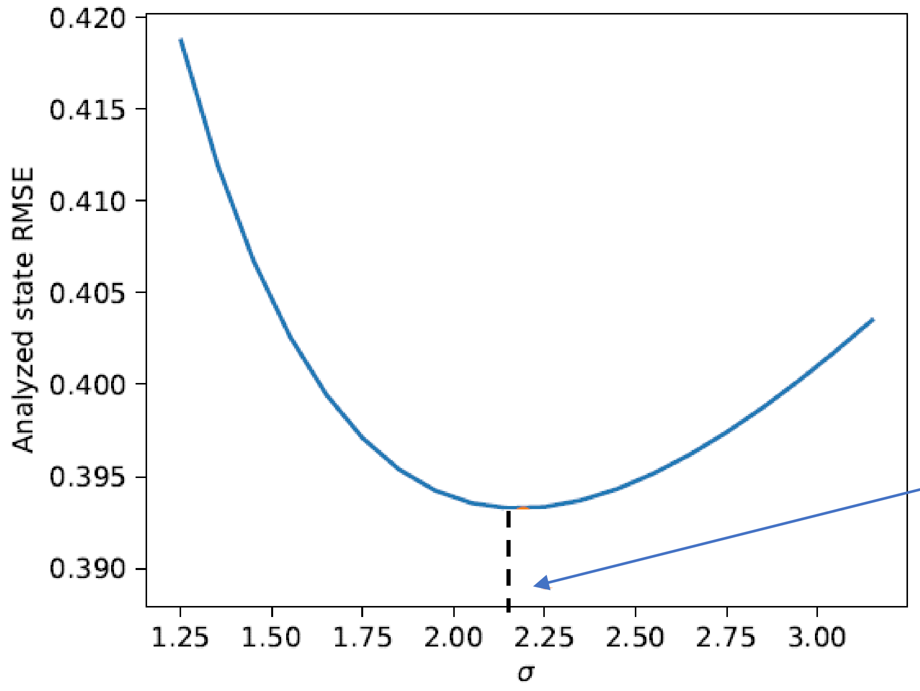
Sophisticated algorithms like Expectation-Maximization are able to estimate these type of parameters (i.e. Dreano et al. 2017

- Assuming: $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$

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

ETKF data assimilation experiment repetitions testing different values of σ

Average RMSE over different experiments with independent observation error samples

