State, global and local parameter estimation using local ensemble Kalman filters: applications to online machine learning of chaotic dynamics

Marc Bocquet, Quentin Malartic, Alban Farchi

and interaction with J. Brajard, A. Carrassi, L. Bertino, M. Bonavita, P. Laloyaux

CEREA, joint laboratory École des Ponts ParisTech and EDF R&D, France

Goals

From model error to the absence of a model

- At crossroads between:

  Data Assimilation (DA), Machine Learning (ML) and Dynamical Systems (DS)

- Goal: Estimate autonomous chaotic dynamics from partial and noisy observations

  $\longrightarrow$ Surrogate model

- Subgoal 1: Develop a Bayesian framework for this estimation problem. ✓
- Subgoal 2: Estimate and minimize the errors attached to the estimation. ✓
- Subgoal 3: What about more complex models? learning hybrid models. ✓
- Subgoal 4: What about online (sequential) learning?

  $\longrightarrow$ This talk [ensemble methods]!

- References connected to data-driven reconstruction of the dynamics in DA and ML:

[Park et al. 1994; Wang et al. 1998; Paduart et al. 2010; Lguensat et al. 2017; Pathak et al. 2017; Harlim 2018; Dueben et al. 2018; Long et al. 2018; Fablet et al. 2018; Vlachas et al. 2020; Brunton et al. 2016] and many more since the beginning of 2020.

Outline

ODE representation for the surrogate model

▶ Ordinary differential equations (ODEs) representation of the surrogate dynamics

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \boldsymbol{\phi}_{\mathbf{A}}(\mathbf{x}),$$

where $\mathbf{A}$ is a set of N_p coefficients.

▶ We need:

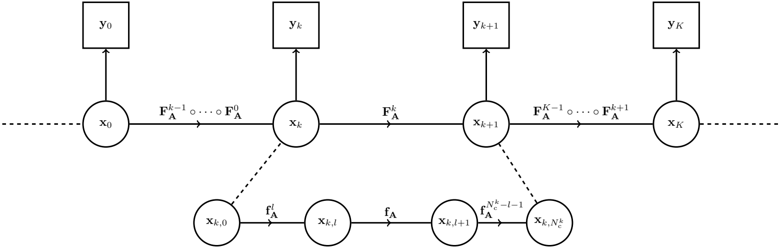
- to specify the tendency $\mathbf{x} \mapsto \boldsymbol{\phi}_{\mathbf{A}}(\mathbf{x})$,
- to choose a numerical scheme to integrate in time the tendency $\boldsymbol{\phi}_{\mathbf{A}}$ and be able to build resolvent of the surrogate dynamics $\mathbf{x}_{k+1} = \mathbf{F}_{\mathbf{A}}(\mathbf{x}_k)$.

▶ Going beyond, we wish to account for (surrogate) model error, so that the surrogate model representation is actually an SDE:

$$\mathrm{d}\mathbf{x} = \boldsymbol{\phi}_{\mathbf{A}}(\mathbf{x})\mathrm{d}t + \sqrt{\mathbf{Q}}\mathrm{d}\mathbf{W}(t),$$

with $\mathbf{W}(t)$ an N_x-dimensional Wiener process.

Integration scheme and cycling



- Choosing a Runge-Kutta method as integration scheme:

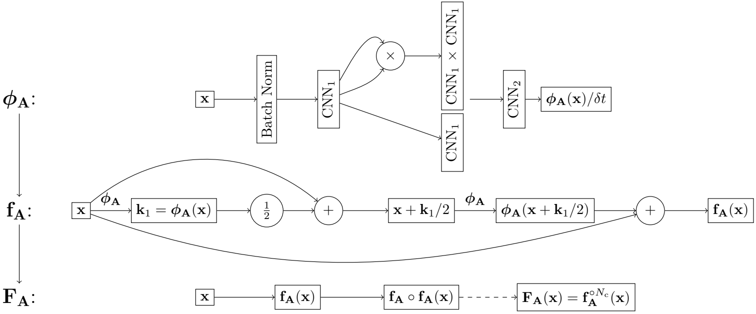
$$\mathbf{f_A}(\mathbf{x}) = \mathbf{x} + h \sum_{i=0}^{N_{\mathrm{RK}}-1} \beta_i \mathbf{k}_i, \qquad \mathbf{k}_i = \boldsymbol{\phi}_\mathbf{A}\left(\mathbf{x} + h \sum_{j=0}^{i-1} \alpha_{i,j} \mathbf{k}_j\right).$$

- Compositions of integration schemes:

$$\mathbf{x}_{k+1} = \mathbf{F}_\mathbf{A}^k(\mathbf{x}_k) \quad \text{where} \quad \mathbf{F}_\mathbf{A}^k \equiv \mathbf{f}_\mathbf{A}^{N_c^k} \equiv \underbrace{\mathbf{f_A} \circ \ldots \circ \mathbf{f_A}}_{N_c^k \text{ times}},$$

Neural network models

- We tested many simple architectures, all following the structure of N_c explicit Runge-Kutta schemes, with linear or nonlinear activation functions:
  - $\phi_{\mathbf{A}}$: minimal representation (as few parameters as possible), or based on a NN (with potentially many parameters) implemented in TensorFlow 2.x
  - Convolutional layers were used for local, homogeneous systems.
  - Locally connected convolutional layers were used for local, heterogeneous systems.



