Combining data assimilation and machine learning

Marc Bocquet[1], Alban Farchi[1], Quentin Malartic[1,2]
Julien Brajard[3,4], Alberto Carrassi[5,6], Laurent Bertino[3]
Massimo Bonavita[7], Patrick Laloyaux[7]

(1) CEREA, École des Ponts and EDF R&D, Île-de-France, France,

(2) LMD/IPSL, ENS, PSL Universite, École Polytechnique, Institut Polytechnique de Paris, Sorbonne Université, CNRS, Paris, France

(3) Nansen Environmental and Remote Sensing Center, Bergen, Norway

(4) Sorbonne University, CNRS-IRD-MNHN, LOCEAN, Paris, France

(5) Department of meteorology, University of Reading, United Kingdom

(6) Mathematical institute, University of Utrecht, The Netherlands

(7) ECMWF, Reading, United Kingdom

Monday, 27 September 2021

43rd EWGLAM and 28th SRNWP Meeting

Outline

From model error to the absence of a model

▶ Data assimilation and model error

Numerical predictions in geophysics based on data assimilation crucially depends on both initial condition and model error [Magnusson et al. 2013]. Mitigation of model error:

- additive stochastic noise (e.g., [Trémolet 2006; Raanes et al. 2015; Sakov et al. 2018])
- estimation of uncertain model parameters (e.g., [Bocquet 2012])
- physically-driven stochastic perturbations (e.g., [Buizza et al. 1999]), stochastic subgrid parameterizations (e.g, [Resseguier et al. 2017]), inflation (e.g., [Raanes et al. 2019])

▶ Data-driven forecast of a physical system [resolvent-based]

One step further: renounce physically-based models and use massive observation

- use data assimilation together with analogues [Lguensat et al. 2017]
- use diffusion maps for a spectral representation of datasets [Harlim 2018]
- use neural networks (NNs), echo states networks, & deep learning [Park et al. 1994; Pathak et al. 2017; Dueben et al. 2018; Vlachas et al. 2020; Bonavita et al. 2020; Arcomano et al. 2020] to represent the resolvent.

▶ Learning the dynamics of a model from its output [tendencies-based]

- more explicit (possibly with NNs) representations of the dynamics using specific regressors e.g., [Paduart et al. 2010; Brunton et al. 2016].
- design NNs that mimic integration schemes [Wang et al. 1998; Fablet et al. 2018; Long et al. 2018]

Objectives

▶ Goal: Estimate chaotic dynamics from partial and noisy observations
$\longrightarrow$ Surrogate model

▶ Unfortunately, basic machine learning requires full, noiseless observations!

▶ But data assimilation techniques naturally account for imperfect observation!

▶ Subgoal 1: Develop a Bayesian framework for this estimation problem.
Estimate and minimize the errors attached to the estimation.

▶ But this surely is an under-determined, hardly scalable problem!

▶ Subgoal 2: What about hybridizing a physical model with a trainable model?

[Bocquet et al. 2019; Brajard et al. 2020; Bocquet et al. 2020a; Brajard et al. 2021; Farchi et al. 2021b; Wikner et al. 2021; Tomizawa et al. 2021].

Objectives

▶ However, data assimilation is sequential as we want to exploit the latest observations. But learning a surrogate model is by essence an offline optimisation problem!

▶ Subgoal 4: What about online (i.e., sequential) learning?

▶ Which data assimilation approach can we use for this task?

▶ Subgoal 4a: What about online learning with variational methods?

▶ Subgoal 4b: What about online learning with ensemble methods?

[Bocquet et al. 2019; Brajard et al. 2020; Bocquet et al. 2020a; Brajard et al. 2021; Farchi et al. 2021b; Malartic et al. 2021].

▶ At crossroads between:
Data Assimilation (DA), Machine Learning (ML) and Dynamical Systems (DS)

Outline

Traditional Bayesian approach to data assimilation

▶ Bayesian justification of the weak-constraint 4D-Var

Application of Bayes' rule over a time window $[t_0, t_K]$ with batches of observations $\mathbf{y}_k$ at each time step t_k. Define $\mathbf{x}_{0:K} = \mathbf{x}_0, \dots, \mathbf{x}_K$ and $\mathbf{y}_{0:K} = \mathbf{y}_0, \dots, \mathbf{y}_K$.
The most general conditional pdf of interest is $p(\mathbf{x}_{0:K}|\mathbf{y}_{0:K})$ and reads:

$$p(\mathbf{x}_{0:K}|\mathbf{y}_{0:K}) \propto p(\mathbf{y}_{0:K}|\mathbf{x}_{0:K})p(\mathbf{x}_{0:K}).$$

