# Development of an offline and online hybrid model for the Integrated Forecasting System

Alban Farchi, Marcin Chrust, Patrick Laloyaux, Massimo Bonavita (ECMWF)
and Marc Bocquet (ENPC)

ECMWF

# Outline

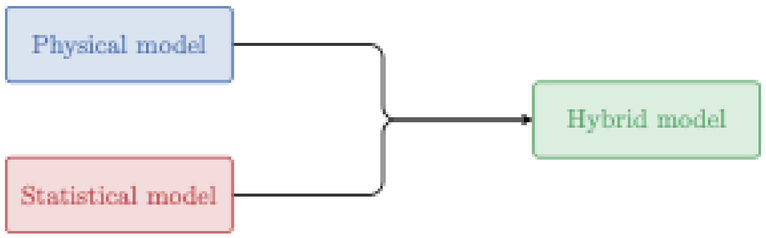
- Methodology
- Offline pre-training
- Online forecast and training experiments

# Why is model error so important?

- NWP models are affected by errors, e.g. unresolved small-scale processes
- Model error is one of the main limitations of forecast accuracy
- Growing interest of the DA community in weak-constraint methods:
  - Iterative EnKF in the presence of additive noise (Sakov et al. 2018)
  - Forcing formulation of weak-constraint 4D-Var (Laloyaux et al. 2020)
- More recent approaches to the model error issue rely on ML techniques to learn the model from data (Bocquet et al 2019, 2020; Brajard et al 2020).

# Hybrid modelling

- We propose to tackle model errors by using hybrid modelling
- The statistical part of the model is trained to learn and correct the error of the physical model.




- A typical NWP model rely on a set of ODEs or PDEs which define the tendencies: $\frac{\partial x}{\partial t} = \phi(x)$
- A numerical scheme is used to integrate the tendencies from $t$ to $t + \delta t$: $x(t + \delta t) = \mathcal{I}\big(x(t)\big)$
- Several integration steps are composed to build the resolvent from one analysis (or window) to the next:

$$\mathcal{M}: x_k \mapsto x_{k+1} = \mathcal{I} \circ \cdots \circ \mathcal{I}(x_k)$$

# Correcting the resolvent or the tendencies?

**Resolvent correction**

- Physical model and NN are independent
- NN must predict the analysis increments
- Adding a (potentially large) NN correction to the state can provoke initialisation shocks
- The hybrid model cannot be used for short-term predictions with ad-hoc approximations (e.g. linear growth of errors in time)

**Tendency correction**

- Physical model and NN are entangled
- Need the adjoint of the physical model to train the NN! Requires online training.
- Can be used as is in DA experiments

**Hybrid approach to build the hybrid model**

1. Train the NN offline as a resolvent correction
2. Rescale the correction from 1 DAW to 1 model time step
3. Fine-tune the NN online as a tendency correction

This methodology has been first validated using low-order models (Farchi et al. 2021, 2023)

# Online training with NN 4D-Var

- To train online the parameters of the NN, we propose to use a variant of weak-constraint 4D-Var (Farchi et al. 2021)

$$\mathcal{J}^{nn}(\theta, x_0) = \frac{1}{2}\left\|x_0 - x_0^b\right\|_{B^{-1}}^2 + \frac{1}{2}\left\|\theta - \theta^b\right\|_{P^{-1}}^2 + \frac{1}{2}\sum_{k=0}^{L}\left\|y_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}^{\theta}(x_0)\right\|_{R_k^{-1}}^2$$

- The parameters $\theta$ (e.g. NN weights and biases) are assumed constant over the DAW
- Information is flowing from one DAW to the next using the priors $x_0^b$ and $\theta^b$

- This approach is very similar to classical parameter estimation in DA and can be seen as a NN formulation of weak-constraint 4D-Var
- Similar approaches had already been developed within an EnKF context (Bocquet et al. 2020)

# Implementation of NN 4D-Var within OOPS

- To implement NN 4D-Var, we can reuse most of the framework already in place for WC 4D-Var
- We only need a NN library to compute the forward, tangent linear, and adjoint operators of the NN
- The NN needs to be applied in the model core, which is implemented in Fortran