Outline

Bayesian analysis of the joint problem

- Bayesian view on state and model estimation:

$$p(\mathbf{A},\mathbf{Q}_{1:K},\mathbf{x}_{0:K}|\mathbf{y}_{0:K},\mathbf{R}_{0:K}) = \frac{p(\mathbf{y}_{0:K}|\mathbf{x}_{0:K},\mathbf{A},\mathbf{Q}_{1:K},\mathbf{R}_{0:K})\,p(\mathbf{x}_{0:K}|\mathbf{A},\mathbf{Q}_{1:K})\,p(\mathbf{A},\mathbf{Q}_{1:K})}{p(\mathbf{y}_{0:K},\mathbf{R}_{0:K})}.$$

- Data assimilation cost function assuming Gaussian errors and Markovian dynamics:

$$\mathcal{J}(\mathbf{A},\mathbf{x}_{0:K},\mathbf{Q}_{1:K}) = \frac{1}{2}\sum_{k=0}^{K}\left\{\|\mathbf{y}_k-\mathbf{H}_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}}+\ln|\mathbf{R}_k|\right\}$$
$$+\frac{1}{2}\sum_{k=1}^{K}\left\{\left\|\mathbf{x}_k-\mathbf{F}_{\mathbf{A}}^{k-1}(\mathbf{x}_{k-1})\right\|^2_{\mathbf{Q}_k^{-1}}+\ln|\mathbf{Q}_k|\right\}-\ln p(\mathbf{x}_0,\mathbf{A},\mathbf{Q}_{1:K}).$$

$\longrightarrow$ This is a (4D) variational problem.
$\longrightarrow$ Allows to rigorously handle partial and noisy observations.

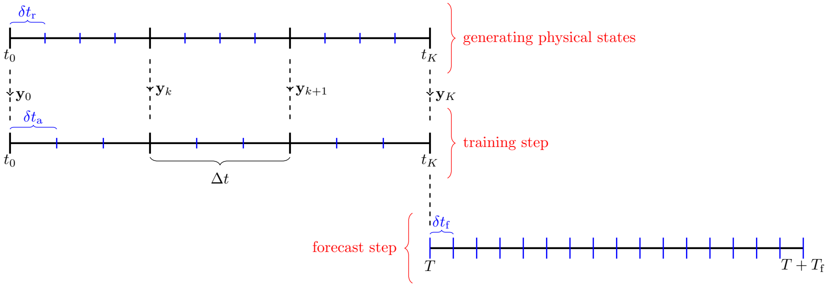
- Typical machine learning cost function with $\mathbf{H}_k=\mathbf{I}_k$ in the limit $\mathbf{R}_k\longrightarrow\mathbf{0}$:

$$\mathcal{J}(\mathbf{A})\approx\frac{1}{2}\sum_{k=1}^{K}\left\|\mathbf{y}_k-\mathbf{F}_{\mathbf{A}}^{k-1}(\mathbf{y}_{k-1})\right\|^2_{\mathbf{Q}_k^{-1}}-\ln p(\mathbf{y}_0,\mathbf{A}).$$

Similar outcome or improved upon [Hsieh et al. 1998; Abarbanel et al. 2018].

Experiment plan

- The reference model, the surrogate model and the forecasting system



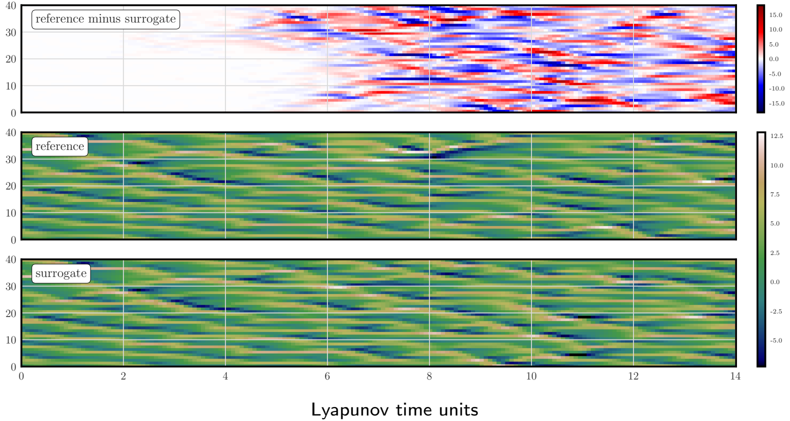
- Metrics of comparison:
  - Forecast skill [FS]: Normalized RMSE (NRMSE) between the reference and surrogate forecasts as a function of lead time (averaged over initial conditions).
  - Lyapunov spectrum [LS].
  - Power spectrum density [PSD].

Almost identifiable model and perfect observations

- Inferring the dynamics from dense & noiseless observations of a non-identifiable model

The Lorenz 96 model (40 variables). Surrogate model based on an RK2 scheme.

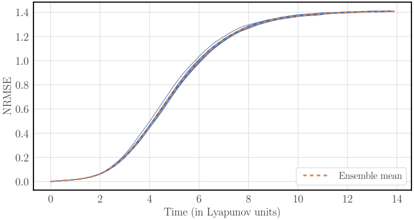
$$\frac{\mathrm{d}x_n}{\mathrm{d}t} = (x_{n+1} - x_{n-2})x_{n-1} - x_n + F,$$

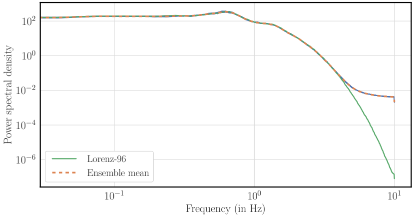


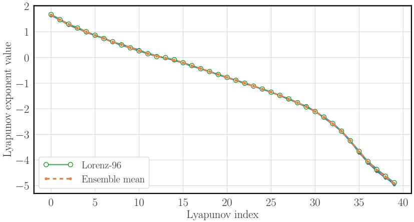
Application to the one-scale Lorenz-96 model

- Very good reconstruction of the long-term properties of the model (L96 model).