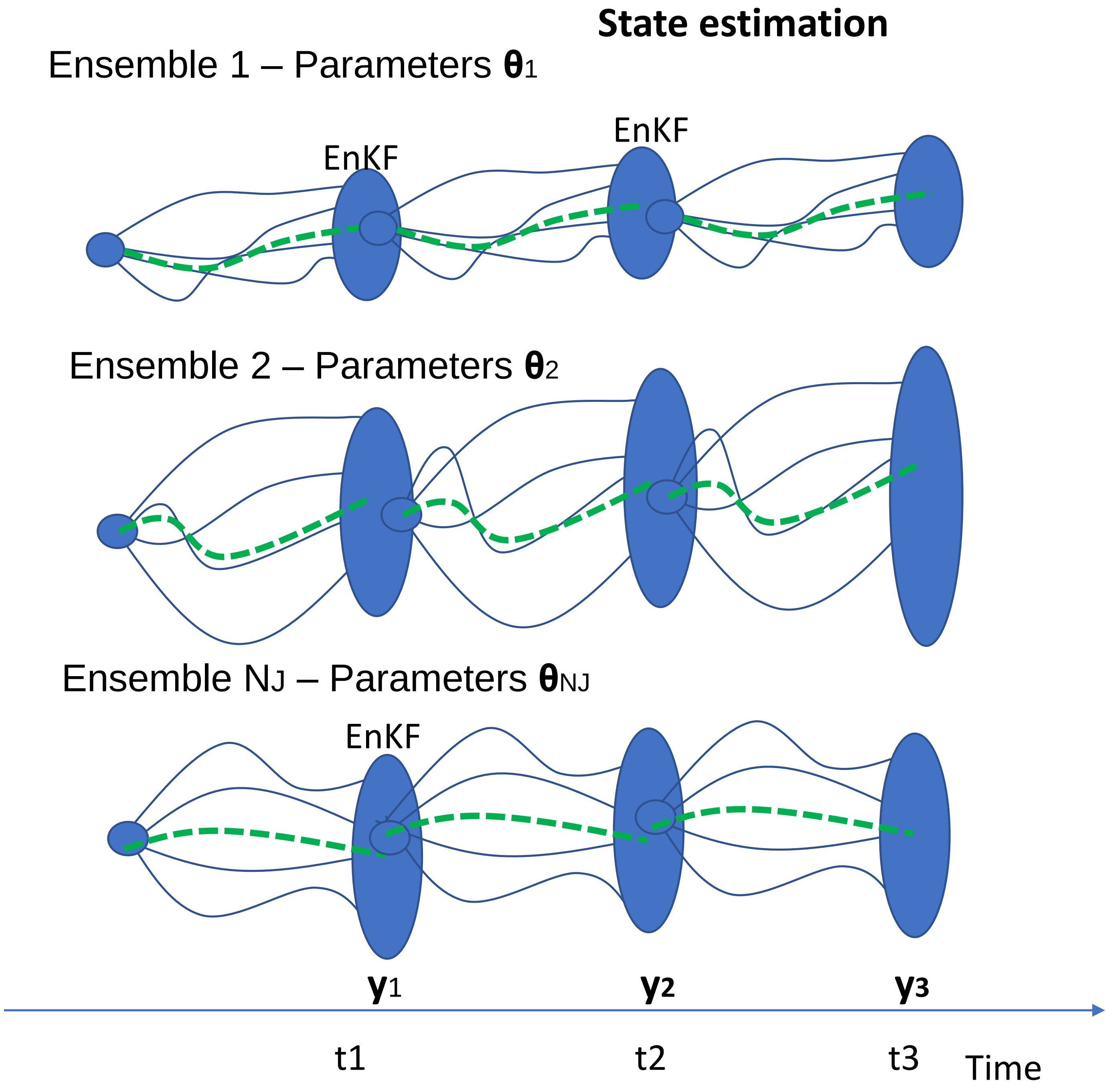


ETKF:

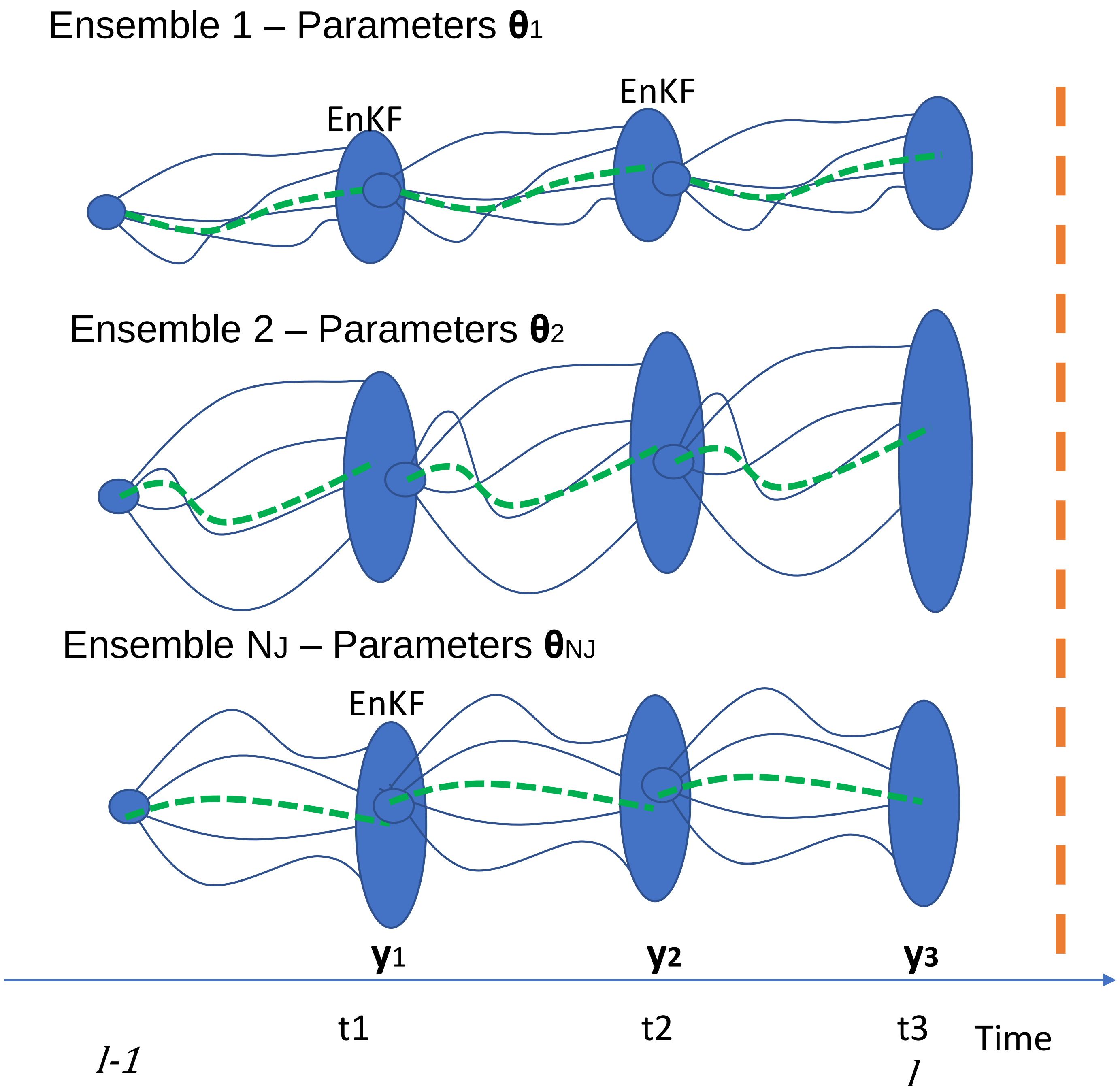
2000 assimilation cycles
30 member ensemble
R=I
H=I
NO covariance inflation

A minimum was identified.

(Analysis RMSE is quite convex!)


