

Stochastic surrogate models, data assimilation, and uncertainty quantification

Marc Bocquet,*

Alban Farchi,^{†,*} Tobias Finn,* Charlotte Durand,* Sibö Cheng,* Wenbo Yu,*
Yumeng Chen,^{||} Ivo Pasmans,^{†,||} Alberto Carrasi[§]

* CEREA, ENPC, EDF R&D, Institut Polytechnique de Paris, Île-de-France, France

^{||} Department of Meteorology and National Centre for Earth Observation, University of Reading, United-Kingdom

[§] Department of Physics and Astronomy, University of Bologna, Italy

[†] ECMWF, Reading, United Kingdom



- 1 Machine learning for data assimilation: blitz literature overview & classification
 - Learning machine emulators, model error through data assimilation
 - Machine learning-aided data assimilation
 - End-to-end approaches to data assimilation
- 2 Learning data assimilation from artificial intelligence: first results
 - Sequential data assimilation for chaotic dynamics
 - Learning data assimilation
 - Investigation and interpretation
- 3 Learning data assimilation from artificial intelligence: advanced results
 - Scalability and locality
 - Learning nonlinearity
- 4 Conclusions

Outline

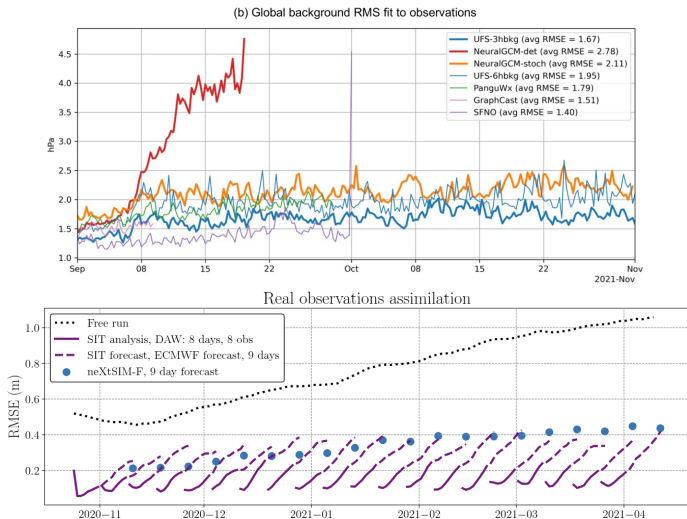
- 1 Machine learning for data assimilation: blitz literature overview & classification
 - Learning machine emulators, model error through data assimilation
 - Machine learning-aided data assimilation
 - End-to-end approaches to data assimilation
- 2 Learning data assimilation from artificial intelligence: first results
 - Sequential data assimilation for chaotic dynamics
 - Learning data assimilation
 - Investigation and interpretation
- 3 Learning data assimilation from artificial intelligence: advanced results
 - Scalability and locality
 - Learning nonlinearity
- 4 Conclusions

Outline

- 1 Machine learning for data assimilation: blitz literature overview & classification
 - Learning machine emulators, model error through data assimilation
 - Machine learning-aided data assimilation
 - End-to-end approaches to data assimilation
- 2 Learning data assimilation from artificial intelligence: first results
 - Sequential data assimilation for chaotic dynamics
 - Learning data assimilation
 - Investigation and interpretation
- 3 Learning data assimilation from artificial intelligence: advanced results
 - Scalability and locality
 - Learning nonlinearity
- 4 Conclusions

Using ML emulators into classical data assimilation cycle

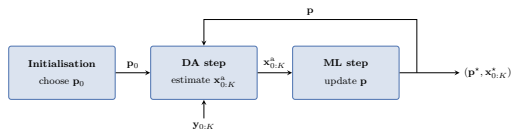
Hatfield et al. 2021; Chattopadhyay et al. 2022; Li et al. 2024; Slivinski et al. 2025; Durand et al. 2025



Learning/correcting dynamics through data assimilation

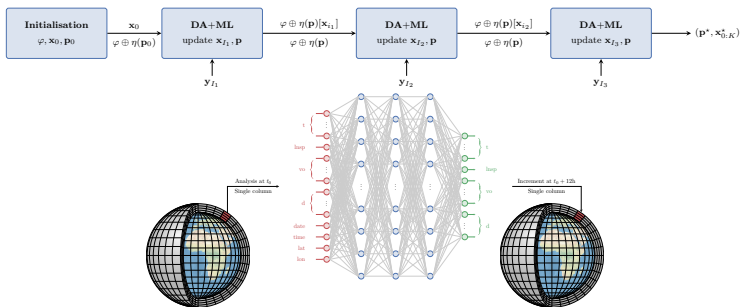
► Learning dynamics through Bayesian data assimilation

Hsieh et al. 1998; Abarbanel et al. 2018; Bocquet et al. 2019; Brajard et al. 2020; Bocquet et al. 2020; Brajard et al. 2021; Farchi et al. 2021b; G. Revach et al. 2022.



► Learning model error within hybrid models through data assimilation

Brajard et al. 2021; Farchi et al. 2021b; Farchi et al. 2021a; Legler et al. 2022; Farchi et al. 2023; Farchi et al. 2025.

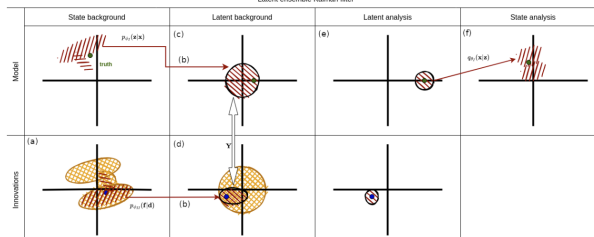
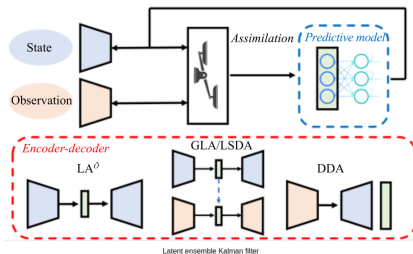


Outline

- 1 Machine learning for data assimilation: blitz literature overview & classification
 - Learning machine emulators, model error through data assimilation
 - **Machine learning-aided data assimilation**
 - End-to-end approaches to data assimilation
- 2 Learning data assimilation from artificial intelligence: first results
 - Sequential data assimilation for chaotic dynamics
 - Learning data assimilation
 - Investigation and interpretation
- 3 Learning data assimilation from artificial intelligence: advanced results
 - Scalability and locality
 - Learning nonlinearity
- 4 Conclusions

Machine learning-induced latent space data assimilation

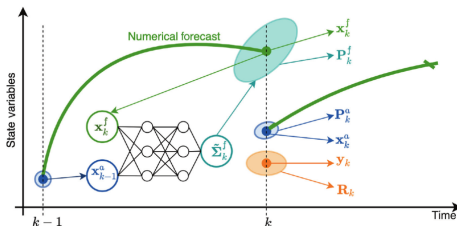
Haarnoja et al. 2016; Becker et al. 2019; Mack et al. 2020; Peyron et al. 2021; Penny et al. 2022; Cheng et al. 2022; Cheng et al. 2023a; Melinc et al. 2024; Pasmans et al. 2025



Machine learning-assisted data assimilation

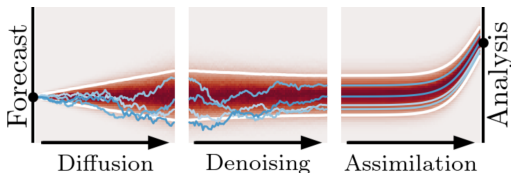
► For uncertainty quantification

Grooms 2021; Yang et al. 2021; Sacco et al. 2022; Sacco et al. 2024; Lu 2025.



► Using generative artificial intelligence (diffusion, flow matching)

Chung et al. 2023; Rozet et al. 2023; Silva et al. 2023; Finn et al. 2024; Rozet et al. 2024; Huang et al. 2024.



► For physical balances, tuning hybrid data assimilation systems

Ruckstuhl et al. 2021; Dong et al. 2023.

