(Parametric) Model Order Reduction in Data Assimilation

SFB 1294 DATA ASSIMILATION COLLOQUIUM
DEPARTMENT OF MATHEMATICS . UNIVERSITÄT POTSDAM
14 JUL 2023 . POTSDAM

H. Bansal, N. Cvetkovic, M. Grepl, H.C. Lie, C. Pagliantini, F. Silva, K. Veroy

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (Grant Agreement No. 818473).

(Parametric) Model Order Reduction in Data Assimilation

SFB 1294 DATA ASSIMILATION COLLOQUIUM
DEPARTMENT OF MATHEMATICS . UNIVERSITÄT POTSDAM
14 JUL 2023 . POTSDAM

H. Bansal, N. Cvetkovic, M. Grepl, H.C. Lie, C. Pagliantini, F. Silva, K. Veroy

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 818473).

MOTIVATION

(1) To develop parametric model order reduction methods that accelerate simulations of parametrized microstructures in the multiscale materials setting …

(2) … such that they can be deployed in the context of control, optimization, inverse problems, and data assimilation …

with a view towards eventually being able to design and control production processes.

MOTIVATION

(1) To develop **parametric model order reduction** methods that accelerate simulations of parametrized microstructures in the multiscale materials setting …

(2) … such that they can be deployed in the context of control, optimization, **inverse problems and data assimilation** …

with a view towards eventually being able to design and control production processes.

OUTLINE

In the parametrized PDE setting:

(1) How can we account for the error introduced by the use of a parametric reduced order model in a data assimilation framework (EnKI)?

(2) How can we construct efficient (both offline *and* online) parametric reduced order models for data assimilation (EnKF)?

(3) How can we mitigate the effect of using an approximate model on parameter inference (via experimental design)?

Parametric Model Order Reduction

Linear PDE: Given the parameter $\mu \in \mathcal{D} \subset \mathbb{R}^P$, we seek $\boldsymbol{u}(\mu)$ s.t.

$$\boldsymbol{A}(\mu)\boldsymbol{u}(\mu) = \boldsymbol{f}(\mu)$$

$\boldsymbol{x}$-dependence omitted

Approximation:

$$\boldsymbol{u}(\mu) \approx \sum_{i=1}^{N} \alpha_i(\mu)\,\boldsymbol{\zeta}_i = \mathbf{Z}_N \boldsymbol{\alpha}_N(\mu) \qquad \text{where} \qquad \mathbf{Z}_N = [\,\boldsymbol{\zeta}_1 \cdots \boldsymbol{\zeta}_N]$$

and $\mathbf{Z}_N$ is obtained from training snapshots $\{\boldsymbol{u}(\mu_i), i = 1, \ldots, N_{\text{train}}\}$ either via a greedy algorithm ($N = N_{\text{train}}$) or singular value decomposition ($N \ll N_{\text{train}}$)

Parametric Model Order Reduction

Projection-Based Model Order Reduction:

$$
\begin{aligned}
\mathbf{Z}_N^T \boldsymbol{A}(\mu)\mathbf{Z}_N \boldsymbol{\alpha}_N(\mu) &= \mathbf{Z}_N^T \boldsymbol{f}(\mu) \\
\boldsymbol{A}_N(\mu)\boldsymbol{\alpha}_N(\mu) &= \boldsymbol{f}_N(\mu)
\end{aligned}
$$

Parametric Model Order Reduction

Projection-Based Model Order Reduction:

$$
\begin{aligned}
\mathbf{Z}_N^T \boldsymbol{A}(\mu) \mathbf{Z}_N \boldsymbol{\alpha}_N(\mu) &= \mathbf{Z}_N^T \boldsymbol{f}(\mu) \\
\boldsymbol{A}_N(\mu) \boldsymbol{\alpha}_N(\mu) &= \boldsymbol{f}_N(\mu)
\end{aligned}
$$

Assume that $\boldsymbol{A}(\mu)$ permits an affine decomposition, $\boldsymbol{A}(\mu) = \sum_{q=1}^{Q} \theta^q(\mu) \boldsymbol{A}^q$, then

$$
\boldsymbol{A}_N(\mu) = \sum_{q=1}^{Q} \theta^q(\mu) \mathbf{Z}_N^T \boldsymbol{A}^q \mathbf{Z}_N^T
$$

and the computation can be decomposed into an (expensive, cost($\mathcal{N}_{\mathrm{FE}}$)) **offline stage** and an (inexpensive, cost(N)) **online stage**.

Parametric Model Order Reduction

General Approach:

(1) Reduce the order or dimension of the problem from $\mathcal{N}_{\mathrm{FE}}$ to $N \ll \mathcal{N}_{\mathrm{FE}}$

(2) Decompose the computation into an

OFFLINE training phase at cost($\mathcal{N}_{\mathrm{FE}}$)

ONLINE deployment phase at cost(N) for any new $\mu \in \mathcal{D}$.

Parametric Model Order Reduction

General Approach:

(1) Reduce the order or dimension of the problem from $\mathcal{N}_{\mathrm{FE}}$ to $N \ll \mathcal{N}_{\mathrm{FE}}$

(2) Decompose the computation into an

OFFLINE training phase at cost($\mathcal{N}_{\mathrm{FE}}$)

ONLINE deployment phase at cost(N) for any new $\mu \in \mathcal{D}$.

Issues: Nonlinearities

Non-affine problems

Parametric Model Order Reduction

Nonlinear PDE: Given the parameter $\mu \in \mathcal{D} \subset \mathbb{R}^P$, we seek $\boldsymbol{u}(\mu)$ s.t.

$$F(\boldsymbol{u}(\mu); \mu) = 0$$

At the k-th Newton iteration, given $\boldsymbol{u}^{k-1}$, find $\delta\boldsymbol{u}$ s.t.

$$J(\boldsymbol{u}^{k-1}; \mu)\, \delta\boldsymbol{u} = -F(\boldsymbol{u}^{k-1}; \mu)$$

and update $\boldsymbol{u}^k = \boldsymbol{u}^{k-1} + \delta\boldsymbol{u}$ until convergence.

Issue: In nonlinear (and nonaffine) problems, $J(\boldsymbol{u}^{k-1}; \mu)$ (or $\boldsymbol{A}(\mu)$) does not admit an offline/online computational decomposition.

$\Rightarrow$ approximate affine decomposition

or hyperreduction

Non-intrusive Model Order Reduction

Linear Methods: $\boldsymbol{u}(\mu) \approx \boldsymbol{Z}_N \, \boldsymbol{\alpha}_N(\mu)$

- **RB/POD + Interpolation** — [Bui-Thanh, Damodaran & Willcox, '03], [Demo, Tezzele & Rozza, '19], …
- **RB/POD + Regression** — [M. Guo & Hesthaven, '18], [M. Guo & Hesthaven, '19], …
- **RB/POD + Neural Networks** — [Wang, Hesthaven & Ray, '19], [Barnett, Farhat & Maday, '22], [Pichi, Ballarin, Rozza & Hesthaven, '23], …
- **Physics-Reinforced NN** — [Chen, Wang, Hesthaven & Zhang, '19]

Nonlinear Methods: $\boldsymbol{u}(\mu) \approx \boldsymbol{\Psi}\left(\boldsymbol{\alpha}_N(\mu)\right)$

- **Autoencoders** — [Lee & Carlberg, '20], [Fresca, Dede & Manzoni, '21], [Maulik, Lusch & Balapakrash '21], [Vinuesa, Eivazi, Le Clainche & Hoyas, '22], [Nikolopoulos, Kalogeris, Papadopoulos, '22], [Romor, Stabile & Rozza '23], …
- **Neural Operators** — [Lu, Jin, Pang, Zhang, Karniadakis, '19], [Cai, Wang, Lu, Zaki, Karniadakis, '21], [De Hoop, Huang, Qian & Stuart, 21], [Li, Zheng, Kovachki, et al. '22], …

 Adapted from a slide by Federico Pichi (EPFL; presented at International Workshop on MOR at NUS, Singapore, May 2023)

MOTIVATION

In the parametrized PDE setting:

(1) How can we account for the error introduced by the use of a parametric reduced order model in a data assimilation framework (EnKI)?

(2) How can we construct efficient (both offline *and* online) parametric reduced order models for data assimilation (EnKF)?

(3) How can we mitigate the effect of using an approximate model on parameter inference (via experimental design)?

joint work with **F. Silva**, C. Pagliantini, M. Grepl

PARAMETER ESTIMATION : UNREGULARIZED VARIATIONAL APPROACH

$$\min_{\boldsymbol{\mu}\in\mathcal{P}} \mathcal{J}(\boldsymbol{\mu}\,|\boldsymbol{y}) := \frac{1}{2}\underbrace{\boxed{\|\boldsymbol{y}-\mathbf{L}\,u\|^2_{\Sigma^{-1}}}}_{\text{DATA MISFIT}} \quad \text{such that} \quad \underbrace{\boxed{\left(\mathcal{M}_{\boldsymbol{\mu}}u,\psi\right)=0 \quad \forall\psi\in\mathcal{Y}}}_{\text{WEAK MODEL}}$$

where:

$\boldsymbol{y} = \mathbf{L}\,u_{\text{TRUE}} + \boldsymbol{\epsilon}$ *with noise* $\boldsymbol{\epsilon}\sim\mathcal{N}(0,\Sigma)$

PARAMETER ESTIMATION : UNREGULARIZED VARIATIONAL APPROACH

$$\min_{\boldsymbol{\mu}\in\mathcal{P}} \mathcal{J}(\boldsymbol{\mu}\,|\boldsymbol{y}) := \frac{1}{2}\boxed{\|\boldsymbol{y}-\mathbf{L}\,u_N\|^2_{\Sigma^{-1}}} \quad \text{such that} \quad \boxed{(\mathcal{M}_{\boldsymbol{\mu}}u_N,\psi)=0 \quad \forall\psi\in\mathcal{Y}_N}$$

DATA MISFIT **WEAK MODEL**

where:

$$\boldsymbol{y} = \mathbf{L}\,u(\boldsymbol{\mu}_{\mathrm{TRUE}}) + \boldsymbol{\epsilon} \quad \textit{with noise} \quad \boldsymbol{\epsilon}\sim\mathcal{N}(0,\Sigma)$$

PARAMETER ESTIMATION : UNREGULARIZED VARIATIONAL APPROACH

$$\min_{\boldsymbol{\mu}\in\mathcal{P}} \mathcal{J}(\boldsymbol{\mu}\,|\boldsymbol{y}) := \frac{1}{2}\boxed{\left\|(\boldsymbol{y}-\overline{\boldsymbol{\delta}}_N)-\mathbf{L}\,u_N\right\|^2_{(\Sigma+\Gamma_N)^{-1}}} \quad \text{such that} \quad \boxed{\left(\mathcal{M}_{\boldsymbol{\mu}}u_N,\psi\right)=0 \quad \forall\psi\in\mathcal{Y}_N}$$

DATA MISFIT **WEAK MODEL**

where:

$$\boldsymbol{y} = \mathbf{L}\,u(\boldsymbol{\mu}_{\mathrm{TRUE}})+\boldsymbol{\epsilon}$$ *with noise* $\boldsymbol{\epsilon}\sim\mathcal{N}(0,\Sigma)$

$$\boldsymbol{\delta}_N(\boldsymbol{\mu}_{\mathrm{TRUE}}) = \mathbf{L}\,(u_N(\boldsymbol{\mu}_{\mathrm{TRUE}})-u(\boldsymbol{\mu}_{\mathrm{TRUE}}))$$ *approximated by* $\boldsymbol{\delta}_N\sim\mathcal{N}(\overline{\boldsymbol{\delta}}_N,\Gamma_N)$

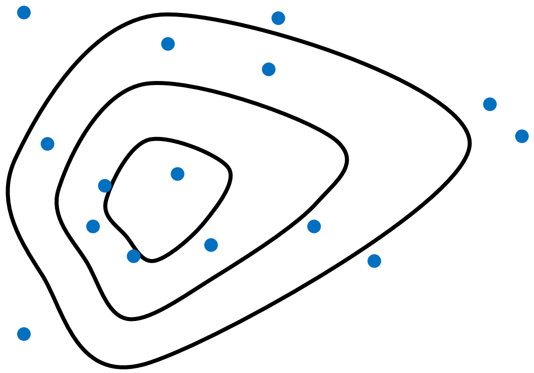
REDUCED BASIS ENSEMBLE KALMAN INVERSION

We sample a particle ensemble of size J from a prior distribution π_0 and update their positions as follows:

For $n = 0,1,\ldots$

i) Compute the model solution for each particle $\boldsymbol{\mu}_n^{(j)}$:

$$u_{N,n}^{(j)} \in \mathcal{X}_N \quad \text{such that} \quad \left(\mathcal{M}_{\boldsymbol{\mu}_n^{(j)}} u_{N,n}^{(j)}, \psi_i\right) = 0 \quad \forall \psi_i \in \mathcal{Y}_N$$



Iglesias, Law, and Stuart
"Ensemble Kalman methods for inverse problems"
(2013)

REDUCED BASIS ENSEMBLE KALMAN METHOD

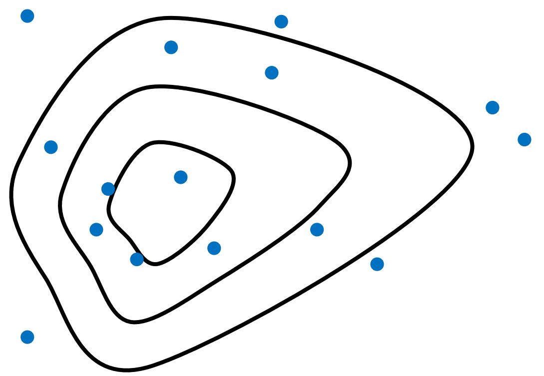
We sample a particle ensemble of size J from a prior distribution π_0 and update their positions as follows:

For $n = 0,1, \ldots$

ii) Compute the covariance matrices :

$$P_{N,n} := \mathrm{sum}\left(\mathbf{L}\, u_{N,n}^{(j)} \otimes \mathbf{L}\, u_{N,n}^{(j)} - \mathbf{L}\bar{u}_{N,n} \otimes \mathbf{L}\bar{u}_{N,n}\right) \cdot (J-1)^{-1}$$

$$Q_{N,n} := \mathrm{sum}\left(\boldsymbol{\mu}_n^{(j)} \otimes \mathbf{L}\, u_{N,n}^{(j)} - \bar{\boldsymbol{\mu}}_n \otimes \mathbf{L}\bar{u}_{N,n}\right) \cdot (J-1)^{-1}$$



THE REDUCED BASIS ENSEMBLE KALMAN METHOD

We sample a particle ensemble of size J from a prior distribution π_0 and update their positions as follows:

For $n = 0,1, \ldots$

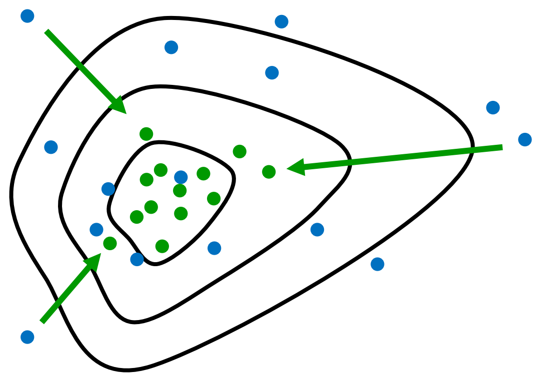
iii) Update each particle $\boldsymbol{\mu}_n^{(j)}$ in the ensemble:

$$\boldsymbol{\mu}_{n+1}^{(j)} = \boldsymbol{\mu}_n^{(j)} + Q_{N,n}\left(\Sigma + \Gamma_\varepsilon + P_{N,n}\right)^{-1}\left(\boldsymbol{y} - \overline{\boldsymbol{\delta}}_\varepsilon - \mathbf{L}\, u_{N,n}^{(j)}\right)$$

where

$$\overline{\boldsymbol{\delta}}_N := \frac{1}{J} \cdot \mathrm{sum}\left(\mathbf{L}\left(u_{N,n}^{(j)} - u_n^{(j)}\right)\right)$$

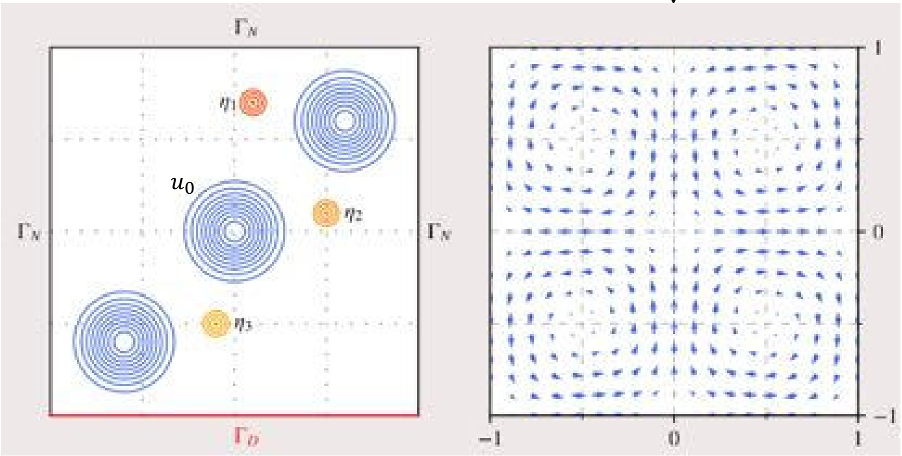
$$\Gamma_N := \frac{1}{J-1} \cdot \mathrm{sum}\left(\mathbf{L}\left(u_{N,n}^{(j)} - u_n^{(j)}\right) \otimes \mathbf{L}\left(u_{N,n}^{(j)} - u_n^{(j)}\right) - \overline{\boldsymbol{\delta}}_N \otimes \overline{\boldsymbol{\delta}}_N\right)$$



ADVECTION-DISPERSION PROBLEM

$$\frac{\partial u}{\partial t} - \boldsymbol{\mu} \cdot \Delta u(t) + \boldsymbol{v} \cdot \nabla u(t) = 0 \quad \text{on } \Omega \coloneqq (-1,+1)^2 \qquad \text{with} \qquad \boldsymbol{v} = \begin{bmatrix} +\sin(\pi x_1)\cos(\pi x_2) \\ -\cos(\pi x_1)\sin(\pi x_2) \end{bmatrix}$$

$$u(0) = u_0$$



we consider:

- 3 sensor locations
- 40 time-activations per sensor
- $t \in (0, 2.4)$
- $\boldsymbol{\mu} \in [1/50, 1/10]$

ADVECTION-DISPERSION PROBLEM : MODEL ORDER REDUCTION

Employing the weak-greedy-POD approach we construct a Reduced Basis space of size 42 :

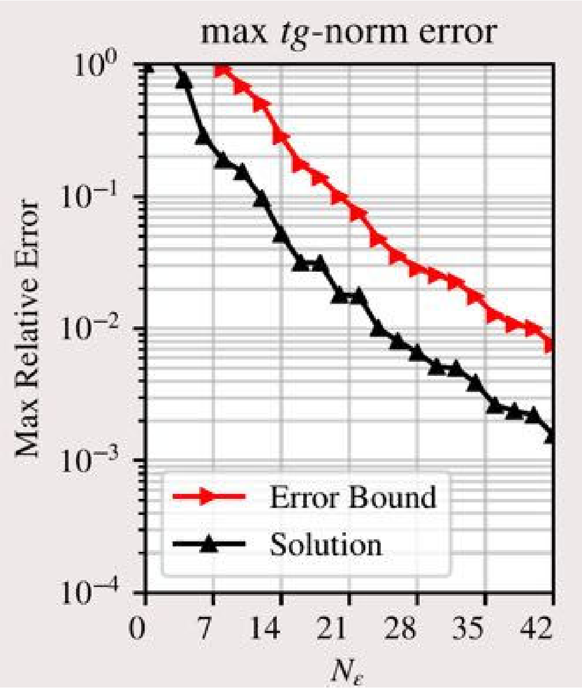
dofs spatial discretization = 10100 (P2-P2 G)

dofs time discretization = 241 (P1-P0 PG)

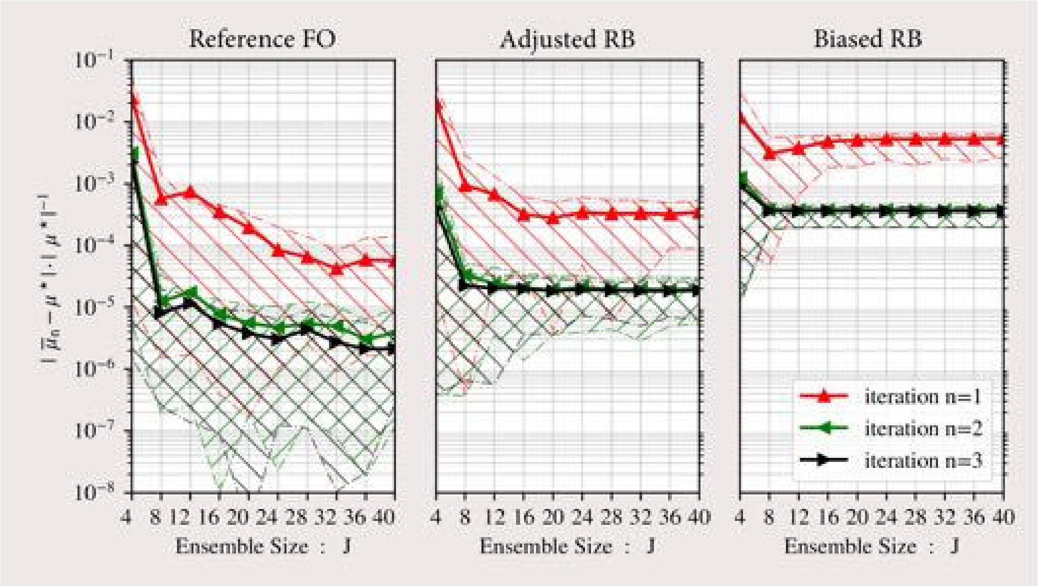
training set size = 81 parameter values

training time = 47s to construct basis (18 evals)

The effectivity of the bound is independent from the space dimension and doesn’t exceed a factor 10



PARAMETER ESTIMATION : ENSEMBLE SIZE

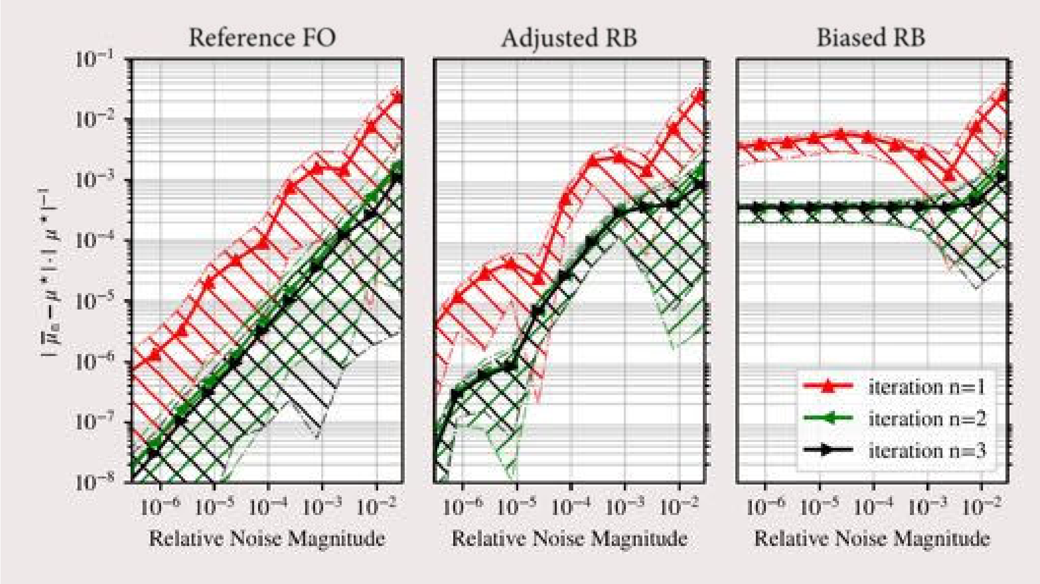


results show a faster convergence to the 'large ensemble behavior' for the reduced basis methods

the adjusted method exhibits a better behavior than the biased method

a rapid convergence of the algorithm to a stable parameter estimation is observed for all cases

PARAMETER ESTIMATION : NOISE MAGNITUDE

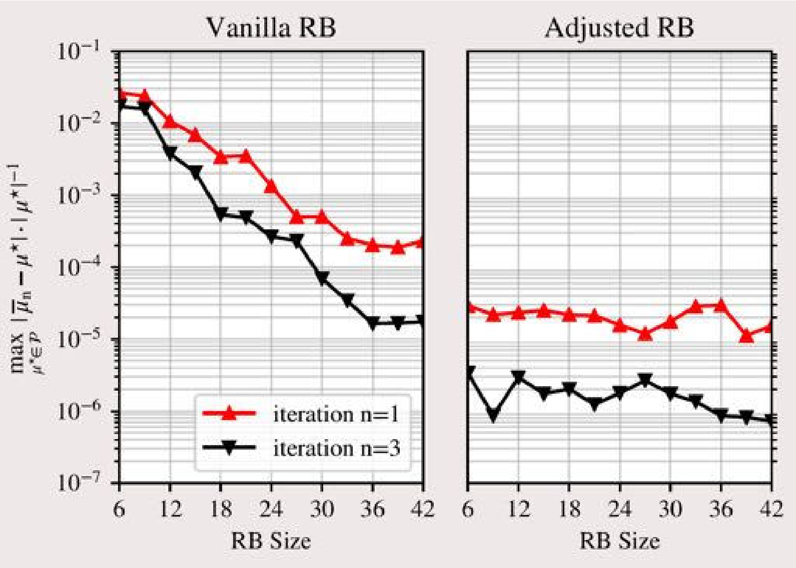


results show a linear convergence when the exact FO model is employed

the error stagnates when the model bias is not corrected in the RB-EnKM

the adjusted RB-EnKM shows an error decay comparable with the FO one

PARAMETER ESTIMATION : REDUCED BASIS SIZE



when the measurements bias is not corrected, the relative error is strictly dependent on the RB model accuracy

with the bias correction, the performances of the method are made independent on the RB.

with 42 bases, one parameter estimation takes 8’’ (55’’ considering the offline cost); one standard EnKM estimation takes 5’ 58’’

