Learning data assimilation

Vincent Chabot (Météo-France) Selime Gürol (CERFACS)

ECMWF Annual Seminar 2026

Outline

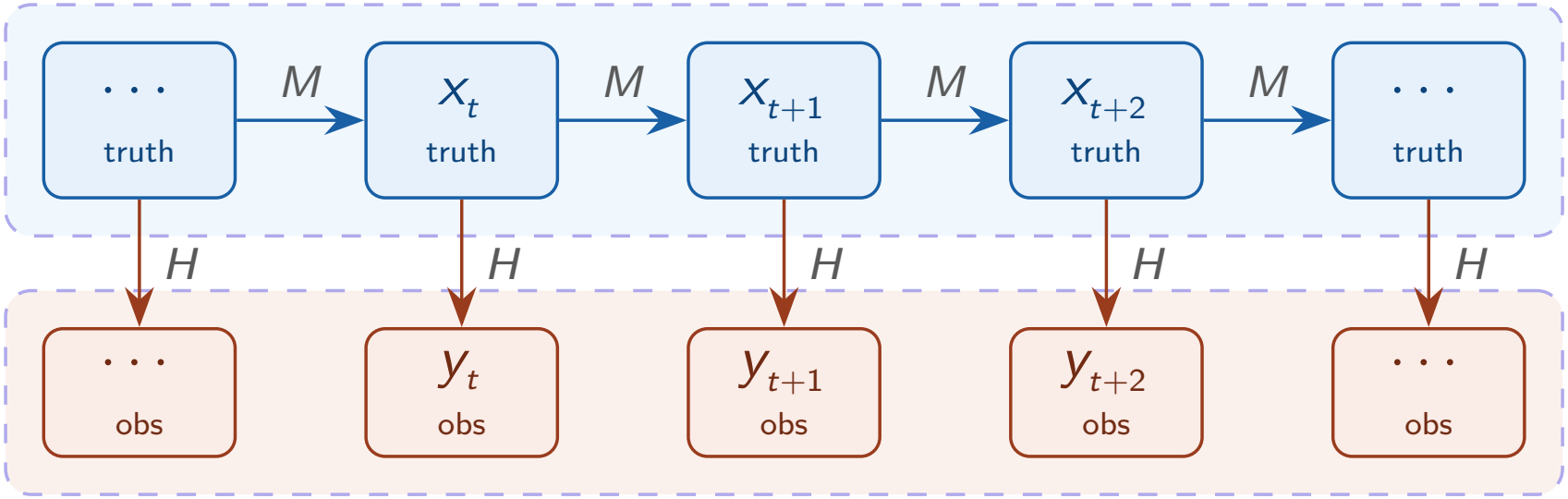
Introduction

Learning DA on idealized, low-dimensional testbeds

Learning DA from analysis, realistic high-dimensional applications

Conclusions & Perspectives

Inverse Problem



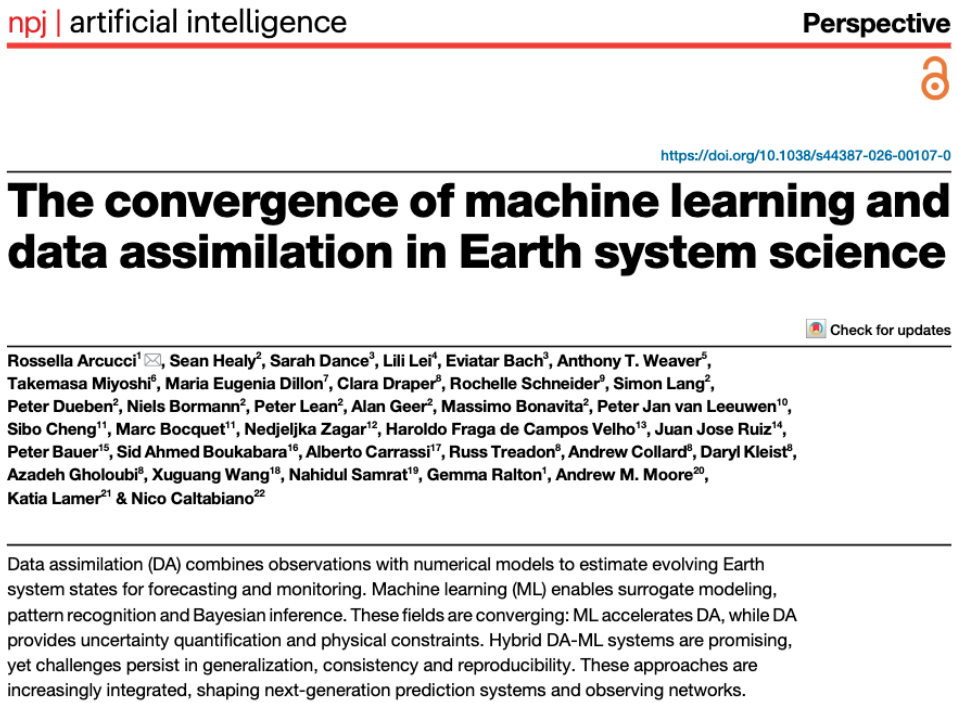
Inverse problem: Given observations **y**, infer the state **x**

- Data Assimilation: Physics + obs. (noisy, sparse) + uncertainties
  $\rightarrow$ Best estimate

- Machine Learning: Obs. (dense)
  $\rightarrow$ Learn patterns/functions

Why use machine learning for DA? [1]

- Improve accuracy
- Reduce computational cost
- Learn complex nonlinear relationships
- Enable new data assimilation paradigms



npj | artificial intelligence

Perspective

https://doi.org/10.1038/s44387-026-00107-0

The convergence of machine learning and data assimilation in Earth system science

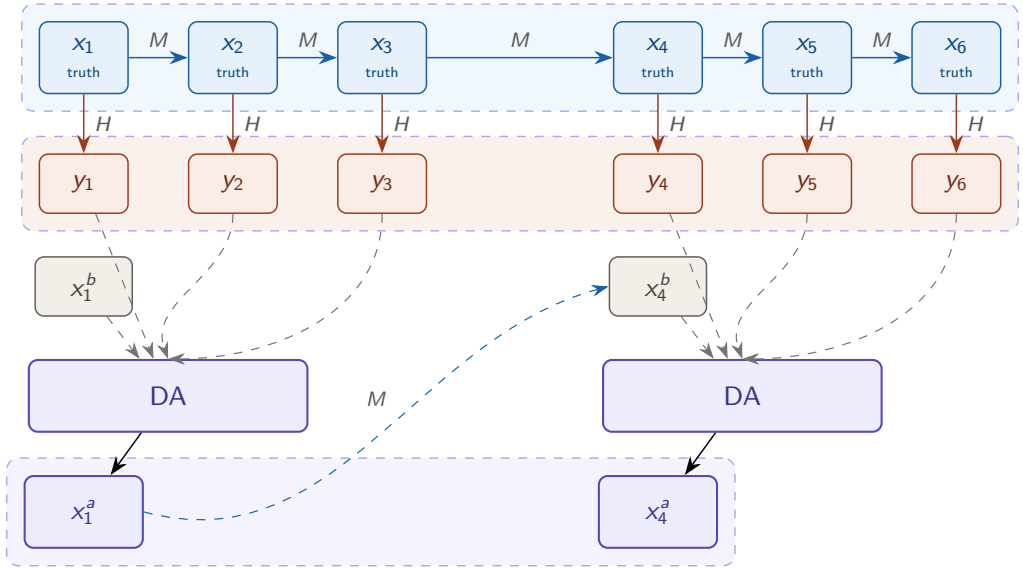
Check for updates

Rossella Arcucci[1] ✉, Sean Healy[2], Sarah Dance[3], Lili Lei[4], Eviatar Bach[3], Anthony T. Weaver[5], Takemasa Miyoshi[6], Maria Eugenia Dillon[7], Clara Draper[8], Rochelle Schneider[9], Simon Lang[2], Peter Dueben[2], Niels Bormann[2], Peter Lean[2], Alan Geer[2], Massimo Bonavita[2], Peter Jan van Leeuwen[10], Sibo Cheng[11], Marc Bocquet[11], Nedjeljka Zagar[12], Haroldo Fraga de Campos Velho[13], Juan Jose Ruiz[14], Peter Bauer[15], Sid Ahmed Boukabara[16], Alberto Carrassi[17], Russ Treadon[8], Andrew Collard[8], Daryl Kleist[8], Azadeh Gholoubi[8], Xuguang Wang[18], Nahidul Samrat[19], Gemma Ralton[1], Andrew M. Moore[20], Katia Lamer[21] & Nico Caltabiano[22]

Data assimilation (DA) combines observations with numerical models to estimate evolving Earth system states for forecasting and monitoring. Machine learning (ML) enables surrogate modeling, pattern recognition and Bayesian inference. These fields are converging: ML accelerates DA, while DA provides uncertainty quantification and physical constraints. Hybrid DA-ML systems are promising, yet challenges persist in generalization, consistency and reproducibility. These approaches are increasingly integrated, shaping next-generation prediction systems and observing networks.

