Building a performant hybrid model framework for the IFS

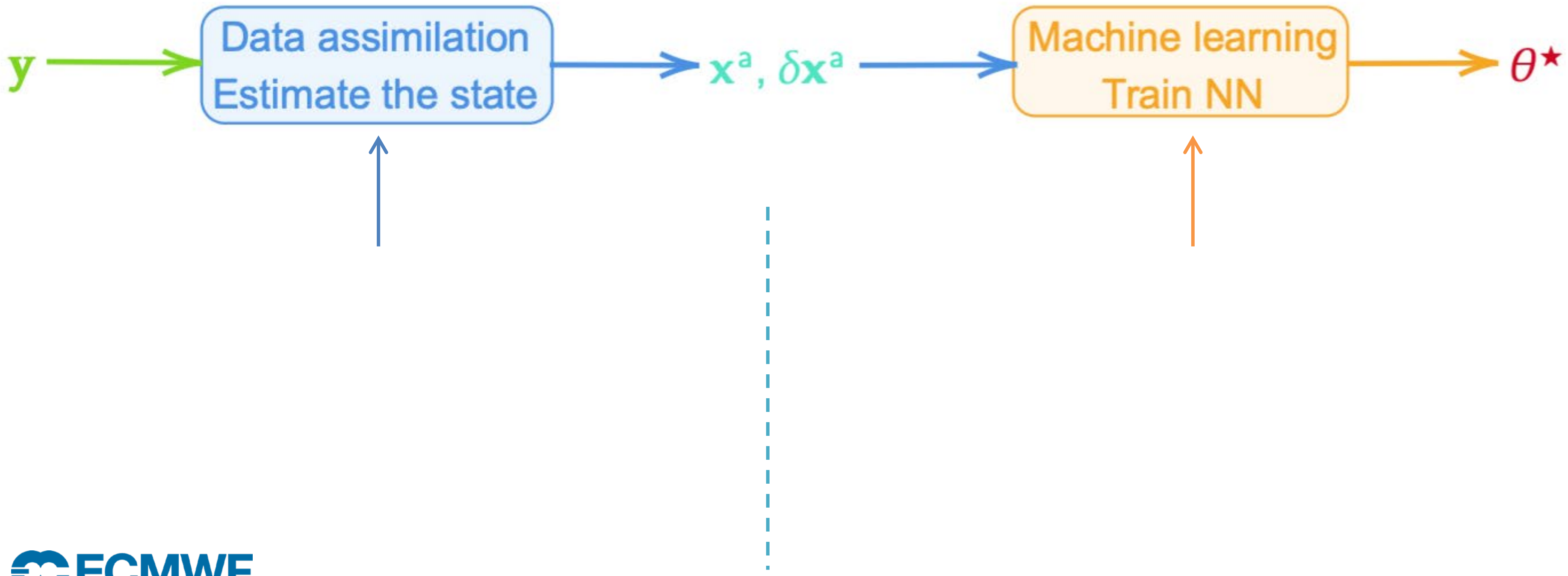
Alban Farchi[1], Marcin Chrust[1]
Marc Bocquet[2], Massimo Bonavita[1], Patrick Laloyaux[1], Wei Pan[1]

[1]ECMWF
[2]CEREA, ENPC

5th ECMWF-ESA Machine Learning Workshop, 13-17 Apr 2026

The early days of hybrid modelling

- The predictable part of analysis increments can be attributed to model errors
- First version of the hybrid model, a.k.a. offline learning (Farchi et al, 2021a):
  - Train a NN to predict the analysis increments.
  - Include the NN in the forecasting system and use the NN output as a forcing throughout the forecast.

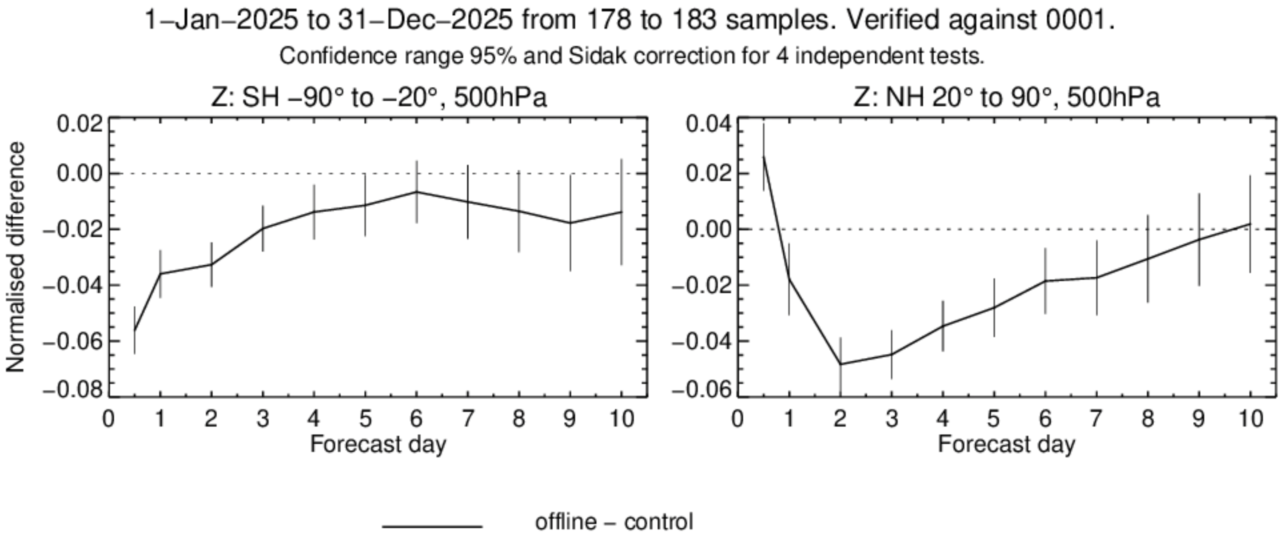


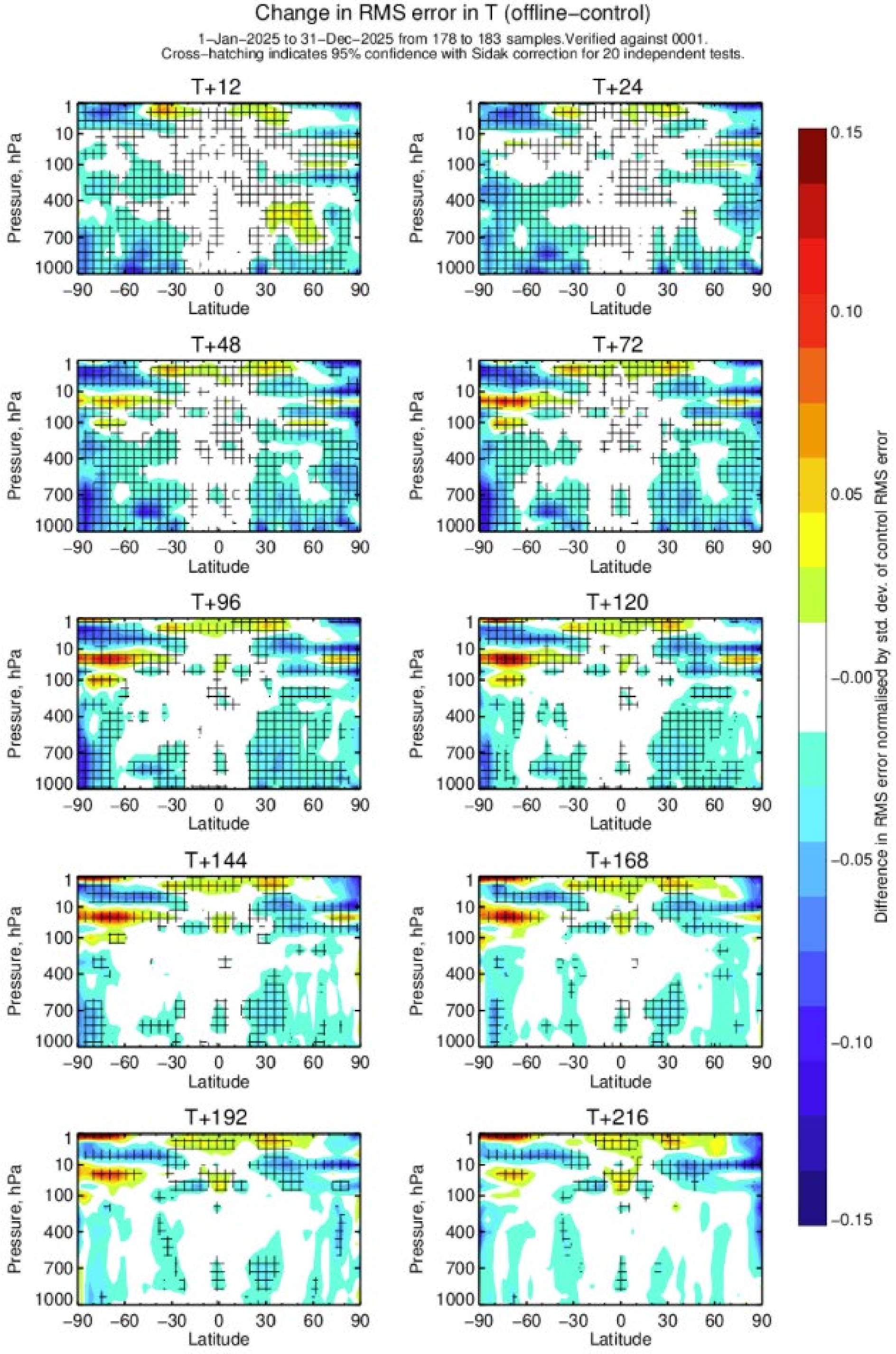
A NN correction for the IFS

- 4 atmospheric variables in the same NN: t, lnsp, vo, d
- Vertical / column architecture: atmospheric columns are processed independently (Bonavita and Laloyaux, 2020, Farchi et al. 2025)