TRACER TRANSPORT PROBLEM

The EnKM is used to estimate the log-conductivity $\boldsymbol{\vartheta}$,
given the observations of the tracer concentration $\hat{\mathbf{L}}(c)$:

$$\frac{\partial c}{\partial t} - \nabla \cdot \big((d_l \boldsymbol{v} \otimes \boldsymbol{v} + d_m \boldsymbol{I}) \cdot \nabla c \big) - \boldsymbol{v} \cdot \nabla c = f_t$$

$$-\nabla \cdot \big(e^{\boldsymbol{\vartheta}} h \nabla h \big) = f_h$$

$$-e^{\boldsymbol{\vartheta}} \nabla h = \boldsymbol{v}_h$$

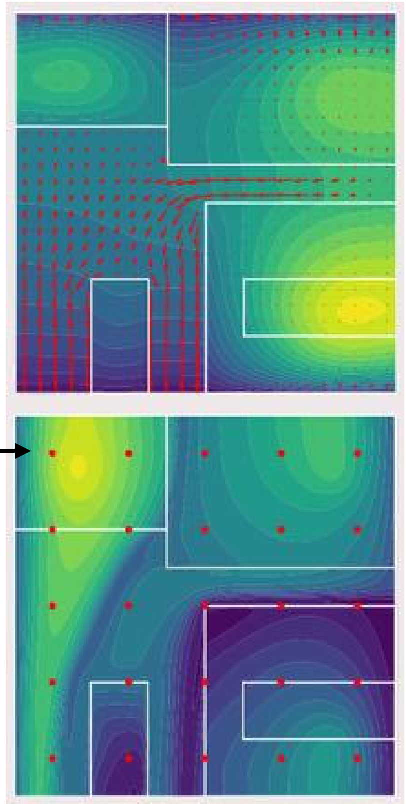
Conrad, Davis, Marzouk,
Pillai, Smith. JUQ (2018)

A POD approximation has been used to efficiently solve
the system of equations. The prior distribution π_0 is chosen to be

$$\pi_0 \coloneqq \times_{i=1}^{6} U(\vartheta_i^{\min}, \vartheta_i^{\max}) \longleftarrow$$

multidimensional
uniform distribution

$\hat{\mathbf{L}}(c) \longrightarrow$



MODEL ORDER REDUCTION

Employing POD to construct RB spaces of size 40 (hydr. head) and 320 (conc.) :

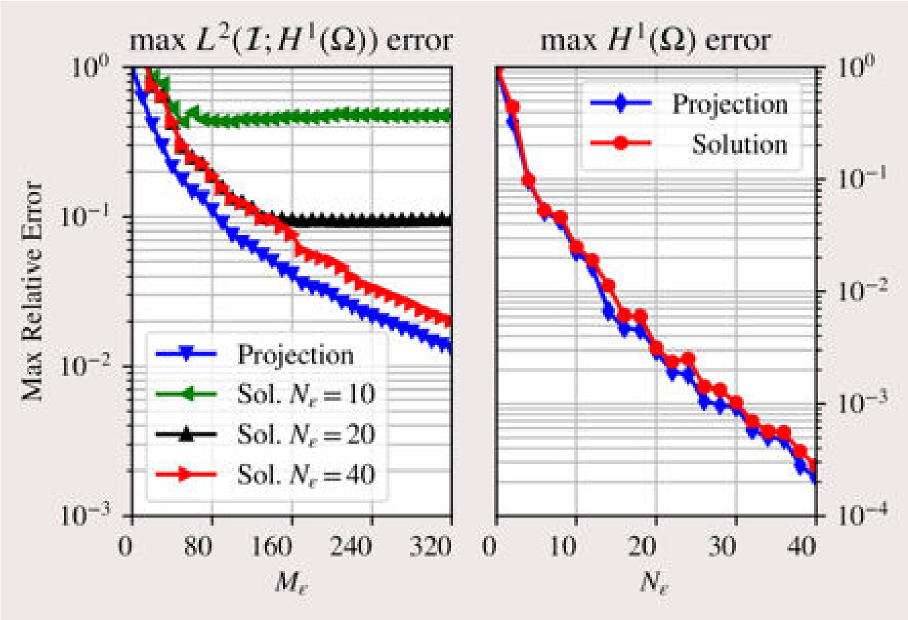
dofs spatial discretization = 44,972

dofs time discretization = 50

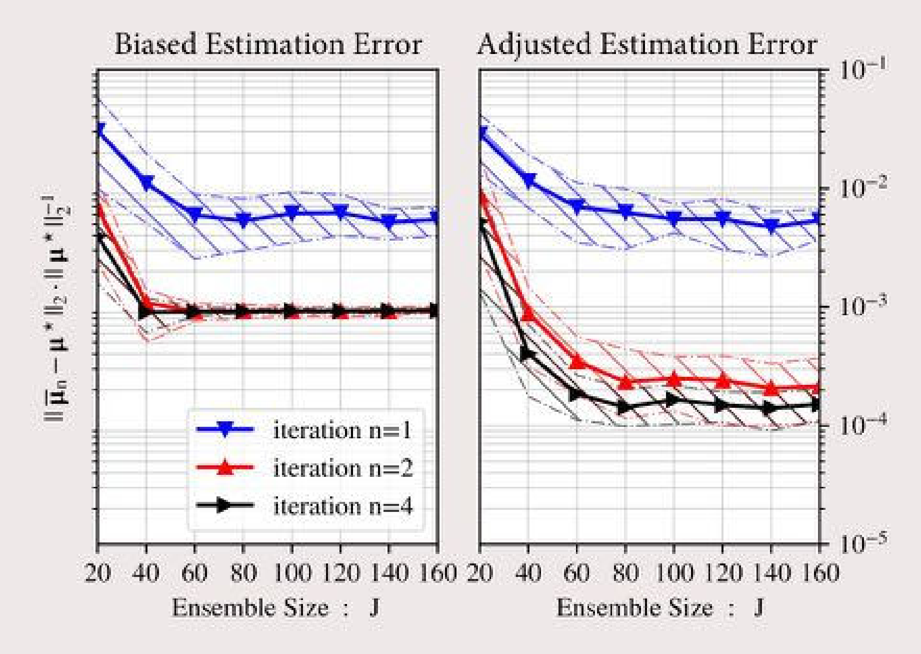
training set size = 2000 x 50

training time = 75h

An accurate reconstruction of the hydraulic head field is essential for the concentration



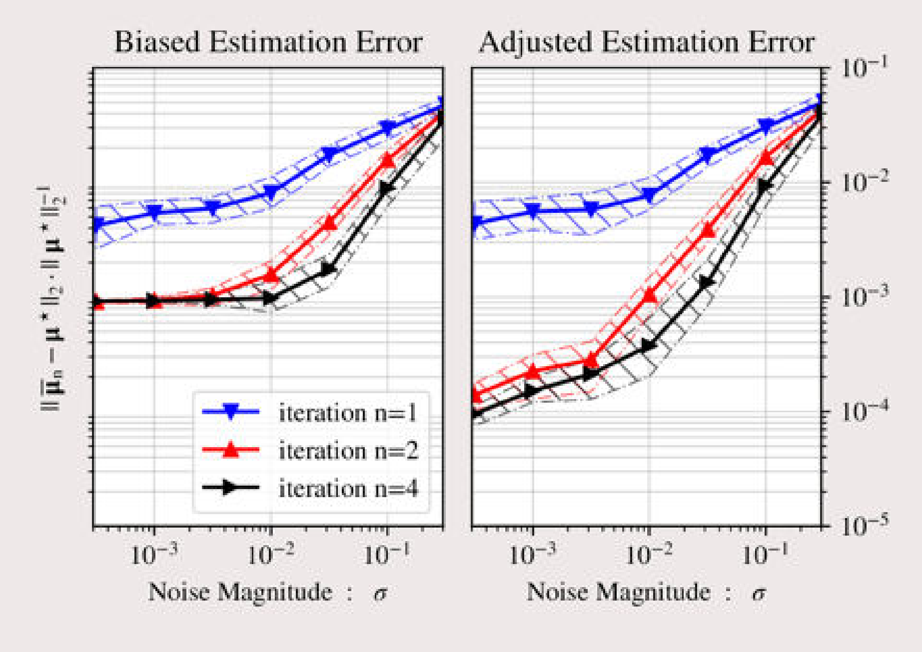
PARAMETER ESTIMATION : ENSEMBLE SIZE



as expected, a lower estimation error is reached with the adjusted method than with the biased method

the higher dimensional parameter space requires larger ensembles if compared to the previous study case

PARAMETER ESTIMATION : NOISE MAGNITUDE



the error stagnates when the model bias is not corrected in the RB-EnKM

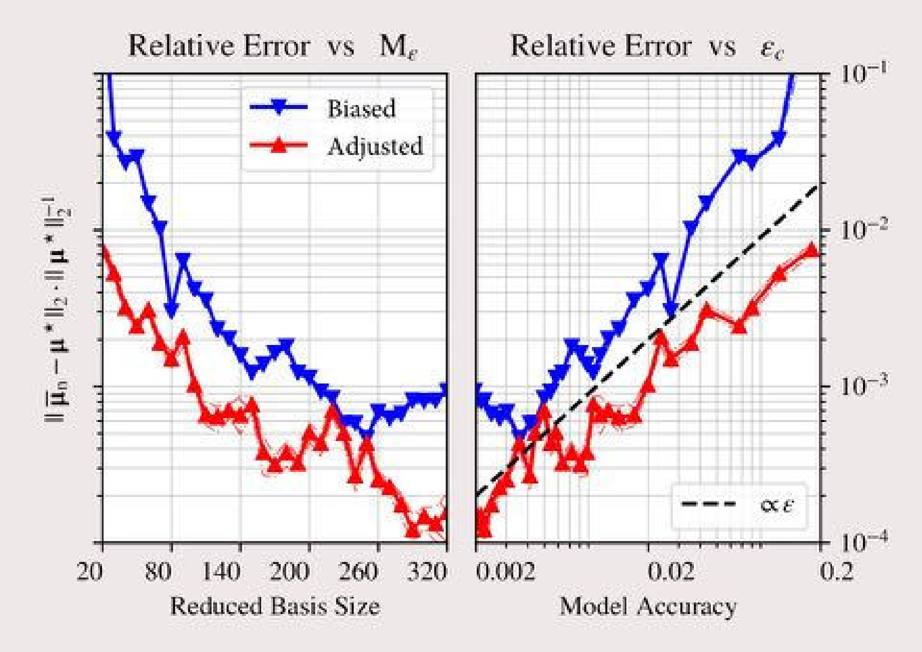
the adjusted RB-EnKM preserves the error decay at low noise magnitudes

PARAMETER ESTIMATION : REDUCED BASIS SIZE

the data bias correction is essential
to reduce the estimation error

the performance of the adjusted method are
not independent of the RB size

the estimation error - model accuracy
relationship seems nearly linear in both cases



MOTIVATION

In the parametrized PDE setting:

(1) How can we account for the error introduced by the use of a parametric reduced order model in a data assimilation framework (EnKI)?

(2) How can we construct efficient (both offline *and* online) parametric reduced order models for data assimilation (EnKF)?

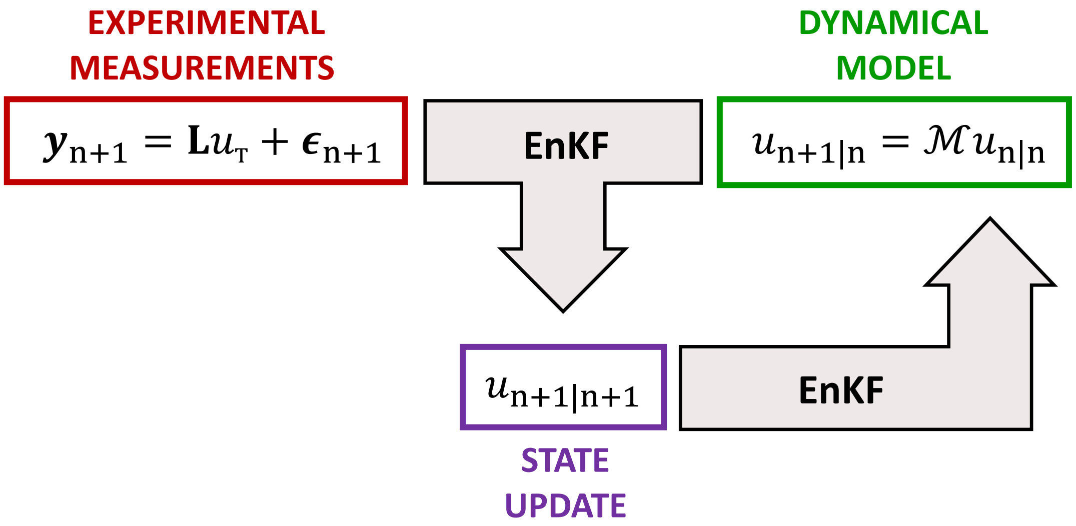
(3) How can we mitigate the effect of using an approximate model on parameter inference (via experimental design)?

joint work with **F. Silva**, C. Pagliantini

[Kal60]

R. E. Kalman. "A new approach to linear filtering and prediction problems". (1960)

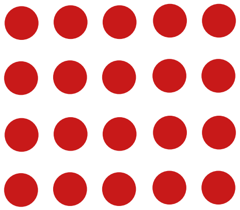
DATA ASSIMILATION



THE ENSEMBLE KALMAN FILTER

PREDICT : $u^{(j)}_{n+1|n} = \mathcal{M} u^{(j)}_{n|n}$

(j)