[1] Arcucci et al. (2026). The convergence of machine learning and data assimilation in Earth System Science. npj Artificial Intelligence

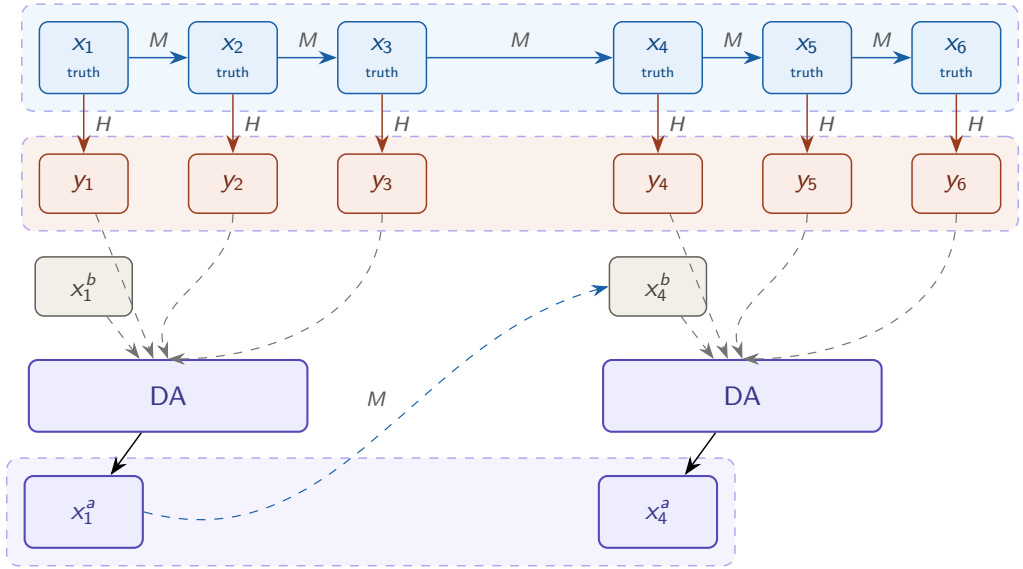
Learning Data Assimilation



- What do we want to learn?
  - analysis (uncertainty-aware)
  - forecast (uncertainty-aware)
  - analysis + forecast jointly (over short or long-time windows)

Goal: improve accuracy and/or reduce computational cost.

Learning Data Assimilation



- What do we want to learn?
  - analysis (uncertainty-aware)
  - forecast (uncertainty-aware)
  - analysis + forecast jointly (over short or long-time windows)

  Goal: improve accuracy and/or reduce computational cost.
- What data can we learn from?
  - Learn from the ground truth data (supervised setting)
  - Learn from the analysis
  - Learn from observations (unsupervised setting)

Outline

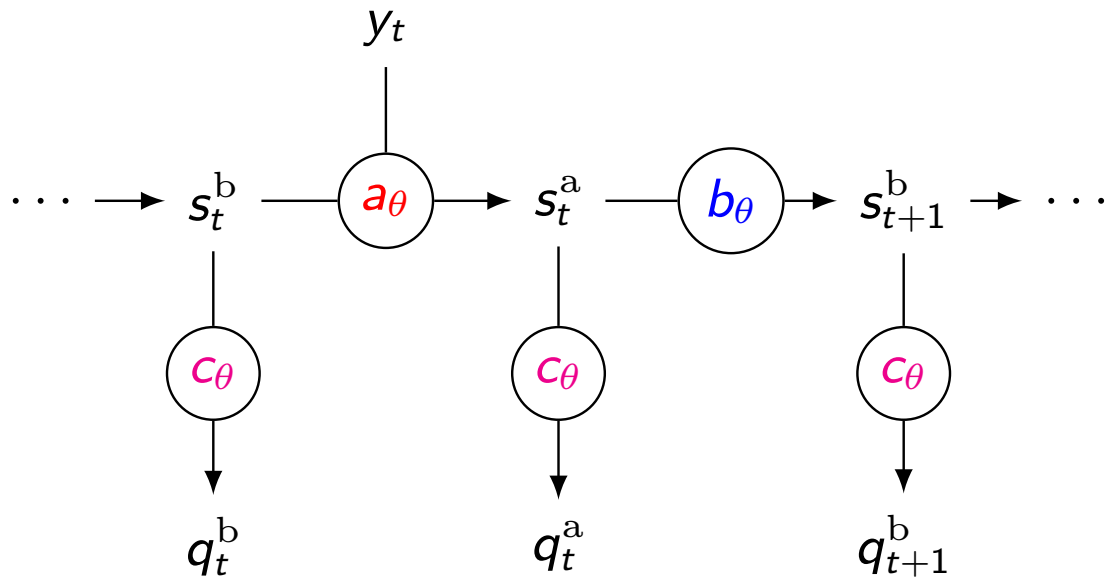
Introduction

Learning DA on idealized, low-dimensional testbeds

Learning DA from analysis, realistic high-dimensional applications

Conclusions & Perspectives

Data Assimilation Networks (DAN) [2]

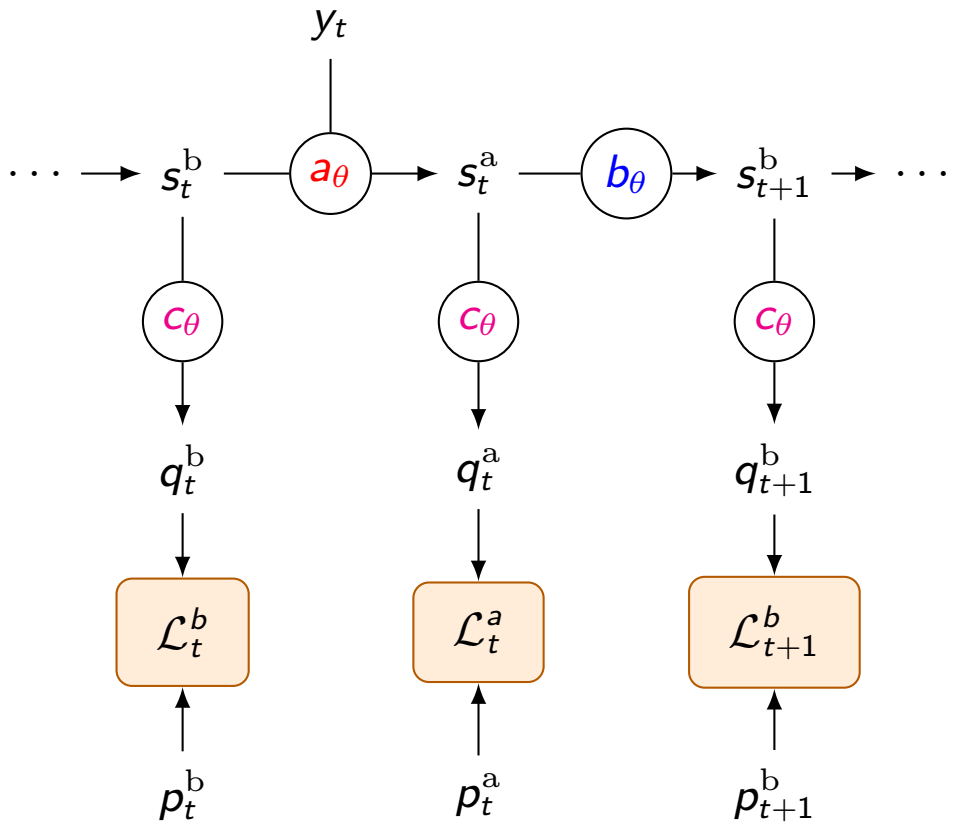


- b_θ: propagator, a_θ: analyzer, s: memory.
- c_θ: the probability coder, transforms an internal representation into an actual pdf.
- Parametrize b_θ, a_θ, c_θ by using Recurrent Neural Networks (RNNs).
- DAN learns probabilistic forecasts and analyses through RNNs. A general framework with no Gaussian assumptions.