Assuming that the observation errors are Gaussian and uncorrelated in time, with error covariance matrices $\mathbf{R}_0, \dots, \mathbf{R}_K$, so that:

$$p(\mathbf{y}_{0:K}|\mathbf{x}_{0:K}) = \prod_{k=0}^{K} p(\mathbf{y}_k|\mathbf{x}_k) \propto \exp\left(-\frac{1}{2}\sum_{k=0}^{K} \|\mathbf{y}_k - H_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}}\right).$$

Next, we assume that the prior pdf $p(\mathbf{x}_{0:K})$ is Markovian, i.e. the state $\mathbf{x}_k$ conditional on the previous state $\mathbf{x}_{k-1}$ does not depend on all other previous past states:

$$p(\mathbf{x}_{0:K}) = p(\mathbf{x}_0)\prod_{k=1}^{K} p(\mathbf{x}_k|\mathbf{x}_{0:k-1}) = p(\mathbf{x}_0)\prod_{k=1}^{K} p(\mathbf{x}_k|\mathbf{x}_{k-1}).$$

Traditional Bayesian approach to data assimilation

▶ Bayesian justification of the weak-constraint 4D-Var

Now, we assume Gaussian statistics for the model error which are uncorrelated in time, with zero bias and error covariance matrices $\mathbf{Q}_1, \dots, \mathbf{Q}_K$ so that:

$$p(\mathbf{x}_{0:K}) \propto p(\mathbf{x}_0) \exp\left(-\frac{1}{2} \sum_{k=1}^{K} \|\mathbf{x}_k - M_k(\mathbf{x}_{k-1})\|^2_{\mathbf{Q}_k^{-1}} \right).$$

We can assemble the likelihood and prior pieces to obtain the cost function associated to the conditional pdf $p(\mathbf{x}_{0:K}|\mathbf{y}_{0:K})$:

$$\mathcal{J}(\mathbf{x}_{0:K}) = -\ln p(\mathbf{x}_{0:K}|\mathbf{y}_{0:K}) \tag{1}$$

$$= -\ln p(\mathbf{x}_0) + \frac{1}{2} \sum_{k=0}^{K} \|\mathbf{y}_k - H_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \frac{1}{2} \sum_{k=1}^{K} \|\mathbf{x}_k - M_k(\mathbf{x}_{k-1})\|^2_{\mathbf{Q}_k^{-1}} \tag{2}$$

Unsurprisingly, this is the cost function of the weak-constraint 4D-Var. The associated statistical assumptions explicitly assume that the model is flawed.

Towards learning complex model error

- Bayesian justification of the weak-constraint 4D-Var

With this type of weak-constraint 4D, one believes that the model can be corrected with some stochastic noise to be added to the state vector.

- More general model error

Instead of considering a known model $\mathbf{x}_k = M_k(\mathbf{x}_{k-1})$, one could assume a parametric form of the model $\mathbf{x}_k = M_k(\mathbf{p}, \mathbf{x}_{k-1})$, that depends on unknow time-independent parameters $\mathbf{p}$.

Bayesian inference of state trajectory and model

▶ Bayesian analysis with model parameters

We can piggyback on the previous Bayesian analysis, but now adding the model parameter vector **p**:

$$p(\mathbf{x}_{0:K},\mathbf{p}|\mathbf{y}_{0:K}) \propto p(\mathbf{y}_{0:K}|\mathbf{x}_{0:K},\mathbf{p})p(\mathbf{x}_{0:K},\mathbf{p}) \propto p(\mathbf{y}_{0:K}|\mathbf{x}_{0:K},\mathbf{p})p(\mathbf{x}_{0:K}|\mathbf{p})p(\mathbf{p}),$$

which requires to introduce a prior pdf $p(\mathbf{p})$ on the parameters. In the language of Bayesian statistics, this is called a hierarchical decomposition of the conditional pdf. As a consequence, the cost function for the state and model parameters problem is

$$\begin{aligned}\mathcal{J}(\mathbf{x}_{0:K},\mathbf{p}) &= -\ln p(\mathbf{x}_{0:K},\mathbf{p}|\mathbf{y}_{0:K}) \\ &= -\ln p(\mathbf{x}_0) + \frac{1}{2}\sum_{k=0}^{K}\|\mathbf{y}_k - H_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \frac{1}{2}\sum_{k=1}^{K}\|\mathbf{x}_k - M_k(\mathbf{p},\mathbf{x}_{k-1})\|^2_{\mathbf{Q}_k^{-1}} \\ &\quad - \ln p(\mathbf{p}).\end{aligned}$$

This cost function is again similar to the weak-constraint 4D-var, but (i) **p** is now part of the control variables, and (ii) there is a background term on **p** that may or may not play a role depending on the importance of the data set.