State estimation
Ensemble 1 – Parameters θ1
EnKF
EnKF
Ensemble 2 – Parameters θ2
Ensemble NJ – Parameters θNJ
EnKF
y1
y2
y3
t1
t2
t3
Time

State estimation



In every state assimilation cycle, each ensemble is filtered independently

$$\{\bar{\boldsymbol{x}}^a_{t_1}(\theta^1), \bar{\boldsymbol{x}}^a_{t_2}(\theta^1), \bar{\boldsymbol{x}}^a_{t_3}(\theta^1), \dots\}$$

$$\{\bar{\boldsymbol{x}}^a_{t_1}(\theta^2), \bar{\boldsymbol{x}}^a_{t_2}(\theta^2), \bar{\boldsymbol{x}}^a_{t_3}(\theta^2), \dots\}$$

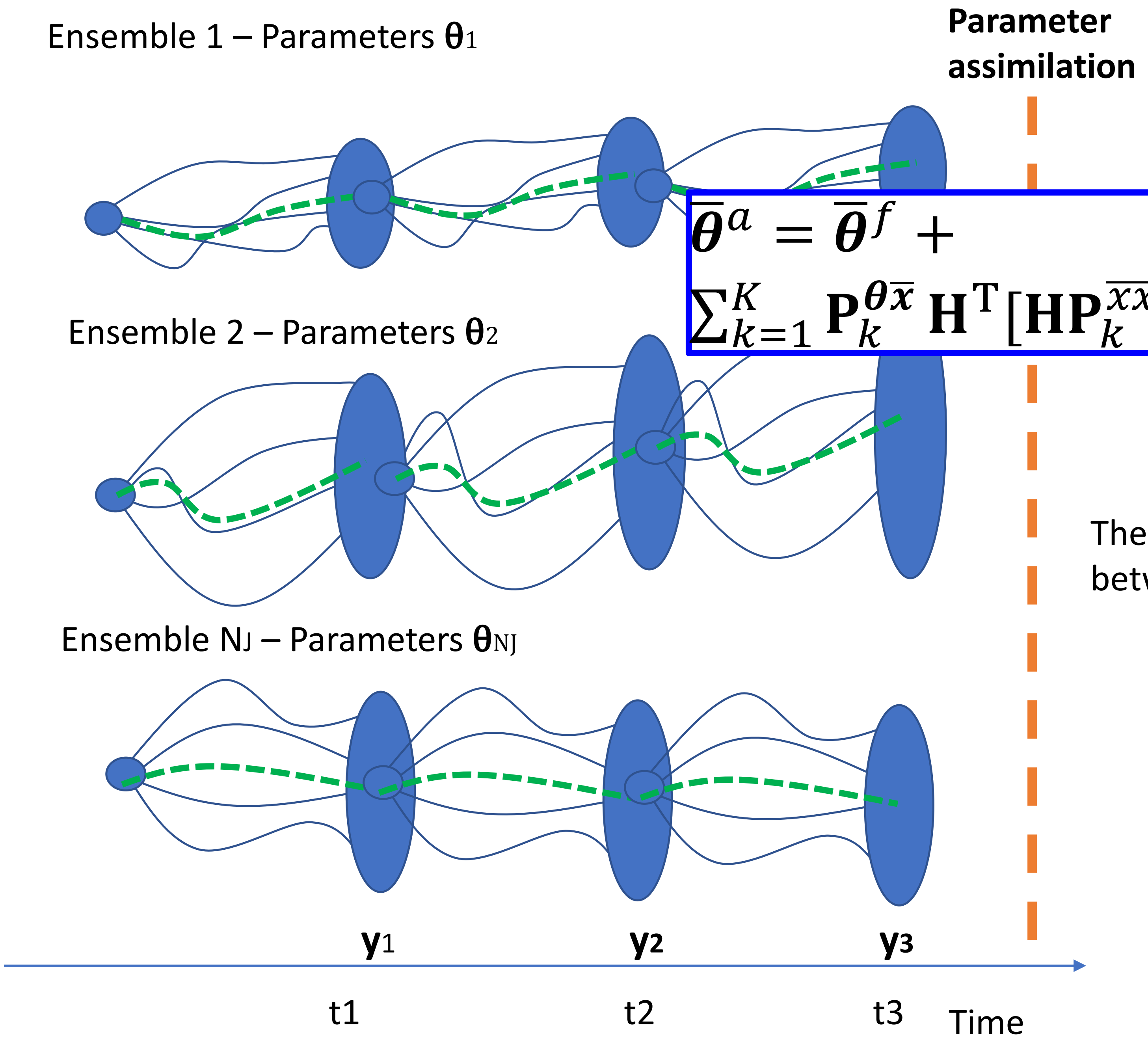
$$\{\bar{\boldsymbol{x}}^a_{t_1}(\theta^{N_J}), \bar{\boldsymbol{x}}^a_{t_2}(\theta^{N_J}), \bar{\boldsymbol{x}}^a_{t_3}(\theta^{N_J}), \dots\}$$