[2] Boudier, P., Fillion, A., Gratton, S., Gürol, S., Zhang, S. (2023). Data assimilation networks. Journal of Advances in Modeling Earth Systems, 15

DAN - Loss Function

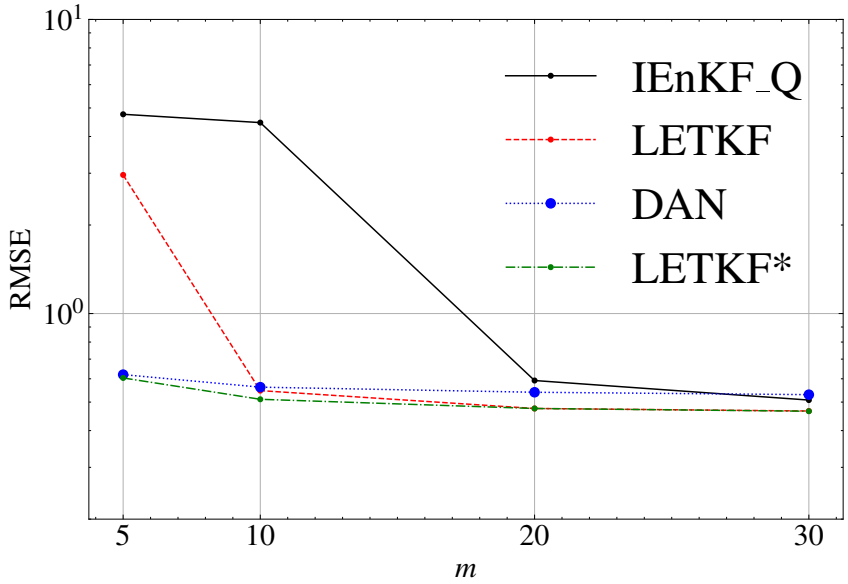
- Supervised setting: Training minimizes the discrepancy between predicted and true probability distributions throughout the DA cycle.



- Loss function based on cross entropy:

$$\mathcal{L}(\theta) = \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} KL(p_t^a|q_{t,\theta}^a) + KL(p_t^b|q_{t,\theta}^b)\right]$$

Numerical experiments with Lorenz-95

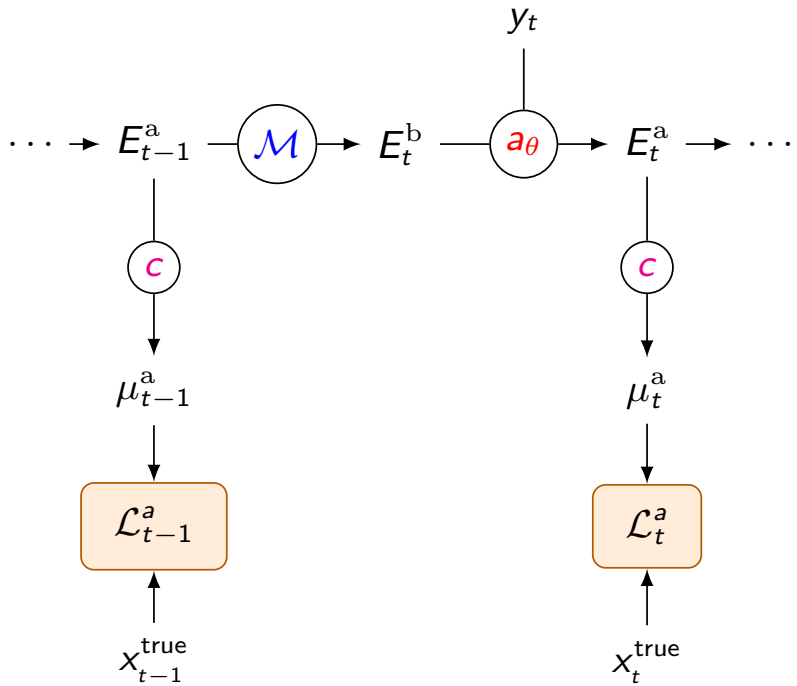


- Performance comparable to tuned LETKF, which includes regularization techniques such as localization and inflation.
- Robust forecast and initialization.
- Demonstrates the potential of ML-based DA.

Next Steps:

- Extension to an unsupervised setting (ongoing PhD, CERFACS/IRIT)
- Adaptation to large-scale systems (post-doc will start in autumn, Météo-France/IRIT)

Learning analysis of EnKF [3] [4]



- Idea: Learn the EnKF analysis increment directly. Parametrize a_θ by using a simple residual CNN,

$$E^a_t = E^b_t + a_\theta(E^b_t, H_t^T R_t^{-1}(y_t - H_t(\mu^b_t)))$$

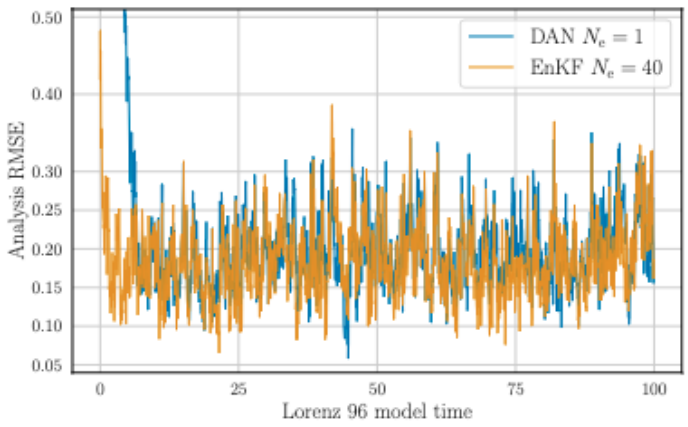
- Loss function is based on MSE through cycles:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \|x^{\mathrm{true}}_t - \mu^a_t(\theta)\|_2^2$$

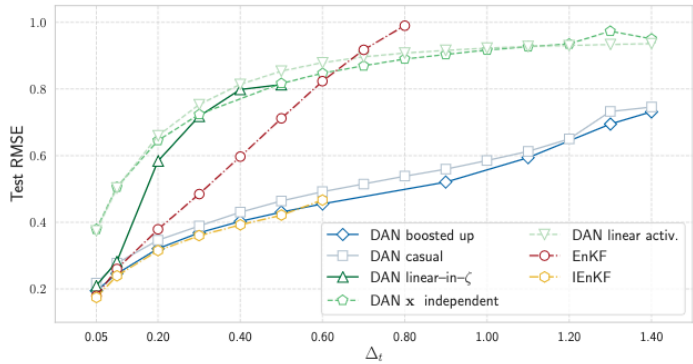
[3] M. Bocquet et al. Accurate deep learning-based filtering for chaotic dynamics by identifying instabilities without an ensemble. Chaos, 2024.

[4] M. Bocquet et al. Elucidating the performance of data assimilation neural networks for chaotic dynamics. EGUsphere, 2026.

Numerical experiments with Lorenz-95



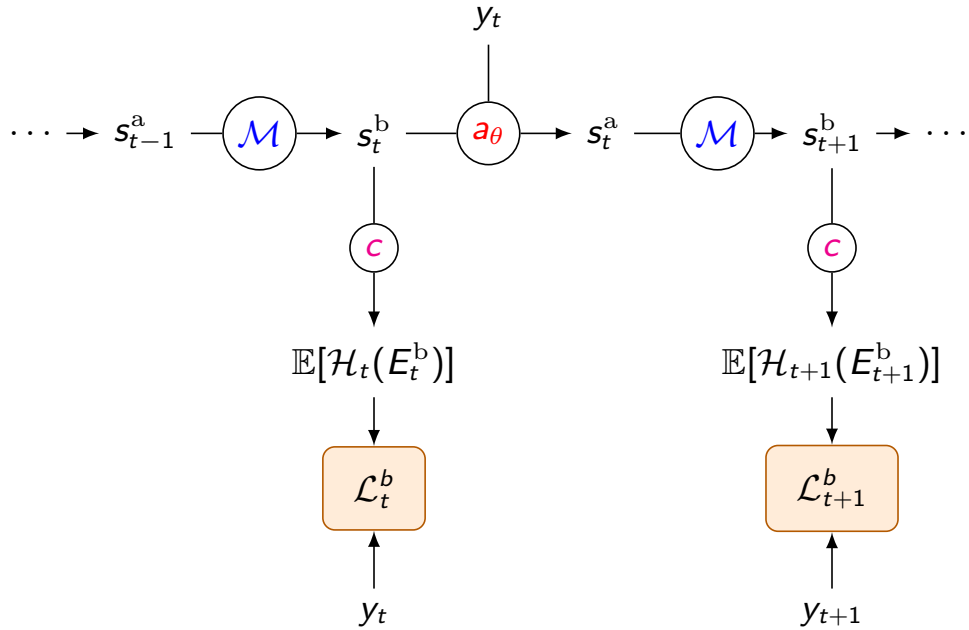
Ref: Bocquet et al., 2024



Ref: Bocquet et al., 2026

- $H = R = I$.
- The performance of DAN approaches that of the best possibly tuned EnKF (same conclusion as before).
- The performance of a_θ barely depends on the ensemble size. DAN implicitly learns the error covariance matrix associated with the analysis.
- Performance of DAN in stronger nonlinear regimes are tested, DAN handles nonlinearity as good as the IEnKF.