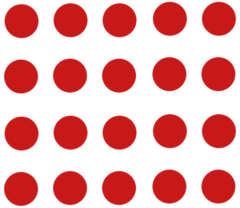


THE ENSEMBLE KALMAN FILTER

PREDICT : $u^{(j)}_{n+1|n} = \mathcal{M} u^{(j)}_{n|n}$

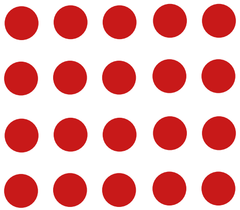
ESTIMATE : $\hat{\mathbf{C}}_{n+1|n} = \mathrm{cov}\left\{u^{(j)}_{n+1|n}\right\}$

(j)



THE ENSEMBLE KALMAN FILTER

(j)



PREDICT : $$u^{(j)}_{n+1|n} = \mathcal{M} u^{(j)}_{n|n}$$

ESTIMATE : $$\hat{\mathbf{C}}_{n+1|n} = \operatorname{cov}\left\{u^{(j)}_{n+1|n}\right\}$$

ANALYSE : $$u^{(j)}_{n+1|n+1} = u^{(j)}_{n+1|n} + \hat{\mathbf{K}}_{n+1}\left(\boldsymbol{y}_{n+1} - \mathbf{L}u^{(j)}_{n+1|n}\right)$$

employing the empirical Kalman gain

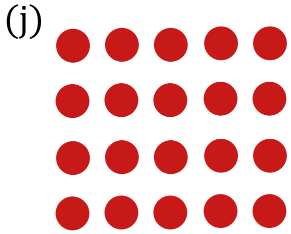
$$\hat{\mathbf{K}}_{n+1} = \hat{\mathbf{C}}_{n+1|n}\mathbf{L}^{*}\left(\mathbf{L}\hat{\mathbf{C}}_{n+1|n}\mathbf{L}^{*} + \boldsymbol{\Sigma}\right)^{-1}$$

THE ENSEMBLE KALMAN FILTER

PREDICT : $$u^{(j)}_{n+1|n} = \mathcal{M} u^{(j)}_{n|n}$$ **EXPENSIVE!**

ESTIMATE : $$\hat{\mathbf{C}}_{n+1|n} = \operatorname{cov}\left\{u^{(j)}_{n+1|n}\right\}$$

ANALYSE : $$u^{(j)}_{n+1|n+1} = u^{(j)}_{n+1|n} + \widehat{\mathbf{K}}_{n+1}\left(\boldsymbol{y}_{n+1} - \mathbf{L}u^{(j)}_{n+1|n}\right)$$



employing the empirical Kalman gain

$$\widehat{\mathbf{K}}_{n+1} = \hat{\mathbf{C}}_{n+1|n}\mathbf{L}^*\left(\mathbf{L}\hat{\mathbf{C}}_{n+1|n}\mathbf{L}^* + \boldsymbol{\Sigma}\right)^{-1}$$

[PMS21]

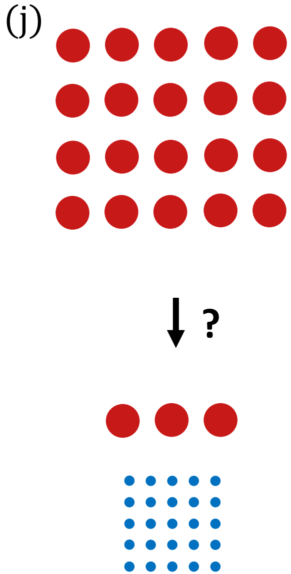
A. Popov, et all. "A multifidelity ensemble Kalman filter with reduced order control variates." (2021)

THE ENSEMBLE KALMAN FILTER

PREDICT : $u^{(j)}_{n+1|n} = \mathcal{M} u^{(j)}_{n|n}$

ESTIMATE : $\hat{\mathbf{C}}_{n+1|n} = \operatorname{cov}\left\{u^{(j)}_{n+1|n}\right\}$

ANALYSE : $u^{(j)}_{n+1|n+1} = u^{(j)}_{n+1|n} + \widehat{\mathbf{K}}_{n+1}\left(\boldsymbol{y}_{n+1} - \mathbf{L} u^{(j)}_{n+1|n}\right)$



employing the empirical Kalman gain

$$\widehat{\mathbf{K}}_{n+1} = \hat{\mathbf{C}}_{n+1|n}\mathbf{L}^*\left(\mathbf{L}\hat{\mathbf{C}}_{n+1|n}\mathbf{L}^* + \boldsymbol{\Sigma}\right)^{-1}$$

THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER

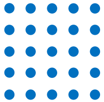
[PMS21]

A. Popov, et all. "A multifidelity ensemble Kalman filter with reduced order control variates." (2021)

Principal Ensemble

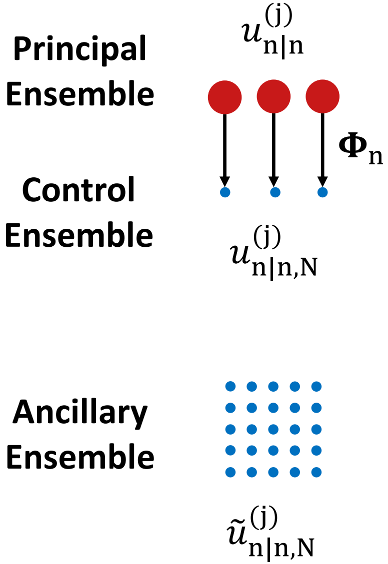
$u_{n|n}^{(j)}$

Ancillary Ensemble



$\tilde{u}_{n|n,N}^{(j)}$

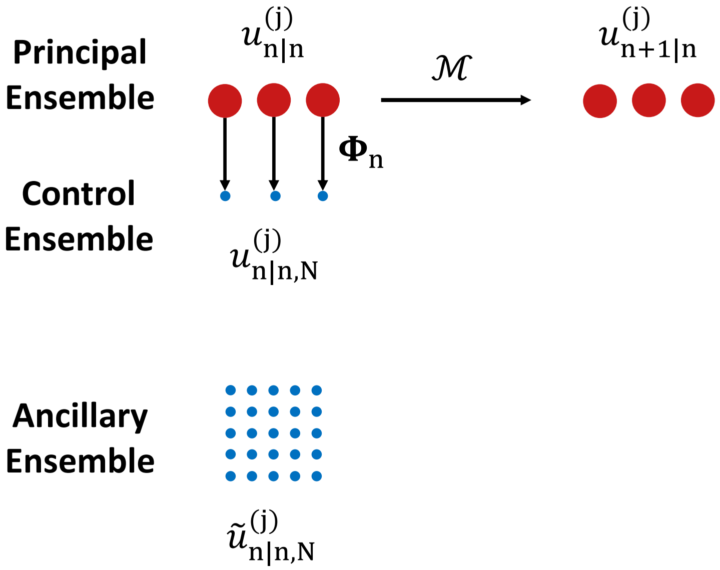
THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER



[PMS21]

A. Popov, et all. "A multifidelity ensemble Kalman filter with reduced order control variates." (2021)

THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER



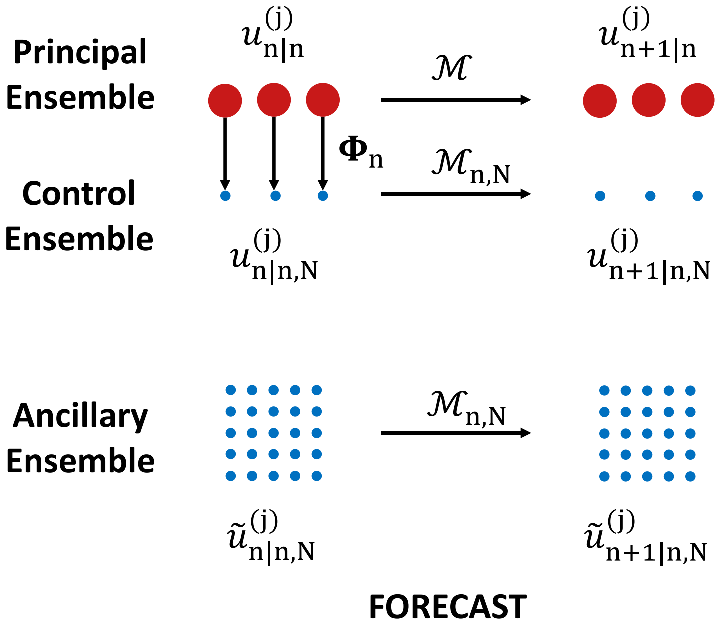
[PMS21]

A. Popov, et all. "A multifidelity ensemble Kalman filter with reduced order control variates." (2021)

THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER

[PMS21]

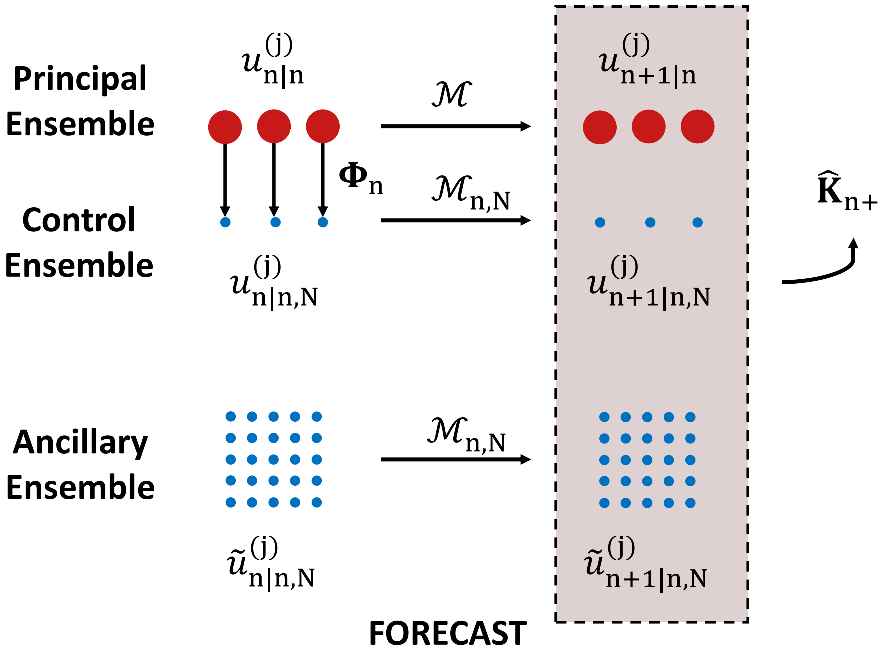
A. Popov, et all. "A multifidelity ensemble Kalman filter with reduced order control variates." (2021)



THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER

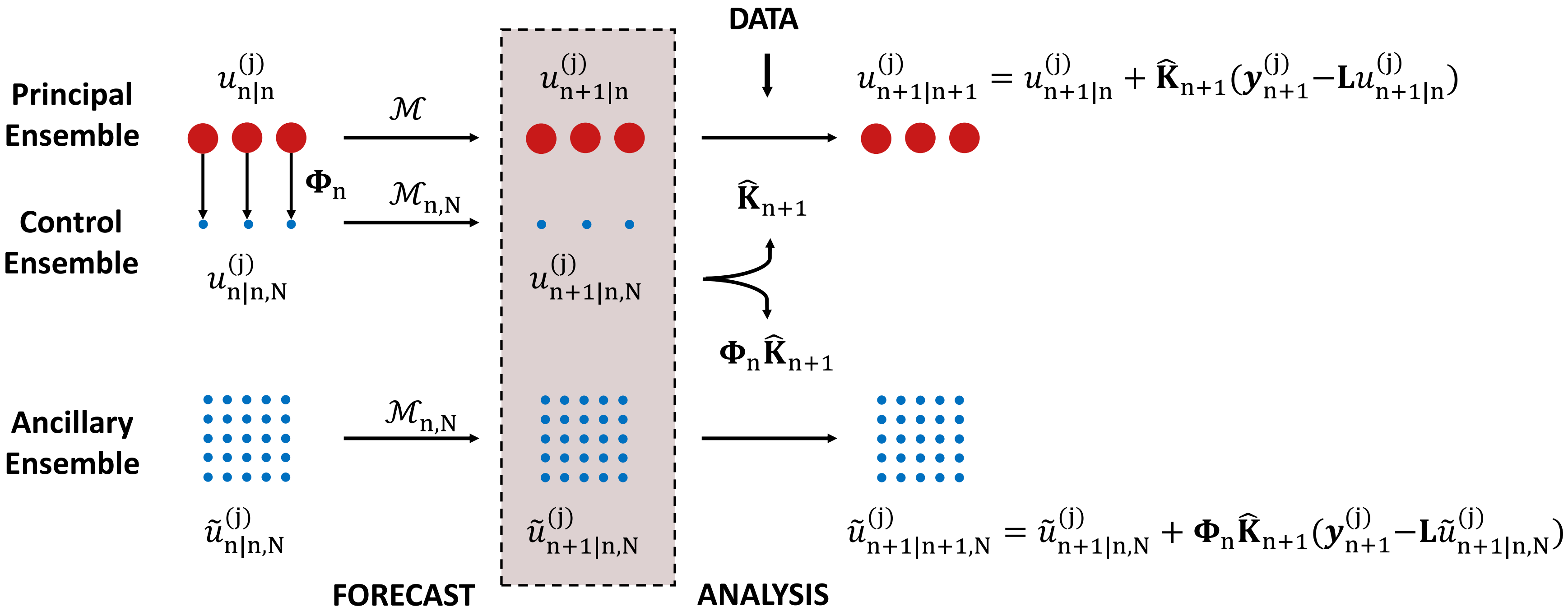
[PMS21]

