Machine learning for data assimilation

Marc Bocquet,[*]

Alban Farchi,[†,*] Tobias Finn,[*] Charlotte Durand,[*] Sibo Cheng,[*]
Yumeng Chen,[‖] Ivo Pasmans,[‖] Alberto Carrassi,[§]
Massimo Bonavita,[†] Patrick Laloyaux,[†] Marcin Chrust,[†]
Julien Brajard,[‡] Laurent Bertino.[‡]

[*] 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,
[‡] NERSC, Bergen, Norway,

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
   - Sequential data assimilation for chaotic dynamics
   - Learning data assimilation
   - Investigation and interpretation
   - Conclusion

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
    - Sequential data assimilation for chaotic dynamics
    - Learning data assimilation
    - Investigation and interpretation
    - Conclusion

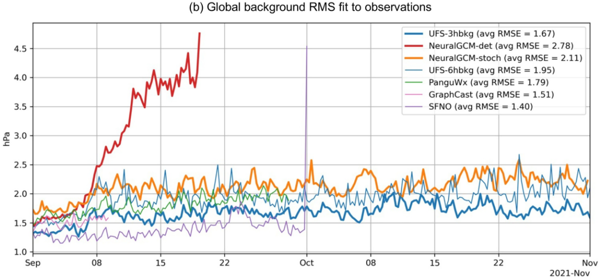
Outline

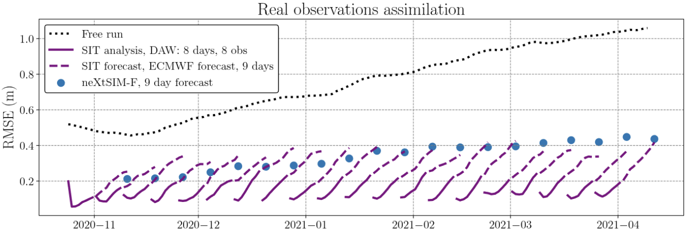
1. Machine learning for data assimilation: blitz literature overview & classification
   - Learning machine emulators, model error through data assimilation
   - Machine learning-aided data assimilation
   - End-to-end approaches to data assimilation

2. Learning data assimilation from artificial intelligence
   - Sequential data assimilation for chaotic dynamics
   - Learning data assimilation
   - Investigation and interpretation
   - Conclusion

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

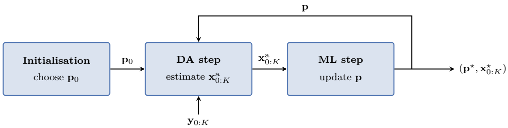




Learning/correcting dynamics through data assimilation

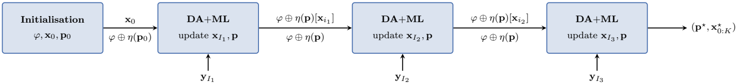
▶ Learning dynamics through Bayesian data assimilation

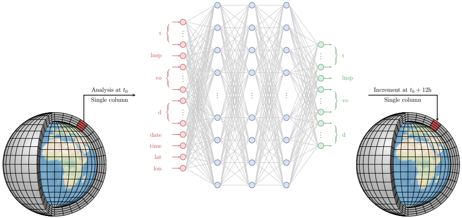
Hsieh et al. 1998; Abarbanel et al. 2018; Bocquet et al. 2019; Brajard et al. 2020; Bocquet et al. 2020; Brajard et al. 2021; Farchi et al. 2021b; G. Revach et al. 2022.



▶ 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.





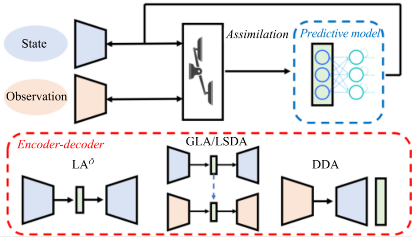
Outline

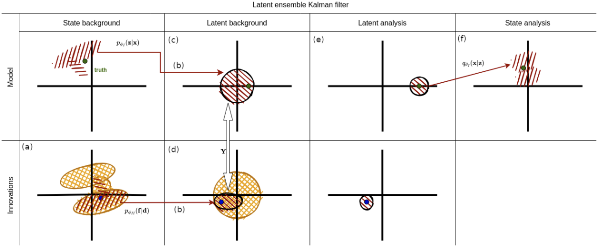
1. Machine learning for data assimilation: blitz literature overview & classification
   - Learning machine emulators, model error through data assimilation
   - Machine learning-aided data assimilation
   - End-to-end approaches to data assimilation

2. Learning data assimilation from artificial intelligence
   - Sequential data assimilation for chaotic dynamics
   - Learning data assimilation
   - Investigation and interpretation
   - Conclusion

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

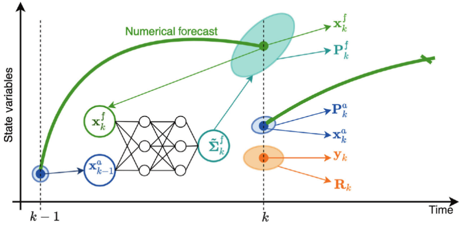




Machine learning-assisted data assimilation

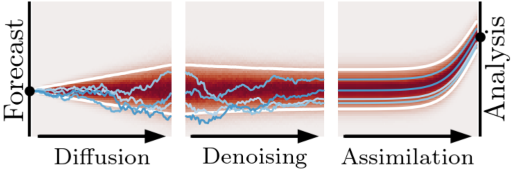
▶ For uncertainty quantification

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



▶ Using generative artificial intelligence

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.