Learning the analysis of EnKF [5]



- $s = (E, Z)$ contains ensemble members and recurrent memory
- $c(s) = \mathbb{E}[\mathcal{H}(E^b)]$
- Parametrize a_θ by using multilayer neural network with recurrent memory:

$$(E^a_t, Z^a_t) = a_\theta \left(E^b_t, \Sigma^b_t, y_t, g(E^b_t), Z^b_{t-1} \right)$$

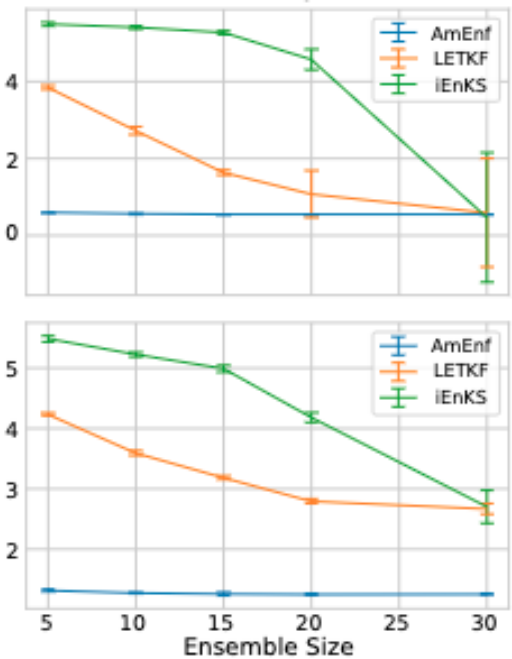
  $\to$ Local dynamics tendency has been included.

- Loss function is calculated through cycles:

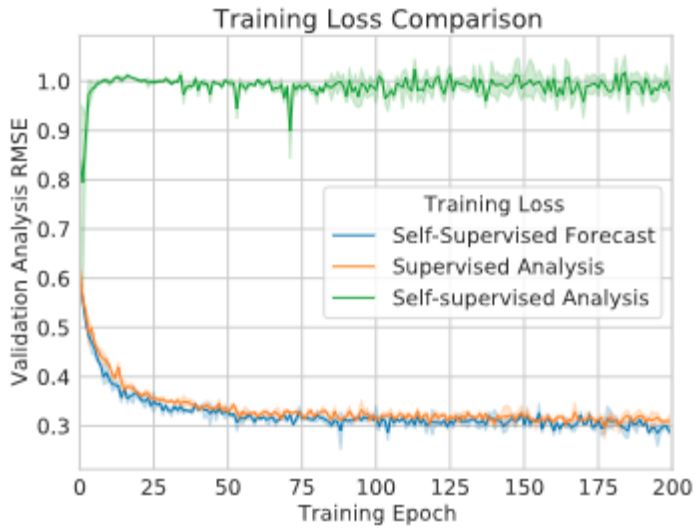
$$\mathcal{L}(\theta) = \frac{1}{T-1} \sum_{t=1}^{T-1} \| y_{t+1} - c(\mathcal{M}(s^a_t(\theta))) \|_2^2$$

[5] McCabe, M. and Brown, J. (2021). Learning to assimilate in chaotic dynamical systems. *NeurIPS*

Numerical experiments with Vissio-Lucarini 20



Test analysis RMSE for $\sigma_o = 1$ and $\sigma_o = 2.5$



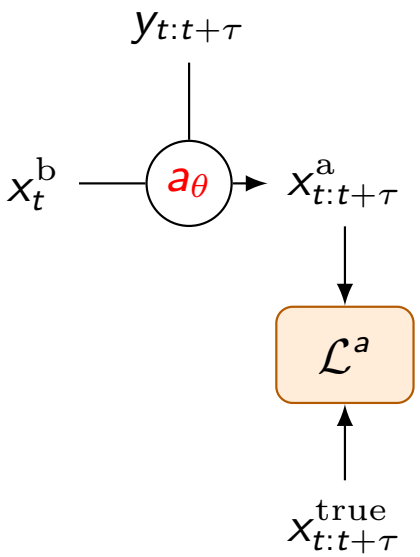
Analysis RMSE, $\sigma_o = 1$

$$\mathcal{L}(\theta) = \frac{1}{T-1} \sum_{t=1}^{T-1} \|x_t^{\text{true}} - \mathbb{E}[E_t^a]\|_2^2$$

$$\mathcal{L}(\theta) = \frac{1}{T-1} \sum_{t=1}^{T-1} \|y_t - c(s_t^a)]\|_2^2$$

- The NN model match or outperform the common used algorithms with more significant improvements at smaller ensemble sizes and higher noise levels.
- Including a differentiable simulation is a key component of the loss function!
- Limitation: It assumes the availability of adjoints and result in costly training process.

Learning increment of VarDA[5]



- a_θ is based on weak-constraint 4DVar with diagonal covariance matrices. No background term.
- a_θ uses the gradients of the weak-constraint 4DVar to update the initial state via LSTM or CNN:

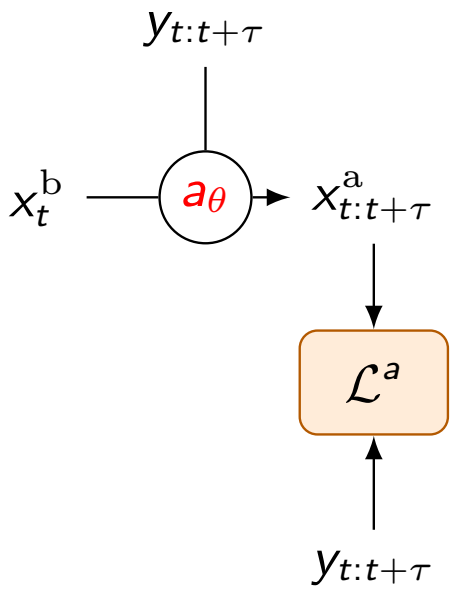
$$x^{(k+1)} = x^{(k)} - a_\theta(\nabla_x \mathcal{J}_{var}(x^{(k)}))$$

- a_θ can use NN-based $\mathcal{M}_\alpha(\cdot)$.

- Supervised setting: $\mathcal{L}^a(\theta) = \|x^a(\theta) - x^{\mathrm{true}}\|_2^2$
- Joint learning of the physical model and analysis step improves the analysis estimate.
- *"Numerical ODE-based representations optimal in terms of forecasting performance may not be optimal for reconstruction purpose".*

[5] R. Fablet et al. Learning variational data assimilation models and solvers. Journal of Advances in Modeling Earth Systems, 2021.

Learning increment of VarDA [6]



- Unsupervised setting:

$$\mathcal{L}^a(\theta) = \lambda_1 \|\mathcal{H}(x^a(\theta)) - y\|^2 + \lambda_2 \|x^a(\theta) - \mathcal{M}_\alpha(x^a(\theta))\|_2^2$$

- Joint learning in this setting does not provide good results! We need a physical model $\mathcal{M}$.
- Faster than steepest descent, but not compared with Gauss-Newton.

[6] R. Fablet et al. Learning variational data assimilation models and solvers. Journal of Advances in Modeling Earth Systems, 2021.

Summary

- Literature is quite rich with proof-of-concept using toy problems based on supervised setting (ground truth).
  - Theoretical insights (convergence results, different look to the existing algorithms, coupling effect between model and DA parameters).
  - The loss function is a key ingredient for successful learning.
- Learning over DA cycles consistently improves performance.
- Semi-supervised learning (observations + physical model) is promising.
- Fully unsupervised DA remains an open research challenge.

Key questions:

- What can we transfer from these theoretical insights to practical applications?
- How these results can guide the practical applications?