A. Popov, et all. "A multifidelity ensemble Kalman filter with reduced order control variates." (2021)

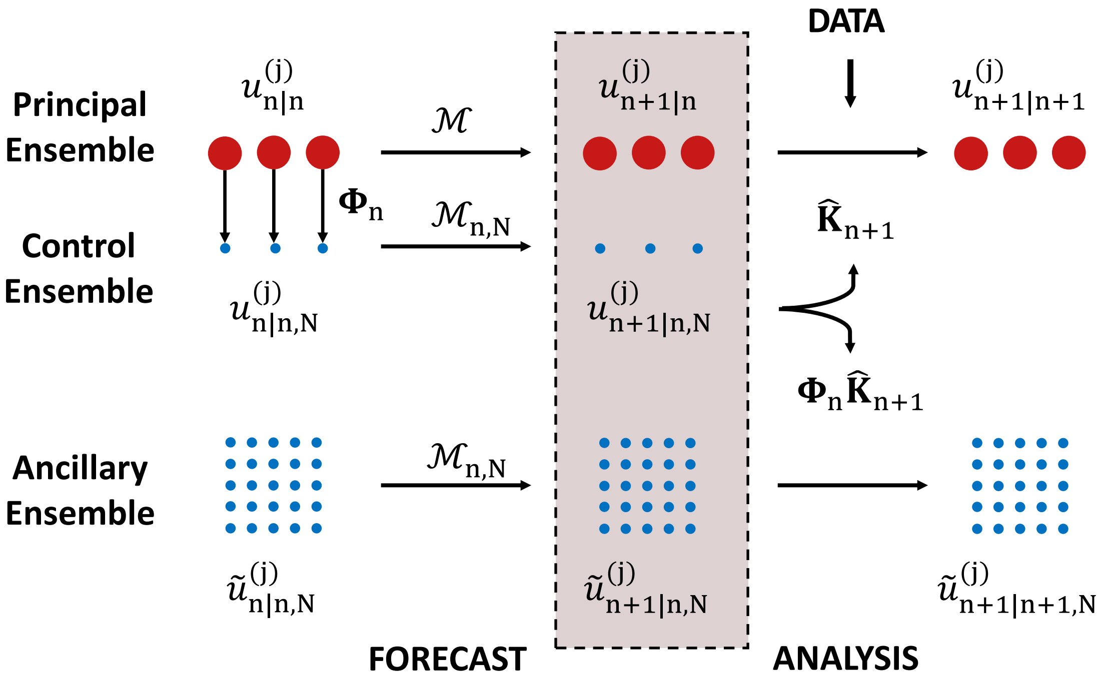


$$\widehat{\mathbf{K}}_{n+1} = \widehat{\mathbf{C}}_{n+1|n}\mathbf{H}^*(\mathbf{H}\widehat{\mathbf{C}}_{n+1|n}\mathbf{H}^* + \boldsymbol{\Sigma})^{-1}$$

THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER : SOME COMPLICATION

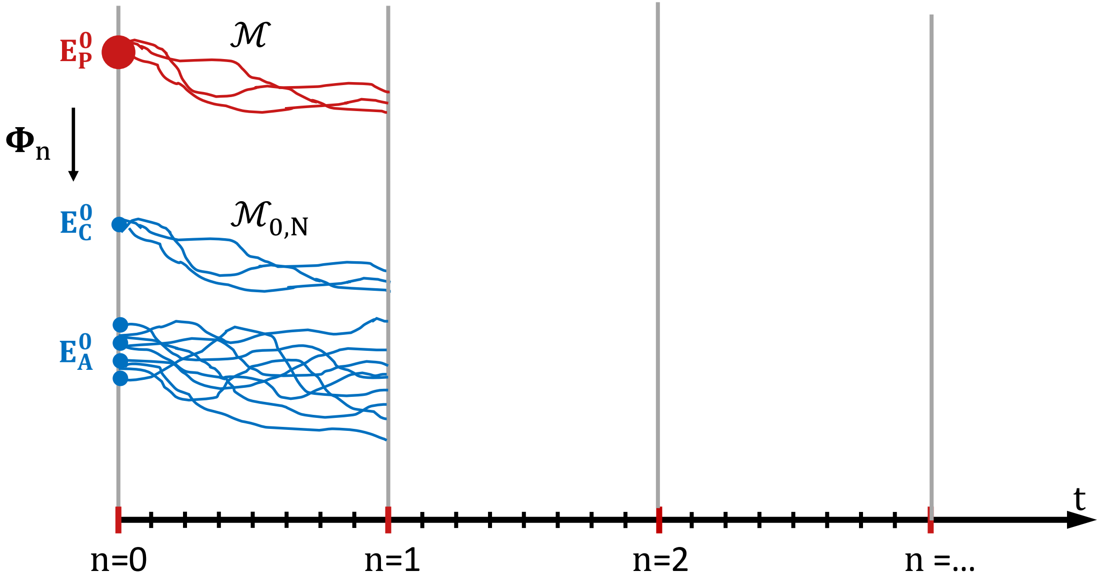


THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER : SOME COMPLICATION



? $\mathcal{M}_{n,N}$?

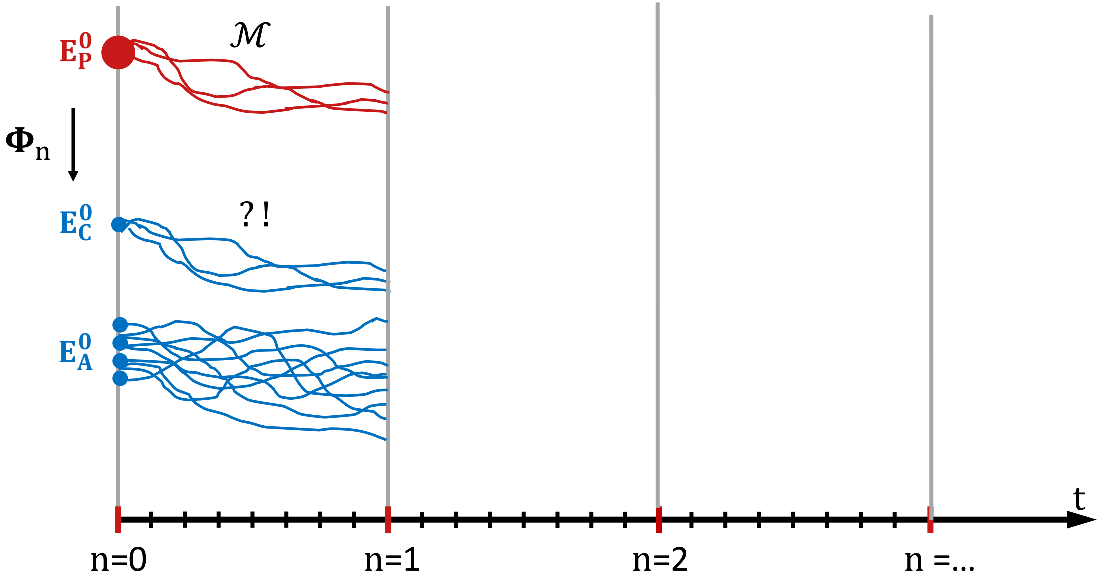
THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER



[PMS21]

A. Popov, et all. "A multifidelity ensemble Kalman filter with reduced order control variates." (2021)

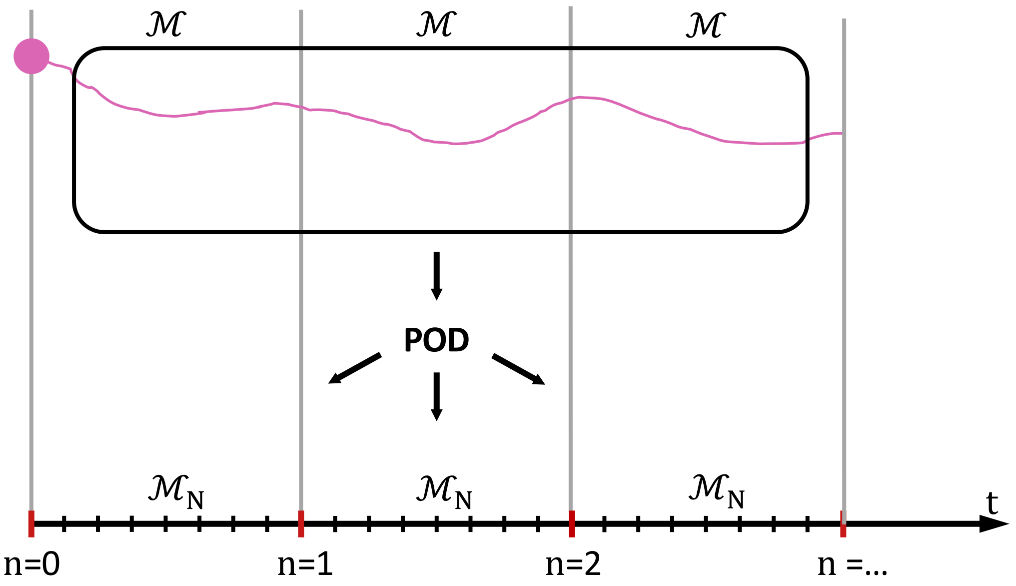
THE MULTI-FIDELITY ENSEMBLE KALMAN FILTER



[PMS21]

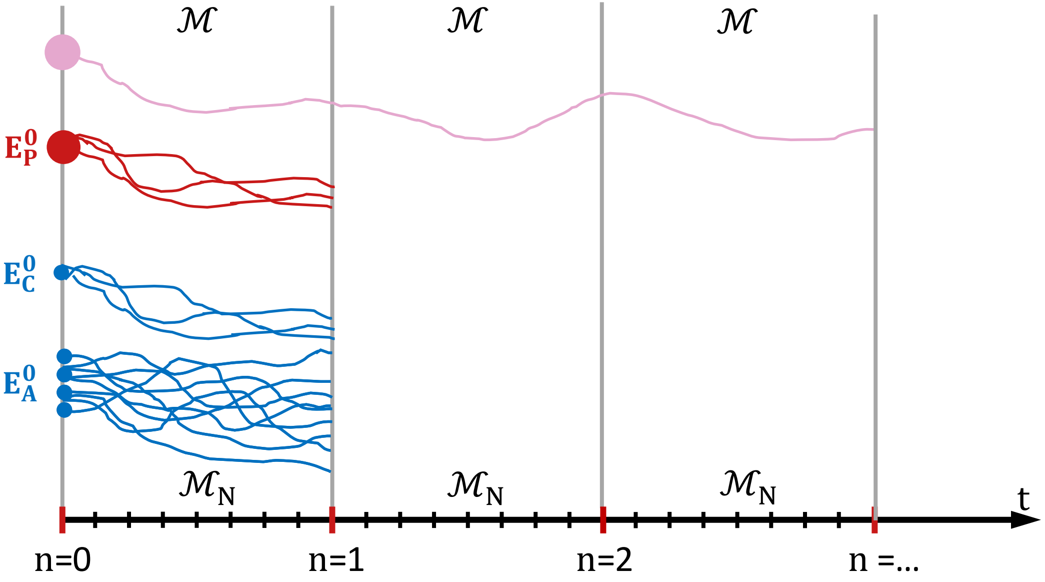
A. Popov, et all. "A multifidelity ensemble Kalman filter with reduced order control variates." (2021)

RB MODEL CONSTRUCTION : [PMS21]



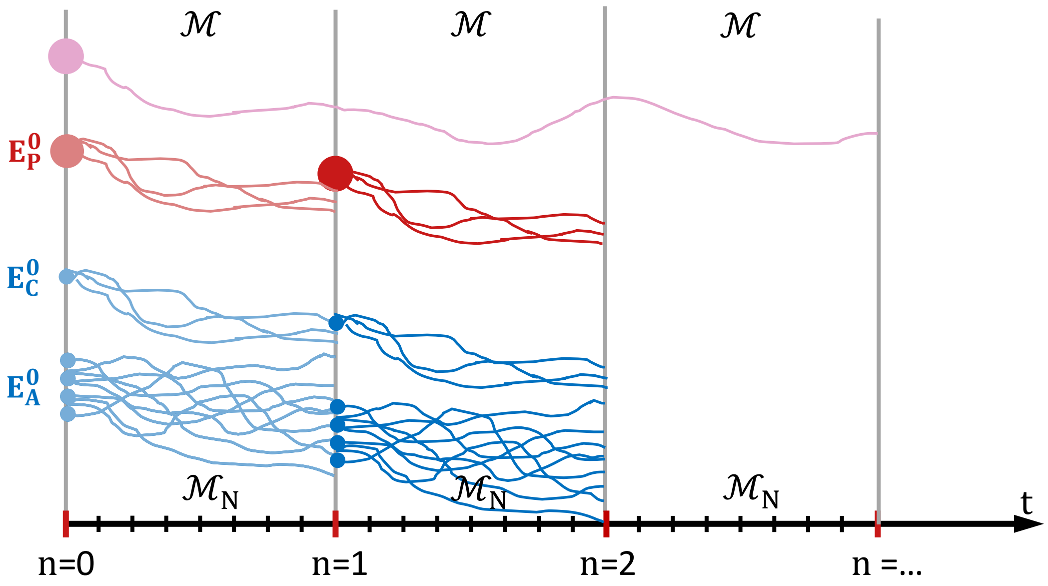
one long trajectory is used to build offline a global RB model

RB MODEL CONSTRUCTION : [PMS21]