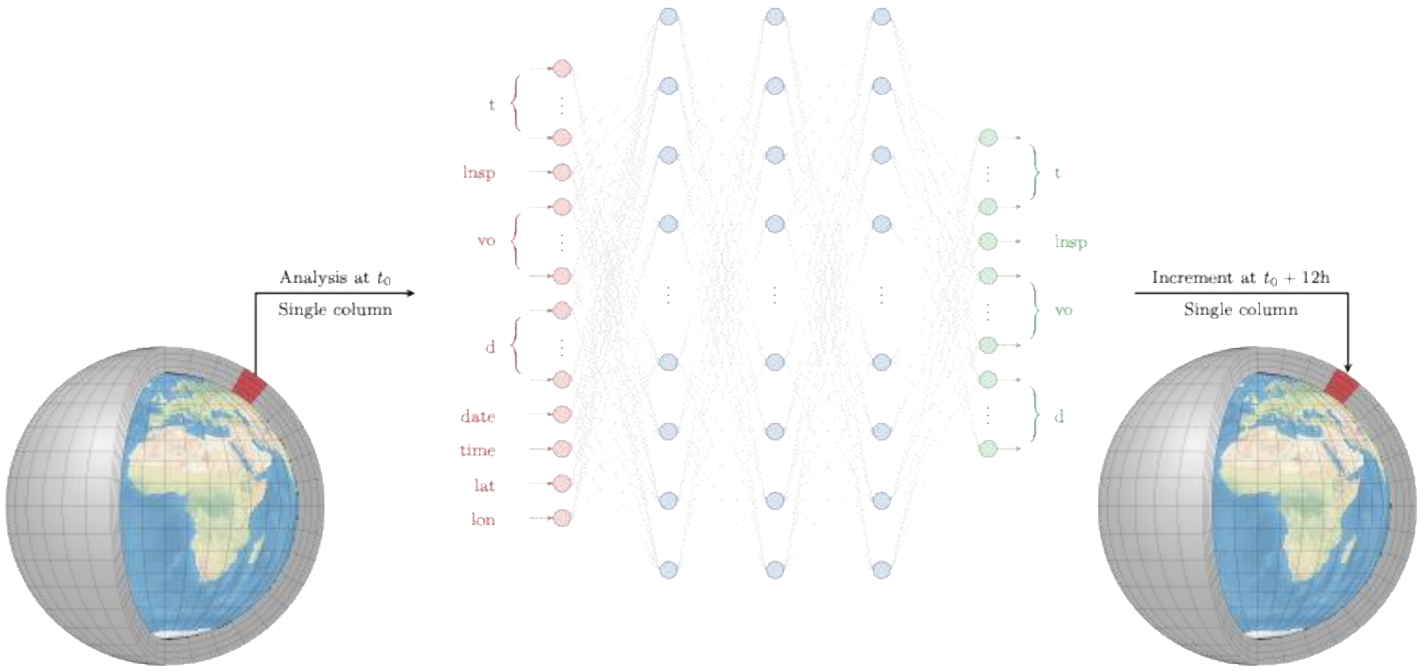
- We implemented our own NN library in Fortran: FNN
- FNN has been interfaced and included in OOPS

# Outline

- Methodology
- Offline pre-training
- Online forecast and training experiments

# A NN correction for the IFS

- We extend the preliminary work of Bonavita and Laloyaux (2020)
- We compute a correction for 4 atmospheric variables in the same NN: temperature (t), logarithm of surface pressure (lnsp), vorticity (vo), and divergence (d)
- We use a vertical architecture, where the NN processes independently each atmospheric column
- We include all 137 model levels in both input and output, for a total of 412 channels




- Once trained, the NN can be applied on any grid
- The num. of parameters is relatively small compared to the size of the control vector (<2%) and to the size of the training dataset (<1%)
- Horizontal information is partially lost