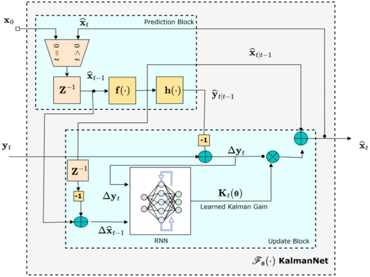
Parametric learning of data assimilation

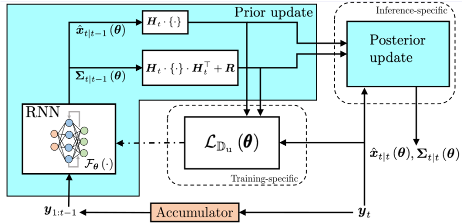
▶ Learning the (static) gain of the (ensemble) Kalman filter

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

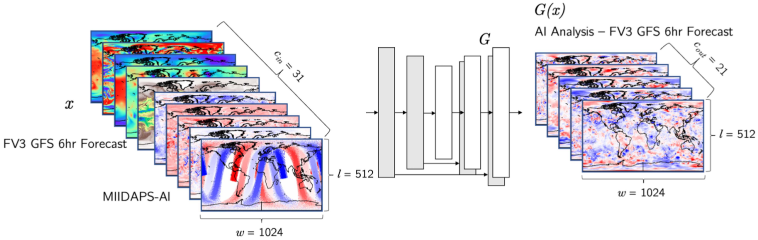
Outline

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
   - Sequential data assimilation for chaotic dynamics
   - Learning data assimilation
   - Investigation and interpretation
   - Conclusion

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



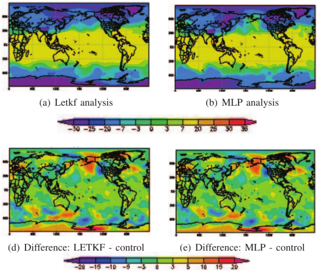
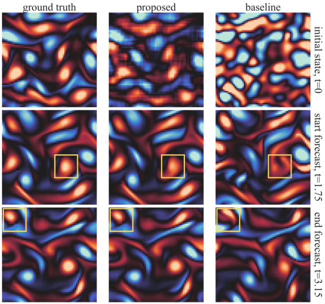
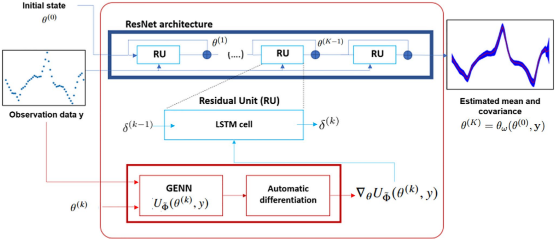


Fig. 6. Temperature (°C) Fields at layer 500 hPa to 08/01/2004 at 12 UTC. (a) LETKF analysis (b) MLP-DA analysis (d) diferrences between LETKF and MLP-DA analyses.

Learning variational solvers

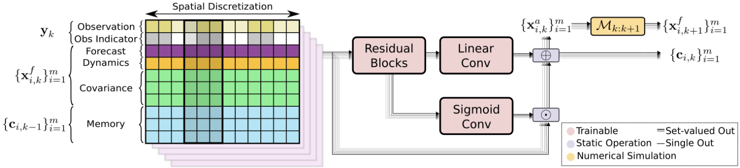
Fablet et al. 2021; Frerix et al. 2021; Lafon et al. 2023; Filoche et al. 2023; Keller et al. 2024

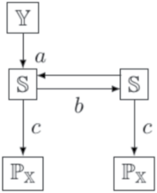


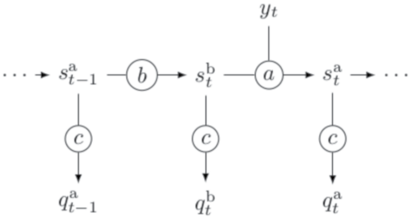


Learning sequential data assimilation through the cycles

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

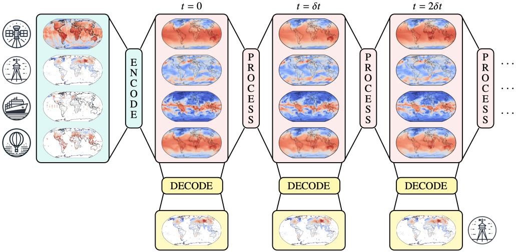


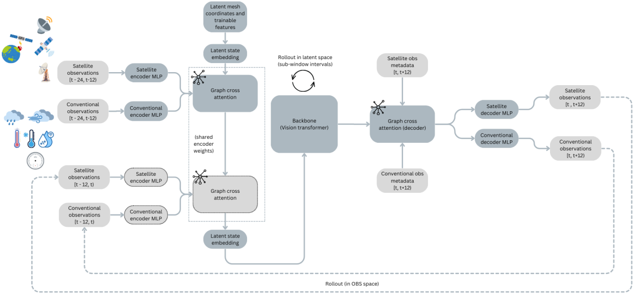




Bypassing data assimilation: end-to-end observation-driven NWP

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





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
- Sequential data assimilation for chaotic dynamics
- Learning data assimilation
- Investigation and interpretation
- Conclusion