Accuracy of the offline hybrid model

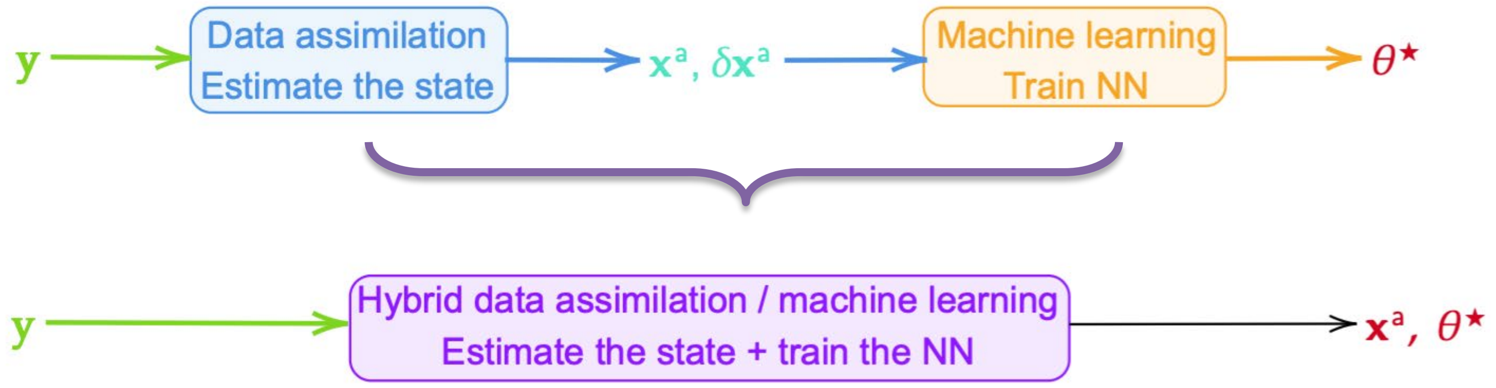
- NN trained using 3 years (2021-2023) of operational increments at large scale (T15)
- Forecast-only experiments in 2025 (unseen data) at TCo399





Limitations of offline learning

- Major limitations of offline learning:
  - Analysis increments are taken as proxy for forecast errors that develop over one DAW.
  - Can significantly differ from the instantaneous model errors.
- Move towards online learning (Farchi et al. 2021, 2023):
  - Interactions between the forecast model and the NN can be taken into account

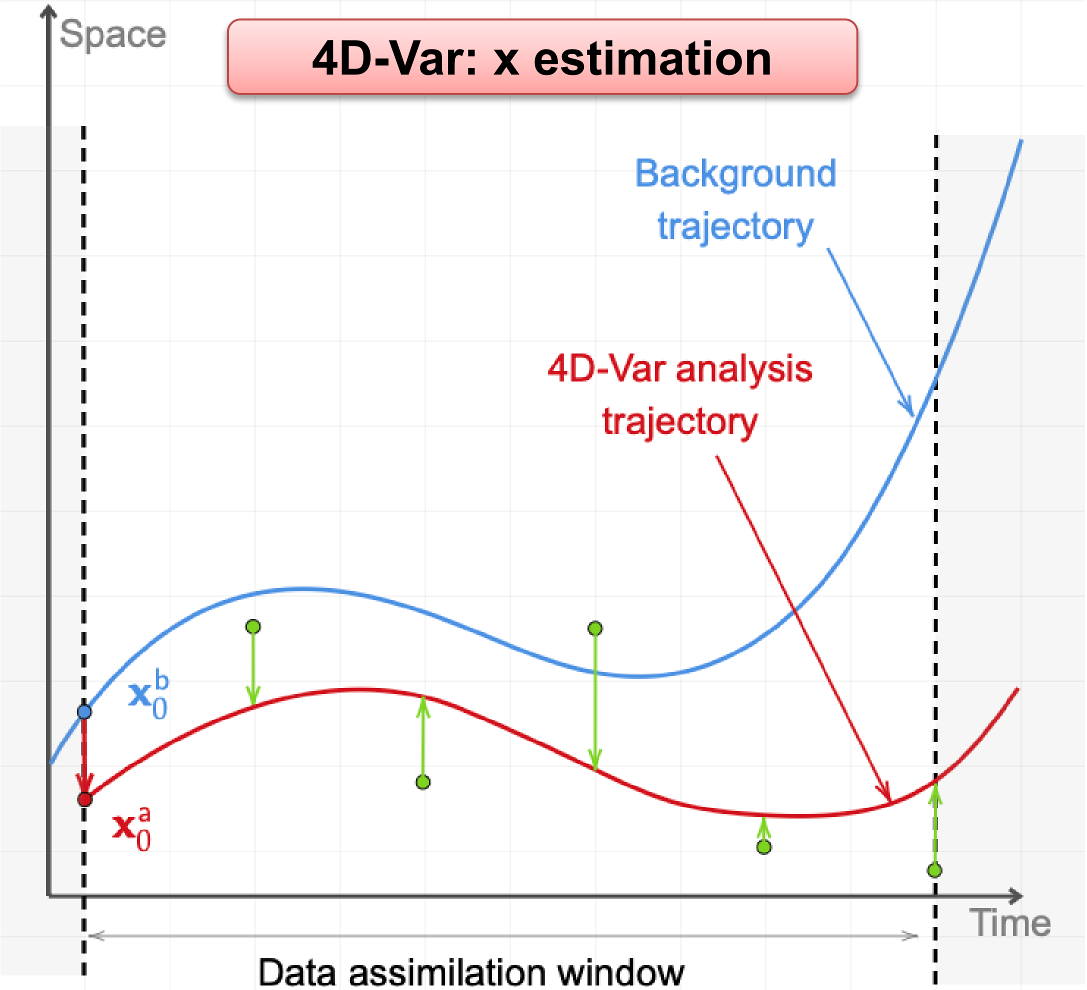


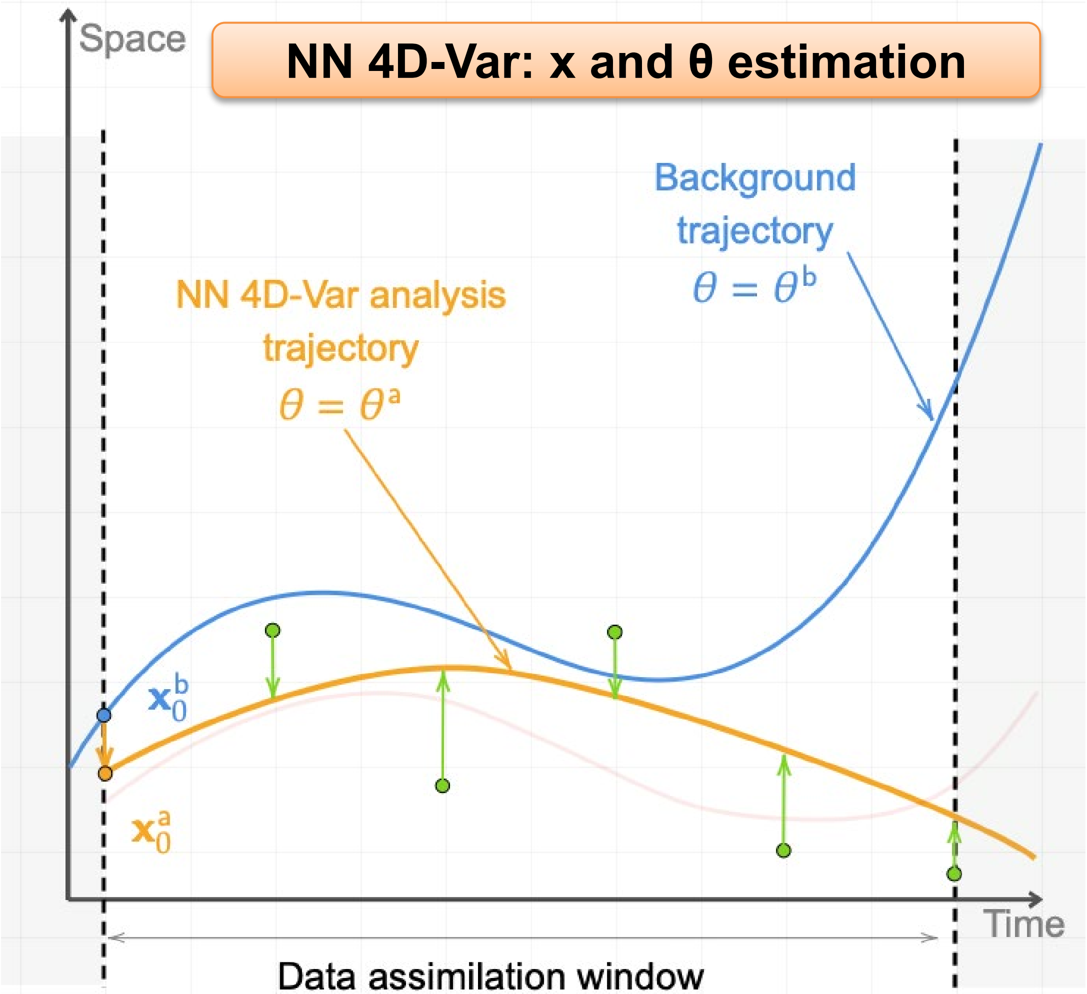
NN 4D-Var – a rigorous formalism to train NN online