[Hsieh et al. 1998; Abarbanel et al. 2018; Bocquet et al. 2019]

Connecting data assimilation and machine learning

▶ Discussion

We note that, to be effective, a data assimilation analysis based on this cost function would require not only the gradient of the cost function with respect to the whole state trajectory, i.e. $\nabla_{\mathbf{x}_{0:K}} \mathcal{J}$, but also the gradient of the cost function with respect to the model parameters, i.e. $\nabla_{\mathbf{p}} \mathcal{J}$.
$\longrightarrow$ Need for the adjoint with respect to the model parameters!

▶ Machine learning limit

This (Bayesian) data assimilation standpoint on the problem of estimating the model (together with the state trajectory) is remarkable as it allows for noisy and partial observations on the physical system, as in traditional data assimilation. Classical and simple machine learning approach of the problem would rather use a dataset which is a complete observation of the physical system with minimal noise, using a simple least-square loss function.

Connecting data assimilation and machine learning

▶ Machine learning limit

Let us assume that the physical system is fully and directly observed, i.e. $\mathbf{H}_k \equiv \mathbf{I}$, and that the observation errors tend to zero, i.e. $\mathbf{R}_k \to \mathbf{0}$. Then the observation term in the cost function is completely frozen and imposes that $\mathbf{x}_k \simeq \mathbf{y}_k$, so that, in this limit, $\mathcal{J}(\mathbf{x}_{0:K}, \mathbf{p})$ becomes

$$\mathcal{J}(\mathbf{p}) = \frac{1}{2}\sum_{k=0}^{K} \left\| \mathbf{y}_k - M_k\left(\mathbf{p}, \mathbf{y}_{k-1}\right) \right\|^2_{\mathbf{Q}_k^{-1}} - \ln p(\mathbf{p}).$$

This coincides with the tyical machine learning loss function with $\mathbf{Q}_k \equiv \mathbf{I}$.

[Bocquet et al. 2019; Bocquet et al. 2020a]

Data assimilation and machine learning unification: Summary

- Bayesian view on state and model estimation:

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

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

$$\begin{aligned}\mathcal{J}(\mathbf{p},\mathbf{x}_{0:K},\mathbf{Q}_{1:K}) =& \frac{1}{2}\sum_{k=0}^{K}\left\{\|\mathbf{y}_k - H_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \ln|\mathbf{R}_k|\right\} \\ &+ \frac{1}{2}\sum_{k=1}^{K}\left\{\|\mathbf{x}_k - \mathbf{M}_k(\mathbf{p},\mathbf{x}_{k-1})\|^2_{\mathbf{Q}_k^{-1}} + \ln|\mathbf{Q}_k|\right\} \\ &- \ln p(\mathbf{x}_0,\mathbf{p},\mathbf{Q}_{1:K}).\end{aligned}$$

$\longrightarrow$ Allows to rigorously handle partial and noisy observations.

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

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

Bayesian analysis of the joint problem: Assuming $\mathbf{Q}_{1:K}$ is known

▶ If the $\mathbf{Q}_{1:K}$ are known, we look for minima of

$$\mathcal{J}(\mathbf{p}, \mathbf{x}_{0:K}|\mathbf{Q}_{1:K}) = -\ln p(\mathbf{p}, \mathbf{x}_{0:K}|\mathbf{y}_{0:K}, \mathbf{R}_{0:K}, \mathbf{Q}_{1:K}).$$

▶ Numerical solution through optimization

(1) $\mathcal{J}(\mathbf{p}, \mathbf{x}_{0:K}|\mathbf{Q}_{1:K})$ can be optimized using a full variational approach:

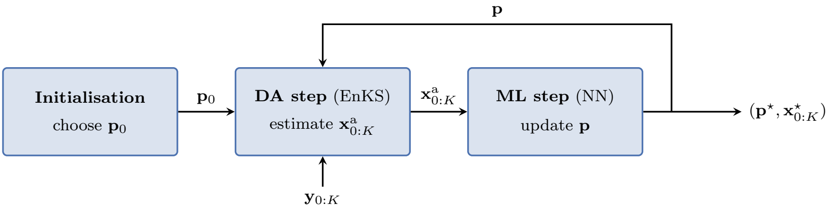
▶ In [Bocquet et al. 2019], $\mathcal{J}(\mathbf{p}, \mathbf{x}_{0:K}|\mathbf{Q}_{1:K})$ is minimized using a full weak-constraint 4D-Var where both $\mathbf{x}_{0:K}$ and $\mathbf{p}$ are control variables.

Bayesian analysis of the joint problem: Assuming $\mathbf{Q}_{1:K}$ is known

(2) $\mathcal{J}(\mathbf{p}, \mathbf{x}_{0:K}|\mathbf{Q}_{1:K})$ is minimized using a coordinate descent:

- using a weak constraint 4D-Var for $\mathbf{x}_{0:K}$ and a variational subproblem for $\mathbf{p}$ [Bocquet et al. 2019].
- using a (higher-dimensional) strong constraint 4D-Var for $\mathbf{x}_{0:K}$ and a variational subproblem for $\mathbf{p}$ [Bocquet et al. 2019].
- using an EnKF/EnKS for $\mathbf{x}_{0:K}$ and a variational subproblem for $\mathbf{p}$ [Brajard et al. 2020; Bocquet et al. 2020a].