Sequential data assimilation for chaotic dynamics

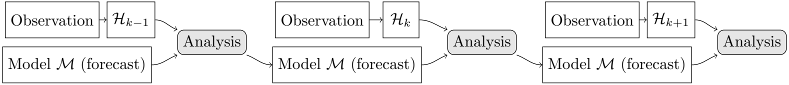
- Here, data assimilation (DA) methods are formulated from

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

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

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.

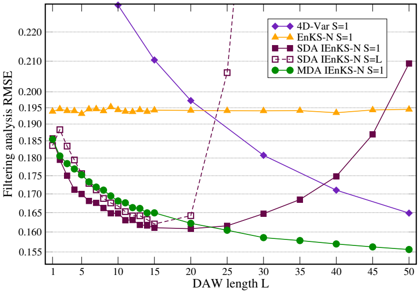


The edge of ensemble filtering methods

▶ 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 . . . but requires regularisation of the error covariances estimate.

▶ 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]:



▶ 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 $a_{\boldsymbol{\theta}}$*, typically a neural network parametrised by $\boldsymbol{\theta}$.

- If $\mathbf{E}_k^{\mathrm{a}}, \mathbf{E}_k^{\mathrm{f}} \in \mathbb{R}^{N_{\mathrm{x}} \times N_{\mathrm{e}}}$ are the analysis and forecast ensemble matrices at time τ_k, $a_{\boldsymbol{\theta}}$ is defined via the (ensemble) update:

$$\mathbf{E}_k^{\mathrm{a}} = \mathbf{E}_k^{\mathrm{f}} + a_{\boldsymbol{\theta}} \left(\mathbf{E}_k^{\mathrm{f}}, \mathbf{H}_k^{\intercal} \mathbf{R}_k^{-1} \boldsymbol{\delta}_k \right), \tag{2a}$$

where $\boldsymbol{\delta}_k$, the innovation at time τ_k, is defined by

$$\boldsymbol{\delta}_k \triangleq \mathbf{y}_k - \mathcal{H}_k \left(\bar{\mathbf{x}}_k^{\mathrm{f}} \right), \quad \bar{\mathbf{x}}_k^{\mathrm{f}} \triangleq \frac{1}{N_{\mathrm{e}}} \sum_{i=1}^{N_{\mathrm{e}}} \mathbf{x}_k^{\mathrm{f},i}. \tag{2b}$$

$\longrightarrow$ Notice our trick: $a_{\boldsymbol{\theta}} \left(\mathbf{E}_k^{\mathrm{f}}, \boldsymbol{\delta}_k \right) \longrightarrow a_{\boldsymbol{\theta}} \left(\mathbf{E}_k^{\mathrm{f}}, \mathbf{H}_k^{\intercal} \mathbf{R}_k^{-1} \boldsymbol{\delta}_k \right)$, i.e., *uplift of observation information in state space.*

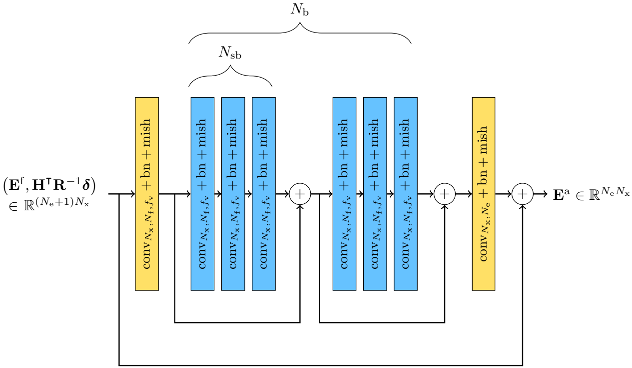
- The DA forecast step propagates the analysis ensemble, member-wise:

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

- The $a_{\boldsymbol{\theta}}$–based sequential DA will be called *DAN* in the following.

Neural network architecture

- We choose $a_{\boldsymbol{\theta}}$ to have a simple residual convolutional neural network (CNN) architecture.