- A new variant of weak-constraint 4D-Var:

(Farchi et al, 2021, 2023)

$$\mathcal{J}(\theta, \mathbf{x}_0) = \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\left\|\theta - \theta^{\mathrm{b}}\right\|^2_{\mathbf{P}^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\left\|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\theta, \mathbf{x}_0)\right\|^2_{\mathbf{R}^{-1}}$$



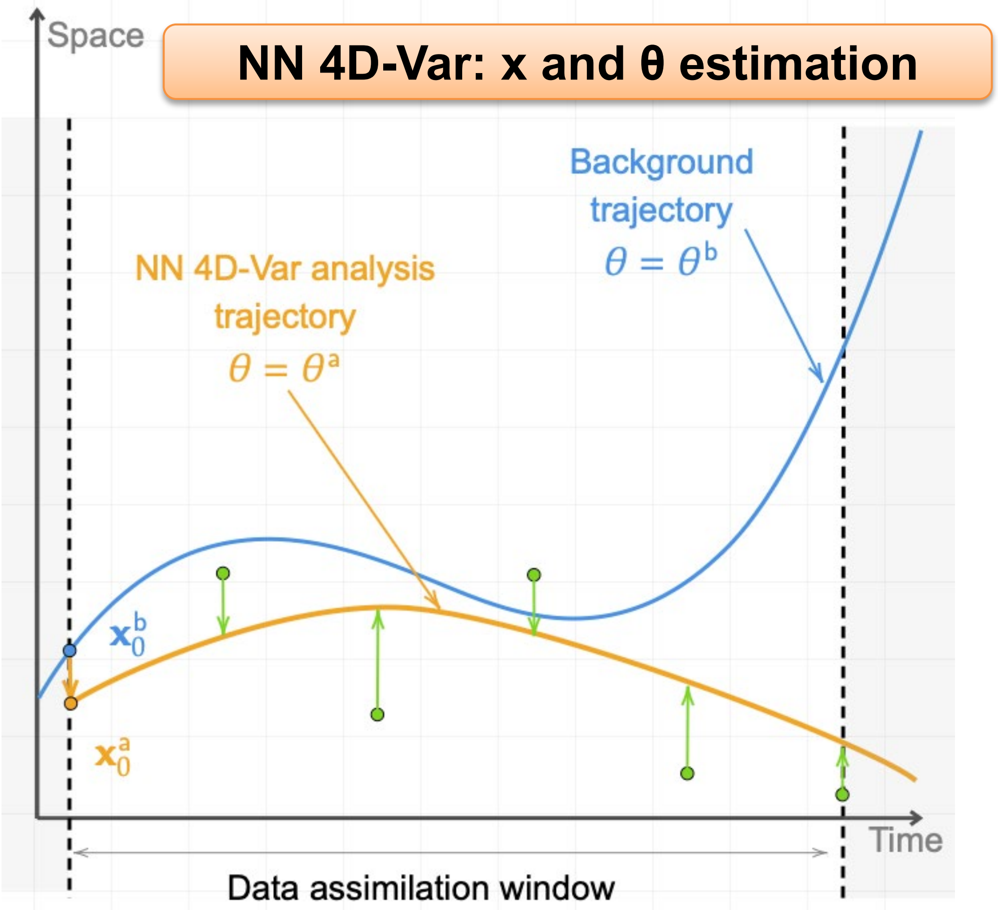


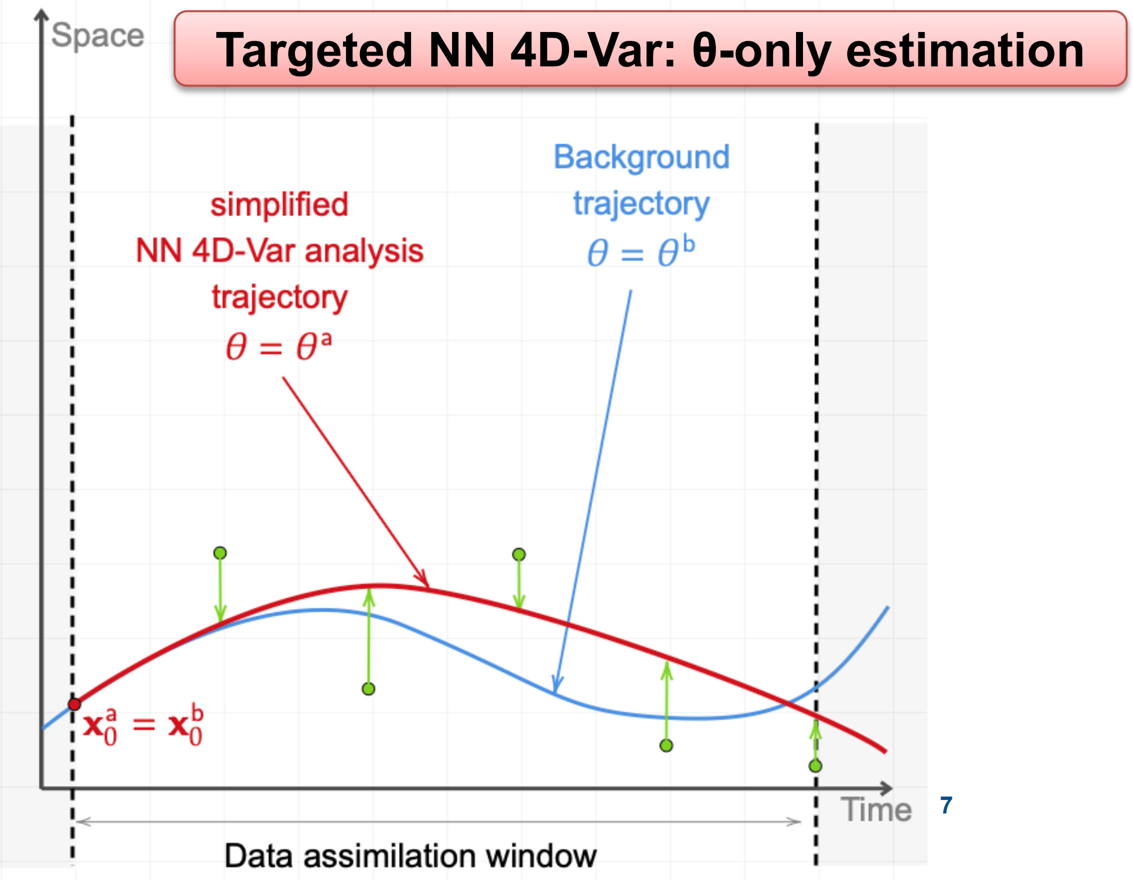
Simplifying NN 4D-Var to focus on the NN parameters

- To further increase the impact of online training, we need 4D-Var to focus on the NN parameters:

$$\mathcal{J}(\theta, \cancel{\mathbf{x}_0}) = \cancel{\frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right\|_{\mathbf{B}^{-1}}^2} + \frac{1}{2}\left\|\theta - \theta^{\mathrm{b}}\right\|_{\mathbf{P}^{-1}}^2 + \frac{1}{2}\sum_{k=0}^{L}\left\|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}\left(\theta,\ \mathbf{x}_0^{\mathrm{pa}}\right)\right\|_{\mathbf{R}^{-1}}^2$$