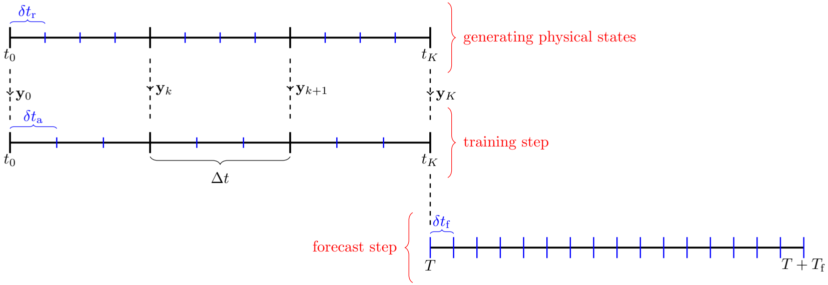
$\longrightarrow$ Combine data assimilation and machine learning techniques in a coordinate descent



Outline

Experiment plan

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



- Metrics of comparison:
  - Model: ODE coefficients norm $\|\mathbf{p}_\mathrm{a} - \mathbf{p}_\mathrm{r}\|_\infty$.
  - Forecast skill [FS]: Normalized RMSE (NRMSE) between the reference and the surrogate forecasts as a function of the lead time (averaged over many initial conditions).
  - Lyapunov spectrum [LS].
  - Power spectrum density [PSD].

Identifiable model and perfect observations

▶ Inferring the dynamics from dense & noiseless observations of identifiable models

- The Lorenz 63 model (L63, 3 variables):

$$\frac{\mathrm{d}x_0}{\mathrm{d}t} = \sigma(x_1 - x_0),$$
$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = \rho x_0 - x_1 - x_0 x_2,$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = \rho x_0 x_1 - \beta x_2,$$

$\longrightarrow \|\mathbf{p}_\mathrm{a} - \mathbf{p}_\mathrm{r}\|_\infty \sim 10^{-13}$ close to perfect reconstruction at machine precision.

- The Lorenz 96 model (L96, 40 variables)

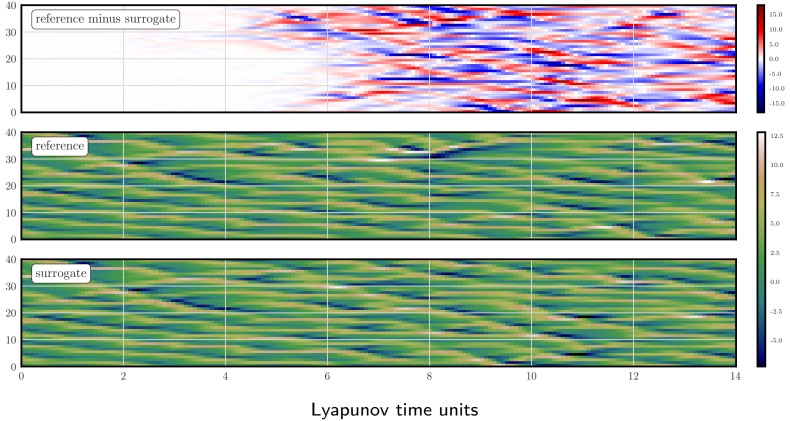
$$\frac{\mathrm{d}x_n}{\mathrm{d}t} = (x_{n+1} - x_{n-2})x_{n-1} - x_n + F,$$

$\longrightarrow \|\mathbf{p}_\mathrm{a} - \mathbf{p}_\mathrm{r}\|_\infty \sim 10^{-13}$ close to perfect reconstruction at machine precision.

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. Analysis of the modeling depth as a function of N_{c}.

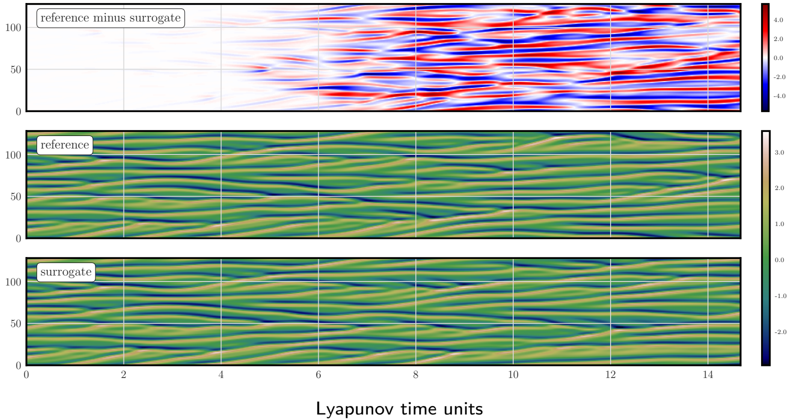


Un-identifiable model and perfect observations

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

The Kuramoto-Sivashinski (KS) model (continuous, 128 variables).