  - CNN+RK4
  - Approximate scheme
  - Fully observed
  - Significantly noisy observations $\mathbf{R} = \mathbf{I}$
  - Long window $K = 5000$, $\Delta t = 0.05$
  - EnKS with $L = 4$
  - 30 EM iterations





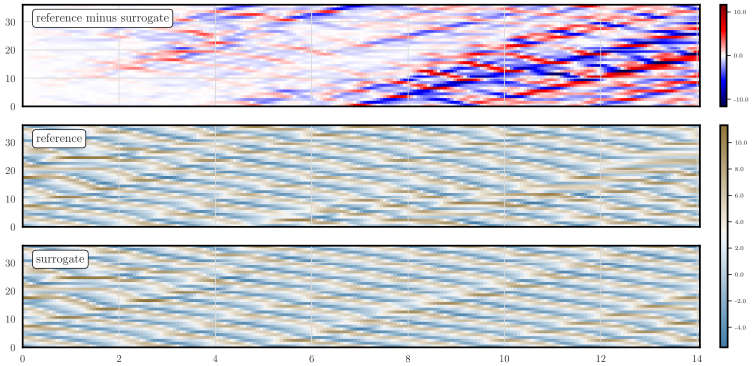


Non-identifiable model and imperfect observations

- The Lorenz 05III (two-scale) model (36 slow & 360 fast variables).

$$\frac{\mathrm{d}x_n}{\mathrm{d}t} = \psi_n^{+}(\mathbf{x}) + F - h\frac{c}{b}\sum_{m=0}^{9} u_{m+10n},$$

$$\frac{\mathrm{d}u_m}{\mathrm{d}t} = \frac{c}{b}\psi_m^{-}(b\mathbf{u}) + h\frac{c}{b}x_{m/10}, \quad \text{with} \quad \psi_n^{\pm}(\mathbf{x}) = x_{n\mp1}(x_{n\pm1} - x_{n\mp2}) - x_n,$$

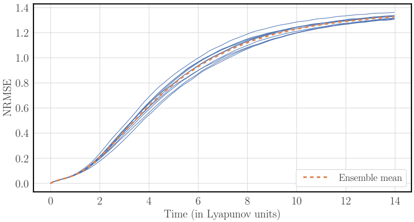


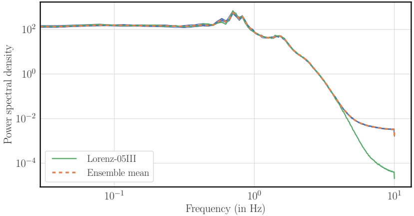
Lyapunov time units

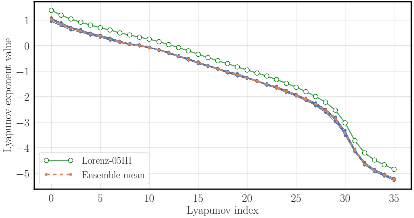
Application to the two-scale Lorenz-05III model

- Good reconstruction of the long-term properties of the model (L05III model).

  - CNN+RK4
  - Approximate scheme
  - Observation of the coarse modes only
  - Significantly noisy observations $\mathbf{R} = \mathbf{I}$
  - Long window $K = 5000$, $\Delta t = 0.05$
  - EnKS with $L = 4$
  - 30 EM iterations







Outline

Online learning scheme: Principle

▶ So far, learning was offline, i.e. based on variational technique using all available data. Can one design a sequential (online) scheme that progressively updates both the state and the model as data are collected?

▶ In the following, we make the assumptions:
(i) *autonomous* and *local* dynamics,
(ii) *homogeneous* dynamics or *heterogeneous* dynamics, or *mixed* dynamics.

▶ All parameters of the model are hereafter noted:

$$\mathbf{A} \longrightarrow \mathbf{p} \in \mathbb{R}^{N_\mathrm{p}} \text{ [Global parameters]}, \quad \mathbf{q} \in \mathbb{R}^{N_\mathrm{q}} \text{ [local parameters]}.$$

▶ Augmented state formalism [Jazwinski 1970; Ruiz et al. 2013]:

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \\ \mathbf{q} \end{bmatrix} \in \mathbb{R}^{N_\mathrm{z}}, \qquad \text{with} \quad N_\mathrm{z} = N_\mathrm{x} + N_\mathrm{p} + N_\mathrm{q}.$$

Just a (more ambitious) parameter estimation problem!?

Online learning scheme: The problem

- We use the augmented state formalism with ensemble Kalman filters (EnKFs):
  1. global deterministic EnKFs (EnSRF, ETKF),
  2. global iterative EnKF (IEnKF), key for nonlinearity.
  3. local EnKFs (LEnSRF, LETKF), key for scalability.

- Adequacy and inadequacy between the main LEnKF classes and the estimation of local and global parameters.

	Global parameters	Local parameters	Mixed set of parameters
LEnSRF	well suited	suited	unclear
CL	localization in parameter space?	numerically costly	solution proposed here
LETKF	only approximate (average)	well suited	unclear
DL	solution proposed here		solution proposed here

- We assume that the observbations are local whenever DL is used.

- Nonlocal observations require CL.