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

Training scheme – 1/2

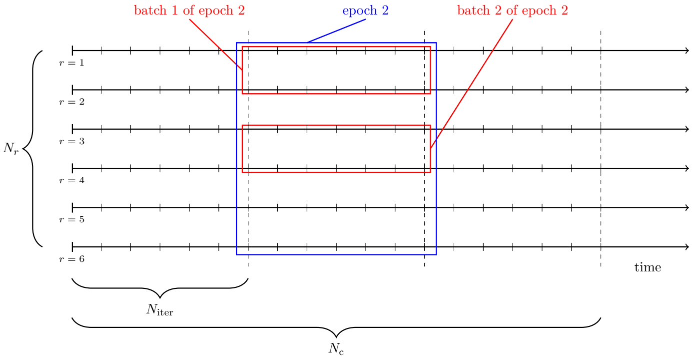
- Literature (focused on *sequential data assimilation*):
  - Learning the analysis of sequential DA is not new [T. P. Härter et al. 2012; R. S. Cintra et al. 2018], though barely explored.
  - Learning key components of the analysis in the (En)KF [H. Hoang et al. 1994; S. Hoang et al. 1998] possibly 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):

We consider N_r, N_c cycle-long ensemble DA runs, based on as many independent concurrent trajectories of the dynamics $\mathbf{x}_k^{\mathrm{t},r}$ and as many sequences of observation vectors $\mathbf{y}_k^r$.
The analysis ensemble is $\mathbf{x}_k^{\mathrm{a},i,r} \in \mathbb{R}^{N_\mathrm{x} \times N_\mathrm{e}}$. The *loss function* is defined by

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{r=1}^{N_\mathrm{r}} \sum_{k=1}^{N_\mathrm{c}} \left\| \mathbf{x}_k^{\mathrm{t},r} - \bar{\mathbf{x}}_k^{\mathrm{a},r}(\boldsymbol{\theta}) \right\|^2, \quad \bar{\mathbf{x}}_k^{\mathrm{a},r} \triangleq \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} \mathbf{x}_k^{\mathrm{a},i,r}. \tag{4}$$

Training scheme – 2/2



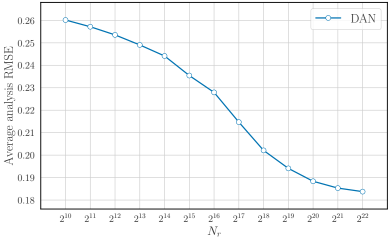
Structure of the dataset organised as a function of time, trajectory sample, batches and epochs.

▶ Like [McCabe et al. 2021; Boudier et al. 2023], we use *truncated backpropagation through time TBPTT* [Tang et al. 2018; Aicher et al. 2020].

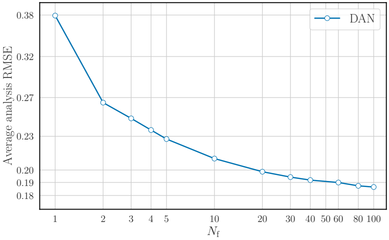
▶ For numerical efficiency, we choose to generate the samples *online*, as the training progresses, i.e. an *infinite training dataset!*

Hyperparameter sensitivity analysis

▶ Sensitivity analysis on key hyperparameters such as the number of trajectories $N_{\rm r}$ in the dataset, and the architecture parameters ($N_{\rm f}$, $N_{\rm b}$, $N_{\rm sb}$) using the standard Lorenz 96 DA configuration ($\mathcal{H} = \mathbf{I}_{\rm x}$, $\mathbf{R} = \mathbf{I}_{\rm x}$).



Filtering performance
vs the number of trajectories $N_{\rm r}$



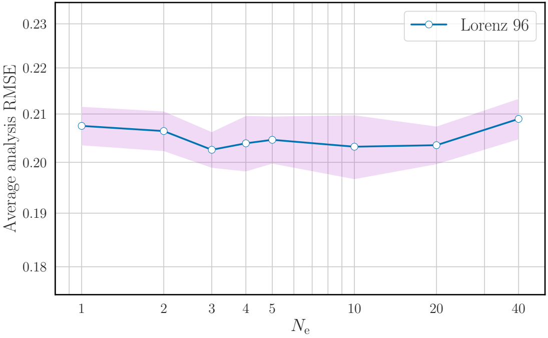
Filtering performance
vs the number of filters $N_{\rm f}$

▶ The learned DA scheme *yields EnKF-like accuracy!*

▶ Compromise between $a_{\boldsymbol{\theta}}$'s size and its accuracy: $N_{\rm r} = 2^{18}$, $N_{\rm f} = 40$, $N_{\rm b} = 5$, $N_{\rm sb} = 5$.

Sensitivity to the ensemble size

▶ First key observation: The performance of $a_{\boldsymbol{\theta}}$ barely depends on the ensemble size N_{e}. Hence localisation is irrelevant and unnecessary.



▶ Second key observation: $a_{\boldsymbol{\theta}}$ 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 $a_{\boldsymbol{\theta}}$ with respect to N_{e} in the training. Potential better solution when $N_{\mathrm{e}} > 1$, but $a_{\boldsymbol{\theta}}$ with $N_{\mathrm{e}} = 1$ is as accurate as the EnKF!

Consequences and further checks

- Hence, from now on, we will focus on the mode: $\boxed{N_\mathrm{e} = 1}$.