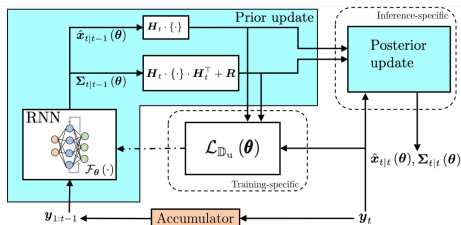
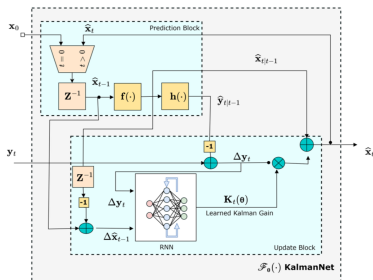
Parametric learning of data assimilation

► Learning the (static) gain of the (ensemble) Kalman filter

H. Hoang et al. 1994; S. Hoang et al. 1998; Luk et al. 2024.

► Tuning (hyper-)parameters through auto-differentiable (ensemble) Kalman filters

Haarnoja et al. 2016; Chen et al. 2022; Luk et al. 2024; Shlezinger et al. 2024.



► Signal processing literature: Kalman filter/low dimensional oriented – still very impressive!

Coskun et al. 2017; Shlezinger et al. 2023; Garcia Satorras et al. 2019; Klushyn et al. 2021; Pratik et al. 2021; Gedon et al. 2021; Shlezinger et al. 2022; Guy Revach et al. 2022; Cheng et al. 2023b; Choi et al. 2023; A. et al. 2024; Buchnik et al. 2024; Imbiriba et al. 2024; Ghosh et al. 2024

Outline

- 1 Machine learning for data assimilation: blitz literature overview & classification
 - Learning machine emulators, model error through data assimilation
 - Machine learning-aided data assimilation
 - End-to-end approaches to data assimilation
- 2 Learning data assimilation from artificial intelligence: first results
 - Sequential data assimilation for chaotic dynamics
 - Learning data assimilation
 - Investigation and interpretation
- 3 Learning data assimilation from artificial intelligence: advanced results
 - Scalability and locality
 - Learning nonlinearity
- 4 Conclusions

Supervised learning of the analysis from data assimilation runs

R. S. Cintra et al. 2012; F. P. Härter et al. 2008; T. P. Härter et al. 2012; R. Cintra et al. 2016; R. S. Cintra et al. 2018; Maddy et al. 2024; Xu et al. 2025

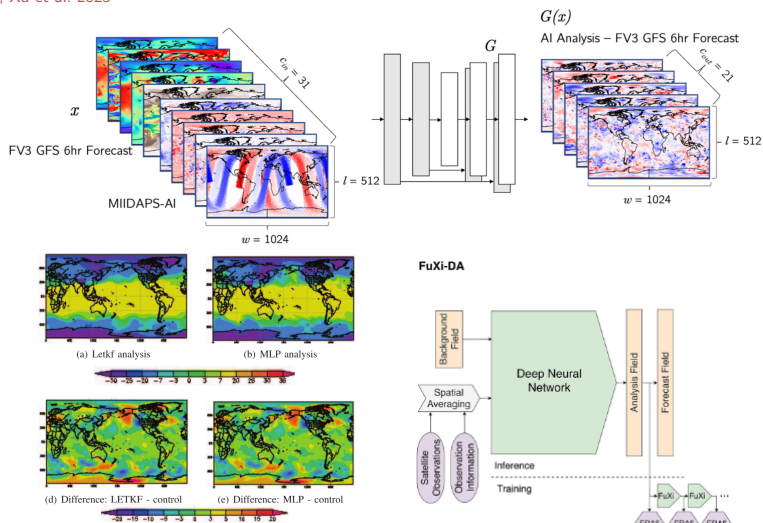
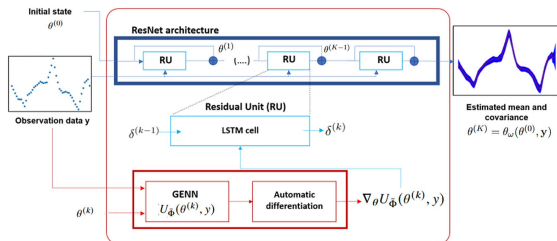
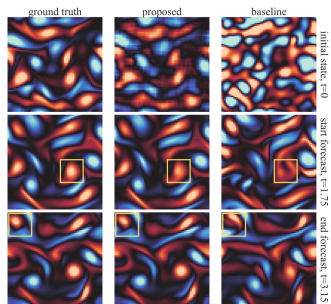


Fig. 6. Temperature ($^{\circ}\text{C}$) Fields at layer 500 hPa on 08/01/2004 at 12 UTC. (a) LETKF analysis (b) MLP-DA analysis (d) differences between LETKF and MLP-DA analyses.

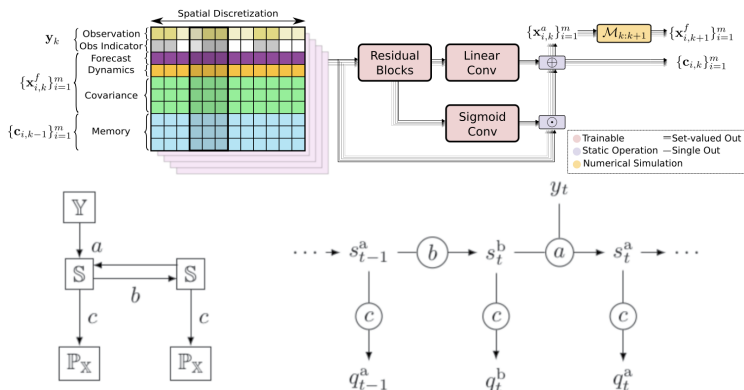
Learning variational solvers

Fablet et al. 2021; Frerix et al. 2021; Lafon et al. 2023; Filoche et al. 2023; Keller et al. 2024



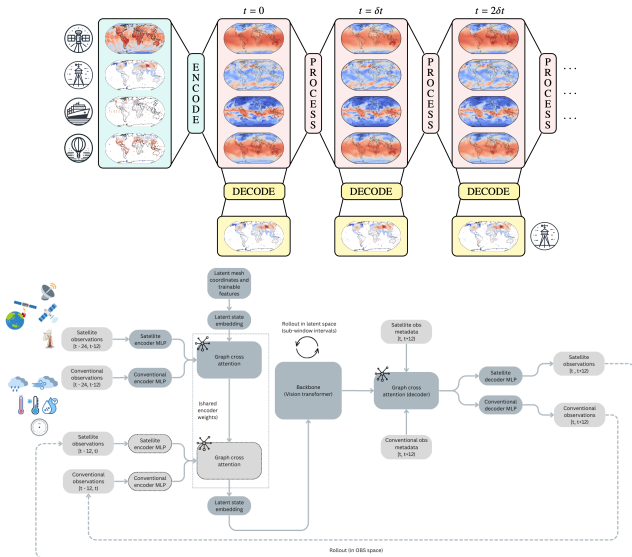
Learning sequential data assimilation through the cycles

McCabe et al. 2021; Boudier et al. 2023; Bocquet et al. 2024; Ghosh et al. 2024



Bypassing data assimilation: end-to-end observation-driven NWP

McNally et al. 2024; Vaughan et al. 2024; Sun et al. 2024; Alexe et al. 2024; Allen et al. 2025; Lean et al. 2025



Outline

- 1 Machine learning for data assimilation: blitz literature overview & classification
 - Learning machine emulators, model error through data assimilation
 - Machine learning-aided data assimilation
 - End-to-end approaches to data assimilation
- 2 Learning data assimilation from artificial intelligence: first results
 - Sequential data assimilation for chaotic dynamics
 - Learning data assimilation
 - Investigation and interpretation
- 3 Learning data assimilation from artificial intelligence: advanced results
 - Scalability and locality
 - Learning nonlinearity
- 4 Conclusions