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

$$\{\bar{\boldsymbol{x}}^f_{t_1}(\theta^1), \bar{\boldsymbol{x}}^f_{t_2}(\theta^1), \bar{\boldsymbol{x}}^f_{t_3}(\theta^1), \dots\}$$

$$\{\bar{\boldsymbol{x}}^f_{t_1}(\theta^2), \bar{\boldsymbol{x}}^f_{t_2}(\theta^2), \bar{\boldsymbol{x}}^f_{t_3}(\theta^2), \dots\}$$

$$\{\bar{\boldsymbol{x}}^f_{t_1}(\theta^{N_J}), \bar{\boldsymbol{x}}^f_{t_2}(\theta^{N_J}), \bar{\boldsymbol{x}}^f_{t_3}(\theta^{N_J}), \dots\}$$



$$\boldsymbol{\overline{\theta}}^{a} = \boldsymbol{\overline{\theta}}^{f} + \sum_{k=1}^{K} \mathbf{P}_k^{\theta\bar{x}} \mathbf{H}^{\mathrm{T}} [\mathbf{H}\mathbf{P}_k^{\bar{x}\bar{x}} \mathbf{H}^{\mathrm{T}} + (\mathbf{H}\overline{\mathbf{P}}_k \mathbf{H}^{\mathrm{T}} + \mathbf{R})](\boldsymbol{y}_k - \mathbf{H}\overline{\overline{\boldsymbol{x}}}_k^{f})$$

The “outer” assimilation cycle exploits covariances between parameters and the mean states

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{P}_k^{\bar{x}\bar{x}} & \mathbf{P}_k^{\bar{x}\theta} \\ \mathbf{P}_k^{\theta\bar{x}} & \mathbf{P}_k^{\theta\theta} \end{bmatrix}$$

Nested EnKFs algorithm

1. Given N_J parameters $\boldsymbol{\theta}^{1:NJ}$ and N_J independent ensembles of N_I each: $\boldsymbol{x}^{1:N_I,1:N_J}$, and (LxK) observations $\boldsymbol{y}_{1:L,1:K}$
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 $\boldsymbol{x}_{l,k}^{a(j,i)}$ through N_J-EnKFs independently

      2.1.2 Store the mean predicted observation $\boldsymbol{H}\overline{\boldsymbol{x}}_{l,k}^{(j)}$ for each ensemble and the average of the forecast error covariance matrix in the observational space $\boldsymbol{H}\widetilde{\boldsymbol{P}}_{l,k}\boldsymbol{H}^{\boldsymbol{T}} = \frac{\boldsymbol{1}}{\boldsymbol{N_J}}\sum_{\boldsymbol{j}} \boldsymbol{H}\widetilde{\boldsymbol{P}}_{l,k}^{j}\,\boldsymbol{H}^{\boldsymbol{T}}$

   2.2 Parameter estimation:

      2.2.1 Concatenate the K mean predicted observations to construct an (n_x x K)-dimensional ensemble

      2.2.2 Construct an aggregated observations vector $\boldsymbol{y}_l^* = \left[\boldsymbol{y}_{l,1}, \dots, \boldsymbol{y}_{l,K}\right]^T$.

      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 $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$$

Synthetic truth: σ^2=2

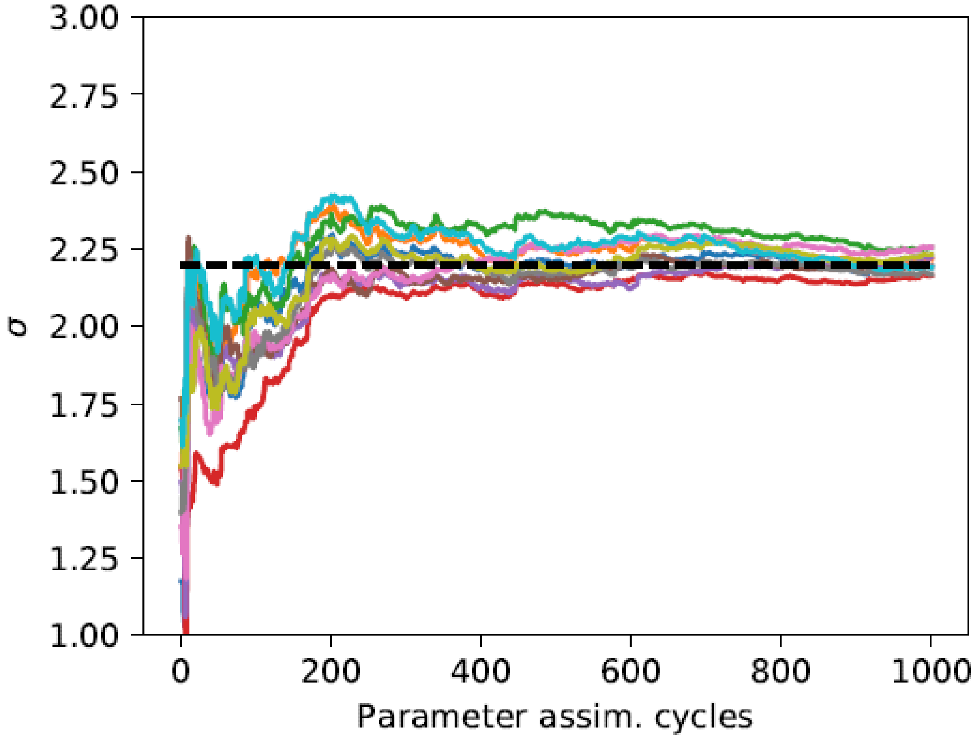
H=I and **R=I**

N_J=15 (number of filters)