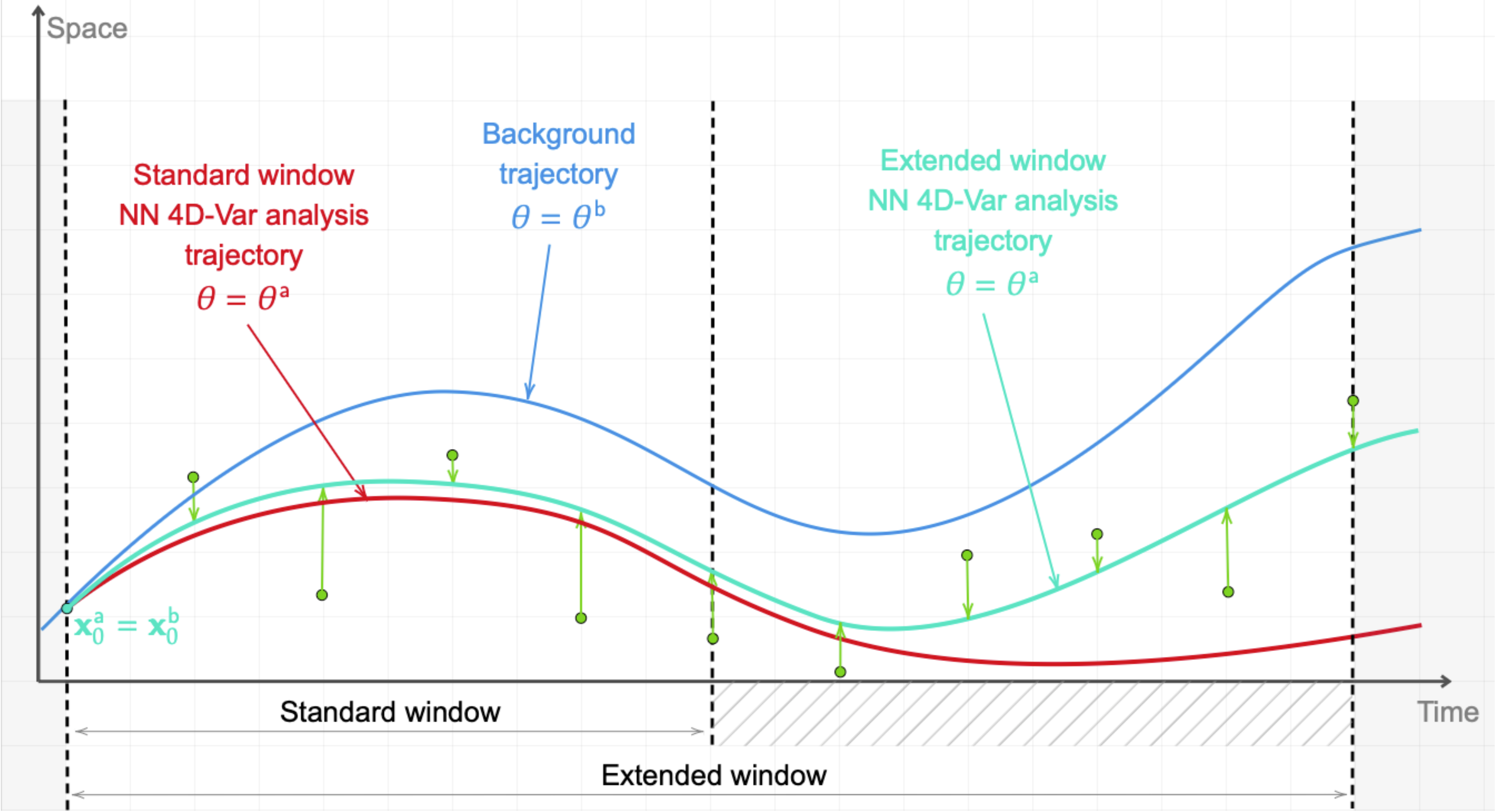
- Where $\mathbf{x}_0^{\mathrm{pa}}$ is a precomputed analysis (e.g. the operational analysis)