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4},$$

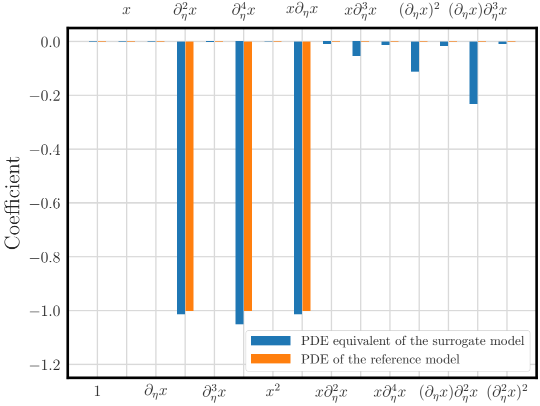


Un-identifiable model and perfect observations

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

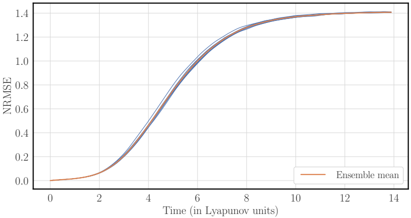
The Kuramoto-Sivashinski (KS) model (continuous, 128 variables).

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4},$$

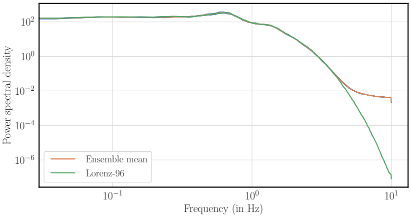


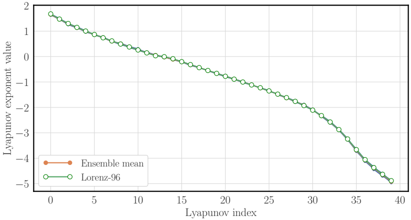
Almost identifiable model and imperfect observations

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



- 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



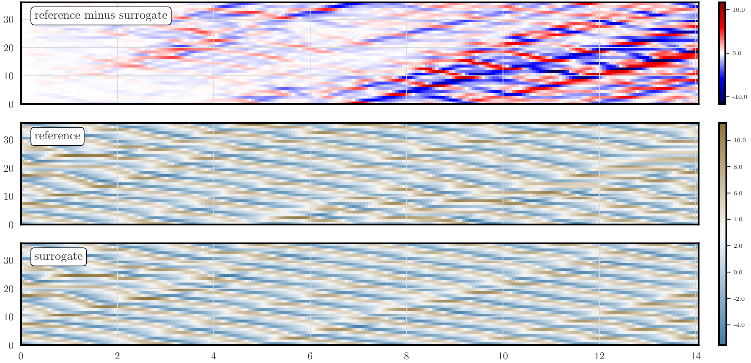


Non-identifiable model and imperfect observations

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

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

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

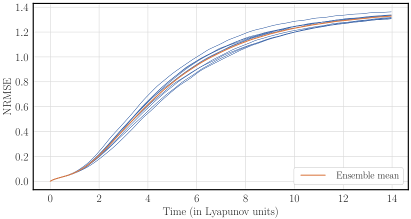


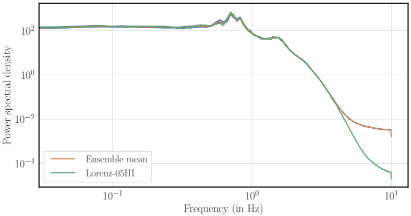
Lyapunov time units

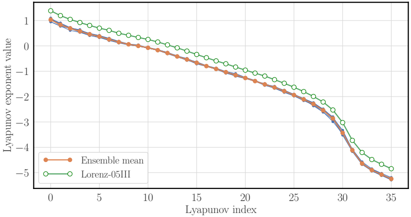
Non-identifiable model and imperfect observations

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

  - 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

Machine learning for model error correction

- We want to use this method to correct the error of a physical model Φ_k.
- In the cost function, we replace $M_k(\mathbf{p}, \mathbf{x}_k)$ with the hybrid model:

$$M_k(\mathbf{p}, \mathbf{x}_{k-1}) \longrightarrow \Phi_k(\mathbf{x}_{k-1}) + M_k(\mathbf{p}, \mathbf{x}_{k-1}).$$