Outline

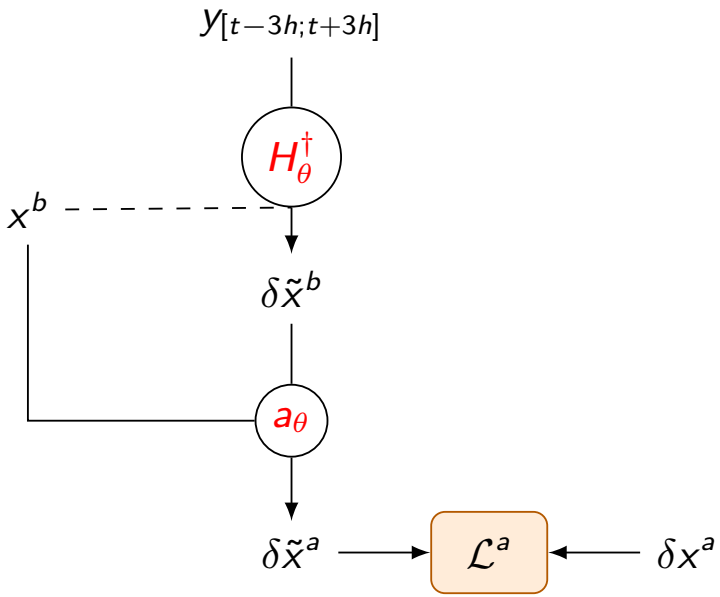
Introduction

Learning DA on idealized, low-dimensional testbeds

Learning DA from analysis, realistic high-dimensional applications

Conclusions & Perspectives

Learning analysis increment [7]



- $\mathcal{H}_\theta^\dagger$: Data-driven retrieval from observations. Fixed during a_θ training.
- Background field is taken at observation time and location.
- $\delta\tilde{x}^b$: Increment between projected observation and background.
- Loss function:

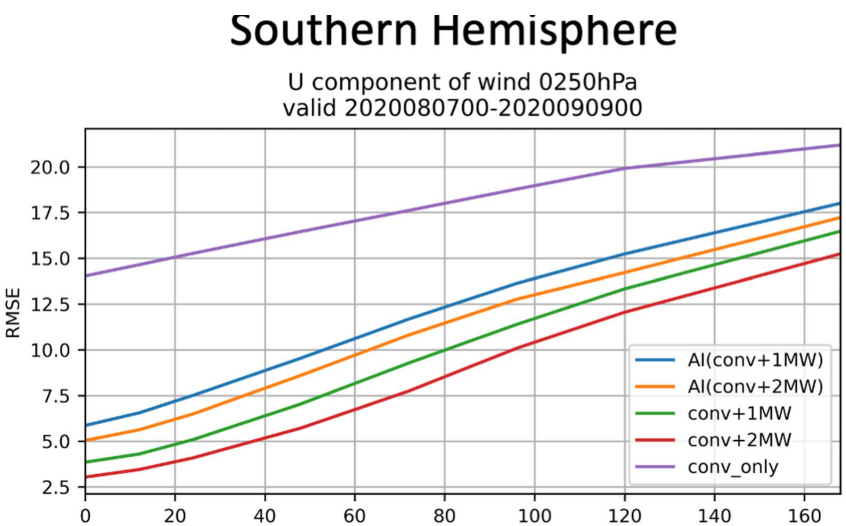
$$\mathcal{L}^a(\theta) = \|\delta\tilde{x}^a - \delta x^a\|_2^2$$

- a_θ is trained to learn the analysis increment δx^a from a background field and a satellite observation-driven predicted-increments.

[7] Maddy et al., Assessing the feasibility of an NWP satellite data assimilation system entirely based on AI techniques, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024.

System performance: pseudo-observations in GDAS

- AI-DA analysis (using one or two sensors) is used as pseudo-observations of u, t, v, r, z (for the initialization).
- Assimilated jointly with conventional observations in the operational GDAS system (analysis, forecast, and ensemble are all cycled).

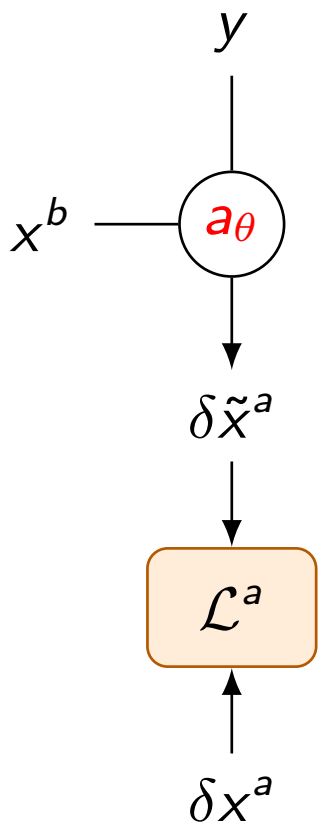


Ref: Maddy et al., 2024

- The observation analysis departure for satellite sensors is decreased when satellite observations are assimilated (wrt to conventional only).
- AI-DA is less performant than classical DA with the same number of sensors.
- Performance is better for Southern Hemisphere.

Learning AROME analysis increment. Preliminary experiments with geostationary observations [8]

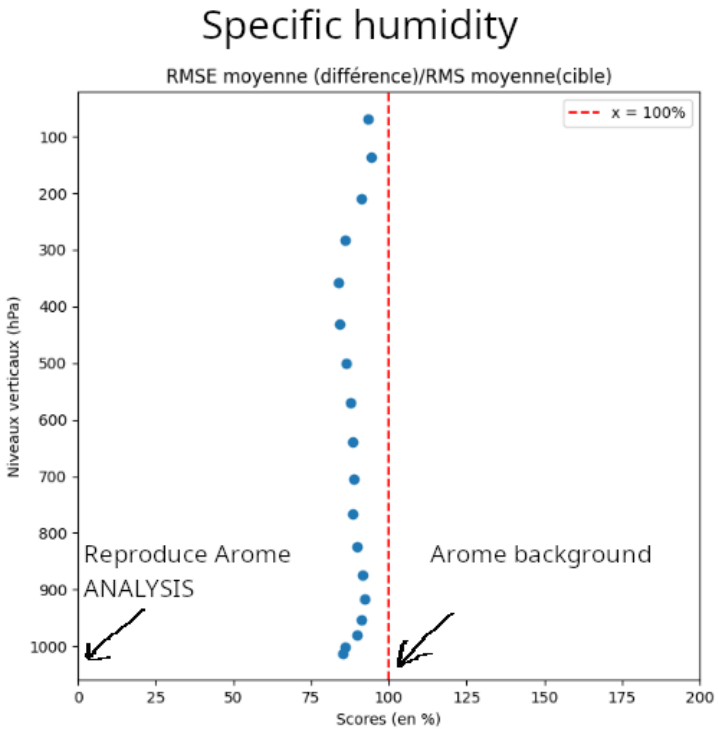
M H B R



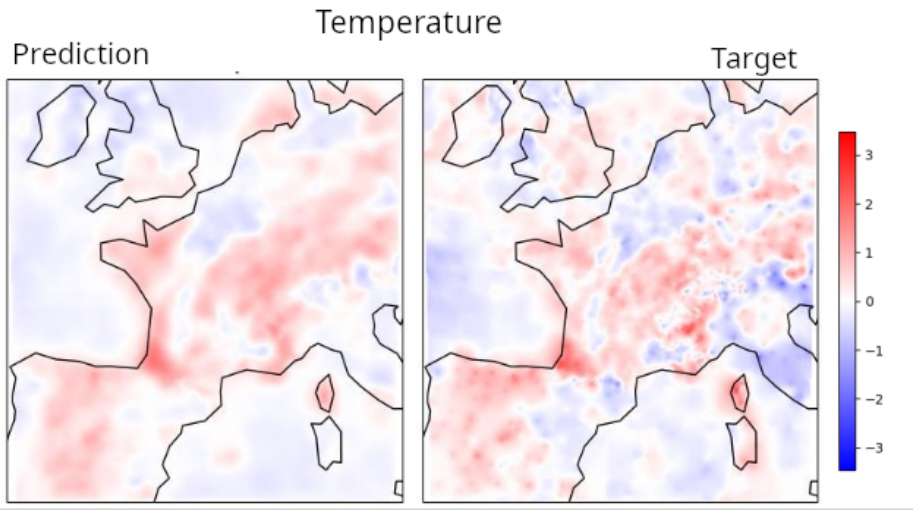
- Operational AROME background and analysis (3DVar)
- Geostationary observations : SEVIRI channels
- On 20 hybrid level at low resolution (5.4 km)

[8] A. Perello-Achfari, Emulation de l'assimilation pour AROME. Etude de cas avec les données SEVIRI ,Intership report, 2025

Learning AROME analysis increment. Preliminary experiments with geostationary observations

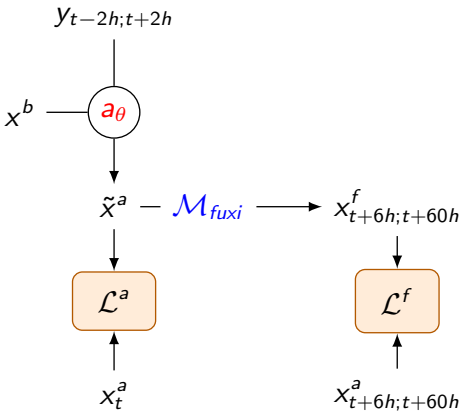


- Analysis increment vertical localisation consistent with the observations used
- Impact on u and v while not observed