Sequential data assimilation for chaotic dynamics

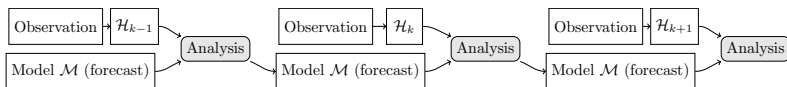
- Here, data assimilation (DA) methods are formulated from

$$\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k), \quad (1a)$$

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \varepsilon_k, \quad \varepsilon_k \sim N(\mathbf{0}, \mathbf{R}_k), \quad (1b)$$

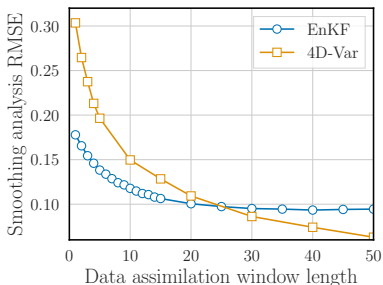
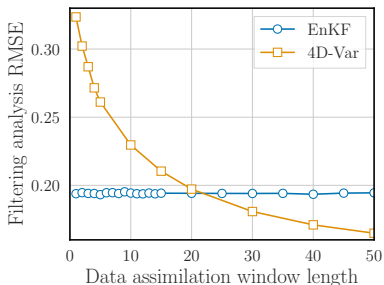
where \mathcal{M} is the *autonomous* evolution model, \mathbf{x}_k is the state vector at time τ_k , \mathbf{y}_k is the observation vector, \mathcal{H}_k is the observation operator, ε_k is the observation error, assumed to be additive, unbiased, white in time, and Gaussian of covariance matrix \mathbf{R}_k .

- DA for geofluids has to be *sequential* in time because (i) observations need to be assimilated *as they arrive* to update the state estimation, (ii) applied to *chaotic dynamics*, typical errors have an exponential growth.



The edge of ensemble filtering methods

- ▶ **The variational methods (3D-Var, 4D-Var):** can handle nonlinearity of the operators, asynchronous observations, but *cannot handle the errors of the day*.
- ▶ **The ensemble filtering methods (EnKFs):** can only handle weak nonlinearity of the operators, cannot handle asynchronous observations, *can handle the errors of the day* through the ensemble.
- ▶ Testing the EnKF ($N_e = 20$), 4D-Var, and IEnKS ($N_e = 20$) variants with the chaotic 40-variable Lorenz 96 model [Bocquet et al. 2013; Asch et al. 2016]:



- ▶ In mild nonlinear regime, the EnKF significantly outperforms the (basic) 4D-Var with moderately large DA windows because it captures the *errors of the day*.

Our focus: learning the analysis

► Let us assume that \mathcal{M} is known, that the Jacobian of \mathcal{H}_k is \mathbf{H}_k , and that we wish to learn an *incremental analysis operator* a_θ , typically a neural network parametrised by θ .

► If $\mathbf{E}_k^a, \mathbf{E}_k^f \in \mathbb{R}^{N_x \times N_e}$ are the analysis and forecast ensemble matrices at time τ_k , a_θ is defined via the (ensemble) update:

$$\mathbf{E}_k^a = \mathbf{E}_k^f + a_\theta \left(\mathbf{E}_k^f, \mathbf{H}_k^\top \mathbf{R}_k^{-1} \delta_k \right), \quad (2a)$$

where δ_k , the innovation at time τ_k , is defined by

$$\delta_k \triangleq \mathbf{y}_k - \mathcal{H}_k(\bar{\mathbf{x}}_k^f), \quad \bar{\mathbf{x}}_k^f \triangleq \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_k^{f,i}. \quad (2b)$$

→ Notice our trick: $a_\theta(\mathbf{E}_k^f, \delta_k) \rightarrow a_\theta(\mathbf{E}_k^f, \mathbf{H}_k^\top \mathbf{R}_k^{-1} \delta_k)$, i.e., *uplift of observation information in state space*.

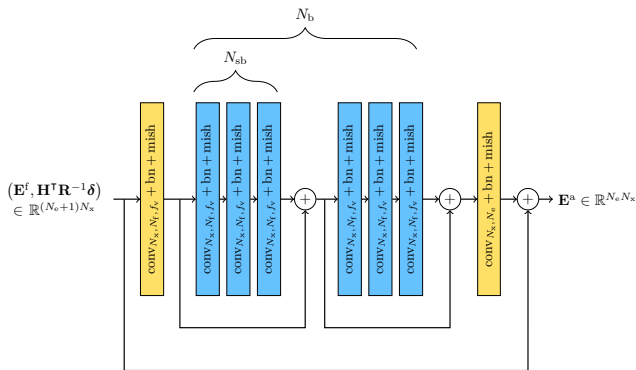
► The DA forecast step propagates the analysis ensemble, member-wise:

$$\mathbf{E}_{k+1}^f = \mathcal{M}(\mathbf{E}_k^a). \quad (3)$$

► The a_θ -based sequential DA will be called *DAN* in the following.

Neural network architecture

- We choose a_θ to have a simple residual convolutional neural network (CNN) architecture.



Architecture of the residual convolutional network, where $N_b = 2$, $N_{sb} = 3$. $\text{conv}_{N_1, N_2, f}$ is a generic one-dimensional convolutional layer of dimension N_1 , with N_2 filters of kernel size f .

Training scheme

- ▶ Full end-to-end, models and DA [Allen et al. 2025; Alexe et al. 2024; Lean et al. 2025]: **not our objective here!**
- ▶ Literature (focused on *sequential data assimilation*):
 - ▶ Learning the analysis of sequential DA is not new [T. P. Härter et al. 2012; R. S. Cintra et al. 2018], though barely explored.
 - ▶ Learning key components of the analysis in the (En)KF [H. Hoang et al. 1994; S. Hoang et al. 1998] possibly leveraging auto-differentiable structure [Haarnoja et al. 2016; Chen et al. 2022; Luk et al. 2024] was also investigated.
 - ▶ Only two key papers so far focused on a *non-parametrised* analysis using backpropagation *through the DA cycles*: [McCabe et al. 2021; Boudier et al. 2023].
- ▶ Our training loss (supervised learning):¹

We consider N_r , N_c cycle-long ensemble DA runs, based on N_r independent concurrent trajectories of the dynamics $\mathbf{x}_k^{t,r}$ and as many sequences of observation vectors \mathbf{y}_k^r .

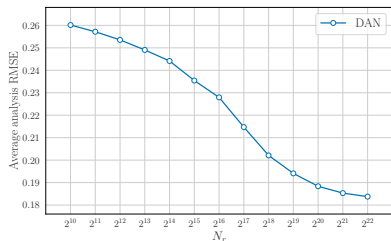
The analysis ensemble is $\mathbf{x}_k^{a,i,r} \in \mathbb{R}^{N_x \times N_e}$. The *loss function* is defined by

$$\mathcal{L}(\theta) = \sum_{r=1}^{N_r} \sum_{k=1}^{N_c} \left\| \mathbf{x}_k^{t,r} - \bar{\mathbf{x}}_k^{a,r}(\theta) \right\|^2, \quad \bar{\mathbf{x}}_k^{a,r} \triangleq \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_k^{a,i,r}. \quad (4)$$

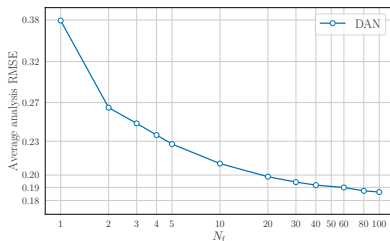
¹[Bocquet et al. 2024]

Hyperparameter sensitivity analysis

- Sensitivity analysis on key hyperparameters such as the number of trajectories N_T in the dataset, and the architecture parameters (N_f , N_b , N_{sb}) using the standard Lorenz 96 DA configuration ($\mathcal{H} = \mathbf{I}_x$, $\mathbf{R} = \mathbf{I}_x$).