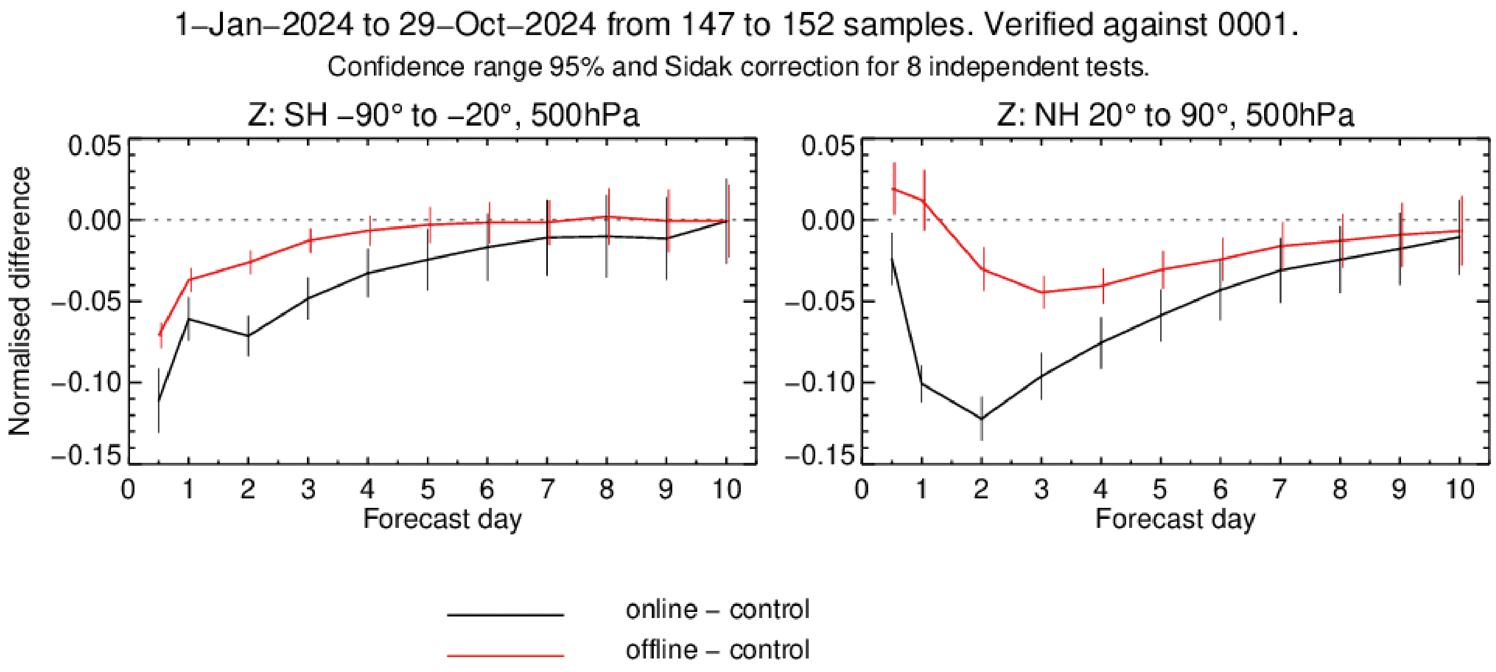
Extending the DA window from 12h to 48h

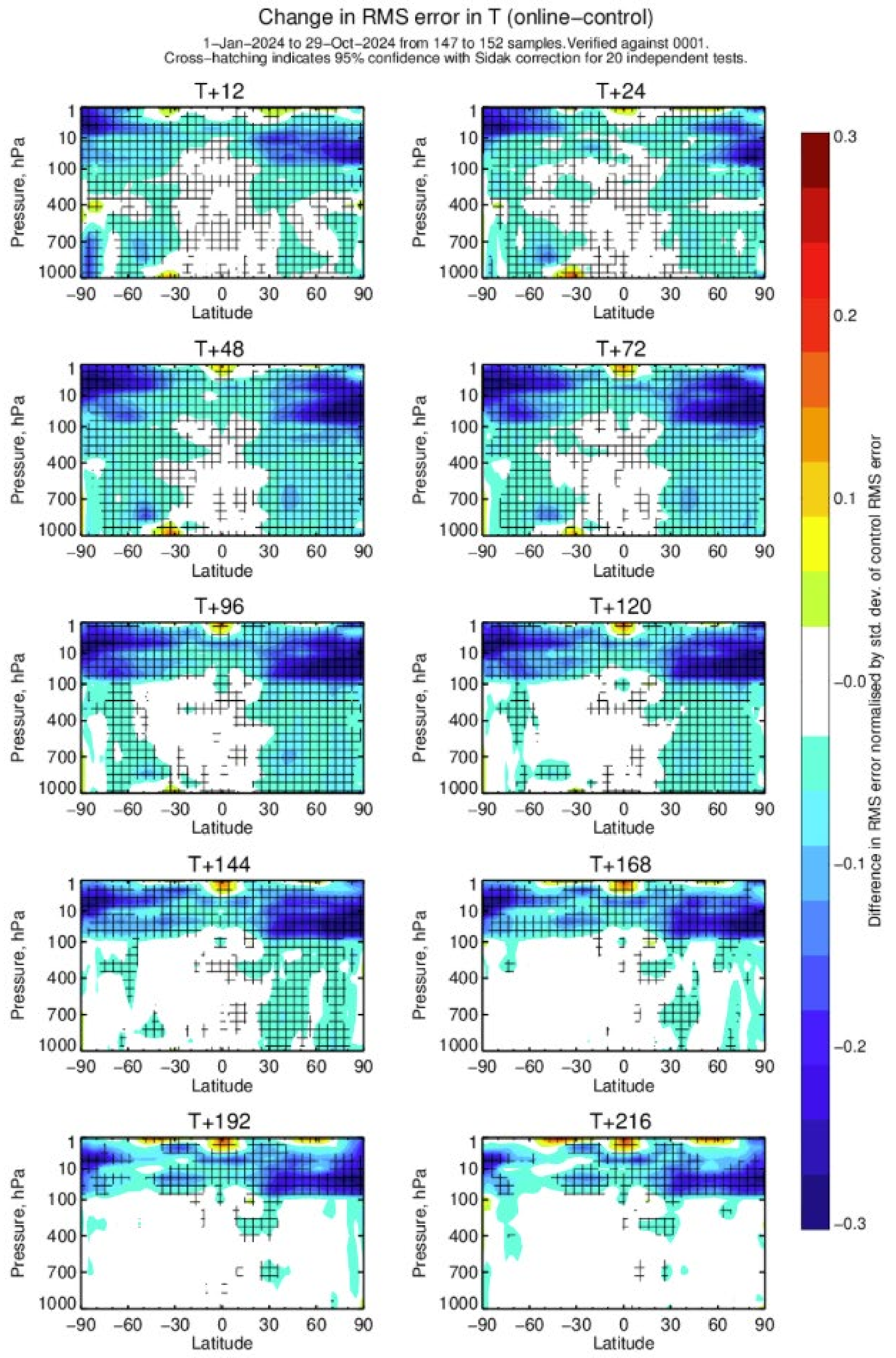
- Furthermore, we decided to extend the length of the DAW from 12h to 48h:
  - More observations are available to constrain the parameters of the NN
  - It helps capture model errors that develop over longer time-scales



Training online with 48h windows

- Online training (fine-tuning) in 2024 at TCo399
- Much stronger impact than with offline training
- Caveats:
  - 48h windows means 36h of future observations
  - Does not correspond to 1 NN, but a collection of NNs

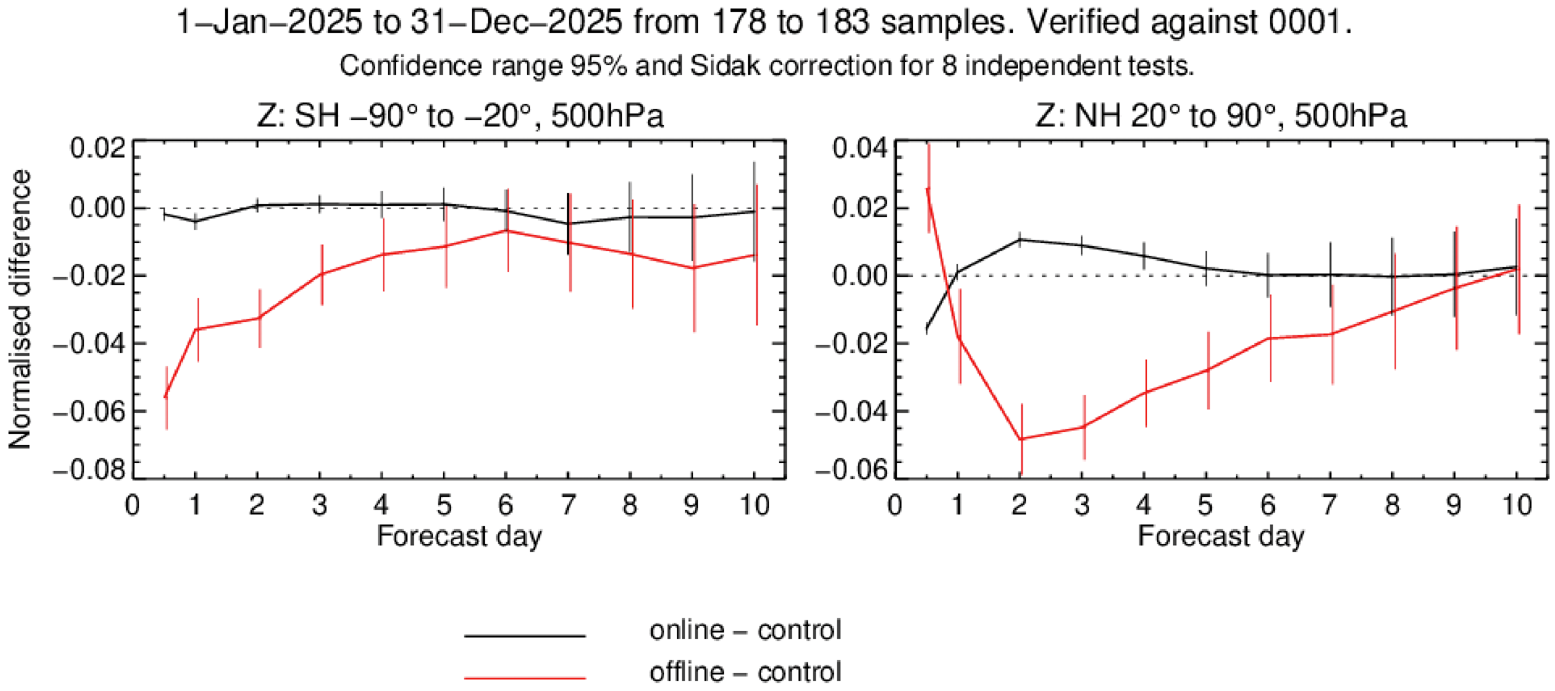


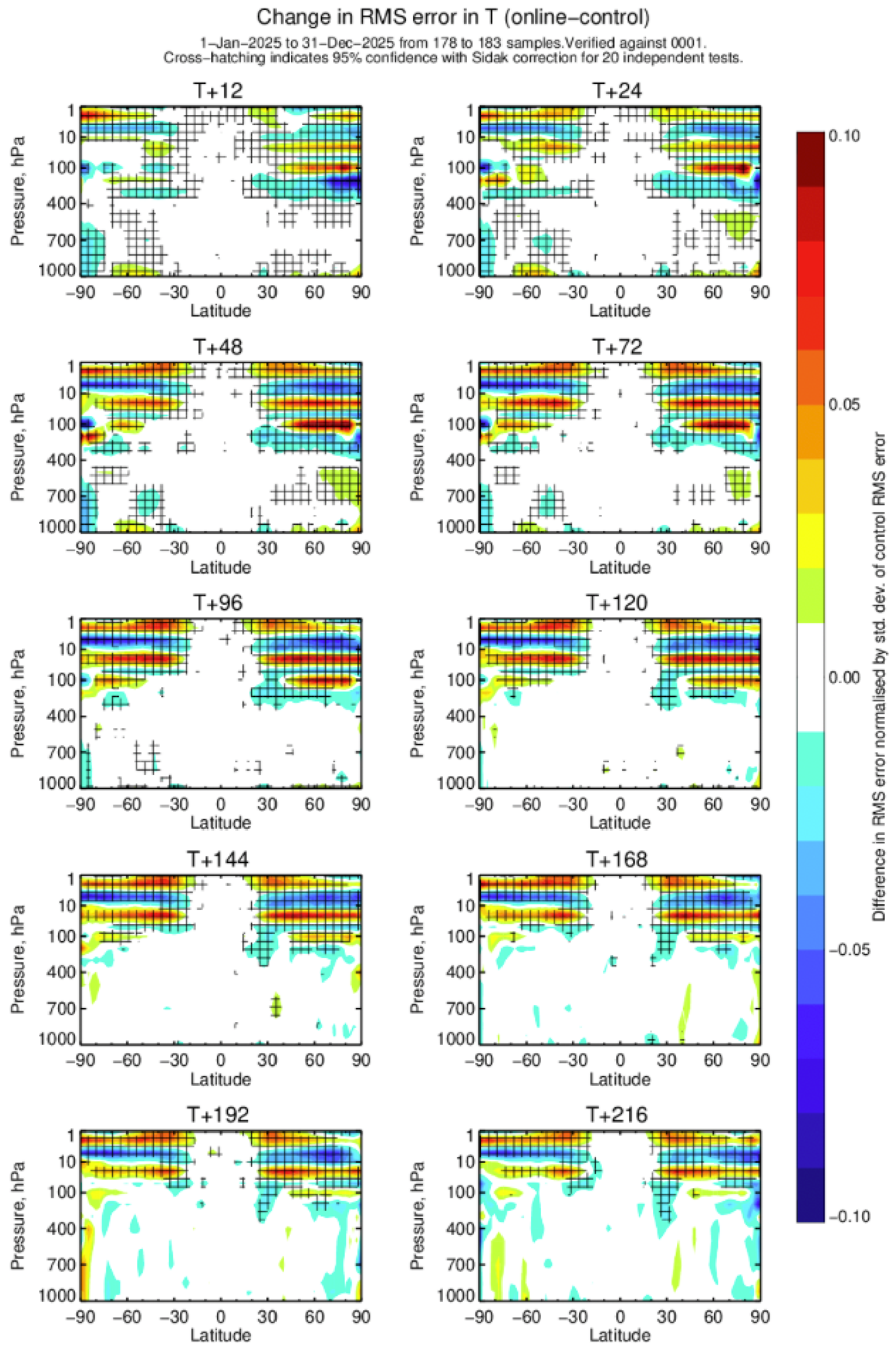


ECMWF

Accuracy of the online hybrid model

- Forecast-only experiments in 2025 (unseen data) at TCo399
- The online NN correction does not generalise to unseen data