Online learning scheme: Notation

- Augmented ensemble matrix: $\mathbf{E} \in \mathbb{R}^{N_\mathrm{z} \times N_\mathrm{e}}$
- Ensemble means and anomalies:

$$\bar{\mathbf{z}} \triangleq \mathbf{E}\mathbf{1}/N_\mathrm{e},$$
$$\mathbf{X} \triangleq \left(\mathbf{E} - \bar{\mathbf{z}}\mathbf{1}^\top\right)/\sqrt{N_\mathrm{e}-1},$$

- Splitting state/global/local:

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_\mathrm{x} \\ \mathbf{E}_\mathrm{p} \\ \mathbf{E}_\mathrm{q} \end{bmatrix}, \quad \bar{\mathbf{z}} = \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{p}} \\ \bar{\mathbf{q}} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_\mathrm{x} \\ \mathbf{X}_\mathrm{p} \\ \mathbf{X}_\mathrm{q} \end{bmatrix}, \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_\mathrm{xx} & \mathbf{A}_\mathrm{xp} & \mathbf{A}_\mathrm{xq} \\ \mathbf{A}_\mathrm{px} & \mathbf{A}_\mathrm{pp} & \mathbf{A}_\mathrm{pq} \\ \mathbf{A}_\mathrm{qx} & \mathbf{A}_\mathrm{qp} & \mathbf{A}_\mathrm{qq} \end{bmatrix}.$$

- Observation operator (key assumption!):

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_\mathrm{x} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

Online learning scheme: EnKF two-step update

- The ensemble update (analysis) can be decomposed into a two-step scheme:
  1. Update the state part of the ensemble $\mathbf{E}_{\mathrm{x}}^{\mathrm{f}} \longrightarrow \mathbf{E}_{\mathrm{x}}^{\mathrm{a}}$ using an EnKF.
  2. Update the parameter part of the ensemble:

$$\mathbf{E}_{\mathrm{p}}^{\mathrm{a}} = \mathbf{E}_{\mathrm{p}}^{\mathrm{f}} + \mathbf{B}_{\mathrm{px}}\mathbf{B}_{\mathrm{xx}}^{-1}\left(\mathbf{E}_{\mathrm{x}}^{\mathrm{a}} - \mathbf{E}_{\mathrm{x}}^{\mathrm{f}}\right),$$

which can be computed:
(i) solving the linear system $\mathbf{B}_{\mathrm{xx}}\boldsymbol{\Delta} = \mathbf{E}_{\mathrm{x}}^{\mathrm{a}} - \mathbf{E}_{\mathrm{x}}^{\mathrm{f}}$, and
(ii) updating $\mathbf{E}_{\mathrm{p}}^{\mathrm{a}} = \mathbf{E}_{\mathrm{p}}^{\mathrm{f}} + \mathbf{B}_{\mathrm{px}}\boldsymbol{\Delta}$ (linear regression!).

- This can actually be proven for any statistical assumption provided the parameters are not directly observed. Should also remain valid for the update of local EnKFs.

[Bocquet et al. 2020b]

Covariance localization

The EnSRF-ML update (global parameters)

▶ Convenient reformulation of the EnSRF update in the observation space, follows [Andrews 1968; Whitaker et al. 2002; Bocquet 2016; Malartic et al. 2021]:

▶ Ancillary matrices:

$$\mathbf{T} \triangleq \mathbf{I} + \mathbf{R}^{-1/2}\mathbf{H}_\mathrm{x}\mathbf{B}_\mathrm{xx}\mathbf{H}_\mathrm{x}^\top \mathbf{R}^{-1/2} \in \mathbb{R}^{N_\mathrm{y}\times N_\mathrm{y}},$$

$$\mathbf{u}_\mathrm{x} \triangleq \mathbf{H}_\mathrm{x}^\top \mathbf{R}^{-1/2}\mathbf{T}^{-1}\mathbf{R}^{-1/2}\left(\mathbf{y} - \mathbf{H}_\mathrm{x}\bar{\mathbf{x}}^\mathrm{f}\right) \in \mathbb{R}^{N_\mathrm{x}},$$

$$\mathbf{U}_\mathrm{x} \triangleq -\mathbf{H}_\mathrm{x}^\top \mathbf{R}^{-1/2}\left(\mathbf{T} + \mathbf{T}^{1/2}\right)^{-1}\mathbf{R}^{-1/2}\mathbf{H}_\mathrm{x}\mathbf{X}_\mathrm{x}^\mathrm{f} \in \mathbb{R}^{N_\mathrm{x}\times N_\mathrm{e}}.$$

▶ Updates:

$$\Delta\bar{\mathbf{x}} = \mathbf{B}_\mathrm{xx}\mathbf{u}_\mathrm{x}, \qquad \Delta\mathbf{X}_\mathrm{x} = \mathbf{B}_\mathrm{xx}\mathbf{U}_\mathrm{x},$$

$$\Delta\bar{\mathbf{p}} = \mathbf{B}_\mathrm{px}\mathbf{u}_\mathrm{x} \qquad \Delta\mathbf{X}_\mathrm{p} = \mathbf{B}_\mathrm{px}\mathbf{U}_\mathrm{x}.$$

[Malartic et al. 2021]

The LEnSRF-ML update (global parameters)