Filtering performance
vs the number of trajectories N_T

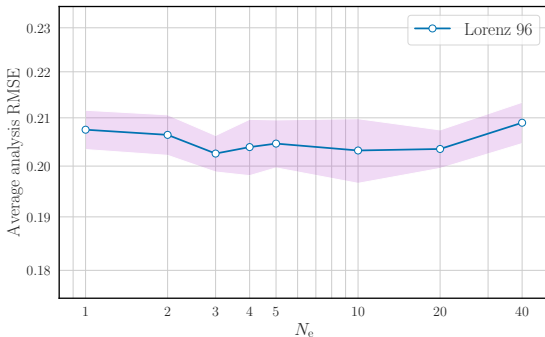


Filtering performance
vs the number of filters N_f

- The learned DA scheme *yields EnKF-like accuracy!*
- Compromise between a_θ 's size and its accuracy: $N_T = 2^{18}$, $N_f = 40$, $N_b = 5$, $N_{sb} = 5$.

Sensitivity to the ensemble size

► **First key observation:** The performance of a_θ barely depends on the ensemble size N_e . Hence localisation is irrelevant and unnecessary.



► **Second key observation:** a_θ does not require inflation and is incredibly robust to noise (as we shall see it applies its own inflation).

► **Explanation from the optimisation standpoint:** *feature collapse* of a_θ with respect to N_e in the training. Potential better solution when $N_e > 1$, but a_θ with $N_e = 1$ is as accurate as the EnKF!

Consequences and further checks

► Hence, from now on, we will focus on the mode: $N_e = 1$.

► Recall

$$\mathbf{E}_k^a = \mathbf{E}_k^f + a_\theta \left(\mathbf{E}_k^f, \mathbf{H}_k^\top \mathbf{R}_k^{-1} \delta_k \right). \quad (5)$$

► Performance of a_θ compared to baselines such as optimally tuned 3D-Var, the learned optimal linear filter, optimally tuned EnKF:

DA method	well-tuned classical	DL-based	aRMSE
EnKF-N, $N_e = 20$	yes		0.191
EnKF-N, $N_e = 40$	yes		0.179
3D-Var	yes		0.40
a_θ , $N_e = 1$, $N_f = 40$		yes	0.191
a_θ , $N_e = 1$, $N_f = 100$		yes	0.185
linear a_θ , $N_e = 1$, $N_f = 40$		yes	0.384
simplified \hat{a}_θ , $N_e = 1$, $N_f = 40$		yes	0.382

where (simplified Ansatz)

$$\mathbf{E}_k^a = \mathbf{E}_k^f + \hat{a}_\theta \left(\mathbf{H}_k^\top \mathbf{R}_k^{-1} \delta_k \right). \quad (6)$$

Operator expansion of the analysis

► We look for *a classical Kalman update that would be a good match to a_θ* seen as a mathematical map, at least for small analysis increments.

► To that end, we define the time-dependent normalised scalar anomalies

$$b_k = \frac{1}{\sqrt{N_x}} \|a_\theta(\mathbf{x}_k, \mathbf{0})\|, \quad (7)$$

along with the associated mean bias b and the standard deviation s of b_k in time.

→ We obtain $b \simeq 5 \times 10^{-3}$ and $s \simeq 10^{-3}$, which are indeed very small compared to the typical aRMSE of an either DAN or EnKF run, i.e., 0.20.

► Next, expanding with respect to the innovation, the following functional form for a_θ is assumed:

$$a_\theta(\mathbf{x}, \mathbf{H}^T \mathbf{R}^{-1} \delta) \approx \mathbf{K}(\mathbf{x}) \cdot \delta, \quad (8)$$

owing to the fact that no state update is needed when the innovation vanishes, and only keeping the leading order term in δ .

Identifying the operators in the expansion

► Innovations $\{\delta_j\}_{j=1,\dots,N_p}$ are sampled from $\delta_j \sim N(\mathbf{0}, \mathbf{R})$.

This yields a set of corresponding incremental updates $\{\mathbf{a}_j = a_\theta(\mathbf{x}, \mathbf{H}^T \mathbf{R}^{-1} \delta_j)\}_{j=1,\dots,N_p}$.

$\mathbf{K}(\mathbf{x})$ is then estimated with the least squares problem

$$\mathcal{L}_{\mathbf{x}}(\mathbf{K}) = \sum_{j=1}^{N_p} \left\| \mathbf{a}_j - \bar{\mathbf{a}} - \mathbf{K}(\mathbf{x}) \cdot (\delta_j - \bar{\delta}) \right\|^2, \quad (9)$$

where $\bar{\mathbf{a}} = N_p^{-1} \sum_{j=1}^{N_p} \mathbf{a}_j$ and $\bar{\delta} = N_p^{-1} \sum_{j=1}^{N_p} \delta_j$.

► Within the *best linear unbiased estimator* framework, \mathbf{K} is related to \mathbf{P}^a through $\mathbf{K} = \mathbf{P}^a \mathbf{H}^T \mathbf{R}^{-1}$ so that from Eq. (8),

$$a_\theta(\mathbf{x}, \mathbf{H}^T \mathbf{R}^{-1} \delta) \approx \mathbf{P}^a \mathbf{H}^T \mathbf{R}^{-1} \delta, \quad (10)$$

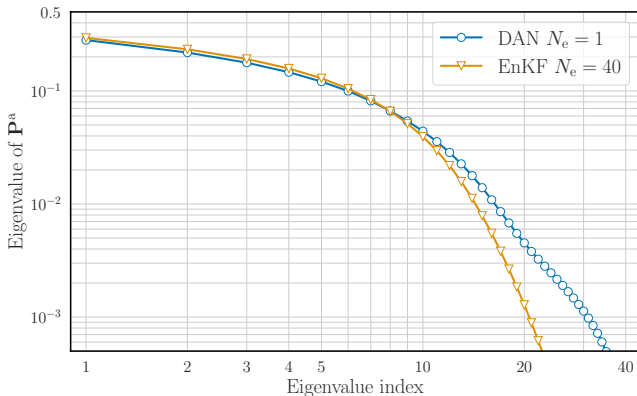
which suggests that an expansion in the second variable $\zeta \in \mathbb{R}^{N_x}$ of a_θ yields

$$a_\theta(\mathbf{x}, \zeta) \approx \mathbf{P}^a(\mathbf{x}) \cdot \zeta. \quad (11)$$

Hence, we can obtain a numerical estimation of an equivalent $\mathbf{P}^a(\mathbf{x})$.

What is learned? Supporting numerical results

- The surrogate \mathbf{P}^a , denoted $\mathbf{P}_{\text{DAN}}^a$ and estimated from Eq. (11), is compared to that of a concurrent well-tuned EnKF with $N_e = 40$, whose analysis error covariance matrix is $\mathbf{P}_{\text{EnKF}}^a$.



- Time-averaged eigenspectra of $\mathbf{P}_{\text{DAN}}^a$ and $\mathbf{P}_{\text{EnKF}}^a$. They are remarkably close to each other for the first 10 modes. Beyond these modes the a_θ operator is likely to selectively apply *some multiplicative inflation*, as one would expect from such stable DA runs.