- If the true trajectory $\mathbf{x}_k^{\mathrm{t}}$ is known (dense, noiseless observations), then the NN would be trained with

$$\mathbf{x}_k^{\mathrm{t}} \mapsto \mathbf{x}_{k+1}^{\mathrm{t}} - \Phi_{k+1}(\mathbf{x}_k^{\mathrm{t}}).$$

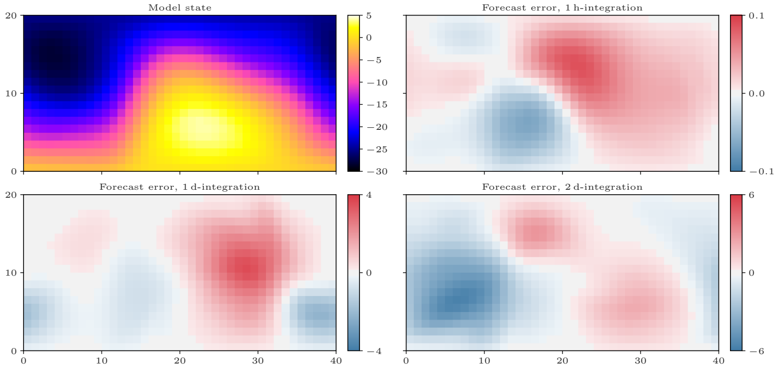
- With sparse and noisy observations, we need to use:
  - the analysis $\mathbf{x}_k^{\mathrm{a}}$ in place of $\mathbf{x}_k^{\mathrm{t}}$;
  - the analysis increment $\mathbf{x}_{k+1}^{\mathrm{a}} - \Phi_{k+1}(\mathbf{x}_k^{\mathrm{a}})$ in place of $\mathbf{x}_{k+1}^{\mathrm{t}} - \Phi_{k+1}(\mathbf{x}_k^{\mathrm{t}})$.

- This corresponds to the first iteration of the coordinate descent!

[Brajard et al. 2021; Farchi et al. 2021b]

Application to the OOPS QG model

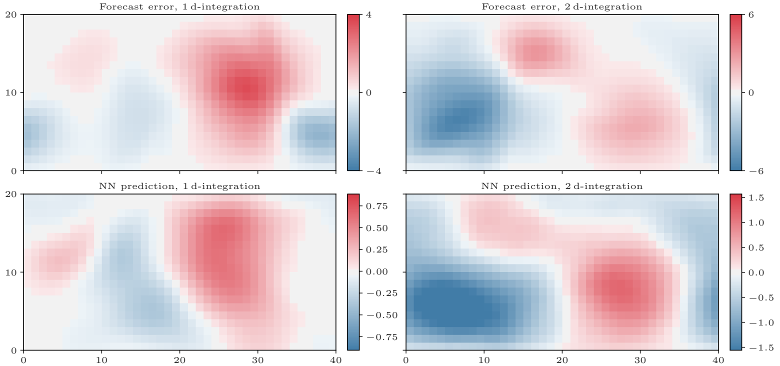
- The method is to be validated using the QG model implemented in OOPS.
- Model error is introduced as perturbed parameters, layer depths and orography, and doubled integration time step.



Stream function of the QG model in the bottom layers. Forecast error of the perturbed model.

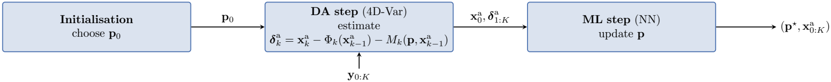
The NN training

- A long cycled 4D-Var experiment is performed with the perturbed QG model.
- Its analysis increments are used to train small NNs.
- Depending on the sampling frequency of the ML step, the NNs are able to explain 80 % to 90 % of the analysis increments variance, but only 30 % to 85 % of the model error variance.

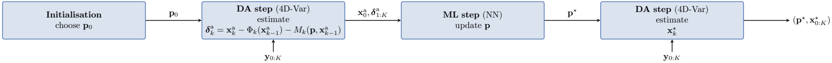


Corrected data assimilation

- One-iteration approximation of the coordinate descent:



- We want to evaluate the potential improvements from the correction in a subsequent 4D-Var experiment.



- We assume a linear error growth in time in the second DA step.
- The model error prediction for a $\delta t = 20\,\mathrm{min}$ forecast (one integration time step) is 1/72 of the model error prediction for a 1 day forecast (one DA window).
- The correction yield a $25\,\%$ reduction in the analysis RMSE.