- Covariance localization in the augmented space:

$$\mathbf{B}_{\mathrm{xx}} = \boldsymbol{\rho}_{\mathrm{xx}} \circ \left[\mathbf{X}_{\mathrm{x}}^{\mathrm{f}} \left(\mathbf{X}_{\mathrm{x}}^{\mathrm{f}} \right)^{\top} \right],$$

$$\mathbf{B}_{\mathrm{px}} = \boldsymbol{\rho}_{\mathrm{px}} \circ \left[\mathbf{X}_{\mathrm{p}}^{\mathrm{f}} \left(\mathbf{X}_{\mathrm{x}}^{\mathrm{f}} \right)^{\top} \right] = \mathbf{B}_{\mathrm{xp}}^{\top},$$

$$\mathbf{B}_{\mathrm{pp}} = \boldsymbol{\rho}_{\mathrm{pp}} \circ \left[\mathbf{X}_{\mathrm{p}}^{\mathrm{f}} \left(\mathbf{X}_{\mathrm{p}}^{\mathrm{f}} \right)^{\top} \right].$$

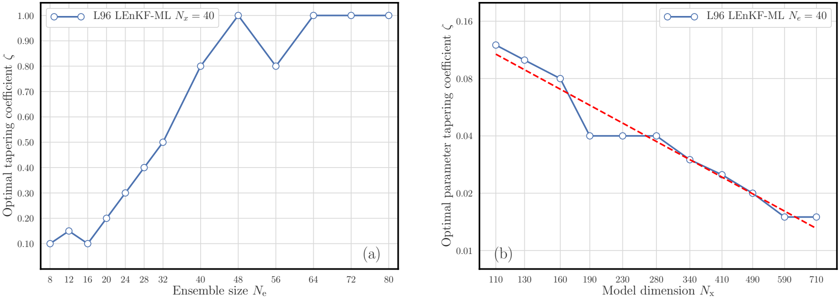
- The localization matrix $\boldsymbol{\rho}_{\mathrm{xx}}$ is the usual localization matrix and almost certainly makes $\mathbf{B}_{\mathrm{xx}}$ positive definite.

- The localization matrix $\boldsymbol{\rho}_{\mathrm{px}}$ has to be uniform because the parameters are global. The most natural choice is to use $\boldsymbol{\rho}_{\mathrm{px}} = \zeta \boldsymbol{\Pi}_{\mathrm{px}}$, where $\boldsymbol{\Pi}_{\mathrm{px}} \in \mathbb{R}^{N_{\mathrm{p}} \times N_{\mathrm{x}}}$ is the matrix full of ones, and where ζ is a tuning parameter [Ruckstuhl et al. 2018].

- For simplicity, we assume $\boldsymbol{\rho}_{\mathrm{px}} = \boldsymbol{\Pi}_{\mathrm{px}}$ and enforce the tapering coefficient ζ in the update:

$$\Delta \bar{\mathbf{p}} = \zeta \mathbf{B}_{\mathrm{px}} \mathbf{u}_{\mathrm{x}}, \qquad \Delta \mathbf{X}_{\mathrm{p}} = \zeta \mathbf{B}_{\mathrm{px}} \mathbf{U}_{\mathrm{x}}.$$

Numerics: Optimal tapering coefficient (global parameters)



- Mathematical constraint: **B** must be positive definite $\longrightarrow$ constraint on ζ.
- Optimal scaling of the tapering consistent with [Ruckstuhl et al. 2018; Bocquet et al. 2020a]:

$$\zeta < \sqrt{\frac{\lambda_{\min}}{N_x}}.$$

[Bocquet et al. 2020b]

The LEnSRF-HML update (mixed parameters)

- Updating both global and local parameters (hybrid updating). We simply add:

$$\Delta\bar{\mathbf{q}} = \mathbf{B}_{\mathrm{qx}}\mathbf{u}_{\mathrm{x}}, \qquad \Delta\mathbf{X}_{\mathrm{q}} = \mathbf{B}_{\mathrm{qx}}\mathbf{U}_{\mathrm{x}}.$$

- Without localization, there is no distinction between local and global parameters.
- With localization we define

$$\mathbf{B}_{\mathrm{qx}} = \boldsymbol{\rho}_{\mathrm{qx}} \circ \left[\mathbf{X}^{\mathrm{f}}_{\mathrm{q}}\left(\mathbf{X}^{\mathrm{f}}_{\mathrm{x}}\right)^{\top}\right] = \mathbf{B}^{\top}_{\mathrm{xq}},$$

$$\mathbf{B}_{\mathrm{qp}} = \boldsymbol{\rho}_{\mathrm{qp}} \circ \left[\mathbf{X}^{\mathrm{f}}_{\mathrm{q}}\left(\mathbf{X}^{\mathrm{f}}_{\mathrm{p}}\right)^{\top}\right] = \mathbf{B}^{\top}_{\mathrm{pq}},$$

$$\mathbf{B}_{\mathrm{pp}} = \boldsymbol{\rho}_{\mathrm{qq}} \circ \left[\mathbf{X}^{\mathrm{f}}_{\mathrm{q}}\left(\mathbf{X}^{\mathrm{f}}_{\mathrm{q}}\right)^{\top}\right].$$

- Neither $\boldsymbol{\rho}_{\mathrm{qp}}$, nor $\boldsymbol{\rho}_{\mathrm{qq}}$ have to be specified.

- As opposed to $\boldsymbol{\rho}_{\mathrm{px}}$, $\boldsymbol{\rho}_{\mathrm{qx}}$ has to reflect the geometry of the local parameters and state variables.