- Recall

$$\mathbf{E}_k^\mathrm{a} = \mathbf{E}_k^\mathrm{f} + a_{\boldsymbol{\theta}} \left(\mathbf{E}_k^\mathrm{f}, \mathbf{H}_k^\intercal \mathbf{R}_k^{-1} \boldsymbol{\delta}_k \right). \tag{5}$$

- Performance of $a_{\boldsymbol{\theta}}$ 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, $N_\mathrm{e} = 20$	yes		0.191
EnKF-N, $N_\mathrm{e} = 40$	yes		0.179
3D-Var	yes		0.40
$a_{\boldsymbol{\theta}}$, $N_\mathrm{e} = 1$, $N_\mathrm{f} = 40$		yes	0.191
$a_{\boldsymbol{\theta}}$, $N_\mathrm{e} = 1$, $N_\mathrm{f} = 100$		yes	0.185
linear $a_{\boldsymbol{\theta}}$, $N_\mathrm{e} = 1$, $N_\mathrm{f} = 40$		yes	0.384
simplified $\hat{a}_{\boldsymbol{\theta}}$, $N_\mathrm{e} = 1$, $N_\mathrm{f} = 40$		yes	0.382

where (simplified Ansatz)

$$\mathbf{E}_k^\mathrm{a} = \mathbf{E}_k^\mathrm{f} + \hat{a}_{\boldsymbol{\theta}} \left(\mathbf{H}_k^\intercal \mathbf{R}_k^{-1} \boldsymbol{\delta}_k \right). \tag{6}$$

Operator expansion of the analysis

- We look for *a classical Kalman update that would be a good match to $a_{\boldsymbol{\theta}}$* seen as a mathematical map, at least for small analysis increments.

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

$$b_k = \frac{1}{\sqrt{N_\mathrm{x}}} \left\| a_{\boldsymbol{\theta}}(\mathbf{x}_k, \mathbf{0}) \right\| , \qquad (7)$$

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

- Next, expanding with respect to the innovation, the following functional form for $a_{\boldsymbol{\theta}}$ is assumed:

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

owing to the fact that no state update is needed when the innovation vanishes, and only keeping the leading order term in $\boldsymbol{\delta}$.

Identifying the operators in the expansion

▶ Innovations $\{\boldsymbol{\delta}_j\}_{j=1,\ldots,N_\mathrm{p}}$ are sampled from $\boldsymbol{\delta}_j \sim N(\mathbf{0}, \mathbf{R})$.
This yields a set of corresponding incremental updates $\left\{\mathbf{a}_j = a_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{H}^\intercal\mathbf{R}^{-1}\boldsymbol{\delta}_j)\right\}_{j=1,\ldots,N_\mathrm{p}}$.
$\mathbf{K}(\mathbf{x})$ is then estimated with the least squares problem

$$\mathcal{L}_{\mathbf{x}}(\mathbf{K}) = \sum_{j=1}^{N_\mathrm{p}} \left\|\mathbf{a}_j - \bar{\mathbf{a}} - \mathbf{K}(\mathbf{x}) \cdot \left(\boldsymbol{\delta}_j - \bar{\boldsymbol{\delta}}\right)\right\|^2, \tag{9}$$

where $\bar{\mathbf{a}} = N_\mathrm{p}^{-1}\sum_{j=1}^{N_\mathrm{p}} \mathbf{a}_j$ and $\bar{\boldsymbol{\delta}} = N_\mathrm{p}^{-1}\sum_{j=1}^{N_\mathrm{p}} \boldsymbol{\delta}_j$.

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

$$a_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{H}^\intercal\mathbf{R}^{-1}\boldsymbol{\delta}) \approx \mathbf{P}^\mathrm{a}\mathbf{H}^\intercal\mathbf{R}^{-1}\boldsymbol{\delta}, \tag{10}$$

which suggests that an expansion in the second variable $\boldsymbol{\zeta} \in \mathbb{R}^{N_\mathrm{x}}$ of $a_{\boldsymbol{\theta}}$ yields

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

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

What is learned? Supporting numerical results – 1/2

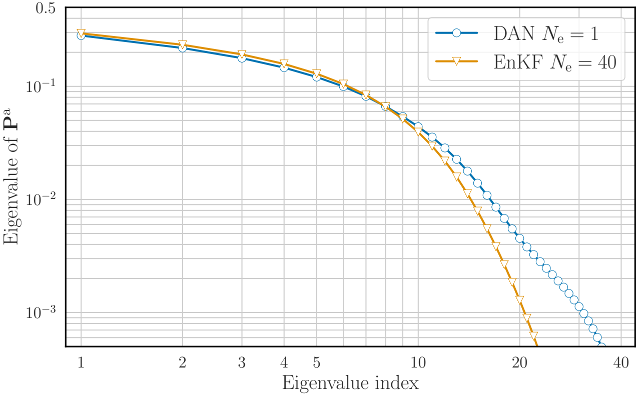
▶ We obtain $b \simeq 5 \times 10^{-3}$ and $s \simeq 10^{-3}$, which are indeed very small compared to the typical aRMSE of an either DAN or EnKF run, i.e., 0.20.

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

$\longrightarrow$ The time-averaged Bures–Wasserstein distance distance between $\mathbf{P}^{\mathrm{a}}_{\mathrm{DAN}}$ and $\mathbf{P}^{\mathrm{a}}_{\mathrm{EnKF}}$ is 0.013 whereas it is 0.048 between $\mathbf{P}^{\mathrm{a}}_{\mathrm{DAN}}$ and $(0.40)^2\mathbf{I}_{\mathrm{x}}$, which approximates $\mathbf{P}^{\mathrm{a}}$ of a well-tuned 3D-Var.