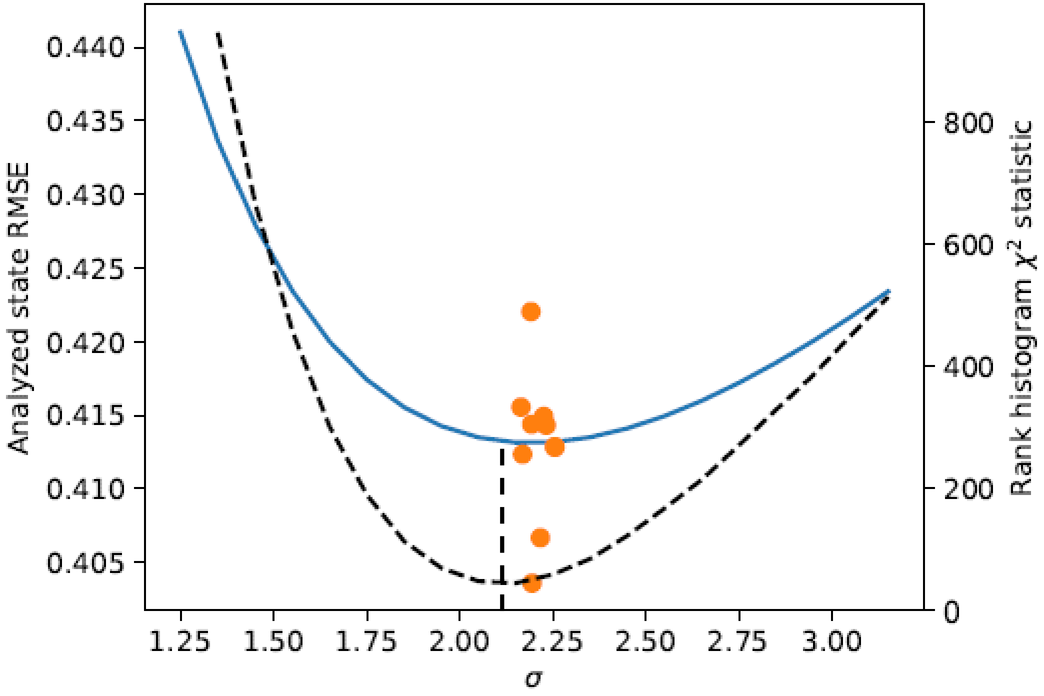
N_I=30 (ens. members)

Δt=0.005

Δt_{obs}=10Δt



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)

Assuming AR(1) process

$$\boldsymbol{e}(t) = \phi \boldsymbol{e}(t - \Delta t) + (1 - \phi^2)^{\frac{1}{2}} \boldsymbol{z}(t); \qquad \boldsymbol{z} \sim N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

Possible stochastic forcing covariance hypothesis

- Isotropic $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$$

- Non-isotropic diagonal $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$$

- Exponential covariance $\boldsymbol{\Sigma}$:

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

- Symmetric circulant $\boldsymbol{\Sigma}$:

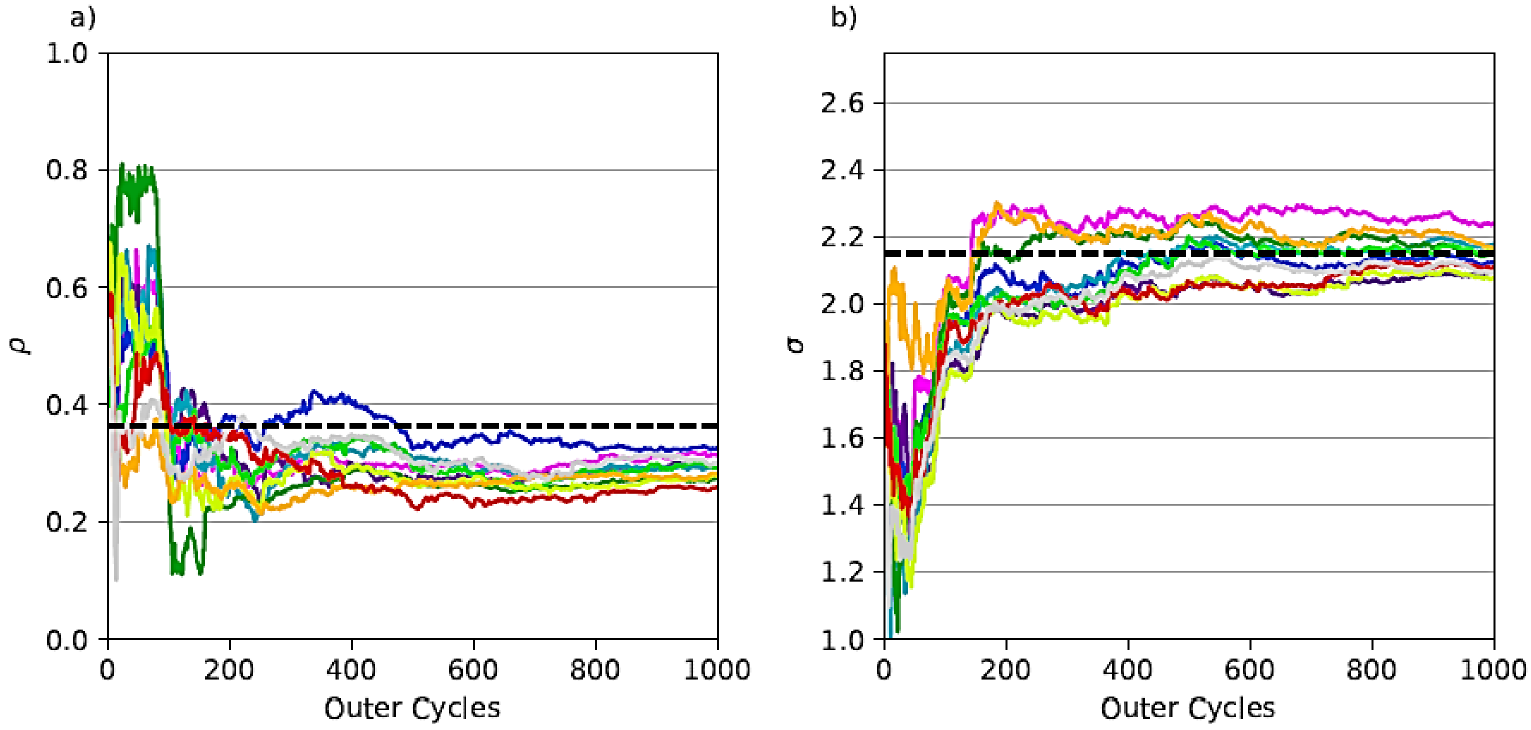
$$\sigma_{i,i-n} = \sigma_{i,i+n}$$

Exponential covariance model $\boldsymbol{\Sigma}$:

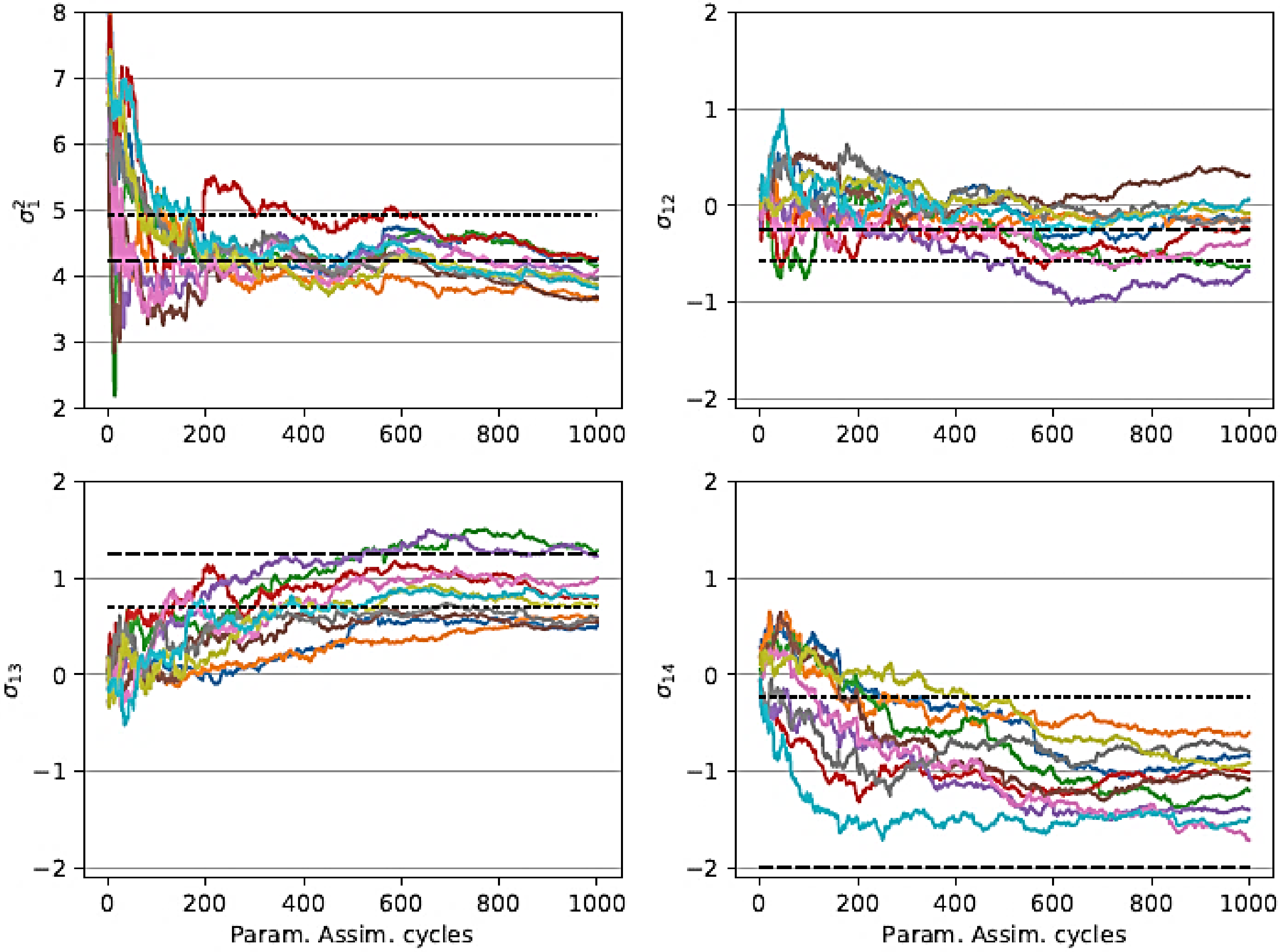
$$\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}_{ij} = \sigma^2 \exp(-\rho|i-j|)$$