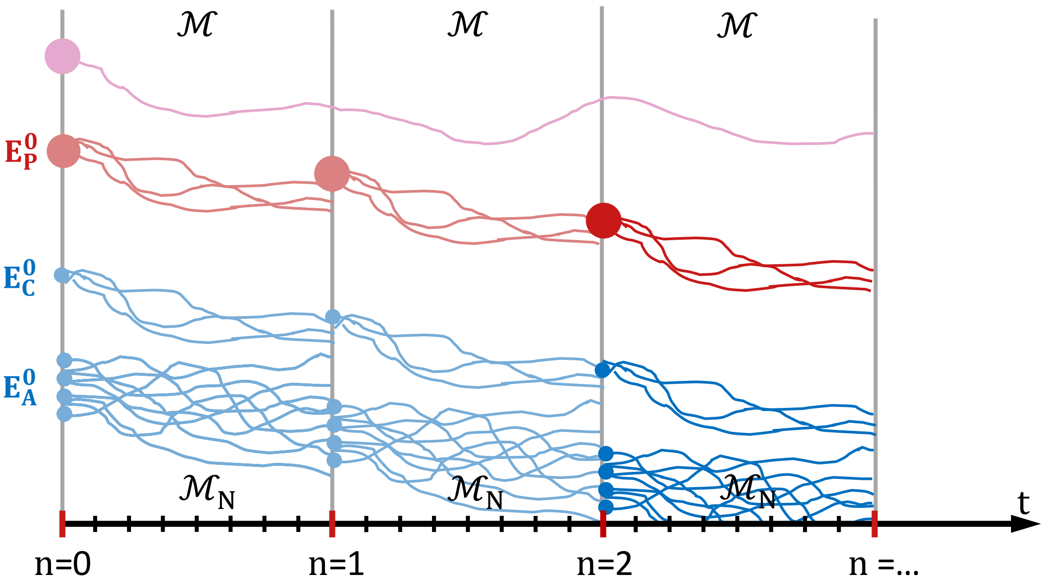
one long trajectory is used to build offline a global RB model

RB MODEL CONSTRUCTION : [PMS21]



one long trajectory is used to build offline a global RB model

RB MODEL CONSTRUCTION : [PMS21]



one long trajectory is used to build offline a global RB model

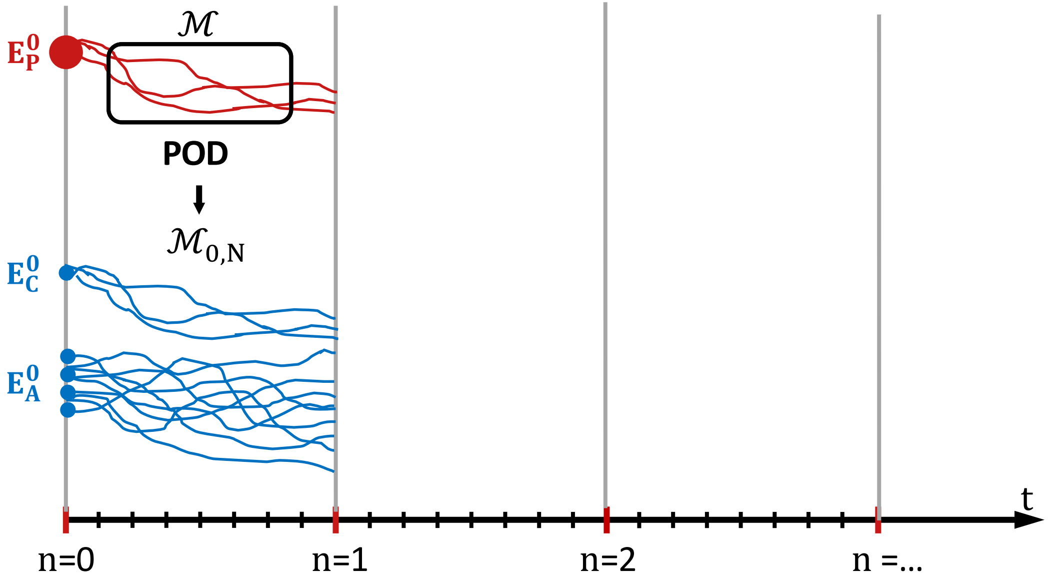
PROs:

- easy to implement
- can incorporate steady states

CONs:

- suffers initial uncertainty
- leads to large RB models
- might introduce biases

RB MODEL CONSTRUCTION : [DY22]

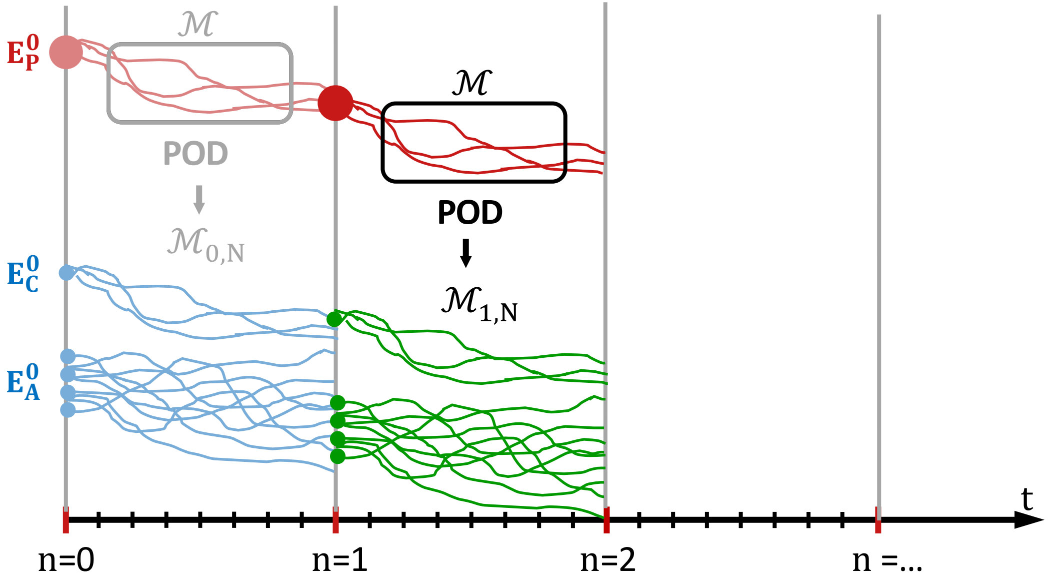


the principal trajectories are used to build RB models on-the-fly

[DY22]

G. Donoghue and M. Yano. "A multi-fidelity ensemble Kalman filter with hyperreduced reduced-order models". (2022)

RB MODEL CONSTRUCTION : [DY22]

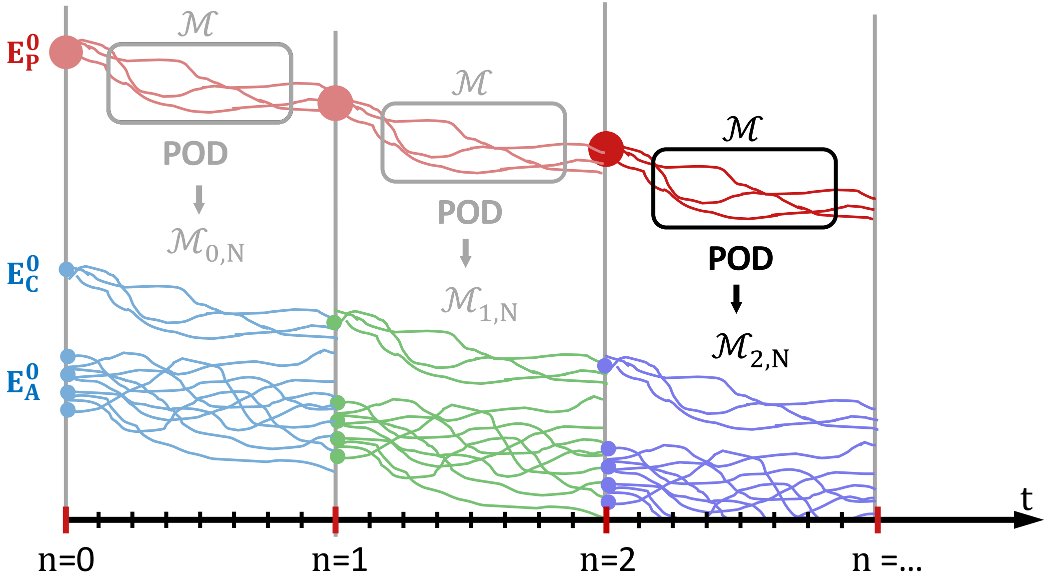


the principal trajectories are used to build RB models on-the-fly

[DY22]

G. Donoghue and M. Yano. "A multi-fidelity ensemble Kalman filter with hyperreduced reduced-order models". (2022)

RB MODEL CONSTRUCTION : [DY22]



the principal trajectories are used to build RB models on-the-fly

PROs:

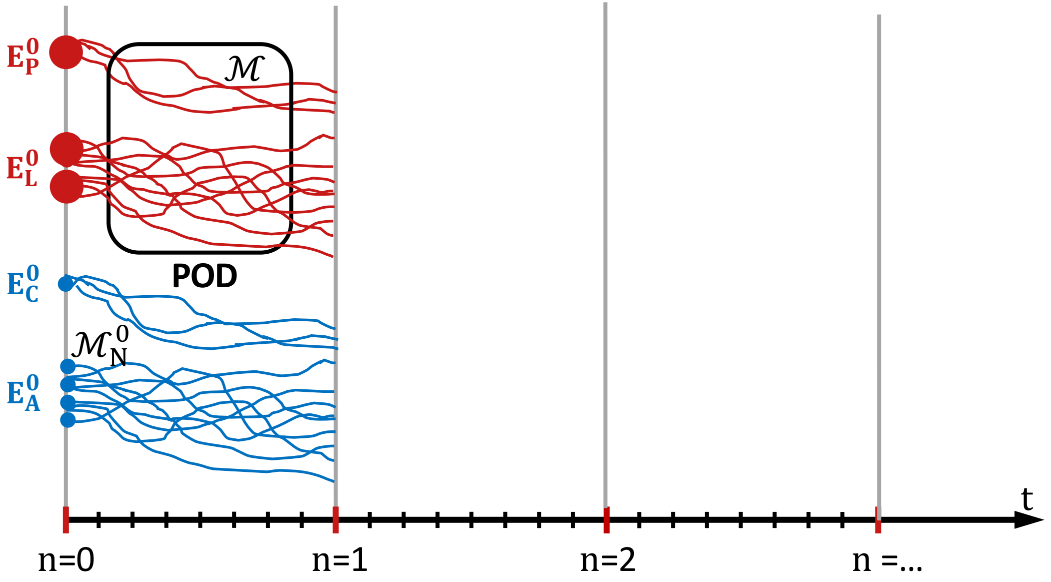
- leads to small RB models
- doesn't introduce biases

CONs:

- requires constant retraining
- too little information extracted (poor accuracy for the RB model)

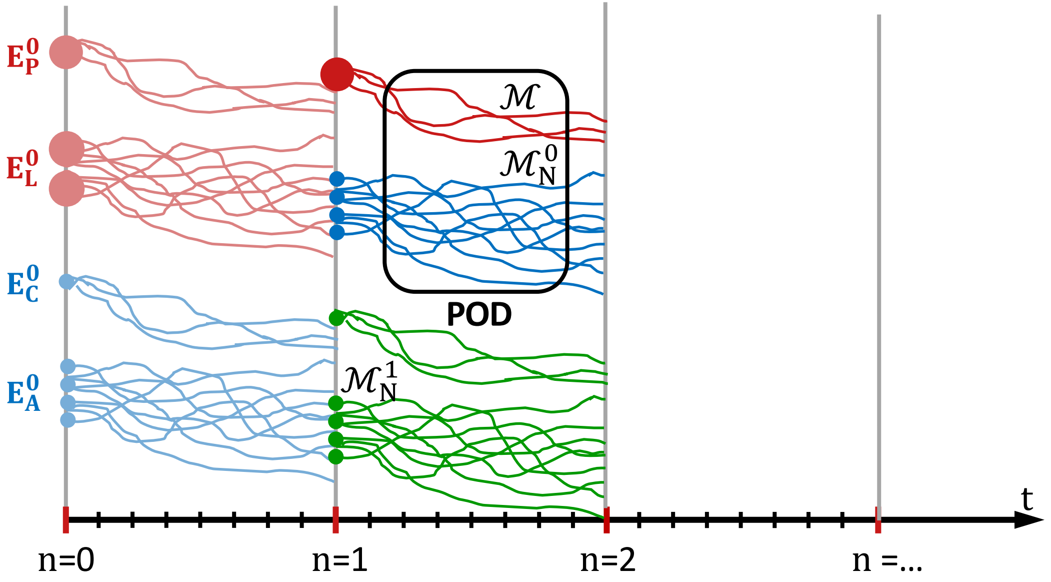
[DY22]