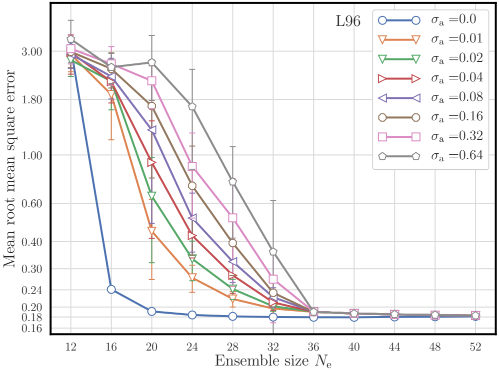
Augmented dynamics and the unstable/neutral subspace

The augmented dynamics

▶ Augmented dynamics (model persistence or Brownian motion):

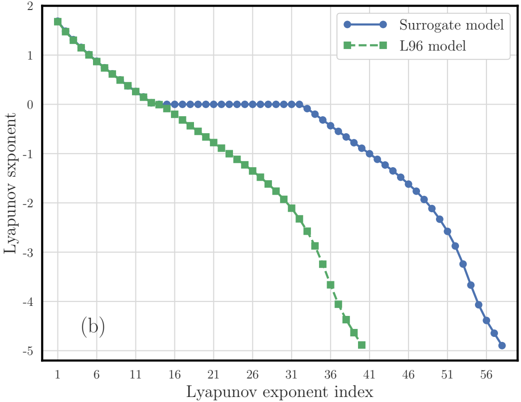
$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{p}_k \end{bmatrix} \mapsto \begin{bmatrix} \mathbf{F}^k(\mathbf{x}_k, \mathbf{p}_k) \\ \mathbf{p}_k \end{bmatrix}$$

▶ Assuming (i) N_0 is the dimension of the *unstable neutral subspace* of the reference dynamics, (ii) N_e is the size of the ensemble, then, in order for the augmented global EnKF (EnKF-ML) to be stable, we must have: $\boxed{N_\mathrm{e} \gtrapprox N_0 + N_\mathrm{p} + 1.}$



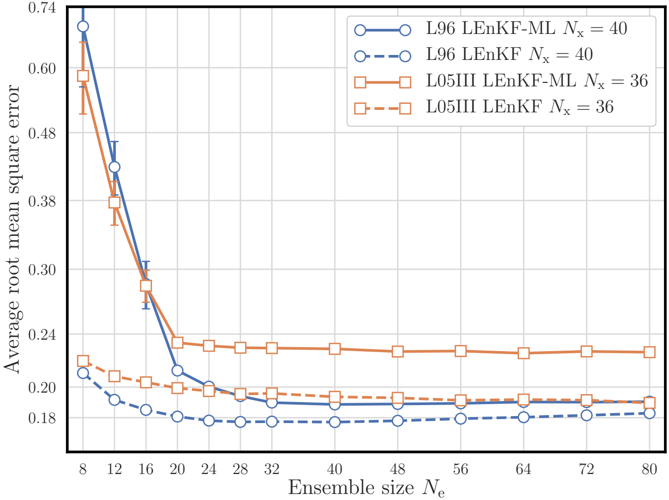
The augmented dynamics: Asymptotic properties

- Lyapunov spectra for the true and augmented L96.



Numerics (global parameters)

- LEnSRF and LEnSRF-ML applied to the L96 and L05III models.



Outline

Conclusions

The main results presented here are from [Bocquet et al. 2020a; Malartic et al. 2021], with preliminaries from [Bocquet et al. 2019]

- Main messages:
  - Bayesian DA view on state and model estimation.
    DA can address goals assigned to ML but with partial & noisy observations.
  - Online EnKFs-ML can also be used to sequentially estimate both state and model.
  - Rigorous ensemble solutions for joint state var./local par./global par. estimation.
  - Theoretical results (backed by numerics):

LEnKF\Parameters	Global	Local	Mixed
CL	clarified	existing	clarified
DL	new and improved	existing	new and improved

  - Successful on 1D low-order models (L96, L05III).

- Open questions and technical hardships (non-exhaustive):
  - Non-autonomous dynamics?
  - More complex models?
  - A 2D case with the mL96 model for radiances like-observations, that mixes CL and DL, local and global parameters and nonlocal observations is under test.

References I

[1] H. D. I. Abarbanel, P. J. Rozdeba, and S. Shirman. "Machine Learning: Deepest Learning as Statistical Data Assimilation Problems". In: *Neural Computation* 30 (2018), pp. 2025–2055.

[2] A. Aksoy, F. Zhang, and J. Nielsen-Gammon. "Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model". In: *Mon. Wea. Rev.* 134 (2006), pp. 2951–2969.

[3] A. Andrews. "A square root formulation of the Kalman covariance equations". In: *AIAA J.* 6 (1968), pp. 1165–1166.

[4] M. Bocquet. "Localization and the iterative ensemble Kalman smoother". In: *Q. J. R. Meteorol. Soc.* 142 (2016), pp. 1075–1089.

[5] M. Bocquet, A. Farchi, and Q. Malartic. "Online learning of both state and dynamics using ensemble Kalman filters". In: *Foundations of Data Science* 0 (2020). Accepted for publication, pp. 00–00.

[6] M. Bocquet and P. Sakov. "Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems". In: *Nonlin. Processes Geophys.* 19 (2012), pp. 383–399.

[7] M. Bocquet et al. "Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization". In: *Foundations of Data Science* 2 (2020), pp. 55–80.

[8] M. Bocquet et al. "Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models". In: *Nonlin. Processes Geophys.* 26 (2019), pp. 143–162.

[9] J. Brajard et al. "Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model". In: *J. Comput. Sci.* 44 (2020), p. 101171.