Main interpretation

- ▶ **Conclusion 1:** a_θ depends on the innovation but also directly on \mathbf{x}_k^f when $N_e = 1$, as opposed to the incremental update of the EnKF: a_θ *extracts important information from \mathbf{x}_k^f* .
- ▶ **Conclusion 2:** a_θ manages to assess a $\mathbf{P}_{\text{DAN}}^a$ with $N_e = 1$ which is very close to $\mathbf{P}_{\text{EnKF}}^a$ with $N_e = 40$, for the dominant axes, and applies multiplicative inflation on the less unstable modes.² We conclude that a_θ *directly learns about the dynamics features*. Hence, for a_θ , critical pieces of information on \mathbf{P}_k^a are encoded, and thus exploitable, in \mathbf{x}_k^f alone.
 → Supported by results from [Sacco et al. 2024](#); [Sakov 2025](#).
- ▶ **Explanation, conclusion 3:** Furthermore, if the DA run (the forecast and analysis cycle) is considered as an ergodic dynamical system of its own,³ the *multiplicative ergodic theorem* guarantees the existence of a mapping between \mathbf{x}_k^f and \mathbf{P}_k^a that a_θ is able to guess. We believe that a generalised variant of *the multiplicative ergodic theorem for non-autonomous random dynamics should be applicable*.⁴

²[Bocquet et al. 2015]

³[Carrassi et al. 2008]

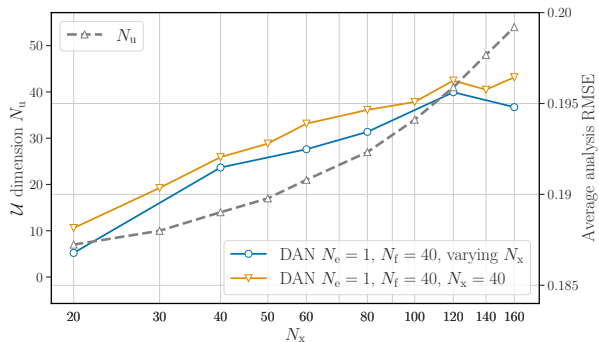
⁴[Arnold 1998; Flandoli et al. 2021]

Outline

- 1 Machine learning for data assimilation: blitz literature overview & classification
 - Learning machine emulators, model error through data assimilation
 - Machine learning-aided data assimilation
 - End-to-end approaches to data assimilation
- 2 Learning data assimilation from artificial intelligence: first results
 - Sequential data assimilation for chaotic dynamics
 - Learning data assimilation
 - Investigation and interpretation
- 3 Learning data assimilation from artificial intelligence: advanced results
 - Scalability and locality
 - Learning nonlinearity
- 4 Conclusions

Locality and scalability – 1/2

- ▶ α_θ is now trained without changing the architecture and the hyperparameters ($N_f = 40$), but with a changing state space dimension $N_x \in [20, 160]$. Almost as good as well tuned EnKFs with changing dimension N_x and $N_e = N_x$!



→ We conjecture that α_θ extracts *local* pieces of information from \mathbf{x}_k^f .

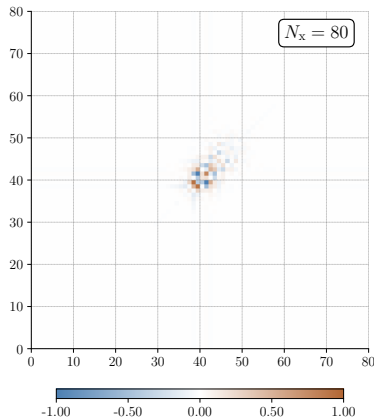
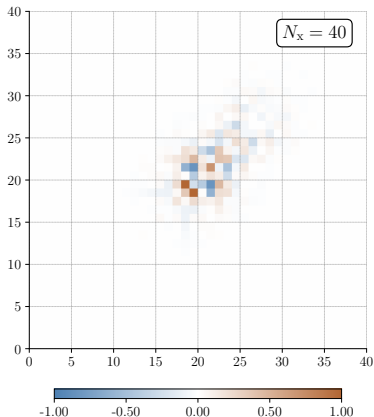
- ▶ α_θ , learned from Lorenz 96 with $N_x = 40$ is now tested on Lorenz 96 models with N_x ranging from 20 to 160 (same weights and biases!). The performance is still on par with retraining! We called this a *transdimensional transfer*.

Locality and scalability – 2/2

► These local patterns (for a_θ , not \mathcal{M}) can be pictured from the sensitivity:

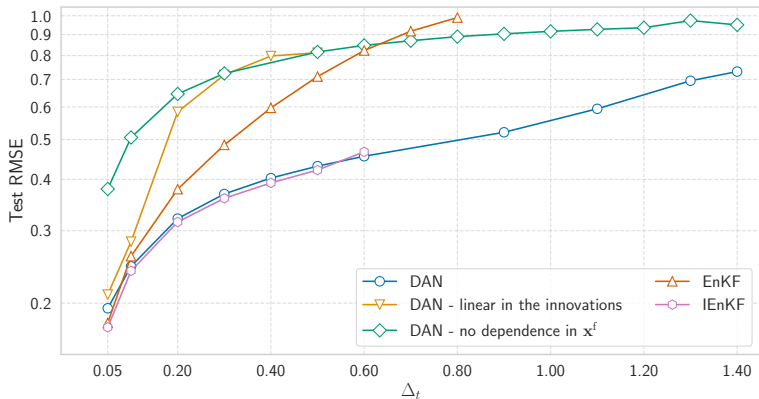
$$\mathbf{S} = \left\langle \mathbf{C} : \left[\nabla_{\mathbf{x}} \nabla_{\zeta} a_\theta(\mathbf{x}, \zeta) \Big|_{\zeta=0} \right] \right\rangle_{\mathbf{x} \in \mathcal{T}} = \left\langle \mathbf{C} : \left[\nabla_{\mathbf{x}} \mathbf{P}^a(\mathbf{x}) \right] \right\rangle_{\mathbf{x} \in \mathcal{T}}, \quad (12)$$

where \mathcal{T} is a long L96 trajectory, and \mathbf{C} is a tensor that leverages translational invariance of the L96 model: $[\mathbf{C}]_{ij}^{nmk} = \frac{1}{N_x} \delta_{n,i+k} \delta_{m,j+k}$.



Data assimilation networks in stronger nonlinear conditions

- ▶ Testing DANs as *the update time-step Δ_t is increased*, with the L96 model.
- ▶ Comparison with well tuned *EnKF*, *IEnKF*, and well-tuned static background DA methods.



- ▶ Performing at least as well as the IEnKF, *without an ensemble, without any nonlinear iterative solver, without inflation and localisation!*
- ▶ In addition to the MET map, DANs also implicitly *learn non-Gaussian priors*.

Outline

- 1 Machine learning for data assimilation: blitz literature overview & classification
 - Learning machine emulators, model error through data assimilation
 - Machine learning-aided data assimilation
 - End-to-end approaches to data assimilation
- 2 Learning data assimilation from artificial intelligence: first results
 - Sequential data assimilation for chaotic dynamics
 - Learning data assimilation
 - Investigation and interpretation
- 3 Learning data assimilation from artificial intelligence: advanced results
 - Scalability and locality
 - Learning nonlinearity
- 4 Conclusions

Conclusions

- ▶ One can learn *robust* DA methods, *without an ensemble*, *without inflation*, that are *as accurate* as the very best baseline methods (EnKF), including in *stronger nonlinear regimes* (IEnKF).
- ▶ We have carried similar numerical experiments with the *Kuramoto–Sivashinski* model and a *single-layer QG model on the sphere*, with similar conclusions.
- ▶ The performance is achieved through (i) implicitly learning the map $\mathbf{x}^f \mapsto \mathbf{P}^f(\mathbf{x}^f)$, and (ii) through implicitly learning non-Gaussian priors!
- ▶ This suggests that end-to-end approaches such as GraphDOP that only have a snapshot of the physical system, can still implicitly rely on dynamical and non-Gaussian priors!
- ▶ Will such *multiplicative ergodic theorem* still be valid in more anisotropic, non-autonomous, forced, multivariate, heterogeneously observed systems?
- ▶ *In any case, this promotes a rethinking of the popular sequential DA schemes for chaotic dynamics.*

Talk mainly based on:

- ▶ M. Bocquet, A. Farchi, T. S. Finn, C. Durand, S. Cheng, Y. Chen, I. Pasmans, and A. Carrassi. "Accurate deep learning-based filtering for chaotic dynamics by identifying instabilities without an ensemble". In: *Chaos* 29 (2024), p. 091104.
- ▶ M. Bocquet, T. S. Finn, S. Cheng, W. Yu, and A. Farchi. "On the performance of data assimilation neural networks in nonlinear conditions". In: (2025). In preparation.