What is learned? Supporting numerical results – 2/2

▶ The time-averaged eigenspectra of $\mathbf{P}^{\mathrm{a}}_{\mathrm{DAN}}$ and $\mathbf{P}^{\mathrm{a}}_{\mathrm{EnKF}}$:



▶ They are remarkably close to each other for the first 10 modes. Beyond these modes the $a_{\boldsymbol{\theta}}$ operator is likely to selectively apply *some multiplicative inflation*, as one would expect from such stable DA runs.

Main interpretation

▶ Conclusion 1: $a_{\boldsymbol{\theta}}$ depends on the innovation but also directly on $\mathbf{x}_k^{\mathrm{f}}$ when $N_{\mathrm{e}} = 1$, as opposed to the incremental update of the EnKF: *$a_{\boldsymbol{\theta}}$ extracts important information from $\mathbf{x}_k^{\mathrm{f}}$.*

▶ Conclusion 2: $a_{\boldsymbol{\theta}}$ manages to assess a $\mathbf{P}^{\mathrm{a}}_{\mathrm{DAN}}$ with $N_{\mathrm{e}} = 1$ which is very close to $\mathbf{P}^{\mathrm{a}}_{\mathrm{EnKF}}$ with $N_{\mathrm{e}} = 40$, for the dominant axes, and applies multiplicative inflation on the less unstable modes.[1] We conclude that *$a_{\boldsymbol{\theta}}$ directly learns about the dynamics features.* Hence, for $a_{\boldsymbol{\theta}}$, critical pieces of information on $\mathbf{P}^{\mathrm{a}}_k$ are encoded, and thus exploitable, in $\mathbf{x}_k^{\mathrm{f}}$ alone.

▶ Explanation, conclusion 3: Furthermore, if the DA run (the forecast and analysis cycle) is considered as an ergodic dynamical system of its own,[2] the *multiplicative ergodic theorem* guarantees the existence of a mapping between $\mathbf{x}_k^{\mathrm{f}}$ and $\mathbf{P}^{\mathrm{a}}_k$ that $a_{\boldsymbol{\theta}}$ is able to guess. We believe that a generalised variant of *the multiplicative ergodic theorem for non-autonomous random dynamics should be applicable*.[3]

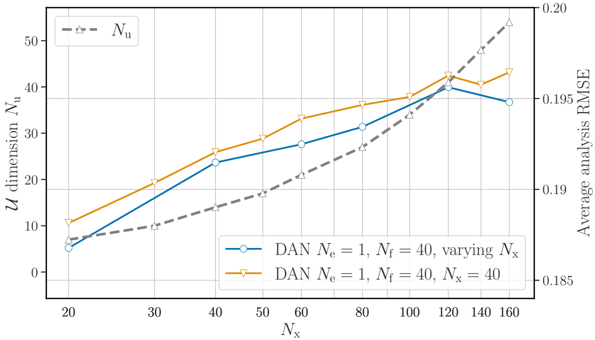
[1] [Bocquet et al. 2015]

[2] [Carrassi et al. 2008]

[3] [Arnold 1998; Flandoli et al. 2021]

Locality and scalability – 1/2

▶ $a_{\boldsymbol{\theta}}$ is now trained without changing the architecture and the hyperparameters ($N_\mathrm{f} = 40$), but with a changing state space dimension $N_\mathrm{x} \in [20, 160]$. Almost as good as well tuned EnKFs with changing dimension N_x and $N_\mathrm{e} = N_\mathrm{x}$!



⟶ We conjecture that $a_{\boldsymbol{\theta}}$ extracts *local* pieces of information from $\mathbf{x}^\mathrm{f}_k$.

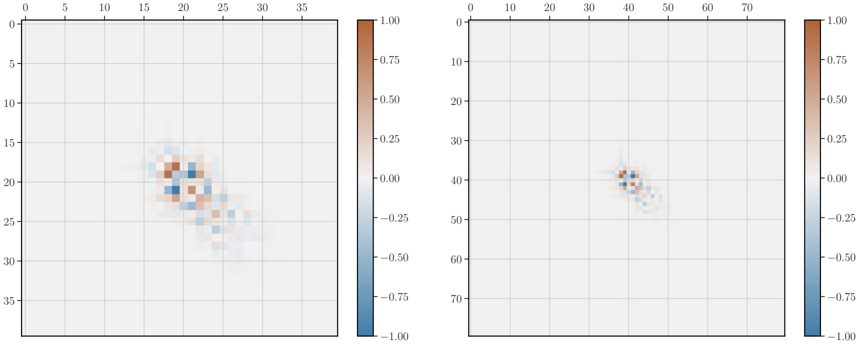
▶ $a_{\boldsymbol{\theta}}$, learned from Lorenz 96 with $N_\mathrm{x} = 40$ is now tested on Lorenz 96 models with N_x ranging from 20 to 160 (same weights and biases!). The performance is still on par with retraining! We called this a *transdimensional transfer*.