[10] J. Brajard et al. "Combining data assimilation and machine learning to infer unresolved scale parametrisation". In: *Phil. Trans. R. Soc. A* 379 (2021), p. 20200086.

[11] S. L. Brunton, J. L. Proctor, and J. N. Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *PNAS* (2016), p. 201517384.

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the Royal Statistical Society. Series B* (1977), pp. 1–38.

[13] P. D. Dueben and P. Bauer. "Challenges and design choices for global weather and climate models based on machine learning". In: *Geosci. Model Dev.* 11 (2018), pp. 3999–4009.

[14] R. Fablet, S. Ouala, and C. Herzet. "Bilinear residual neural network for the identification and forecasting of dynamical systems". In: *EUSIPCO 2018, European Signal Processing Conference*. Rome, Italy, 2018, pp. 1–5.

References II

[15] A. Farchi et al. "Using machine learning to correct model error in data assimilation and forecast applications". In: *Q. J. R. Meteorol. Soc.* 0 (2020). Submitted, p. 0.

[16] E. Fertig et al. "Observation bias correction with an ensemble Kalman filter". In: *Tellus A* 61 (2009), pp. 210–226.

[17] Z. Ghahramani and S. T. Roweis. "Learning nonlinear dynamical systems using an EM algorithm". In: *Advances in neural information processing systems*. 1999, pp. 431–437.

[18] J. Harlim. *Data-driven computational methods: parameter and operator estimations*. Cambridge University Press, Cambridge, 2018, p. 158.

[19] W. W. Hsieh and B. Tang. "Applying Neural Network Models to Prediction and Data Analysis in Meteorology and Oceanography". In: *Bull. Amer. Meteor. Soc.* 79 (1998), pp. 1855–1870.

[20] X.-M. Hu, F. Zhang, and J. W. Nielsen-Gammon. "Ensemble-based simultaneous state and parameter estimation for treatment of mesoscale model error: A real-data study". In: *Geophys. Res. Lett.* 37 (2010), p. L08802.

[21] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New-York, 1970, p. 376.

[22] Y. A. LeCun et al. "Efficient backprop". In: *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.

[23] R. Lguensat et al. "The Analog Data Assimilation". In: *Mon. Wea. Rev.* 145 (2017), pp. 4093–4107.

[24] Z. Long et al. "PDE-Net: Learning PDEs from Data". In: *Proceedings of the 35th International Conference on Machine Learning*. 2018.

[25] Q. Malartic, A. Farchi, and M. Bocquet. "Global and local parameter estimation using local ensemble Kalman filters: applications to online machine learning of chaotic dynamics". In: *SIAM/ASA J. Uncertainty Quantification* 0 (2021). Submitted, pp. 00–00.

[26] V. D. Nguyen et al. "EM-like Learning Chaotic Dynamics from Noisy and Partial Observations". In: *arXiv preprint arXiv:1903.10335* (2019).

[27] J. Paduart et al. "Identification of nonlinear systems using polynomial nonlinear state space models". In: *Automatica* 46 (2010), pp. 647–656.

[28] D. C. Park and Y. Zhu. "Bilinear recurrent neural network". In: *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*. Vol. 3. 1994, pp. 1459–1464.

[29] J. Pathak et al. "Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach". In: *Phys. Rev. Lett.* 120 (2018), p. 024102.

[30] J. Pathak et al. "Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data". In: *Chaos* 27 (2017), p. 121102.

References III

[31] M. Pulido et al. “Stochastic parameterization identification using ensemble Kalman filtering combined with maximum likelihood methods”. In: *Tellus A* 70 (2018), p. 1442099.

[32] P. N. Raanes, A. Carrassi, and L. Bertino. “Extending the square root method to account for additive forecast noise in ensemble methods”. In: *Mon. Wea. Rev.* 143 (2015), pp. 3857–38730.

[33] Y. M. Ruckstuhl and T. Janjić. “Parameter and state estimation with ensemble Kalman filter based algorithms for convective-scale applications”. In: *Q. J. R. Meteorol. Soc.* 144 (2018), pp. 826–841.

[34] J. J. Ruiz, M. Pulido, and T. Miyoshi. “Estimating model parameters with ensemble-based data assimilation: A Review”. In: *J. Meteorol. Soc. Japan* 91 (2013), pp. 79–99.

[35] P. Sakov, J.-M. Haussaire, and M. Bocquet. “An iterative ensemble Kalman filter in presence of additive model error”. In: *Q. J. R. Meteorol. Soc.* 144 (2018), pp. 1297–1309.

[36] P. Sakov, D. S. Oliver, and L. Bertino. “An iterative EnKF for strongly nonlinear systems”. In: *Mon. Wea. Rev.* 140 (2012), pp. 1988–2004.

[37] P. R. Vlachas et al. “Backpropagation algorithms and Reservoir Computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics”. In: *Neural Networks* 126 (2020), pp. 191–217.

[38] Y.-J. Wang and C.-T. Lin. “Runge-Kutta neural network for identification of dynamical systems in high accuracy”. In: *IEEE Transactions on Neural Networks* 9 (1998), pp. 294–307.

[39] G. C. G. Wei and M. A. Tanner. “A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms”. In: *Journal of the American Statistical Association* 85 (1990), pp. 699–704.

[40] J. S. Whitaker and T. M. Hamill. “Ensemble Data Assimilation without Perturbed Observations”. In: *Mon. Wea. Rev.* 130 (2002), pp. 1913–1924.