References |

- [1] Gu. A. and Dao. T. "Mamba: Linear-Time Sequence Modeling with Selective State Spaces". In: *First Conference on Language Modeling*. 2024.
- [2] 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. eprint: https://doi.org/10.1162/neco_a_01094.
- [3] C. Aicher, N. J. Foti, and E. B. Fox. "Adaptively Truncating Backpropagation Through Time to Control Gradient Bias". In: *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*. Ed. by Ryan P. Adams and Vibhav Gogate. Vol. 115. Proceedings of Machine Learning Research. PMLR, 22–25 Jul 2020, pp. 799–808.
- [4] M. Alexe et al. *GraphDOP: Towards skilful data-driven medium-range weather forecasts learnt and initialised directly from observations*. 2024. arXiv: 2412.15687 [physics.ao-ph].
- [5] A. Allen et al. "End-to-end data-driven weather prediction". In: *Nature* (2025).
- [6] L. Arnold. *Random Dynamical Systems*. Springer Berlin, Heidelberg, 1998, p. 586.
- [7] M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation: Methods, Algorithms, and Applications*. Fundamentals of Algorithms. SIAM, Philadelphia, 2016, p. 324.
- [8] P. Becker, H. Pandya, G. Gebhardt, C. Zhao, C. J. Taylor, and G. Neumann. "Recurrent Kalman Networks: Factorized Inference in High-Dimensional Deep Feature Spaces". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 544–552.
- [9] 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.
- [10] 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.
- [11] M. Bocquet, A. Farchi, T. S. Finn, C. Durand, S. Cheng, Y. Chen, I. Pasmans, and A. Carrassi. "Accurate deep learning-based filtering for chaotic dynamics by identifying instabilities without an ensemble". In: *Chaos* 29 (2024), p. 091104.
- [12] M. Bocquet, T. S. Finn, S. Cheng, W. Yu, and A. Farchi. "On the performance of data assimilation neural networks in nonlinear conditions". In: (2025). in preparation.
- [13] M. Bocquet, P. N. Raanes, and A. Hannart. "Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation". In: *Nonlin. Processes Geophys.* 22 (2015), pp. 645–662.
- [14] M. Bocquet and P. Savov. "Joint state and parameter estimation with an iterative ensemble Kalman smoother". In: *Nonlin. Processes Geophys.* 20 (2013), pp. 803–818.

References II

- [15] P. Boudier, A. Fillion, S. Gratton, S. Gürol, and S. Zhang. "Data Assimilation Networks". In: *J. Adv. Model. Earth Syst.* 15 (2023), e2022MS003353.
- [16] 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.
- [17] 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. eprint: arXiv:2009.04318.
- [18] I. Buchnik, G. Revach, D. Steger, R. J. G. van Sloun, T. Routtenberg, and N. Shlezinger. "Latent-KalmanNet: Learned Kalman Filtering for Tracking From High-Dimensional Signals". In: *IEEE Transactions on Signal Processing* 72 (2024), pp. 352–367.
- [19] A. Carrassi, M. Ghil, A. Trevisan, and F. Uboldi. "Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system". In: *Chaos* 18 (2008), p. 023112.
- [20] A. Chattopadhyay, M. Mustafa, P. Hassanzadeh, E. Bach, and K. Kashinath. "Towards physics-inspired data-driven weather forecasting: integrating data assimilation with a deep spatial-transformer-based U-NET in a case study with ERA5". In: *Geosci. Model Dev.* 15 (2022), pp. 2221–2237.
- [21] Y. Chen, D. Sanz-Alonso, and R. Willett. "Autodifferentiable Ensemble Kalman Filters". In: *SIAM J. Math. Data Sci.* 4 (2022), pp. 801–833.
- [22] S. Cheng, J. Chen, C. Anastasiou, P. Angeli, O. K. Matar, Y.-K. Guo, C. C. Pain, and R. Arcucci. "Generalised Latent Assimilation in Heterogeneous Reduced Spaces with Machine Learning Surrogate Models". In: *J. Sci. Comput.* 94 (2023), pp. 1573–7691.
- [23] S. Cheng, I. C. Prentice, Y. Huang, Y. Jin, Y.-K. Guo, and R. Arcucci. "Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting". In: *J. Comp. Phys.* 464 (2022), p. 111302.
- [24] S. Cheng et al. "Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review". In: *IEEE/CAA J. Autom. Sin.* 10 (2023), pp. 1361–1387.
- [25] G. Choi, J. Park, N. Shlezinger, Y. C. Eldar, and N. Lee. "Split-KalmanNet: A Robust Model-Based Deep Learning Approach for State Estimation". In: *IEEE Transactions on Vehicular Technology* 72 (2023), 12326–12331.
- [26] H. Chung, J. Kim, M. Thompson Mccann, M. L. Klasky, and J. C. Ye. "Diffusion Posterior Sampling for General Noisy Inverse Problems". In: *The Eleventh International Conference on Learning Representations*. 2023.
- [27] R. Cintra, H. de Campos Velho, and S. Cocks. "Tracking the model: Data assimilation by artificial neural network". In: *2016 International Joint Conference on Neural Networks (IJCNN)*. 2016, pp. 403–410.
- [28] R. S. Cintra and H. F. de Campos Velho. "Data assimilation by artificial neural networks for an atmospheric general circulation model". In: *Advanced applications for artificial neural networks*. Ed. by A. ElShahat. IntechOpen, 2018. Chap. 17, pp. 265–286.

References III

- [29] R. S. Cintra and H. F. de Campos Velho. "Global data assimilation using artificial neural networks in SPEEDY model". In: *1st International Symposium Uncertainty Quantification and Stochastic Modeling*. Maresias, São Sebastião (SP), Brazil, 2012, pp. 648–654.
- [30] H. Coskun, F. Achilles, R. DiPietro, N. Navab, and F. Tombari. "Long Short-Term Memory Kalman Filters: Recurrent Neural Estimators for Pose Regularization". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 5525–5533.
- [31] R. Dong, H. Leng, C. Zhao, J. Song, J. Zhao, and X. Cao. "A hybrid data assimilation system based on machine learning". In: *Frontiers in Earth Science* 10 (2023).
- [32] C. Durand, T. S. Finn, A. Farchi, M. Bocquet, J. Brajard, and L. Bertino. "Four-dimensional variational data assimilation with a sea-ice thickness emulator". In: *EGU sphere* 2025 (2025), pp. 1–36.
- [33] R. Fablet, B. Chapron, L. Drumetz, E. Mémin, O. Pannekoucke, and F. Rousseau. "Learning Variational Data Assimilation Models and Solvers". In: *J. Adv. Model. Earth Syst.* 13 (2021), e2021MS002572.
- [34] 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.* 55 (2021), p. 101468.
- [35] A. Farchi, M. Chrut, M. Bocquet, and M. Bonavita. "Development of an offline and online hybrid model for the Integrated Forecasting System". In: *Q. J. R. Meteorol. Soc.* (2025), e4934.
- [36] A. Farchi, M. Chrut, M. Bocquet, P. Laloyaux, and M. Bonavita. "Online Model Error Correction With Neural Networks in the Incremental 4D-Var Framework". In: *J. Adv. Model. Earth Syst.* 15 (2023), e2022MS003474.
- [37] 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.
- [38] A. Filoche, J. Brajard, A. Charantonis, and D. Béréziat. "Learning 4DVAR Inversion Directly from Observations". In: *Computational Science – ICCS 2023*. Ed. by Jiří Mikyška, Clélia de Mulatier, Maciej Paszynski, Valeria V. Krzhizhanovskaya, Jack J. Dongarra, and Peter M.A. Sloot. Cham: Springer Nature Switzerland, 2023, pp. 414–421.
- [39] T. S. Finn, L. Disson, A. Farchi, M. Bocquet, and C. Durand. "Representation learning with unconditional denoising diffusion models for dynamical systems". In: *Nonlin. Processes Geophys.* 31 (2024), pp. 409–431.
- [40] F. Flandoli and E. Tonello. *An introduction to random dynamical systems for climate*. 2021.
- [41] T. Frerix, D. Kochkov, J. Smith, D. Cremers, M. Brenner, and S. Hoyer. "Variational Data Assimilation with a Learned Inverse Observation Operator". In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 18–24 Jul 2021, pp. 3449–3458.