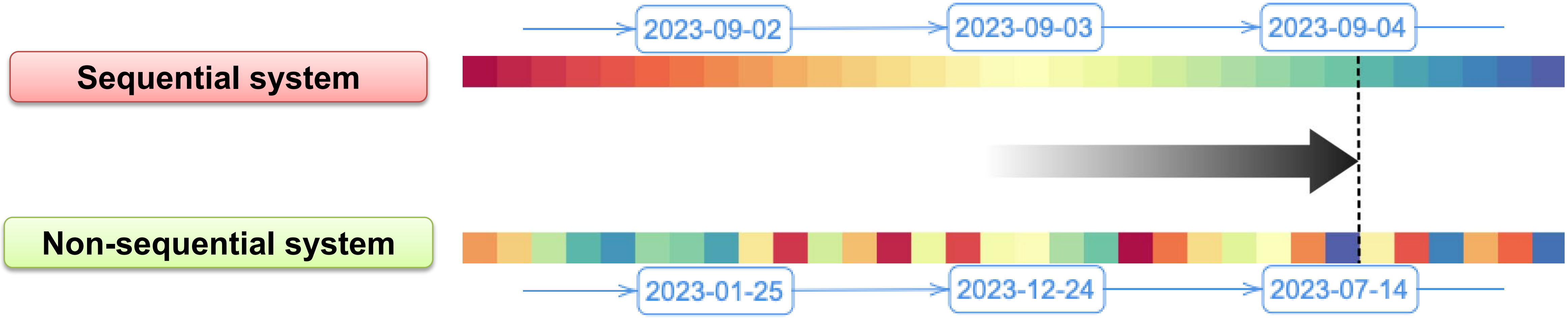




ECMWF

Limitations of a sequential NN 4D-Var system

- NWP models are not autonomous, and model errors have seasonal components
- NN 4D-Var has been tested in a sequential setup
    - NN parameters may only keep in memory information about model errors in the recent past
- We need to relax the sequential assumption of the system



NN 4D-Var – a reimplementation of SGD?

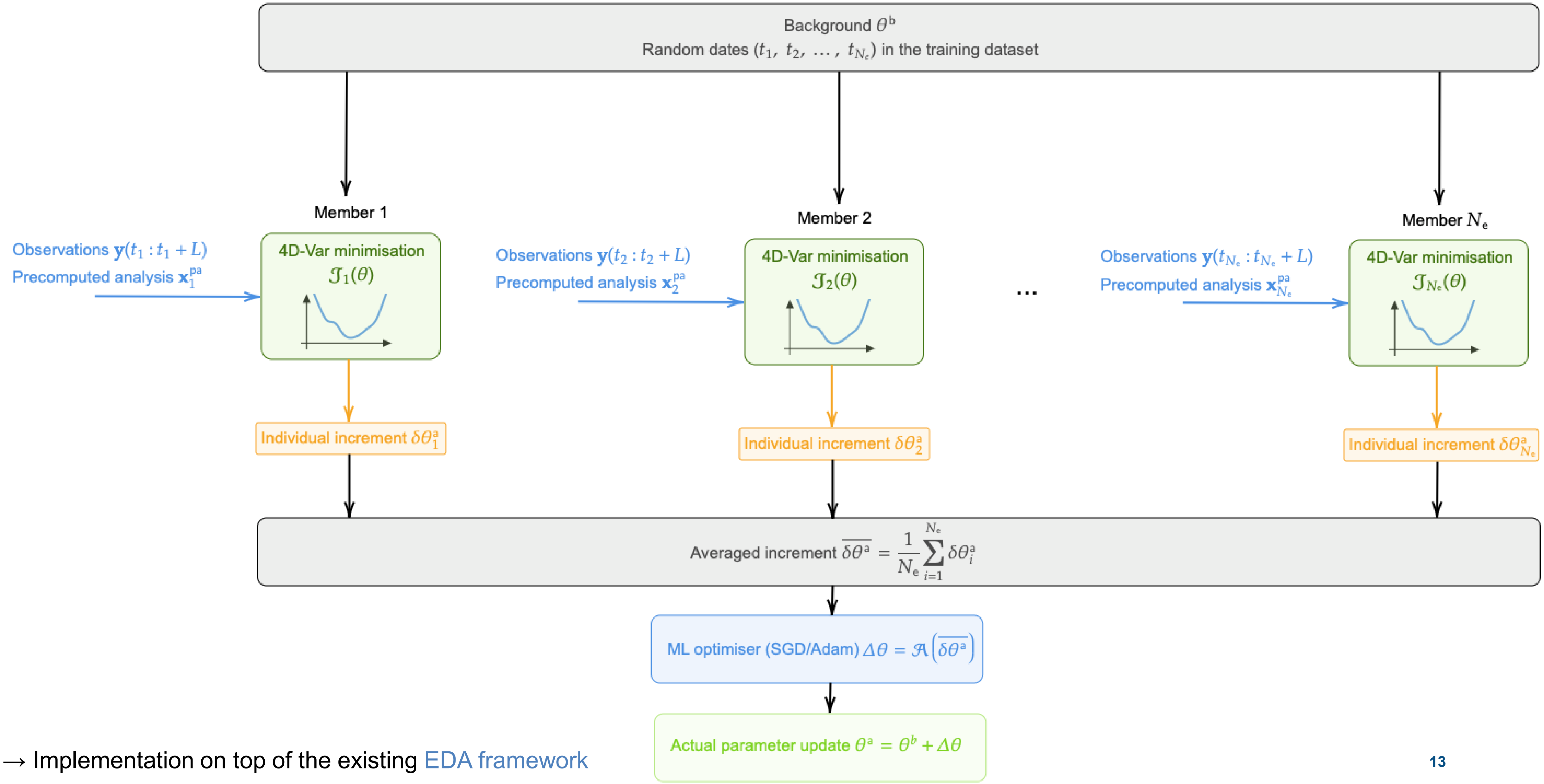
	NN 4D-Var	Targeted NN 4D-Var	SGD
Control Variable	x and θ	Only θ	Only θ
Order of windows / sample	Sequential	Sequential / Random	Random
Parameter updated mechanism	$\theta(t+1) = \theta(t) + \delta\theta^{a}(t)$	$\theta(t+1) = \theta(t) + \delta\theta^{a}(t)$	$\theta(t+1) = \theta(t) - \gamma\nabla_{\theta}\mathcal{L}(\theta(t))$

- There is a clear similarity between one cycle of the targeted NN 4D-Var and one iteration in a SGD algorithm, with

$$\delta\theta^{a}(t) \Leftrightarrow -\gamma\nabla_{\theta}\mathcal{L}(\theta(t))$$

- Pushing the analogy further, we define a batched-variant of NN 4D-Var, where several windows are processed in parallel
- Effectively multiplies the number of observation windows that the NN see in a fixed amount of optimisation steps