Synthetic truth: σ^2=2

ρ =0.3



• Symmetric circulan **Σ**:



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

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

Real model

-------- Estimation through exhaustive exploration

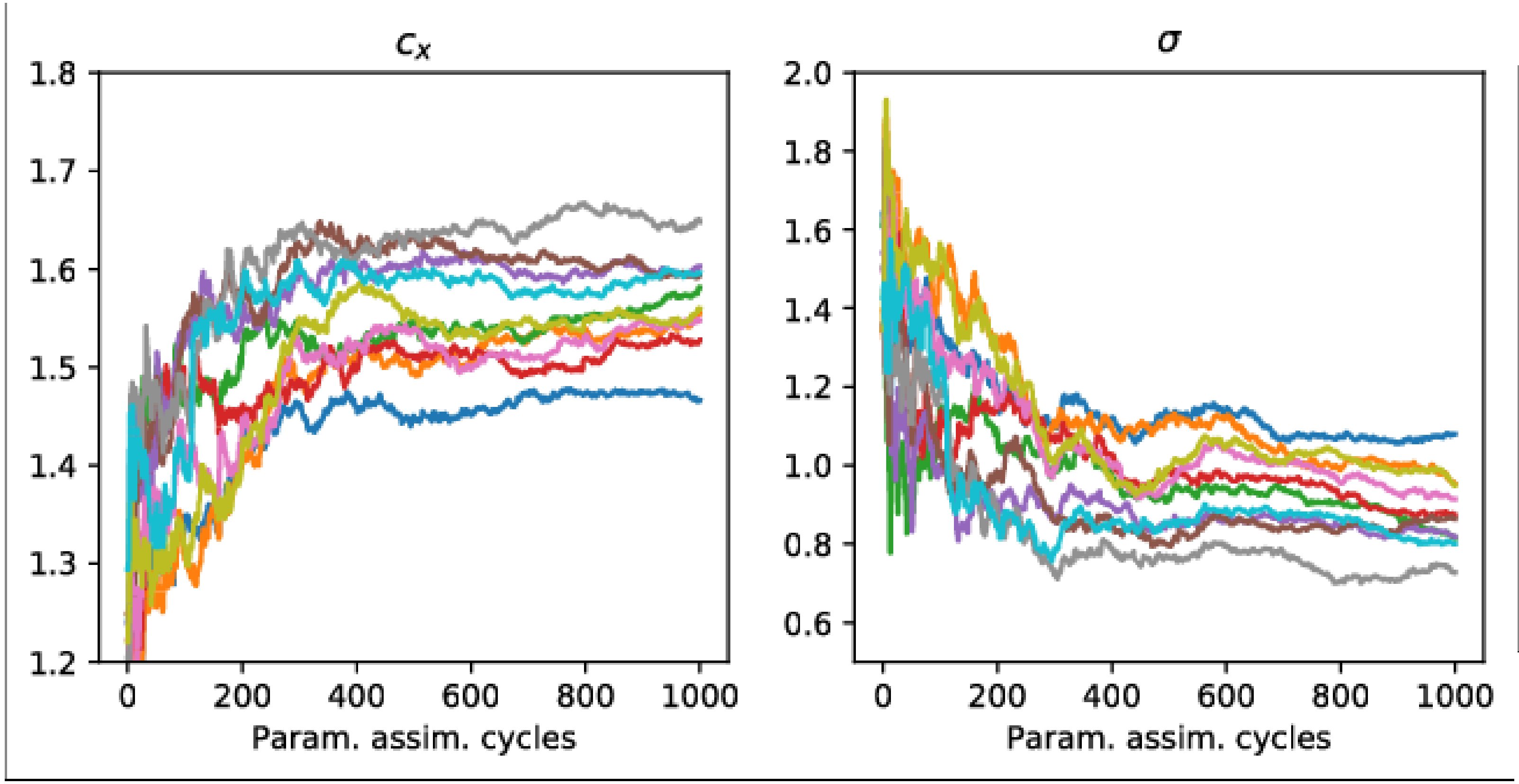
.... Estimation using residuals as in Pulido et al. 2016