References IV

- [42] V. Garcia Satorras, Z. Akata, and M. Welling. "Combining Generative and Discriminative Models for Hybrid Inference". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc., 2019.
- [43] D. Gedon, N. Wahlström, T. B. Schön, and L. Ljung. "Deep State Space Models for Nonlinear System Identification". In: *IFAC-PapersOnLine* 54 (2021). 19th IFAC Symposium on System Identification SYSID 2021, pp. 481–486.
- [44] A. Ghosh, A. Honoré, and S. Chatterjee. "DANSE: Data-Driven Non-Linear State Estimation of Model-Free Process in Unsupervised Learning Setup". In: *IEEE Transactions on Signal Processing* 72 (2024), pp. 1824–1838.
- [45] I. Grooms. "Analog ensemble data assimilation and a method for constructing analogs with variational autoencoders". In: *Q. J. R. Meteorol. Soc.* 147 (2021), pp. 139–149.
- [46] T. Haarnoja, A. Ajay, S. Levine, and P. Abbeel. "Backprop KF: Learning Discriminative Deterministic State Estimators". In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett. Vol. 29. Curran Associates, Inc., 2016.
- [47] F. P. Härter and H. F. de Campos Velho. "New approach to applying neural network in nonlinear dynamic model". In: *Appl. Math. Model.* 32 (2008), pp. 2621–2633.
- [48] T. P. Härter and H. F. de Campos Velho. "Data Assimilation Procedure by Recurrent Neural Network". In: *Eng. Appl. Comput. Fluid Mech.* 6 (2012), pp. 224–233.
- [49] S. Hatfield, M. Chantry, P. Dueben, P. Lopez, A. Geer, and T. Palmer. "Building Tangent-Linear and Adjoint Models for Data Assimilation With Neural Networks". In: *J. Adv. Model. Earth Syst.* 13 (2021), e2021MS002521.
- [50] H.S. Hoang, P. De Mey, and O. Talagrand. "A simple adaptive algorithm of stochastic approximation type for system parameter and state estimation". In: *Proceedings of 1994 33rd IEEE Conference on Decision and Control*. Vol. 1. 1994, 747–752 vol.1.
- [51] S. Hoang, R. Baraille, O. Talagrand, X. Carton, and P. De Mey. "Adaptive filtering: application to satellite data assimilation in oceanography". In: *Dynam. Atmos. Ocean* 27 (1998), pp. 257–281.
- [52] 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.
- [53] L. Huang, L. Gianinazzi, Y. Yu, P. D. Dueben, and T. Hoefler. "DiffDA: a diffusion model for weather-scale data assimilation". In: *Proceedings of the 41st International Conference on Machine Learning*. ICML'24. Vienna, Austria: JMLR.org, 2024.
- [54] T. Imbiriba, O. Straka, J. Duník, and P. Closas. "Augmented Physics-Based Machine Learning for Navigation and Tracking". In: *IEEE Transactions on Aerospace and Electronic Systems* 60 (2024), pp. 2692–2704.

References V

- [55] J. D. Keller and R. Potthast. *AI-based data assimilation: Learning the functional of analysis estimation*. 2024. arXiv: 2406.00390 [physics.ao-ph].
- [56] A. Klushyn, R. Kurle, M. Soelch, B. Cseke, and P. van der Smagt. "Latent Matters: Learning Deep State-Space Models". In: *Advances in Neural Information Processing Systems*. Ed. by A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan. 2021.
- [57] N. Lafon, R. Fablet, and P. Naveau. "Uncertainty Quantification When Learning Dynamical Models and Solvers With Variational Methods". In: *J. Adv. Model. Earth Syst.* 15 (2023), e2022MS003446.
- [58] P. Lean, M. Alexe, E. Boucher, E. Pinnington, S. Lang, P. Laloyaux, N. Bormann, and A. McNally. *Learning from nature: insights into GraphDOP's representations of the Earth System*. 2025. arXiv: 2508.18018 [physics.ao-ph].
- [59] S. Legler and T. Janjić. "Combining data assimilation and machine learning to estimate parameters of a convective-scale model". In: *Q. J. R. Meteorol. Soc.* 148 (2022), pp. 860–874.
- [60] Y. Li, W. Han, H. Li, W. Duan, L. Chen, X. Zhong, J. Wang, Y. Liu, and X. Sun. "FuXi-En4DVar: An Assimilation System Based on Machine Learning Weather Forecasting Model Ensuring Physical Constraints". In: *Geophys. Res. Lett.* 51 (2024), e2024GL111136.
- [61] F. Lu. "U-Net Kalman Filter (UNetKF): An Example of Machine Learning-Assisted Data Assimilation". In: *J. Adv. Model. Earth Syst.* 17 (2025), e2023MS003979.
- [62] E. Luk, E. Bach, R. Baptista, and A. Stuart. *Learning Optimal Filters Using Variational Inference*. 2024. arXiv: 2406.18066 [cs.LG].
- [63] J. Mack, R. Arcucci, M. Molina-Solana, and Y.-K. Guo. "Attention-based Convolutional Autoencoders for 3D-Variational Data Assimilation". In: *Computer Methods in Applied Mechanics and Engineering* 372 (2020), p. 113291.
- [64] E. S. Maddy, S. A. Boukabara, and F. Iturbide-Sanchez. "Assessing the Feasibility of an NWP Satellite Data Assimilation System Entirely Based on AI Techniques". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), pp. 9828–9845.
- [65] M. McCabe and J. Brown. "Learning to Assimilate in Chaotic Dynamical Systems". In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan. Vol. 34. Curran Associates, Inc., 2021, pp. 12237–12250.
- [66] T. McNally, C. Lessig, P. Lean, M. Chantry, M. Alexe, and S. Lang. *Red sky at night... producing weather forecasts directly from observations*. Jan. 2024.
- [67] B. Melinc and Z. Zaplotnik. "3D-Var data assimilation using a variational autoencoder". In: *Q. J. R. Meteorol. Soc.* 150 (2024), pp. 2273–2295.
- [68] I Pasmans, Y. Chen, T. S. Finn, M. Bocquet, and A. Carrassi. *Ensemble Kalman filter in latent space using a variational autoencoder pair*. 2025. arXiv: 2502.12987 [cs.LG].