RB MODEL CONSTRUCTION : PROPOSED APPROACH



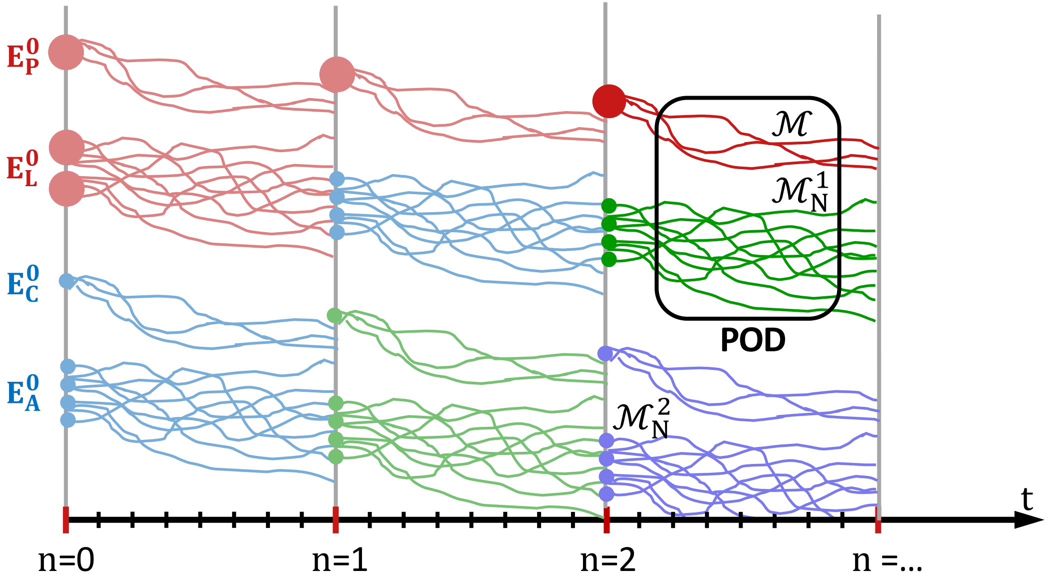
principal and legacy trajectories are used to build RB models on-the-fly

RB MODEL CONSTRUCTION : PROPOSED APPROACH



principal and auxiliary ancillary trajectories are used to build RB models on-the-fly

RB MODEL CONSTRUCTION : PROPOSED APPROACH



principal and auxiliary ancillary trajectories are used to build RB models on-the-fly

PROs:

- retains past model's information
- achieves good accuracy

CONs:

- first step is significantly expensive
- large range of RB model sizes

[MWW20]

C. Mou, Z. Wang, D. Wells, X. Xie and T. Iliescu. "Reduced order models for the quasi-geostrophic equations: A brief survey." (2020)

QUASI-GEOSTROPHIC EQUATIONS

find $\omega = \omega(\mathrm{x},\mathrm{y},\mathrm{t})$, $\psi = \psi(\mathrm{x},\mathrm{y},\mathrm{t})$ such that

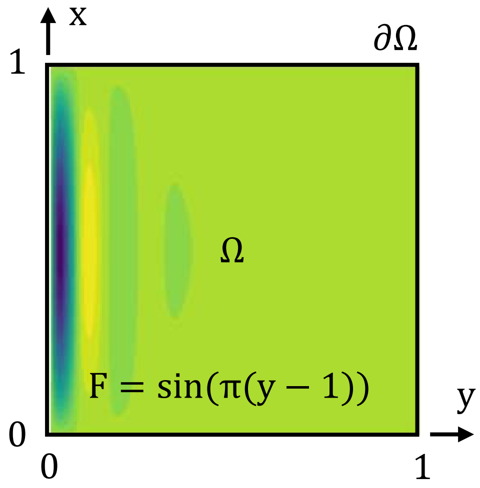
$$\partial_t \omega = \mathrm{Ro}\,\mathrm{J}(\omega,\psi) + \partial_\mathrm{x}\psi + \frac{\mathrm{Ro}}{\mathrm{Re}}\Delta\omega + \mathrm{F}, 0\Delta\psi + \omega = 0 \longleftarrow \mathrm{J}(\omega,\psi) = \partial_\mathrm{x}\psi\partial_\mathrm{y}\omega - \partial_\mathrm{x}\psi\partial_\mathrm{y}\omega$$

given the boundary and initial conditions

$$\omega(\mathrm{x},\mathrm{y},\mathrm{t}) = 0, \quad (\mathrm{x},\mathrm{y}) \in \partial\Omega$$
$$\psi(\mathrm{x},\mathrm{y},\mathrm{t}) = 0, \quad (\mathrm{x},\mathrm{y}) \in \partial\Omega$$
$$\omega(\mathrm{x},\mathrm{y},0) = \omega_0, \quad (\mathrm{x},\mathrm{y}) \in \Omega$$

and

$$\partial_\mathrm{x}\psi_0 + \frac{\mathrm{Ro}}{\mathrm{Re}}\Delta\omega_0 + \mathrm{F} = 0, 0\Delta\psi_0 + \omega_0 = 0$$



QUASI-GEOSTROPHIC EQUATIONS

high-fidelity physical model constructed considering:

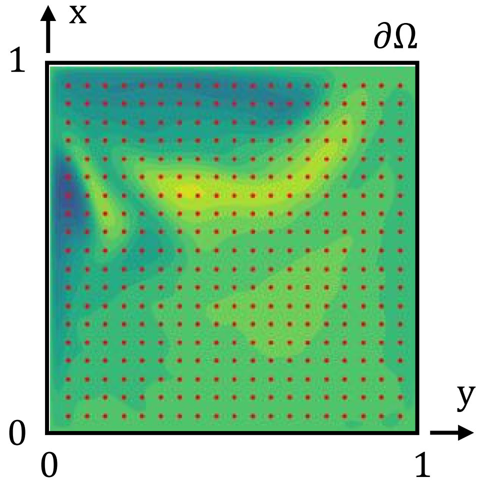
- fully implicit mid-point discretization in time $(\mathrm{dt} = 0.1)$
- P1 finite elements discretization in space (4225 dofs)

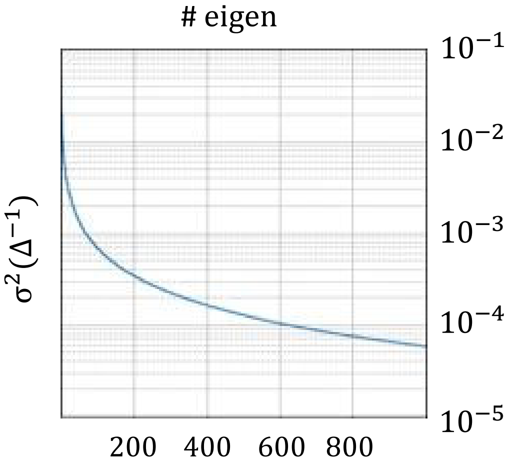
measurement model constructed considering:

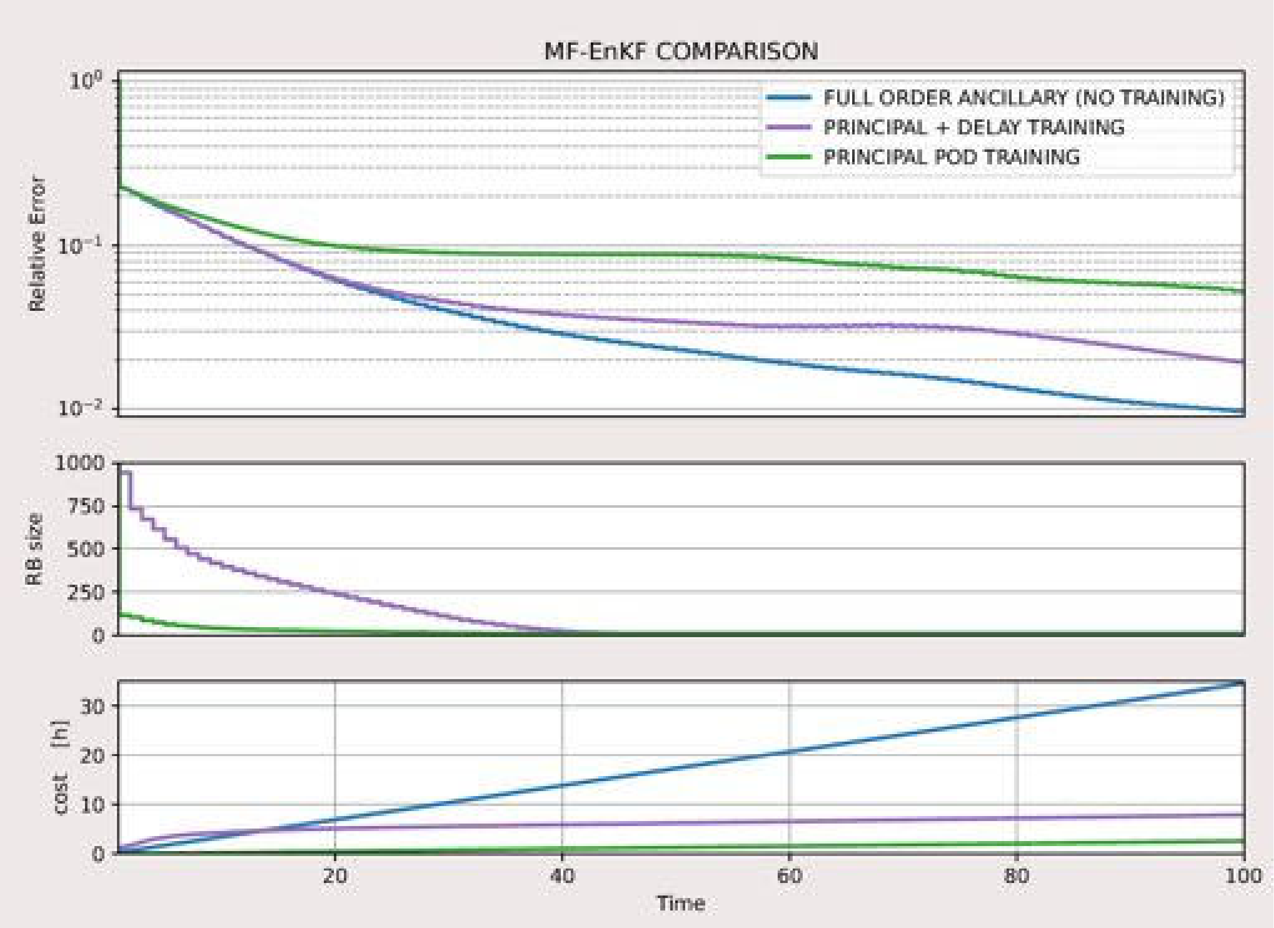
- evenly spaced sensor positions $(19\text{x}19)$
- data collection every 10 time-steps

probabilistic model assumes:

- homoscedastic noise $\boldsymbol{\epsilon}_{\mathrm{n+1}} \sim N(0, \sigma^2 \mathbf{I})$ $(\sigma = 10^{-4})$
- normal initial sample distribution $\omega_{0|0} \sim N(0, \Delta^{-1})$







$\#\mathbf{E_P}$: **50**

$\#\mathbf{E_A}$: **500**

REFERENCES

[ILS13] M. A. Iglesias, K. J. H. Law, and A. M. Stuart. "**Ensemble Kalman methods for inverse problems**". In: Inverse Problems 29.4 (2013)

[PMQ16] S. Pagani, A. Manzoni, and A. Quarteroni. "**A reduced basis ensemble Kalman filter for state/parameter identification in large-scale nonlinear dynamical systems**" (2016)

[SEB10] P. Sakov, G. Evensen, and L. Bertino. "**Asynchronous data assimilation with the EnKF**" (2010)

[Cal+18] D. Calvetti et al. "**Accounting for model error due to unresolved scales within ensemble Kalman filtering**". In: Quarterly Journal of the Royal Met. Society (2018)

[Kär+18] M. Kärcher et al. "**Reduced basis approximation and a posteriori error bounds for 4D-Var data assimilation**", In: Optim. Eng. (2018)

[Hec12] F. Hecht. "**New development in freefem++**". In: J. Numer. Math (2012)

REFERENCES

[DM22] G. Donoghue and M. Yano. "**A multi-fidelity ensemble Kalman filter with hyperreduced reduced-order models**“. (2022)

[PMS21] A. Popov, C. Mou, A. Sandu, and T. Iliescu. "**A multifidelity ensemble Kalman filter with reduced order control variates.**" (2021)

[MWW20] C. Mou, Z. Wang, D. Wells, X. Xie and T. Iliescu. "**Reduced order models for the quasi-geostrophic equations: A brief survey.**" (2020)

[SBM18] M. Strazzullo, F. Ballarin, R. Mosetti, and G. Rozza. "**Model Reduction for Parametrized Optimal Control Problems in Environmental Marine Sciences and Engineering**“ (2018)

[Eve03] G. Evensen. "**The ensemble Kalman filter: Theoretical formulation and practical implementation**". (2003)

[Kal60] R. E. Kalman. "**A new approach to linear filtering and prediction problems**". (1960)

MOTIVATION

In the parametrized PDE setting:

(1) How can we account for the error introduced by the use of a parametric reduced order model in a data assimilation framework?

(2) How can we construct efficient (both offline *and* online) parametric reduced order models for data assimilation?