Batching NN 4D-Var

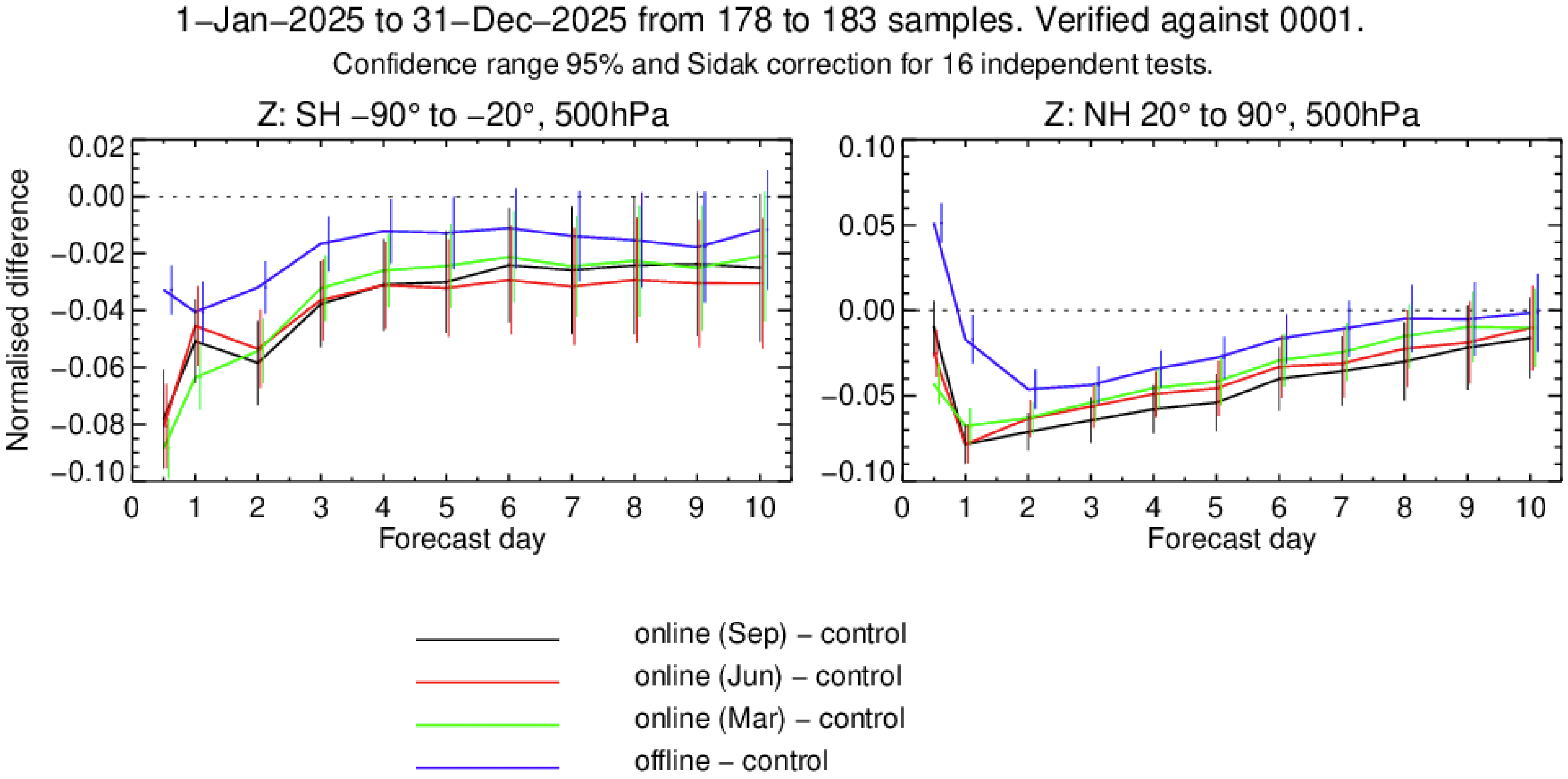


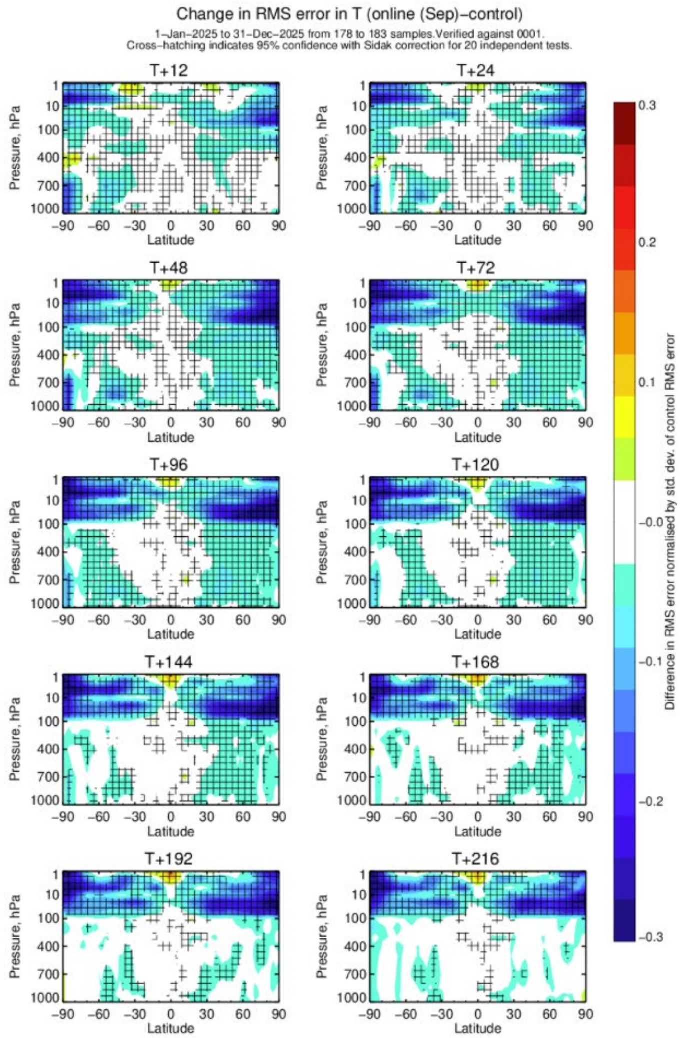
→ Implementation on top of the existing EDA framework

Accuracy of the online hybrid model

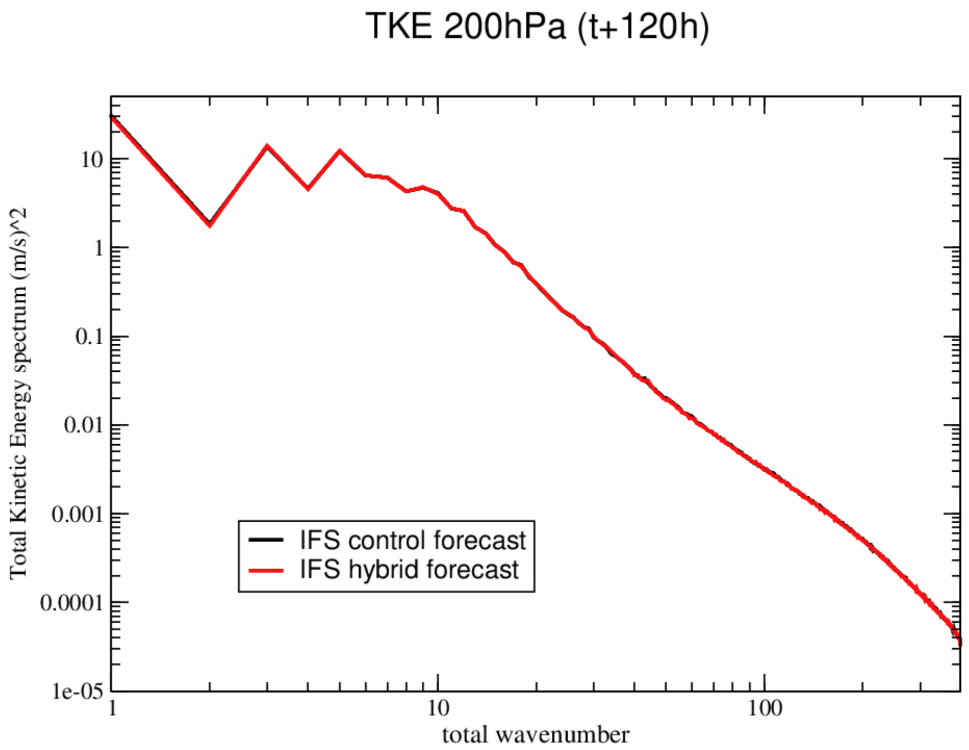
Fictitious training time	Optim. steps	DAWs	Epochs
2024-01-01 to 2024-03-01	31	248	1.36
2024-01-01 to 2024-06-01	77	616	3.38
2024-01-01 to 2024-09-01	123	984	5.41