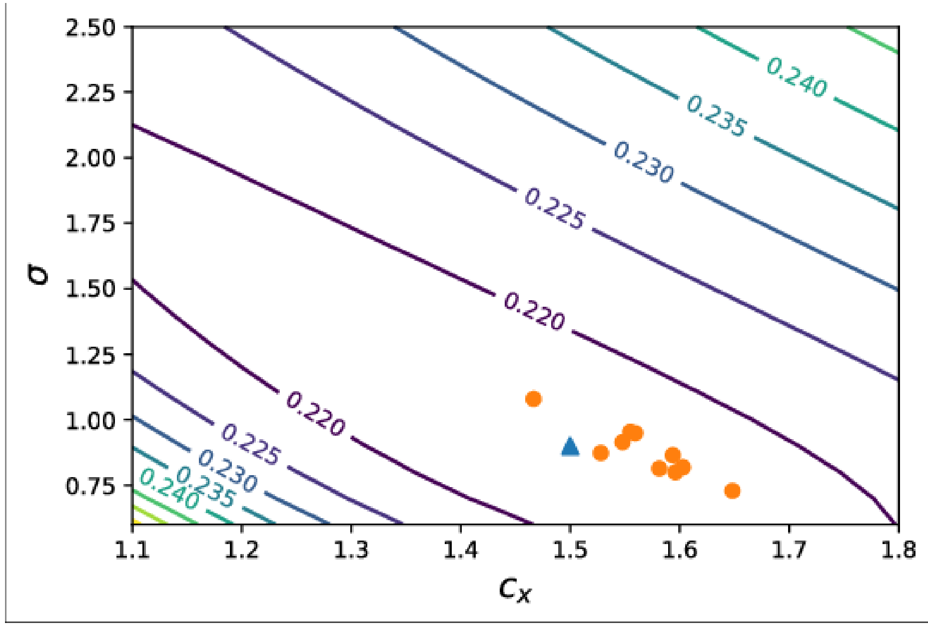
- Simultaneous estimation: Multiplicative inflation + σ

(Isotropic non-correlated case)

H=**I** and **R**=0.15^2**I**
N_J=15 (number of filte
N_I=30 (ens. members)

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$$





Estimations for different observation error samples and initial conditions.

EnKF analysis RMSE as a function of the inflation factor and σ 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 → 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

$$p(\boldsymbol{x}_{l+1},\boldsymbol{\theta}_{l+1}|\boldsymbol{y}_{l,1:K}) = p(\boldsymbol{x}_{l+1}|\boldsymbol{\theta}_{l+1},\boldsymbol{y}_{l,1:K})\,p(\boldsymbol{\theta}_{l+1}|\boldsymbol{y}_{l,1:K})$$

Sequentially along the DA window

$$p(\boldsymbol{\theta}_{l+1}|\boldsymbol{y}_{l,1:K}) \propto p(\boldsymbol{\theta}_{l+1}|\boldsymbol{y}_{l-1,1:K})\prod_{k=1}^{K} p(\boldsymbol{y}_{l,k}|\boldsymbol{y}_{l,1:k-1}|\boldsymbol{y}_{l,1:K})$$

$$p(\mathbf{y}_{l,k}|\mathbf{y}_{l,k-1},\theta_{l+1}) = \int p(\mathbf{y}_{l,k}|\mathbf{x}_{l,k},\theta_{l+1})\,p(\mathbf{x}_{l,k}|\mathbf{y}_{l,1:k-1},\theta_{l+1})\,\mathrm{d}\mathbf{x}_{l,k}$$

Accurate marginalization over the full state $\boldsymbol{x}$ is quite expensive

$$p(\mathbf{y}_{l,k}|\mathbf{y}_{l,k-1},\theta_{l+1}) = \int p(\mathbf{y}_{l,k}|\mathbf{x}_{l,k},\theta_{l+1})p(\mathbf{x}_{l,k}|\mathbf{y}_{l,1:k-1},\theta_{l+1})\mathrm{d}\mathbf{x}_{l,k} \quad (1)$$

Instead we asume (Pulido et al. 2018):

$$p(\mathbf{y}_{l,k}|\mathbf{x}_{l,k},\theta_{l+1}) \propto \exp\left[(\mathbf{y}_{l,k} - \mathcal{H}(\mathbf{x}_{l,k}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}_{l,k} - \mathcal{H}_{l,k}(\mathbf{x}_{l,k}))\right]$$

$$p(\mathbf{x}_{l,k}|\mathbf{y}_{l,1:k-1},\theta_{l+1}) \propto \exp\left[\left(\mathbf{x}_{l,k} - \overline{\mathbf{x}}_{l,k}^{f}(\theta_{l+1})\right)^{\mathrm{T}}\mathbf{P}_{l,k}(\theta_{l+1})^{-1}(\mathbf{x}_{l,k} - \overline{\mathbf{x}}_{l,k}^{f}(\theta_{l+1}))\right]$$

Replacing in (1):

$$\prod_{k=1}^{K} p(\mathbf{y}_{l,k}|\mathbf{y}_{l,k-1},\theta_{l+1}) \propto$$

$$\prod_{k=1}^{K} \exp\left[(\mathbf{y}_{l,k} - \mathcal{H}_{l,k}(\overline{\mathbf{x}}_{l,k}^{f}))^{\mathrm{T}}(\mathbf{H}_{l,k}\mathbf{P}_{l,k}\mathbf{H}_{l,k}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{y}_{l,k} - \mathcal{H}_{l,k}(\overline{\mathbf{x}}_{l,k}^{f}))\right]$$

Likelihood solved through the nested ensemble Kalman filters