Locality and scalability – 2/2

- These local patterns (for $a_{\boldsymbol{\theta}}$, not $\mathcal{M}$) can be pictured from the sensitivity:

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

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



Conclusion: addendum and perspectives

- We have successfully tested learning $a_{\boldsymbol{\theta}}$ in a *semi-supervised* setting (no truth required), i.e. purely from *observations and the dynamics*.

- We have carried similar numerical experiments with the *Kuramoto–Sivashinski* model and a *single-layer QG model on the sphere*, with the same conclusions.

- 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.

$\longrightarrow$ Second part of this talk (mainly) based on Bocquet et al., Chaos, 2024.

References I

[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] P. Becker et al. "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.

[8] 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.

[9] M. Bocquet and P. Sakov. "Joint state and parameter estimation with an iterative ensemble Kalman smoother". In: *Nonlin. Processes Geophys.* 20 (2013), pp. 803–818.

[10] M. Bocquet et al. "Accurate deep learning-based filtering for chaotic dynamics by identifying instabilities without an ensemble". In: *Chaos* 29 (2024), p. 091104.

[11] 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.

[12] 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.

[13] P. Boudier et al. "Data Assimilation Networks". In: *J. Adv. Model. Earth Syst.* 15 (2023), e2022MS003353.

[14] 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.

References II

[15] 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. eprint: arXiv:2009.04318.

[16] I. Buchnik et al. “Latent-KalmanNet: Learned Kalman Filtering for Tracking From High-Dimensional Signals”. In: *IEEE Transactions on Signal Processing* 72 (2024), pp. 352–367.

[17] A. Carrassi et al. “Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system”. In: *Chaos* 18 (2008), p. 023112.

[18] A. Chattopadhyay et al. “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.

[19] Y. Chen, D. Sanz-Alonso, and R. Willett. “Autodifferentiable Ensemble Kalman Filters”. In: *SIAM J. Math. Data Sci.* 4 (2022), pp. 801–833.

[20] S. Cheng et al. “Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting”. In: *J. Comp. Phys.* 464 (2022), p. 111302.

[21] S. Cheng et al. “Generalised Latent Assimilation in Heterogeneous Reduced Spaces with Machine Learning Surrogate Models”. In: *J. Sci. Comput.* 94 (2023), pp. 1573–7691.

[22] 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.

[23] G. Choi et al. “Split-KalmanNet: A Robust Model-Based Deep Learning Approach for State Estimation”. In: *IEEE Transactions on Vehicular Technology* 72 (2023), 12326–=12331.

[24] H. Chung et al. “Diffusion Posterior Sampling for General Noisy Inverse Problems”. In: *The Eleventh International Conference on Learning Representations*. 2023.

[25] R. Cintra, H. de Campos Velho, and S. Cocke. “Tracking the model: Data assimilation by artificial neural network”. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. 2016, pp. 403–410.

[26] 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.

[27] 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.

[28] H. Coskun et al. “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.

References III

[29] R. Dong et al. "A hybrid data assimilation system based on machine learning". In: *Frontiers in Earth Science* 10 (2023).

[30] C. Durand et al. "Four-dimensional variational data assimilation with a sea-ice thickness emulator". In: *EGUsphere* 2025 (2025), pp. 1–36.

[31] R. Fablet et al. "Learning Variational Data Assimilation Models and Solvers". In: *J. Adv. Model. Earth Syst.* 13 (2021), e2021MS002572.

[32] A. Farchi et al. "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.

[33] A. Farchi et al. "Development of an offline and online hybrid model for the Integrated Forecasting System". In: *Q. J. R. Meteorol. Soc.* (2025), e4934.

[34] A. Farchi et al. "Online Model Error Correction With Neural Networks in the Incremental 4D-Var Framework". In: *J. Adv. Model. Earth Syst.* 15 (2023), e2022MS003474.

[35] A. Farchi et al. "Using machine learning to correct model error in data assimilation and forecast applications". In: *Q. J. R. Meteorol. Soc.* 147 (2021), pp. 3067–3084.

[36] A. Filoche et al. "Learning 4DVAR Inversion Directly from Observations". In: *Computational Science – ICCS 2023*. Ed. by Jiří Mikyška et al. Cham: Springer Nature Switzerland, 2023, pp. 414–421.

[37] T. S. Finn et al. "Representation learning with unconditional denoising diffusion models for dynamical systems". In: *Nonlin. Processes Geophys.* 31 (2024), pp. 409–431.

[38] F. Flandoli and E. Tonello. *An introduction to random dynamical systems for climate*. 2021.

[39] T. Frerix et al. "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.

[40] 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 et al. Vol. 32. Curran Associates, Inc., 2019.

[41] D. Gedon et al. "Deep State Space Models for Nonlinear System Identification". In: *IFAC-PapersOnLine* 54 (2021). 19th IFAC Symposium on System Identification SYSID 2021, pp. 481–486.

[42] 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.

[43] 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.

References IV

[44] T. Haarnoja et al. "Backprop KF: Learning Discriminative Deterministic State Estimators". In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee et al. Vol. 29. Curran Associates, Inc., 2016.