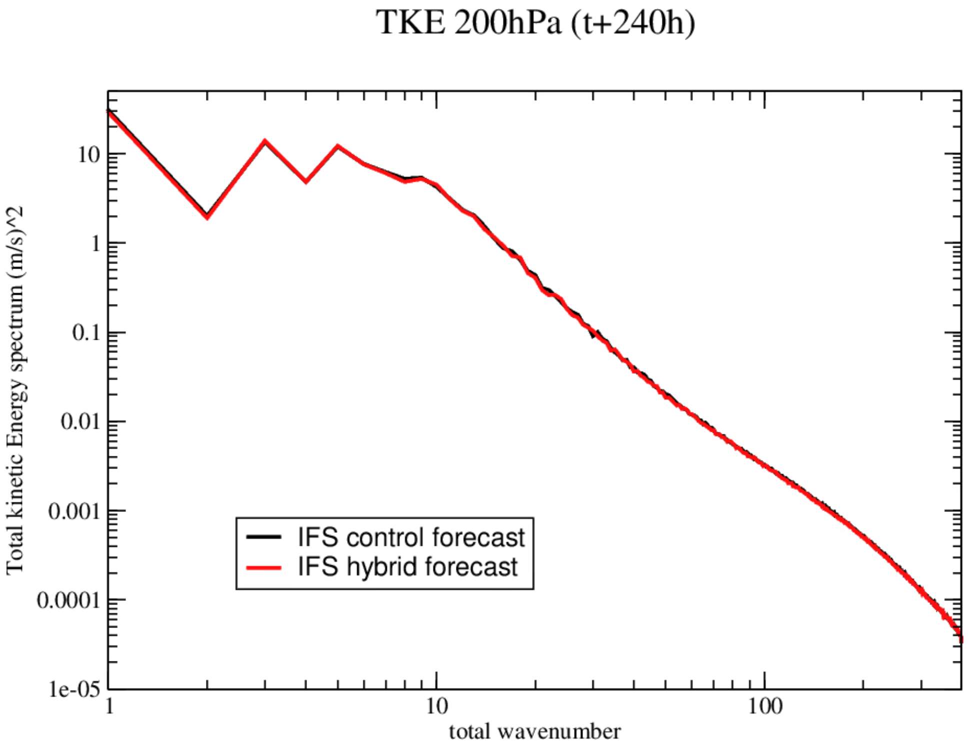
- Online training at TCo399, with 8 members
- Training dataset: 182 48h-windows over 2024
- Hybrid model evaluated in forecast-only experiments in 2025 (unseen data) after approximately 2, 5, and 8 fictitious months of training





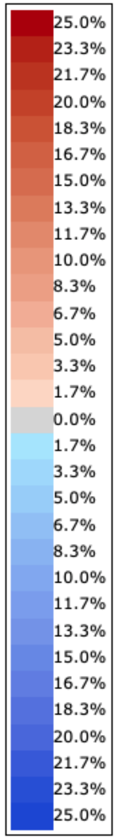
Physical consistency

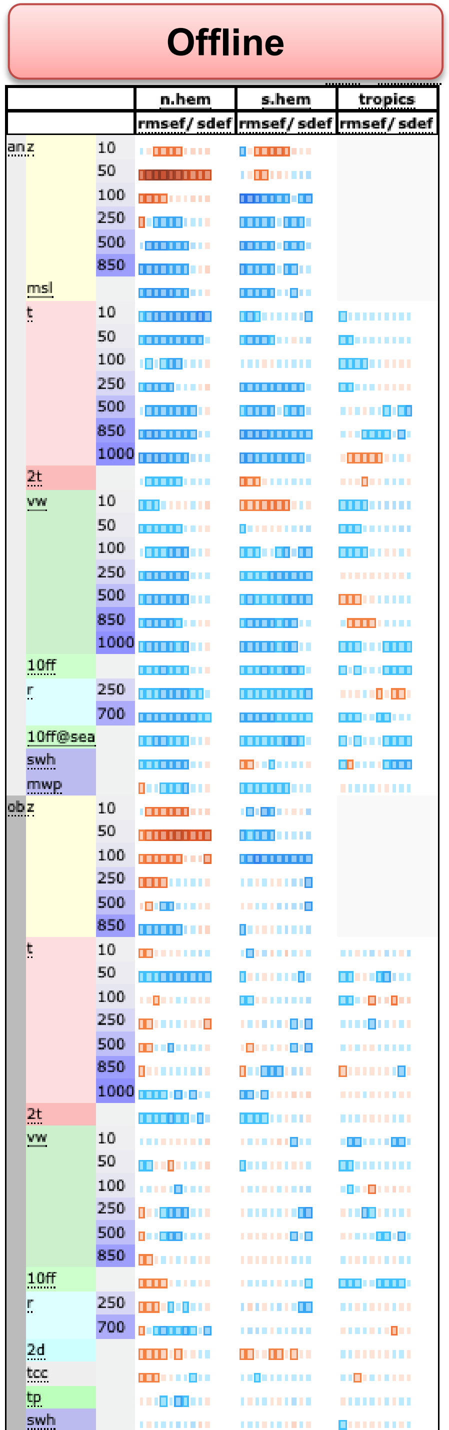


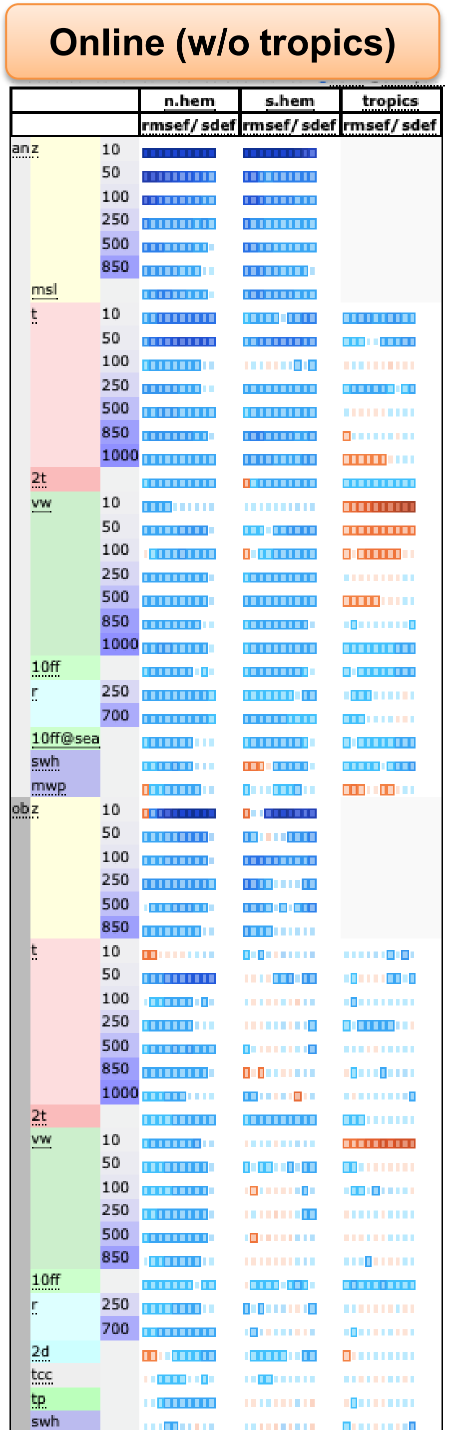


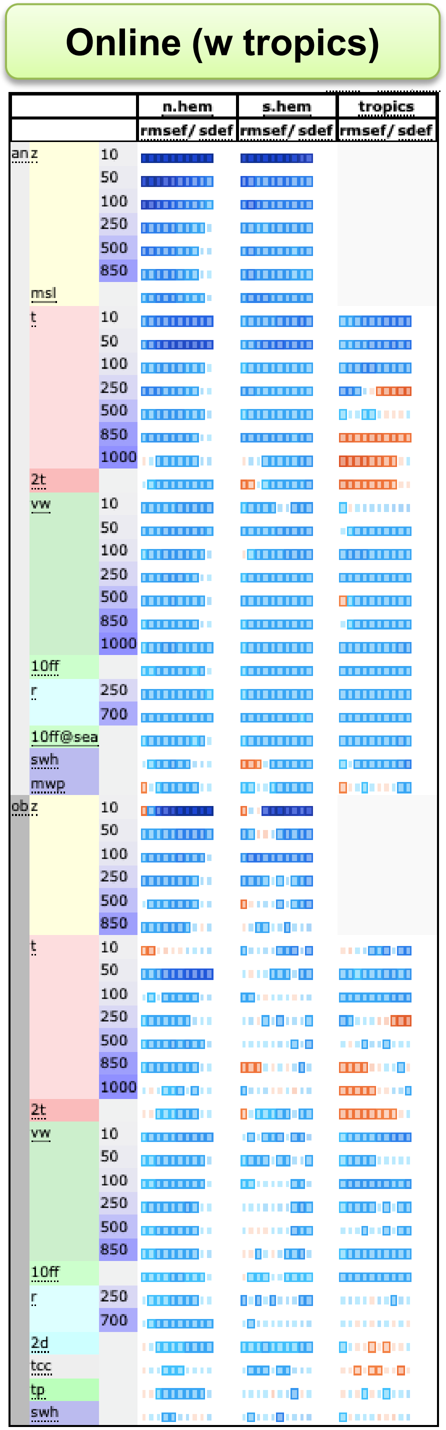
- Dynamically indistinguishable from the IFS forecasts

Final words: scorecards (2025)









ECMWF

Conclusions

- We have developed a robust batched NN 4D-Var system to train a NN-based correction of model error
- This recent breakthrough is currently undergoing comprehensive testing at TCo399, with initial results appearing highly promising (and competitive with other hybrid models)
- Several extensions are already being explored to both further improve the performance of the system and reduce its cost

- The batched NN 4D-Var system is unique:
  - It leverages the existing 4D-Var and the EDA frameworks to substitute stochastic gradient descent, which could only be implemented at ECMWF, where both the tools and expertise are available
  - It is, to our knowledge, the only hybrid model in which model error is learned directly comparing model forecasts to observations
  - The hybrid model forecasts are dynamically indistinguishable from that of the IFS

→ Paper in preparation!