[Farchi et al. 2021b]

Outline

Online model error correction

- So far, the model error has been learned *offline*: the ML (or training) step first requires a long analysis trajectory.
- We now investigate the possibility to make *online* learning, *i.e.* improving the correction as new observations become available.
- To do this, we use the formalism of DA to estimate both the state and the NN parameters (SC-4D-Var + param. est. $\sim$ WC-4D-Var):

$$\mathcal{J}(\mathbf{p},\mathbf{x}) = \frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right\|^2_{\mathbf{B}_{\mathrm{x}}^{-1}} + \frac{1}{2}\left\|\mathbf{p}-\mathbf{p}^{\mathrm{b}}\right\|^2_{\mathbf{B}_{\mathrm{p}}^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\left\|\mathbf{y}_k - H_k \circ \mathcal{M}_{k:0}(\mathbf{p},\mathbf{x})\right\|^2_{\mathbf{R}_k^{-1}}$$

- Information is flowing from one window to the next using the prior for the state $\mathbf{x}^{\mathrm{b}}$ and for the NN parameters $\mathbf{p}^{\mathrm{b}}$.

- This is very similar to classical *parameter estimation* in DA!
- This has been already investigated in an EnKF+ML context [Bocquet et al. 2020a; Malartic et al. 2021], but with scalablity constraints on the ensemble size.

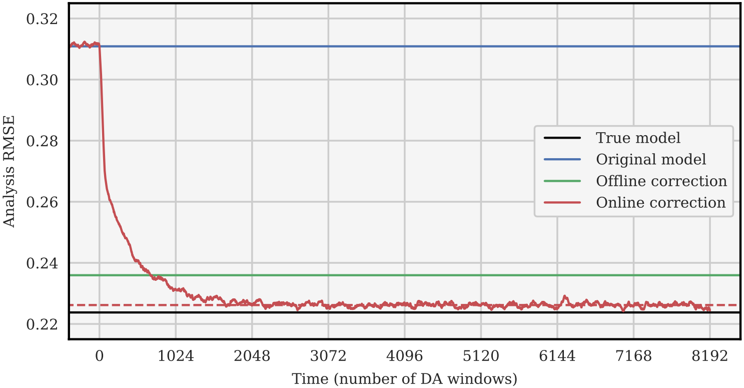
Online or offline model error correction: numerical comparison

- Again with the 2-scale Lorenz model (L05-III).
- We use the *tendency correction approach*; it does not require the assumption of linear growth of errors.

- We start the experiment by using the (non-corrected) physical model Φ_k.
- At some point, we switch on the online correction.
- Starting from a large value, we progressively decrease the parameter background error covariance matrix $\mathbf{B}_\mathrm{p}$ as the model improves.

[Farchi et al. 2021a]

Online or offline model error correction: numerical comparison



- The online correction steadily improves the model.
- At some point, the online correction *gets more accurate* than the offline correction.
- Eventually, the improvement saturates. The analysis error is similar to that obtained with the true model!

[Farchi et al. 2021a]

Conclusions

▶ Main messages:

- Unification of data assimilation and machine learning within a Bayesian framework (familiar to the DA community)
- Surrogate models/model error can theoretically be learned with partial & noisy observations.
- Tested with L63, L96, L05-III, KS, 2-layer OOPS QG model.
- Hybrid models with a known physical part should be considered for realistic high-dimensional systems, with or without a known adjoint, learning tendencies or resolvents.
- Online estimation of the state and surrogate model/model error has a lot of potential. Next generation (WC-)4D-Var?

All results presented here are from [Bocquet et al. 2019; Brajard et al. 2020; Bocquet et al. 2020a; Brajard et al. 2021; Farchi et al. 2021b; Bocquet et al. 2020b; Farchi et al. 2021a; Malartic et al. 2021].

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. DOI: 10.1162/neco_a_01094.

[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. DOI: 10.1175/MWR3224.1.

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

[4] T. Arcomano et al. "A Machine Learning-Based Global Atmospheric Forecast Model". In: *Geophys. Res. Lett.* 47 (2020), e2020GL087776. DOI: 10.1029/2020GL087776.

[5] M. Bocquet. "Localization and the iterative ensemble Kalman smoother". In: *Q. J. R. Meteorol. Soc.* 142 (2016), pp. 1075–1089. DOI: 10.1002/qj.2711.

[6] M. Bocquet. "Parameter field estimation for atmospheric dispersion: Application to the Chernobyl accident using 4D-Var". In: *Q. J. R. Meteorol. Soc.* 138 (2012), pp. 664–681. DOI: 10.1002/qj.961.

[7] M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino. "Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization". In: *Foundations of Data Science* 2 (2020), pp. 55–80. DOI: 10.3934/fods.2020004.

[8] M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino. "Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models". In: *Nonlin. Processes Geophys.* 26 (2019), pp. 143–162. DOI: 10.5194/npg-26-143-2019.

[9] 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. DOI: 10.3934/fods.2020015.

[10] M. Bonavita and P. Laloyaux. "Machine Learning for Model Error Inference and Correction". In: *J. Adv. Model. Earth Syst.* 12 (2020), e2020MS002232. DOI: 10.1029/2020MS002232.

[11] J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. "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. DOI: 10.1016/j.jocs.2020.101171.