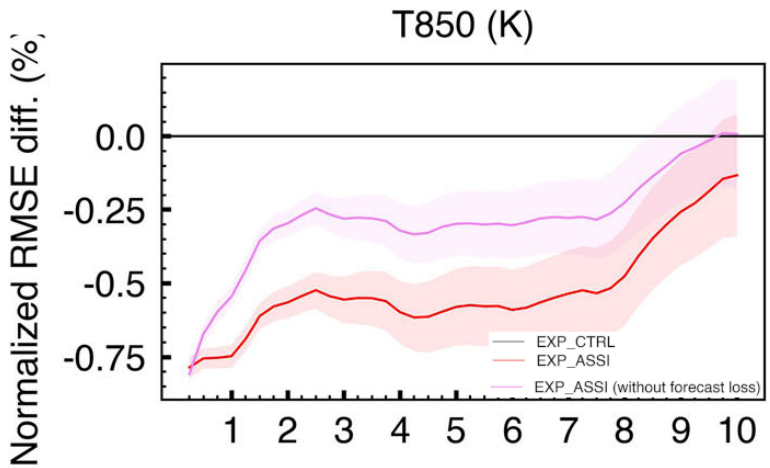
- Increment too smooth
- Large scale structure is learnt.

FuXI-DA : Learning ERA5 analysis increment. Test with Geostationary observation [9]



- background is obtained using Fuxi model, $\mathcal{M}_{fuxi}$ initialized with ERA5 reanalysis.
- 8 observations timestep are integrated of AGRI instrument aboard of Fengyun- 4B satellite.

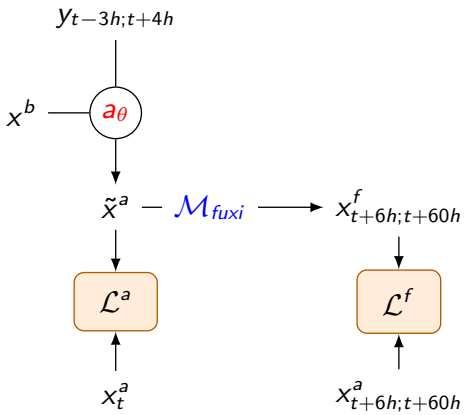
$$\mathcal{L}(\theta) = \|\tilde{x}^a - x^a_t\|_1 + \frac{1}{T}\sum_{k=1}^{T}\left\|\mathcal{M}^k_{fuxi}(\tilde{x}^a) - x^a_{t+k}\right\|_1$$



- Experiments with forecast loss lead to a bigger RMSE decrease than without.

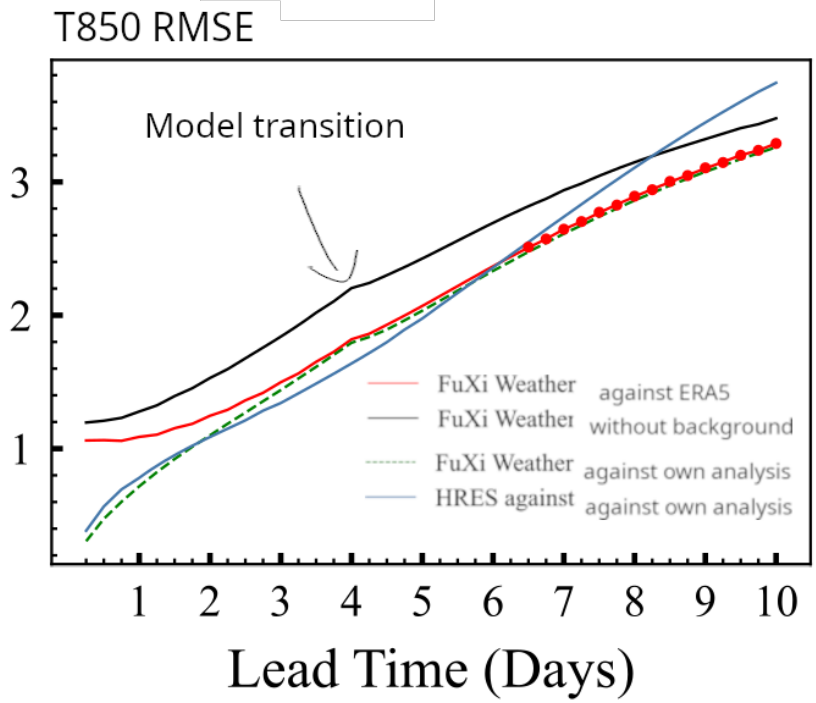
[9] Xu et al., FuXi-DA: A generalized deep learning data assimilation framework for assimilating satellite observations, npj Clim Atmos, 2025.

FuXIWeather : Learning analysis increment [10]



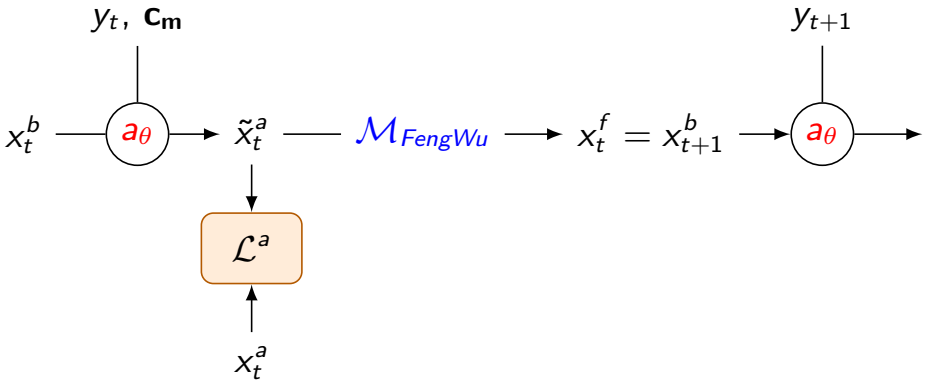
- MW and GNSSRO observations are used
- Training: Background comes from Fuxi forecast (using ERA5) with a $[6h; 120h]$ leadtime
- Forecast and DA had been cycle for a year
- FuxiWeather analysis exhibit error similar to 42h Fuxi forecast starting from ERA5 (compared to ERA5).

Sparse/Missing observation : Masked after normalisation.
Mask is also provided as a supplementary input.



[10] Sun et al., A data-to-forecast machine learning system for global weather, Nature Communications, 2025

FengWu-ADAS [11]

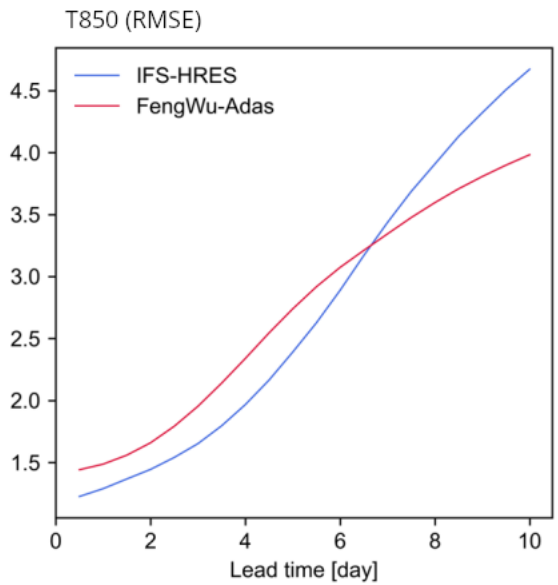


Background (x^b) comes from forecast made with previous analysis.

Sparse/Missing observation: Fill by 0. A confidency map $\mathbf{c_m}$ is provided to the network (range [0;1]). Used as a multiplicative factor in convolutions layers.