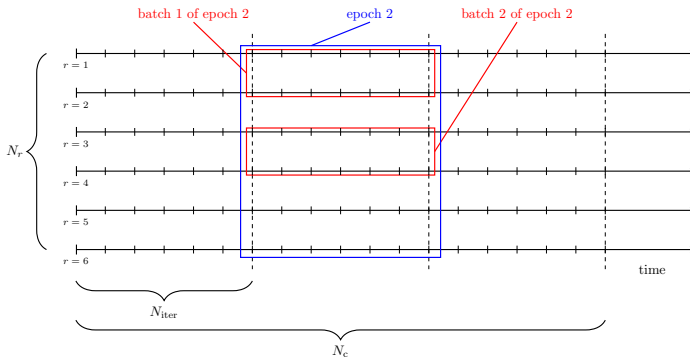
References VI

- [69] S. G. Penny, T. A. Smith, T.-C. Chen, J. A. Platt, H.-Y. Lin, M. Goodliff, and H. D. I. Abarbanel. "Integrating recurrent neural networks with data assimilation for scalable data-driven state estimation". In: *J. Adv. Model. Earth Syst.* 14 (2022), e2021MS002843.
- [70] M. Peyron, A. Fillion, S. Gürol, V. Marchais, S. Gratton, P. Boudier, and G. Goret. "Latent space data assimilation by using deep learning". In: *Q. J. R. Meteorol. Soc.* 147 (2021), pp. 3759–3777.
- [71] K. Pratik, R. A. Amjad, A. Behboodi, J. B. Soriaga, and M. Welling. "Neural Augmentation of Kalman Filter with Hypernetwork for Channel Tracking". In: *2021 IEEE Global Communications Conference (GLOBECOM)*. 2021, pp. 1–6.
- [72] G. Revach, N. Shlezinger, X. Ni, A. López Escoriza, L., R. J. G. van Sloun, and Y. C. Eldar. "KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics". In: *IEEE Trans. Signal Process.* 70 (2022), pp. 1532–1547.
- [73] Guy Revach, Nir Shlezinger, Timur Locher, Xiaoyong Ni, Ruud J. G. van Sloun, and Yonina C. Eldar. "Unsupervised Learned Kalman Filtering". In: *2022 30th European Signal Processing Conference (EUSIPCO)*. 2022, pp. 1571–1575.
- [74] F. Rozet, G. Andry, F. Lanusse, and G. Louppe. "Learning Diffusion Priors from Observations by Expectation Maximization". In: *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. 2024.
- [75] F. Rozet and G. Louppe. "Score-based Data Assimilation". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine. Vol. 36. Curran Associates, Inc., 2023, pp. 40521–40541.
- [76] Y. Ruckstuhl, T. Janjić, and S. Rasp. "Training a convolutional neural network to conserve mass in data assimilation". In: *Nonlin. Processes Geophys.* 28 (2021), pp. 111–119.
- [77] M. A. Sacco, M. Pulido, J. J. Ruiz, and P. Tandeo. "On-line machine-learning forecast uncertainty estimation for sequential data assimilation". In: *Q. J. R. Meteorol. Soc.* 150 (2024), pp. 2937–2954.
- [78] M. A. Sacco, J. J. Ruiz, M. Pulido, and P. Tandeo. "Evaluation of machine learning techniques for forecast uncertainty quantification". In: *Q. J. R. Meteorol. Soc.* 148 (2022), pp. 3470–3490.
- [79] P. Sakov. *On building the state error covariance from a state estimate*. 2025. arXiv: 2411.14809 [nlin.CD].
- [80] N. Shlezinger, Y. C. Eldar, and S. P. Boyd. "Model-Based Deep Learning: On the Intersection of Deep Learning and Optimization". In: *IEEE Access* 10 (2022), pp. 115384–115398.
- [81] N. Shlezinger, G. Revach, Ghosh. A., S. Chatterjee, Tang. S., T. Imbiriba, J. Dunik, O. Straka, P. Closas, and Eldar. Y. C. 2024. eprint: 2410.12289 (cs.LG).
- [82] N. Shlezinger and T. Routtenberg. "Discriminative and Generative Learning for the Linear Estimation of Random Signals [Lecture Notes]". In: *IEEE Signal Processing Magazine* 40 (2023), pp. 75–82.

References VII

- [83] V. L. S. Silva, C. E. Heaney, N. Nenov, and C. C. Pain. *Generative Network-Based Reduced-Order Model for Prediction, Data Assimilation and Uncertainty Quantification*. 2023. arXiv: 2105.13859 [cs.LG].
- [84] L. C. Slivinski, J. S. Whitaker, S. Frolov, T. A. Smith, and N. Agarwal. "Assimilating Observed Surface Pressure into ML Weather Prediction Models". In: *Geophys. Res. Lett.* 52 (2025), e2024GL114396.
- [85] X. Sun et al. *FuXi Weather: A data-to-forecast machine learning system for global weather*. 2024. arXiv: 2408.05472 [cs.LG].
- [86] H. Tang and J. Glass. "On Training Recurrent Networks with Truncated Backpropagation Through time in Speech Recognition". In: *2018 IEEE Spoken Language Technology Workshop (SLT)*. 2018, pp. 48–55.
- [87] A. Vaughan et al. *Aardvark weather: end-to-end data-driven weather forecasting*. 2024. arXiv: 2404.00411 [physics.ao-ph].
- [88] X. Xu, X. Sun, W. Han, X. Zhong, L. Chen, Z. Gao, and H. Li. "FuXi-DA: a generalized deep learning data assimilation framework for assimilating satellite observations". In: *npj Clim Atmos Sci* (2025), p. 156.
- [89] L. M. Yang and I. Grooms. "Machine learning techniques to construct patched analog ensembles for data assimilation". In: *J. Comput. Sci.* 443 (2021), p. 110532.

Training scheme – 2/2



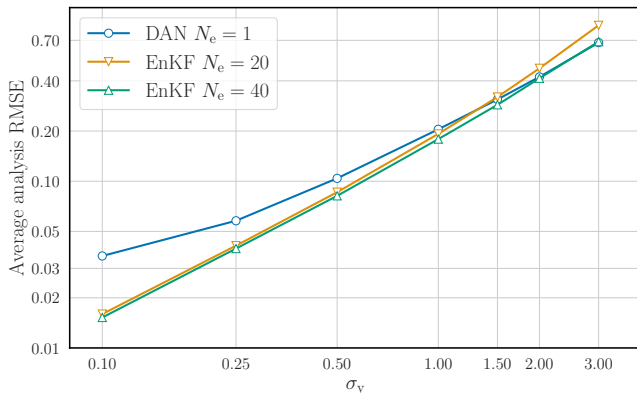
Structure of the dataset organised as a function of time, trajectory sample, batches and epochs.

► Like [McCabe et al. 2021; Boudier et al. 2023], we use *truncated backpropagation through time TBPTT* [Tang et al. 2018; Aicher et al. 2020], with a truncation at $N_{iter} \ll N_c$.

► For numerical efficiency, we choose to generate the samples *online*, as the training progresses, i.e. an *infinite training dataset!*

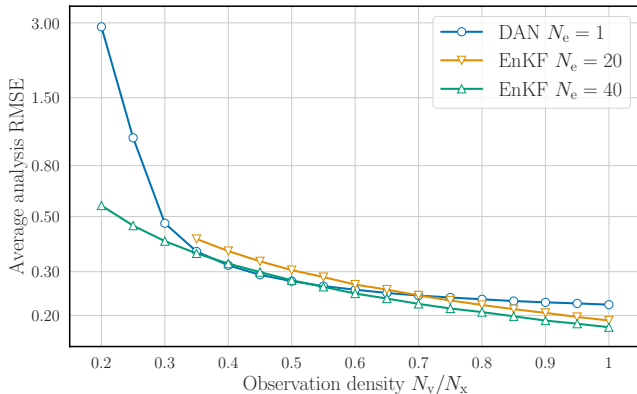
Sensitivity to observation error magnitude

- ▶ Next, we carry out a series of experiments that are not central to our message here but further ground the *viability of such learned a_θ* (assuming here $\mathbf{R}_k \triangleq \mathbf{I}_x$).
- ▶ Impact of the *observation noise magnitude* on the data assimilation tests:



Sensitivity to observation sparsity

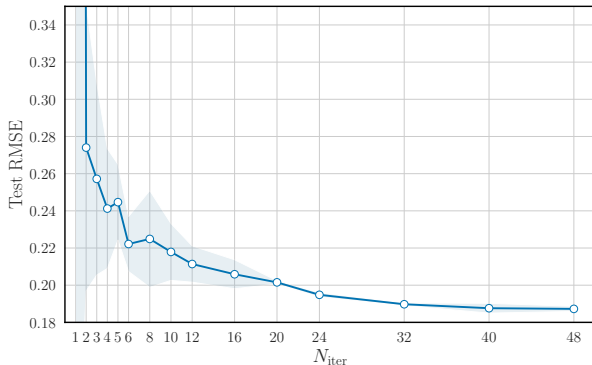
- Impact of the *sparsity of the observation dataset* on the data assimilation tests:



- a_θ trained with time-dependent, random, observation numbers and positions.

Sensitivity to the training (truncation) depth N_{iter}

- ▶ Training through $N_{\text{iter}} = 1$ cycle cannot learn about the direct impact of the dynamics on DA.
- ▶ Training through N_{iter} chained cycles is expected to be crucial to the accuracy and robustness of the learned a_{θ} [Bocquet et al. 2025].



- ▶ Training depth does matter!
As expected, $N_{\text{iter}} \geq 2$ cycles are required to see significant benefits.