[12] J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. "Combining data assimilation and machine learning to infer unresolved scale parametrisation". In: *Phil. Trans. R. Soc. A* 379 (2021), p. 20200086. DOI: 10.1098/rsta.2020.0086.

[13] 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. DOI: 10.1073/pnas.1517384113.

References II

[14] R. Buizza, M. Miller, and T. N. Palmer. "Stochastic representation of model uncertainties in the ECMWF ensemble prediction system". In: *Q. J. R. Meteorol. Soc.* 125 (1999), pp. 2887–2908. DOI: 10.1002/qj.49712556006.

[15] 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. DOI: 10.5194/gmd-11-3999-2018.

[16] 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. URL: https://hal.archives-ouvertes.fr/hal-01686766.

[17] A. Farchi, M. Bocquet, P. Laloyaux, M. Bonavita, and Q. Malartic. "A comparison of combined data assimilation and machine learning methods for offline and online model error correction". In: *J. Comput. Sci.* (2021). Submitted. URL: https://arxiv.org/pdf/2107.11114.pdf.

[18] A. Farchi, P. Laloyaux, M. Bonavita, and M. Bocquet. "Using machine learning to correct model error in data assimilation and forecast applications". In: *Q. J. R. Meteorol. Soc.* 147 (2021), pp. 3067–3084. DOI: 10.1002/qj.4116.

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

[20] 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. DOI: 10.1175/1520-0477(1998)079<1855:ANNMTP>2.0.CO;2.

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

[22] R. Lguensat, P Tandeo, P Ailliot, M. Pulido, and R. Fablet. "The Analog Data Assimilation". In: *Mon. Wea. Rev.* 145 (2017), pp. 4093–4107. DOI: 10.1175/MWR-D-16-0441.1.

[23] Z. Long, Y. Lu, X. Ma, and B. Dong. "PDE-Net: Learning PDEs from Data". In: *Proceedings of the 35th International Conference on Machine Learning*. 2018.

[24] L. Magnusson and E. Källén. "Factors influencing skill improvements in the ECMWF forecasting system". In: *Mon. Wea. Rev.* 141 (2013), pp. 3142–3153. DOI: 10.1175/MWR-D-12-00318.1.

[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. URL: https://arxiv.org/pdf/2006.03859.pdf.

[26] H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. "Dual state–parameter estimation of hydrological models using ensemble Kalman filter". In: *Advances in Water Resources* 28 (2005), pp. 135–147. DOI: 10.1016/j.advwatres.2004.09.002.

References III

[27] J. Paduart et al. “Identification of nonlinear systems using polynomial nonlinear state space models”. In: *Automatica* 46 (2010), pp. 647–656. DOI: 10.1016/j.automatica.2010.01.001.

[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, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott. “Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data”. In: *Chaos* 27 (2017), p. 121102. DOI: 10.1063/1.5010300.

[30] P. N. Raanes, M. Bocquet, and A. Carrassi. “Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures”. In: *Q. J. R. Meteorol. Soc.* 145 (2019), pp. 53–75. DOI: 10.1002/qj.3386.

[31] 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. DOI: 10.1175/MWR-D-14-00375.1.

[32] V. Resseguier, E. Mémin, and B. Chapron. “Geophysical flows under location uncertainty, Part I Random transport and general models”. In: *Geophysical & Astrophysical Fluid Dynamics* 111 (2017), pp. 149–176. DOI: 10.1080/03091929.2017.1310210.

[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. DOI: 10.1002/qj.3257.

[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. DOI: doi:10.2151/jmsj.2013-201.

[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. DOI: 10.1002/qj.3213.

[36] F. Tomizawa and Y. Sawada. “Combining Ensemble Kalman Filter and Reservoir Computing to predict spatio-temporal chaotic systems from imperfect observations and models”. In: *Geosci. Model Dev. Discuss.* 2020 (2021), pp. 1–33. DOI: 10.5194/gmd-2020-211. URL: https://gmd.copernicus.org/preprints/gmd-2020-211/.

[37] Y. Trémolet. “Accounting for an imperfect model in 4D-Var”. In: *Q. J. R. Meteorol. Soc.* 132 (2006), pp. 2483–2504. DOI: 10.1256/qj.05.224.

[38] 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. DOI: 10.1016/j.neunet.2020.02.016.

[39] 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. DOI: 10.1109/72.661124.

References IV

[40] J. S. Whitaker and T. M. Hamill. “Ensemble Data Assimilation without Perturbed Observations”. In: *Mon. Wea. Rev.* 130 (2002), pp. 1913–1924. DOI: 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

[41] A. Wikner et al. “Using data assimilation to train a hybrid forecast system that combines machine-learning and knowledge-based components”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31 (2021), p. 053114. DOI: 10.1063/5.0048050.