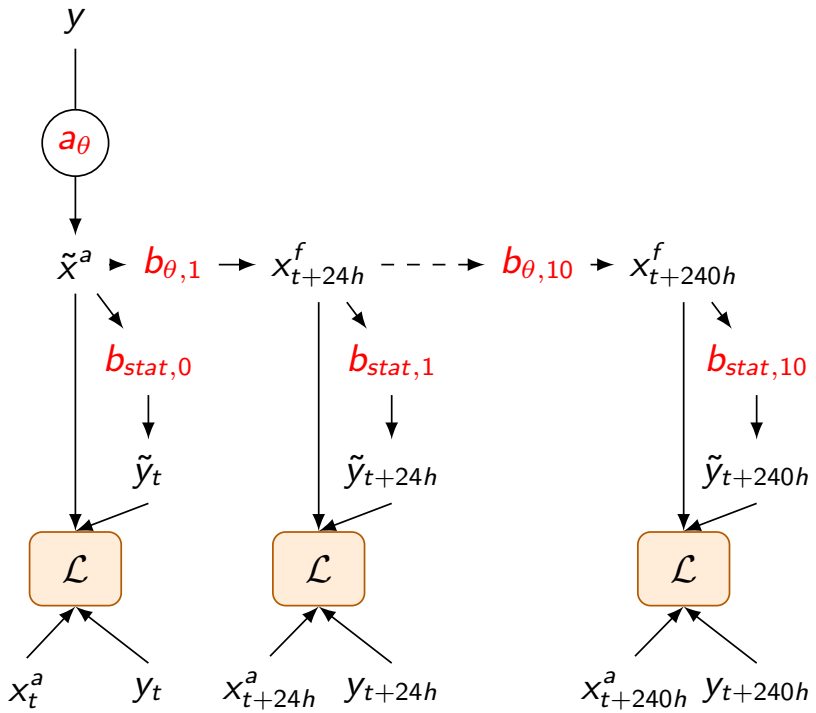
On synthetic test case, cyclic training show higher performance wrt single-step training.

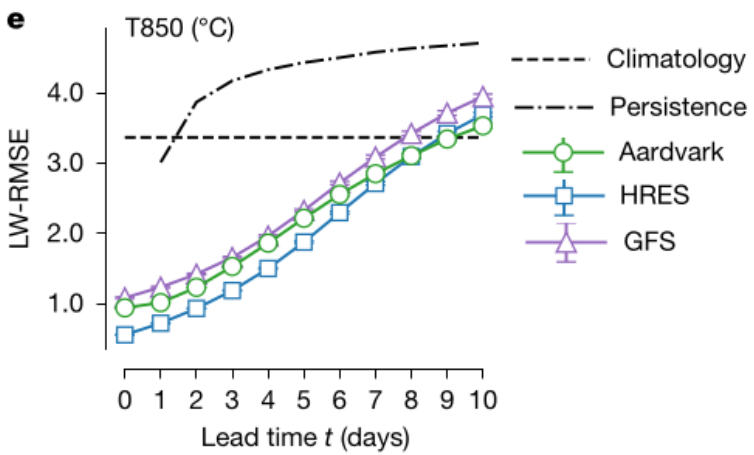
Observation used (real test case) : conventional ones.



[11] K. Chen et al., Towards an end-to-end artificial intelligence driven global weather forecasting system, Preprint, 2025

Aardvark [12]

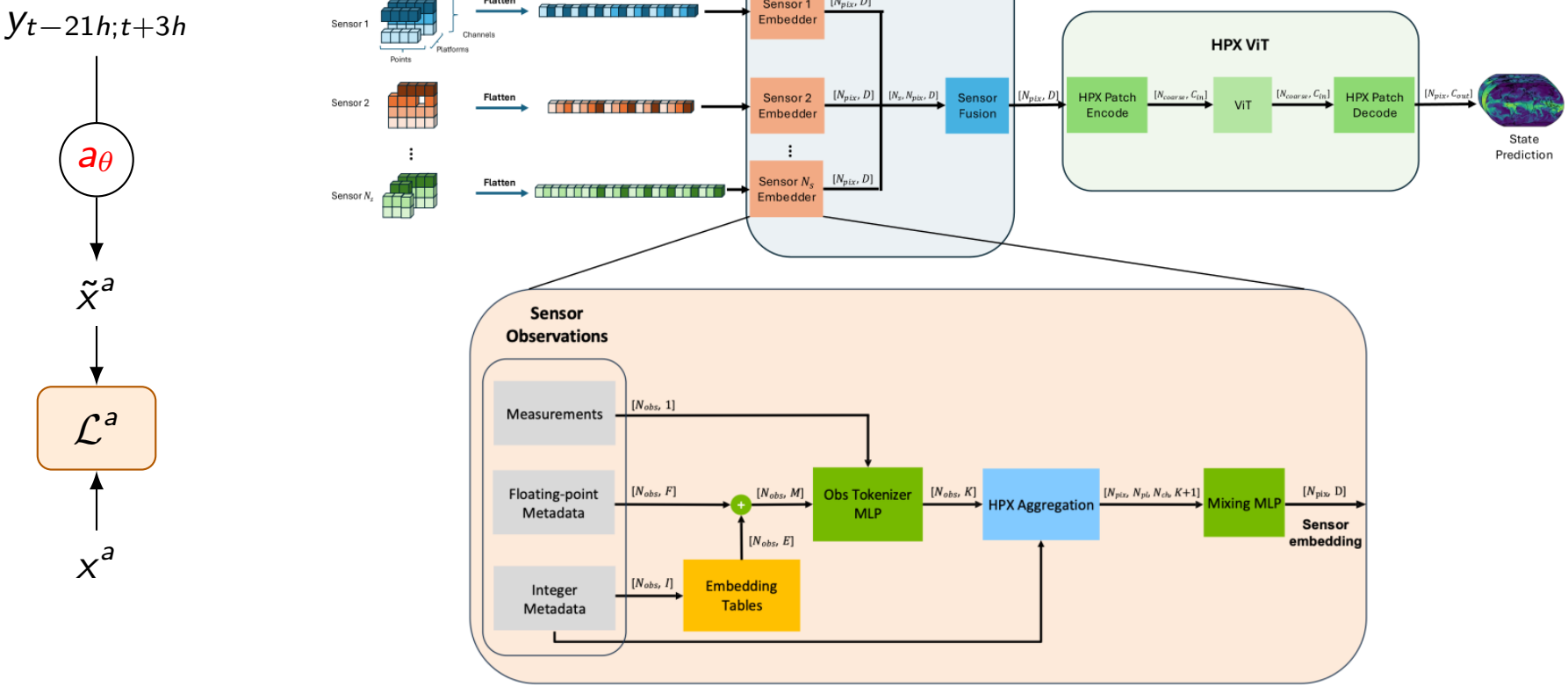




- One observation per "channel" per assimilation window : Aggregation of available observation over a 24h window.
- Observation : Conventional, ASCAT, MW, IR (HIRS, IASI (PCA)), Geostationnary (Gridsat).
- Sparse/Missing data : SetConv layer (Gordon et al. 2020)
- Ablation study has been performed.

[12] A. Allen et al., End-to-end data-driven weather prediction, Nature, 2025

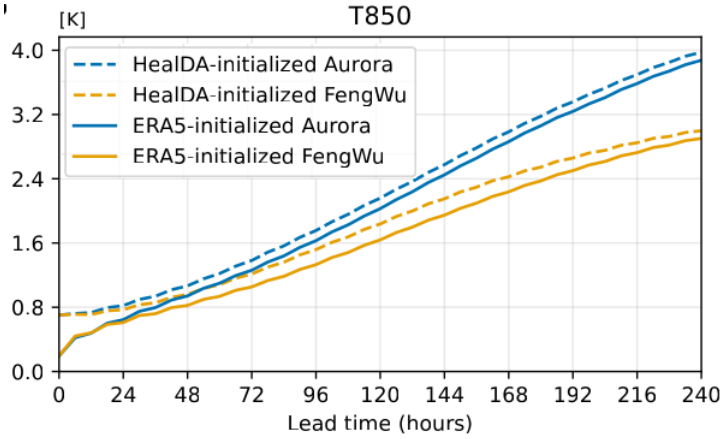
HealDA [13]



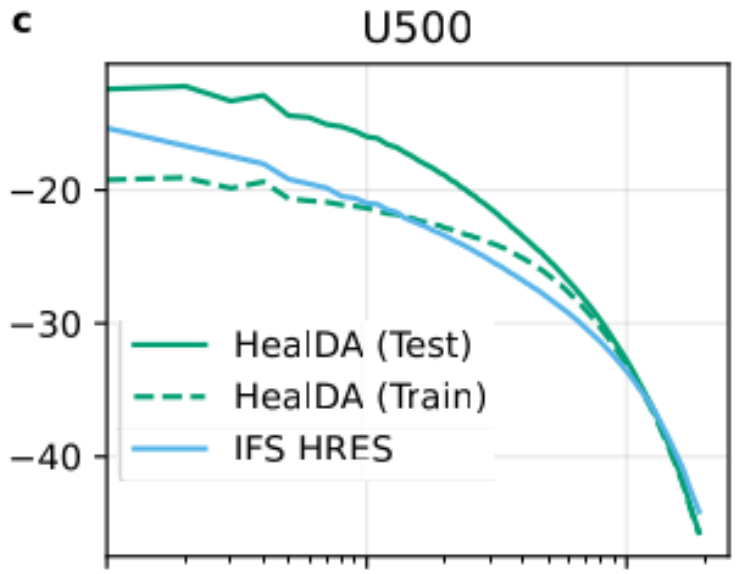
All observations are passed as single measurement with metadata to a specific encoder. The encoder output are latent feature on the HealPix Grid.