Domain localization

The ensemble transform Kalman filter (domain localization)

▶ Generic ETKF update (incremental)

$$\Delta\bar{\mathbf{z}} = \mathbf{X}^{\mathrm{f}}\mathbf{w}^{\mathrm{a}}, \qquad \Delta\mathbf{X} = \mathbf{X}^{\mathrm{f}}\left(\mathbf{T}^{-1/2} - \mathbf{I}\right).$$

with the definitions:

$$\mathbf{Y} \triangleq \mathbf{H}\mathbf{X}^{\mathrm{f}},$$
$$\mathbf{T} \triangleq \mathbf{I} + \mathbf{Y}^{\top}\mathbf{R}^{-1}\mathbf{Y},$$
$$\mathbf{w}^{\mathrm{a}} \triangleq \mathbf{T}^{-1}\mathbf{Y}^{\top}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}\bar{\mathbf{z}}^{\mathrm{f}}\right).$$

▶ Now, assume that the observations are local. The domain localization (DL) of the ETKF (LETKF) uses for each augmented state variable $n \in \{1, \ldots, N_{\mathrm{z}}\}$:

$$\mathbf{R}_n^{-1} \triangleq \boldsymbol{\rho}_n \circ \mathbf{R}^{-1},$$

where $\boldsymbol{\rho}_n \in \mathbb{R}^{N_{\mathrm{y}} \times N_{\mathrm{y}}}$ is the localization matrix in observation space for the n-th variable.

The LETKF-ML (global parameters)

- Update of the ETKF-ML

$$\Delta\bar{\mathbf{x}} = \mathbf{X}^{\mathrm{f}}_{\mathrm{x}}\mathbf{w}^{\mathrm{a}}, \qquad \Delta\mathbf{X}_{\mathrm{x}} = \mathbf{X}^{\mathrm{f}}_{\mathrm{x}}\left(\mathbf{T}^{-1/2} - \mathbf{I}\right),$$

$$\Delta\bar{\mathbf{p}} = \mathbf{X}^{\mathrm{f}}_{\mathrm{p}}\left(\mathbf{X}^{\mathrm{f}}_{\mathrm{x}}\right)^{\top}\mathbf{u}_{\mathrm{x}}, \qquad \Delta\mathbf{X}_{\mathrm{p}} = \mathbf{X}^{\mathrm{f}}_{\mathrm{p}}\left(\mathbf{X}^{\mathrm{f}}_{\mathrm{x}}\right)^{\top}\mathbf{U}_{\mathrm{x}},$$

with the definition (on the right/local version):

$$\mathbf{u}_{\mathrm{x}} = \mathbf{H}_{\mathrm{x}}^{\top}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}_{\mathrm{x}}\bar{\mathbf{x}}^{\mathrm{f}} - \mathbf{Y}\mathbf{w}^{\mathrm{a}}\right),$$

$$\mathbf{U}_{\mathrm{x}} = -\mathbf{H}_{\mathrm{x}}^{\top}\mathbf{R}^{-1}\mathbf{Y}\left(\mathbf{T} + \mathbf{T}^{1/2}\right)^{-1}.$$

- This global parameter update is fully consistent with the DL framework.

This is an improvement over the [Aksoy et al. 2006; Fertig et al. 2009; Hu et al. 2010] average semi-empirical technique!

- With the tapering coefficient:

$$\Delta\bar{\mathbf{p}} = \zeta\mathbf{X}^{\mathrm{f}}_{\mathrm{p}}\left(\mathbf{X}^{\mathrm{f}}_{\mathrm{x}}\right)^{\top}\mathbf{u}_{\mathrm{x}}, \qquad \Delta\mathbf{X}_{\mathrm{p}} = \zeta\mathbf{X}^{\mathrm{f}}_{\mathrm{p}}\left(\mathbf{X}^{\mathrm{f}}_{\mathrm{x}}\right)^{\top}\mathbf{U}_{\mathrm{x}},$$

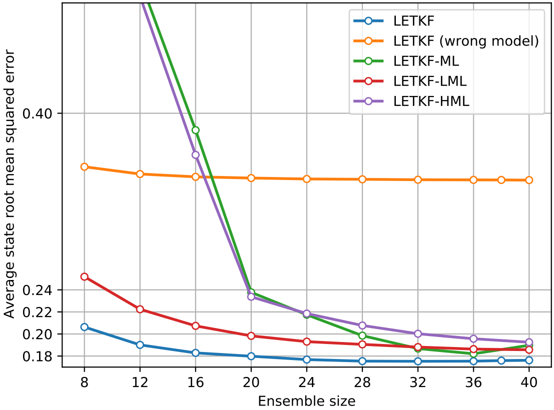
The LETKF-HML (mixed parameters)

▶ To account for local parameters, a very simple addition to the LETKF-ML scheme:

$$\Delta\bar{\mathbf{q}} = \mathbf{X}_{\mathrm{q}}^{\mathrm{f}}\mathbf{w}^{\mathrm{a}},$$
$$\Delta\mathbf{X}_{\mathrm{q}} = \mathbf{X}_{\mathrm{q}}^{\mathrm{f}}\left(\mathbf{T}^{-1/2} - \mathbf{I}\right).$$

Numerics (assorted LETKFs-ML)

- L96 model where the forcing is inhomogeneous: $F = 8 + \sin(2\pi n / N_x)$



- LETKF (wrong model): known dynamics except for $F = 8$.
- LETK-ML: unknown dynamics but known forcing
- LETKF-LML: known dynamics but unknown forcing
- LETKF-HML: unknown dynamics, unknown forcing