(3) How can we mitigate the effect of using an approximate model on parameter inference (via experimental design)?

joint work with H. Bansal, **N. Cvetkovic**, H.C. Lie

Setup

$$Y := \mathcal{O} \circ \mathcal{M}^\dagger(\theta^\dagger) + \varepsilon$$

$\mathcal{M}^\dagger : \Theta \to \mathcal{U}$	true model
$\theta^\dagger \in \Theta$	true parameter
$\mathcal{M}^\dagger(\theta^\dagger) \in \mathcal{U}$	true state
$\mathcal{O} : \mathcal{U} \to \mathbb{R}^n$	observation operator, linear, continuous
$\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$	noise
$y \in \mathbb{R}^n$	realisation of Y - data

Bayesian inverse problem: specify prior $\theta \sim \mu_\theta$ and true misfit

$$\Phi^{y,\dagger}(\theta') := \frac{1}{2}\|y - \mathcal{O} \circ \mathcal{M}^\dagger(\theta')\|^2_{\Sigma_\varepsilon^{-1}}$$

Using Bayes' law the true posterior on θ

$$\mathrm{d}\mu_\theta^{y,\dagger}(\theta') := \frac{\exp(-\Phi^{y,\dagger}(\theta'))}{Z(\Phi^{y,\dagger})}\mathrm{d}\mu_\theta(\theta')$$

Approximate posterior

If $\mathcal{M}^\dagger$ is unknown or expensive it is approximated by $\mathcal{M}$

$\delta^\dagger := \mathcal{M}^\dagger - \mathcal{M}$ model error

$$Y := \mathcal{O} \circ \mathcal{M}(\theta^\dagger) + \varepsilon$$

Given prior $\theta \sim \mu_\theta$, data y, and approximate misfit

$$\Phi^{y,\mathrm{A}}(\theta') := \frac{1}{2}\|y - \mathcal{O} \circ \mathcal{M}(\theta')\|^2_{\Sigma_\varepsilon^{-1}}$$

approximate posterior

$$\mathrm{d}\mu_\theta^{y,\mathrm{A}}(\theta') := \frac{\exp(-\Phi^{y,\mathrm{A}}(\theta'))}{Z(\Phi^{y,\mathrm{A}})}\mathrm{d}\mu_\theta(\theta')$$

Approximate posterior corresponds to the assumption that $\delta^\dagger = 0$

Question: Can we mitigate the error propagation to posterior?

Enhanced noise approach

$$Y = \mathcal{O} \circ \mathcal{M}^\dagger(\theta^\dagger) + \varepsilon = \mathcal{O} \circ \mathcal{M}(\theta^\dagger) + (\mathcal{O} \circ \underbrace{\delta^\dagger(\theta^\dagger)}_{\approx \mathfrak{u}} + \varepsilon)$$

Assumption: $\mathfrak{u} \sim \mathcal{N}(m_\mathfrak{u}, \Sigma_\mathfrak{u})$, independent of $\theta \sim \mu_\theta$ and $\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$.

$$\mathcal{O}\mathfrak{u} + \varepsilon \sim \mathcal{N}(\mathcal{O}m_\mathfrak{u}, \Sigma_\varepsilon + \mathcal{O}\Sigma_\mathfrak{u}\mathcal{O}^*)$$

Enhanced noise misfit and enhanced noise posterior

$$\Phi^{y,\mathrm{E}}(\theta') := \frac{1}{2}\|y - \mathcal{O} \circ \mathcal{M}(\theta') - \mathcal{O}m_\mathfrak{u}\|^2_{(\Sigma_\varepsilon + \mathcal{O}\Sigma_\mathfrak{u}\mathcal{O}^*)^{-1}}$$

$$\mathrm{d}\mu_\theta^{y,\mathrm{E}}(\theta') := \frac{\exp(-\Phi^{y,\mathrm{E}}(\theta'))}{Z(\Phi^{y,\mathrm{E}})}\mathrm{d}\mu_\theta(\theta')$$

Note: the only unknown that we aim to infer is $\theta^\dagger$

Question: How to mitigate the effect of using an approximate model on parameter inference?

Hint:

- Analyse distances between $\mu_\theta^{y,\dagger}$, $\mu_\theta^{y,\mathrm{A}}$ and $\mu_\theta^{y,\mathrm{E}}$
- Answer in terms of selecting an appropriate observation operator $\mathcal{O}$

Given μ and $\Phi \in L^1_\mu(E;\mathbb{R})$, define μ_Φ by

$$\frac{\mathrm{d}\mu_\Phi}{\mathrm{d}\mu}(x') = \frac{\exp(-\Phi(x'))}{Z(\Phi)}, \quad Z(\Phi) := \int_E \exp(-\Phi(x'))\mathrm{d}\mu(x')$$

Theorem: (Sprungk 2020)

Let μ, $\Phi_1 \in L^1_\mu(E;\mathbb{R}_{\geq 0})$, and $\Phi_2 \in L^1_\mu(E;\mathbb{R}_{\geq 0})$. Then

$$\max\{d_{\mathrm{KL}}(\mu_{\Phi_1}\|\mu_{\Phi_2}), d_{\mathrm{KL}}(\mu_{\Phi_2}\|\mu_{\Phi_1})\} \leq C\|\Phi_1 - \Phi_2\|_{L^1_\mu}$$

Error of $\mu_\theta^{y,A}$ with respect to $\mu_\theta^{y,\dagger}$

Proposition: If $\Phi^{y,A}, \Phi^{y,\dagger} \in L^1_{\mu_\theta}(\Theta, \mathbb{R}_{\geq 0})$, then

$$\max\{d_{\mathrm{KL}}(\mu_\theta^{y,A} \| \mu_\theta^{y,\dagger}), d_{\mathrm{KL}}(\mu_\theta^{y,\dagger} \| \mu_\theta^{y,A})\} \leq C \| \|\mathcal{O} \circ \delta^\dagger\|^2_{\Sigma_\varepsilon^{-1}} \|^{1/2}_{L^1_{\mu_\theta}}$$

Interpretation:

- $\mathbb{P}(\mathcal{O} \circ \delta^\dagger(\theta) = 0) = 1 \quad \Longrightarrow \quad \mu_\theta^{y,A} = \mu_\theta^{y,\dagger}$
- Let V be linear space and $\delta^\dagger \in V \subsetneq \mathcal{U}$ then
  $V \subseteq \ker(\mathcal{O}) \quad \Longrightarrow \quad \mu_\theta^{y,A} = \mu_\theta^{y,\dagger}$

Key message: $\mu_\theta^{y,A}$ can be as good as $\mu_\theta^{y,\dagger}$

Error of $\mu_\theta^{y,\mathrm{E}}$ with respect to $\mu_\theta^{y,\dagger}$

Proposition: If $\Phi^{y,\mathrm{E}}, \Phi^{y,\dagger} \in L^1_{\mu_\theta}(\Theta, \mathbb{R}_{\geq 0})$, then

$$\max\{d_{\mathrm{KL}}(\mu_\theta^{y,\dagger} \| \mu_\theta^{y,\mathrm{E}}), d_{\mathrm{KL}}(\mu_\theta^{y,\mathrm{E}} \| \mu_\theta^{y,\dagger})\}$$
$$\leq C\Big(\|\|\mathcal{O} \circ (\delta^\dagger - m_{\mathfrak{u}})\|^2_{\Sigma_\varepsilon^{-1}}\|^{1/2}_{L^1_{\mu_\theta}}$$
$$+ \|\|y - \mathcal{O} \circ \mathcal{M} - \mathcal{O} m_{\mathfrak{u}}\|^2_{\Sigma_\varepsilon^{-1} - (\Sigma_\varepsilon + \mathcal{O}\Sigma_{\mathfrak{u}}\mathcal{O}^*)^{-1}}\|_{L^1_{\mu_\theta}}\Big)$$

Interpretation: $\mu_\theta^{y,\mathrm{E}} = \mu_\theta^{y,\dagger}$ if

1. $\delta^\dagger - m_{\mathfrak{u}} \in \ker(\mathcal{O}) \quad \mu_\theta -$ a. s.
2. $\mathbb{P}\big(y - \mathcal{O} \circ \mathcal{M}(\theta) - \mathcal{O} m_{\mathfrak{u}} \in \ker\big(\Sigma_\varepsilon^{-1} - (\Sigma_\varepsilon + \mathcal{O}\Sigma_{\mathfrak{u}}\mathcal{O}^*)^{-1}\big) = 1$

Special case:

- $\mathcal{O}\Sigma_{\mathfrak{u}}\mathcal{O}^* = 0$ is a sufficient condition for the second term to be zero

Discrepancy between $\mu_\theta^{y,\mathrm{E}}$ and $\mu_\theta^{y,\mathrm{A}}$

Proposition: If $\Phi^{y,\mathrm{A}}, \Phi^{y,\mathrm{E}} \in L^1_{\mu_\theta}(\Theta, \mathbb{R}_{\geq 0})$, then

$$\begin{aligned}\max\{d_{\mathrm{KL}}(\mu_\theta^{y,\mathrm{A}} \| \mu_\theta^{y,\mathrm{E}}), d_{\mathrm{KL}}(\mu_\theta^{y,\mathrm{E}} \| \mu_\theta^{y,\mathrm{A}})\} \\ &\leq C\Big(\|\mathcal{O} m_{\mathfrak{u}}\|_{\Sigma_\varepsilon^{-1}} \\ &+ \big\|\|y - \mathcal{O} \circ \mathcal{M} - \mathcal{O} m_{\mathfrak{u}}\|^2_{\Sigma_\varepsilon^{-1} - (\Sigma_\varepsilon + \mathcal{O}\Sigma_{\mathfrak{u}}\mathcal{O}^*)^{-1}}\big\|_{L^1_{\mu_\theta}}\Big)\end{aligned}$$

Interpretation: $\mu_\theta^{y,\mathrm{E}} = \mu_\theta^{y,\mathrm{A}}$ if

1. $\mathcal{O} m_{\mathfrak{u}} = 0$
2. $\mathbb{P}\big(y - \mathcal{O} \circ \mathcal{M}(\theta) - \mathcal{O} m_{\mathfrak{u}} \in \ker\big(\Sigma_\varepsilon^{-1} - (\Sigma_\varepsilon + \mathcal{O}\Sigma_{\mathfrak{u}}\mathcal{O}^*)^{-1}\big)\big) = 1$

Key message: If $\mu_\theta^{y,\mathrm{E}} = \mu_\theta^{y,\mathrm{A}}$ using $\mu_\theta^{y,\mathrm{A}}$ should be preferred

Example

Consider the IBVP

$$
\begin{aligned}
(\partial_t + \mathcal{L})w(x,t) &= s(x)\theta'(t) && (x,t) \in \mathcal{D} \times \mathcal{T} && (\mathrm{PDE} - \mathcal{M}^\dagger)\\
\nabla w(x,t) \cdot n(x) &= 0 && (x,t) \in \partial\mathcal{D} \times \mathcal{T} && (\mathrm{BC})\\
w(x,0) &= b^\dagger && (x,0) \in \mathcal{D} \times \{0\} && (\mathrm{IC})
\end{aligned}
$$

with

$$\mathcal{L} := -\kappa\Delta + v \cdot \nabla$$

for fixed $\kappa > 0$, $v : \mathcal{D} \to \mathbb{R}^2$, and fixed nonzero $s, b^\dagger : \mathcal{D} \to \mathbb{R}$, and $\theta' : \mathcal{T} \to \mathbb{R}$

Define $\mathcal{M}^\dagger$ as $\mathcal{M}^\dagger(\theta') := w(\theta')$

Define $\mathcal{M}(\theta')$ as a solution $w(\theta')$ of $(\mathrm{PDE} - \mathcal{M}^\dagger)$ with (BC) and IC:

$$w(x,0) = 0 \quad (x,0) \in \mathcal{D} \times \{0\}$$

Since $\delta^\dagger = \mathcal{M}^\dagger - \mathcal{M}$, $\delta^\dagger(\theta') := w$ for

$$(\partial_t + \mathcal{L})w(x,t) = 0 \qquad (x,t) \in \mathcal{D} \times \mathcal{T} \qquad (\text{PDE-}\delta^\dagger)$$

with (BC) and (IC)

Example

Define state space

$$\mathcal{U} := C^{2,1}(\mathcal{D} \times \mathcal{T}) \cap \{w \text{ satisfies (BC)}\},$$

where

$$C^{2,1}(\mathcal{D} \times \mathcal{T}) := \{w \mid w,\ \partial_t w,\ \partial_x^\beta w \in L^\infty,\ \text{for } 0 < |\beta| \le 2\}$$

Define observation operator $\mathcal{O} : \mathcal{U} \to \mathbb{R}^J$

$$\mathcal{O} : w \mapsto ((\partial_t + \mathcal{L})w(\hat{x}, t_j))_{j=1}^J$$

for some $\hat{x} \in \mathcal{D}$ and $t_j \in \mathcal{T}$

$\mathcal{O}$ is linear and continuous

It holds that

$$\{w \mid w \text{ satisfies (PDE-}\delta^\dagger)\} \subset \ker(\mathcal{O})$$

Therefore

$$\mu_\theta^{y,\mathrm{A}} = \mu_\theta^{y,\dagger}$$

Summary

- Analysed common approaches for Bayesian inference in the presence of model error
  - approximate posterior
  - enhanced noise posterior
- Derived positive and negative criteria for selection of observation operator

Drawbacks:

- Designing $\mathcal{O}$ is problem specific and challenging
- Good $\mathcal{O}$ with respect to $\delta^\dagger$ may be bad with respect to θ

Advantages:

- General results: nonlinear models, non-Gaussian priors, KL-divergence
- Mild assumptions: $L^1_{\mu_\theta}$ misfits

References

CVETKOVIĆ, N., LIE, H. C., BANSAL, H., AND VEROY–GREPL, K., *Choosing observation operators to mitigate model error in Bayesian inverse problems*, arXiv:2301.04863, 2023

ALEXANDERIAN, A., PETRA, N., STADLER, G., AND SUNSERI, I., *Optimal design of large-scale Bayesian linear inverse problems under reducible model uncertainty: Good to know what you don't know.*, SIAM-ASA J. Uncertain. Quantif., 2021

SPRUNGK, B., *On the local Lipschitz stability of Bayesian inverse problems*, Inverse Probl, 2020