[13] A. Gupta et al., HealDA: Highlighting the importance of initial errors in end-to-end AI weather forecasts, Preprint, 2026

HealDA



RMSE wrt to ERA5



Analysis spectral error decomposition (with respect to ERA5)

- Able to initiate different state of the art AI emulator (Fourcastnet, Aurora, FengWu).
- A 12-24h shift in predictability is observed using HealDa instead of ERA5 analysis.

Large scale overfitting, especially for upper-air variables.

Comparison

	Conv.	IR	MW	Geo.	GNSSRO	AMV	Scatt.
[Maddy et al., 2024]	X	X	yes	X	X	X	X
Perello internship	X	X	X	yes	X	X	X
[Xu et al., 2025]	X	X	X	yes	X	X	X
[Chen et al., 2025]	yes	X	X	X	X	X	X
[Allen et al., 2025]	yes	yes	yes	yes	X	X	yes
[Sun et al., 2025]	X	X	yes	X	yes	X	X
[Gupta et al., 2026]	yes	X	yes	X	yes	yes	yes

Table: Summary of observation type used in several studies.

	Training	Test
[Maddy et al., 2024]	18 months	3 months (end 2020)
[Sun et al., 2025]	13 month	1 year (July 2023 -)
[Gupta et al., 2026]	2000-2021	2022
[Allen et al., 2025]	Processor : 1979 - 2017 FullModel : 2007-2017	2018
[Chen et al., 2025]	2012-2016	2017

Outline

Introduction

Learning DA on idealized, low-dimensional testbeds

Learning DA from analysis, realistic high-dimensional applications

Conclusions & Perspectives

Conclusions

- Potential to improve analysis accuracy and/or reduce computational cost using AI.

- Still considerable progress needed before large-scale operational deployment.

- **Key challenges:**
  - Generalization, explainability, physical consistency, uncertainty estimation, handling sparse and noisy data
  - Gap between research and operations
  - Growing literature, but not always easily comparable
  - Need for common platforms, dataset, benchmarking ?

Perspectives

- DAN extension to the unsupervised setting
- DAN implementation for ARPEGE
- Learning ARPEGE analysis within ANEMOI framework
- DAN for AROME reanalysis (ClimateNext Project)







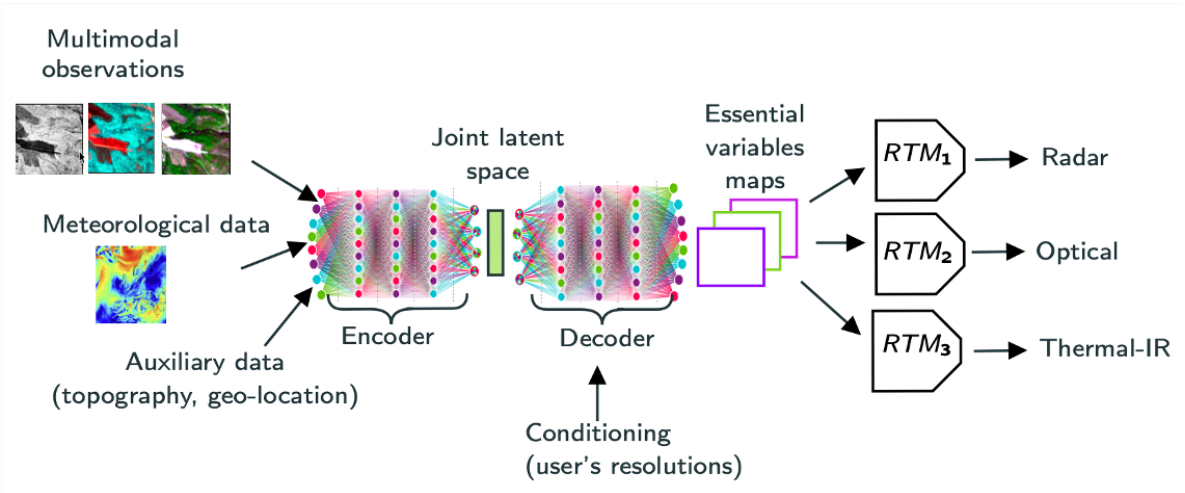






A related effort: REpresentation Learning for Earth Observation (RELEO)

- Industrial chair within ANITI (Artificial and Natural Intelligence Toulouse Institute).
- Academic partners: CNES, CESBIO, IRIT, ISAE
- Industrial partners: CERFACS, CD Sopra Steria, Magellium, Thales Alenia, Thales Services Numériques



- Learns directly from satellite observations to estimate essential climate and biodiversity variables used for land surface monitoring.
- Shared open challenges: multi-modal data, uncertainty estimation, generalization
- Continual Learning: Continuously update models with new EO data using DA.

Thank you for your attention !

References I



Allen, A., Markou, S., Tebbutt, W., Requeima, J., Bruinsma, W. P., Andersson, T. R., Herzog, M., Lane, N. D., Chantry, M., Hosking, J. S., et al. (2025).

End-to-end data-driven weather prediction.

Nature, 641(8065):1172–1179.



Arcucci, R., Healy, S., Dance, S., et al. (2026).

The convergence of machine learning and data assimilation in earth system science.

npj Artificial Intelligence, 2:48.



Bocquet, M., Farchi, A., Finn, T. S., Durand, C., Cheng, S., Chen, Y., Pasmans, I., and Carrassi, A. (2024).

Accurate deep learning-based filtering for chaotic dynamics by identifying instabilities without an ensemble.

Chaos: An Interdisciplinary Journal of Nonlinear Science, 34(9):091104.

References II



Bocquet, M., Finn, T. S., Cheng, S., and Farchi, A. (2026).

Elucidating the performance of data assimilation neural networks for chaotic dynamics.

EGUsphere.



Boudier, P., Fillion, A., Gratton, S., Gürol, S., and Zhang, S. (2023).

Data assimilation networks.

Journal of Advances in Modeling Earth Systems, 15(4):e2022MS003353.



Chen, K., Bai, L., Ling, F., Ye, P., Chen, T., Luo, J.-J., Chen, H., Xiao, Y., Chen, K., Han, T., et al. (2025).

Towards an end-to-end artificial intelligence driven global weather forecasting system.

arXiv preprint arXiv:2312.12462.

References III



Cheng, S., Quilodrán-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard, J., Xiao, D., Janjić, T., Ding, W., Guo, Y., Carrassi, A., Bocquet, M., and Arcucci, R. (2023).

Machine learning with data assimilation and uncertainty quantification for dynamical systems: A review.

IEEE/CAA Journal of Automatica Sinica, 10(6):1361–1387.



Fablet, R., Chapron, B., Drumetz, L., Mémin, E., Pannekoucke, O., and Rousseau, F. (2021).

Learning variational data assimilation models and solvers.

Journal of Advances in Modeling Earth Systems, 13(10):e2021MS002572.



Faucheron, R., Valero, S., Gürol, S., and Poulain, V. (2026).

Probabilistic representation learning of satellite image time series via perceiver-vae.

In *Proceedings of the 2026 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Washington, DC, USA. IEEE.

Accepted for publication.

References IV



Gupta, A., Subramaniam, A., Pritchard, M. S., Kashinath, K., Frolov, S., Lieberman, K., Miller, C., Silverman, N., and Brenowitz, N. D. (2026).

Healda: Highlighting the importance of initial errors in end-to-end ai weather forecasts.

arXiv preprint arXiv:2601.17636.



Maddy, E. S., Boukabara, S. A., and Iturbide-Sanchez, F. (2024).

Assessing the feasibility of an nwp satellite data assimilation system entirely based on ai techniques.

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17:9828–9845.



McCabe, M. and Brown, J. (2021).

Learning to assimilate in chaotic dynamical systems.

In *Advances in Neural Information Processing Systems (NeurIPS)*, NeurIPS '21.

References V



Sun, X., Zhong, X., Xu, X., Huang, Y., Li, H., Neelin, J. D., Chen, D., Feng, J., Han, W., Wu, L., et al. (2025).

A data-to-forecast machine learning system for global weather.

Nature Communications, 16(1):6658.



Xu, X., Sun, X., Han, W., Zhong, X., Chen, L., Gao, Z., and Li, H. (2025).

Fuxi-da: A generalized deep learning data assimilation framework for assimilating satellite observations.

npj Climate and Atmospheric Science, 8(1):156.