[45] 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.

[46] 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.

[47] S. Hatfield et al. "Building Tangent-Linear and Adjoint Models for Data Assimilation With Neural Networks". In: *J. Adv. Model. Earth Syst.* 13 (2021), e2021MS002521.

[48] 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.

[49] S. Hoang et al. "Adaptive filtering: application to satellite data assimilation in oceanography". In: *Dynam. Atmos. Ocean* 27 (1998), pp. 257–281.

[50] 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.

[51] L. Huang et al. "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.

[52] T. Imbiriba et al. "Augmented Physics-Based Machine Learning for Navigation and Tracking". In: *IEEE Transactions on Aerospace and Electronic Systems* 60 (2024), pp. 2692–2704.

[53] J. D. Keller and R. Potthast. *AI-based data assimilation: Learning the functional of analysis estimation*. 2024. arXiv: 2406.00390 [physics.ao-ph].

[54] A. Klushyn et al. "Latent Matters: Learning Deep State-Space Models". In: *Advances in Neural Information Processing Systems*. Ed. by A. Beygelzimer et al. 2021.

[55] 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.

[56] 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.

[57] Y. Li et al. "FuXi-En4DVar: An Assimilation System Based on Machine Learning Weather Forecasting Model Ensuring Physical Constraints". In: *Geophys. Res. Lett.* 51 (2024), e2024GL111136.

References V

[58] F. Lu. "U-Net Kalman Filter (UNetKF): An Example of Machine Learning-Assisted Data Assimilation". In: *J. Adv. Model. Earth Syst.* 17 (2025), e2023MS003979.

[59] E. Luk et al. *Learning Optimal Filters Using Variational Inference*. 2024. arXiv: 2406.18066 [cs.LG].

[60] J. Mack et al. "Attention-based Convolutional Autoencoders for 3D-Variational Data Assimilation". In: *Computer Methods in Applied Mechanics and Engineering* 372 (2020), p. 113291.

[61] 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.

[62] M. McCabe and J. Brown. "Learning to Assimilate in Chaotic Dynamical Systems". In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 12237–12250.

[63] T. McNally et al. *Red sky at night... producing weather forecasts directly from observations*. Jan. 2024.

[64] B. Melinc and Z. Zaplotnik. "3D-Var data assimilation using a variational autoencoder". In: *Q. J. R. Meteorol. Soc.* 150 (2024), pp. 2273–2295.

[65] I Pasmans et al. *Ensemble Kalman filter in latent space using a variational autoencoder pair*. 2025. arXiv: 2502.12987 [cs.LG].

[66] S. G. Penny et al. "Integrating recurrent neural networks with data assimilation for scalable data-driven state estimation". In: *J. Adv. Model. Earth Syst.* 14 (2022), e2021MS002843.

[67] M. Peyron et al. "Latent space data assimilation by using deep learning". In: *Q. J. R. Meteorol. Soc.* 147 (2021), pp. 3759–3777.

[68] K. Pratik et al. "Neural Augmentation of Kalman Filter with Hypernetwork for Channel Tracking". In: *2021 IEEE Global Communications Conference (GLOBECOM)*. 2021, pp. 1–6.

[69] G. Revach et al. "KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics". In: *IEEE Trans. Signal Process.* 70 (2022), pp. 1532–1547.

[70] Guy Revach et al. "Unsupervised Learned Kalman Filtering". In: *2022 30th European Signal Processing Conference (EUSIPCO)*. 2022, pp. 1571–1575.

[71] F. Rozet and G. Louppe. "Score-based Data Assimilation". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh et al. Vol. 36. Curran Associates, Inc., 2023, pp. 40521–40541.

[72] F. Rozet et al. "Learning Diffusion Priors from Observations by Expectation Maximization". In: *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. 2024.

References VI

[73] 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.

[74] M. A. Sacco et al. "Evaluation of machine learning techniques for forecast uncertainty quantification". In: *Q. J. R. Meteorol. Soc.* 148 (2022), pp. 3470–3490.

[75] M. A. Sacco et al. "On-line machine-learning forecast uncertainty estimation for sequential data assimilation". In: *Q. J. R. Meteorol. Soc.* 150 (2024), pp. 2937–2954.

[76] 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.

[77] 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.

[78] N. Shlezinger et al. 2024. eprint: 2410.12289 (cs.LG).

[79] V. L. S. Silva et al. *Generative Network-Based Reduced-Order Model for Prediction, Data Assimilation and Uncertainty Quantification*. 2023. arXiv: 2105.13859 [cs.LG].

[80] L. C. Slivinski et al. "Assimilating Observed Surface Pressure into ML Weather Prediction Models". In: *Geophys. Res. Lett.* 52 (2025), e2024GL114396.

[81] X. Sun et al. *FuXi Weather: A data-to-forecast machine learning system for global weather*. 2024. arXiv: 2408.05472 [cs.LG].

[82] 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.

[83] A. Vaughan et al. *Aardvark weather: end-to-end data-driven weather forecasting*. 2024. arXiv: 2404.00411 [physics.ao-ph].

[84] 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.