# Description of the training dataset

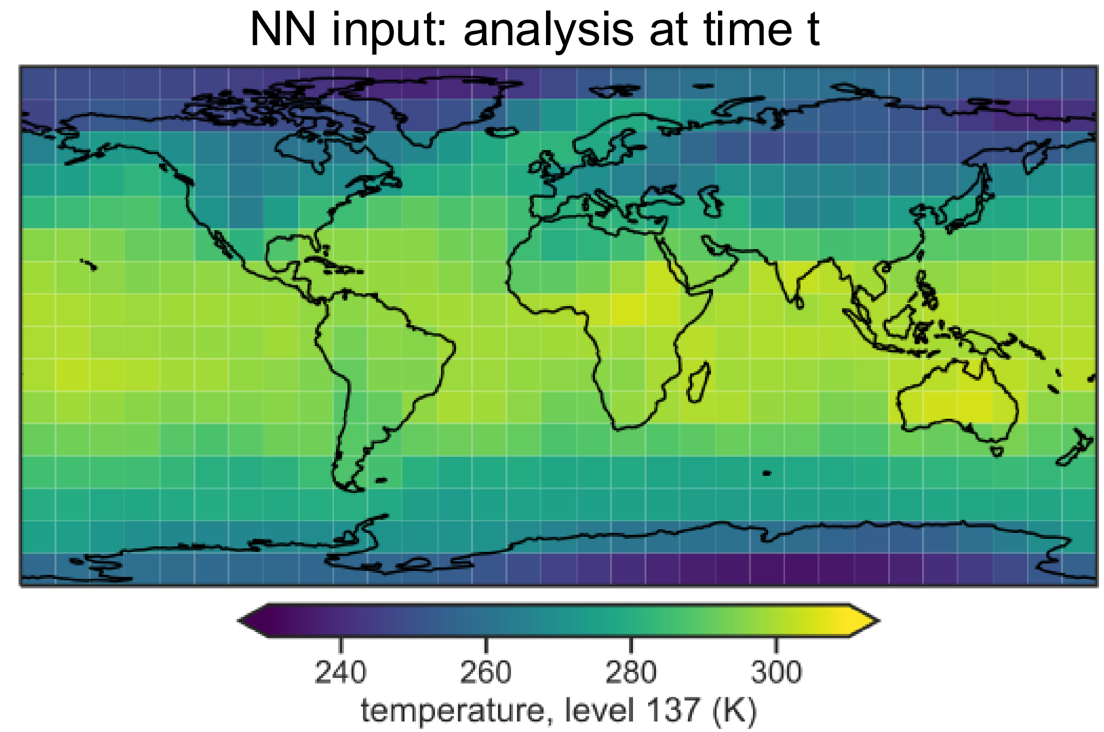
- We use data from the operational archive between 2021 and 2023:
    - Training set: 2021 and 2022 (cycles 47R1 to 47R3)
    - Validation set: 2023 (cycles 47R3 to 48R1)
- 2024 is set aside for online evaluation
- Pre-2021 data is discarded because the implementation of WC 4D-Var (in June 2020 with cycle 47R1) significantly modified the increments

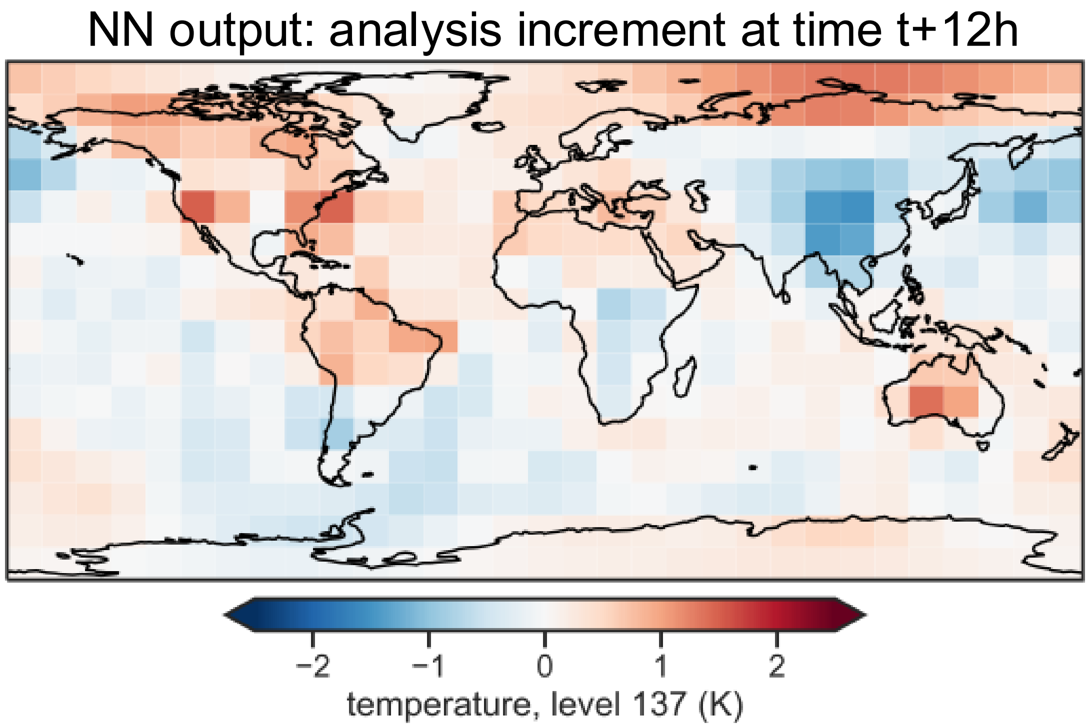
- The NN is trained to predict the sum of the analysis increments and the model error diagnosed by WC 4D-Var (1 sample every 12h)

$$\mathcal{F}_\theta : x_k^a \mapsto x_{k+1}^a - x_{k+1}^f + 12 * w_k^a$$

ECMWF

# Choice of the spectral truncation



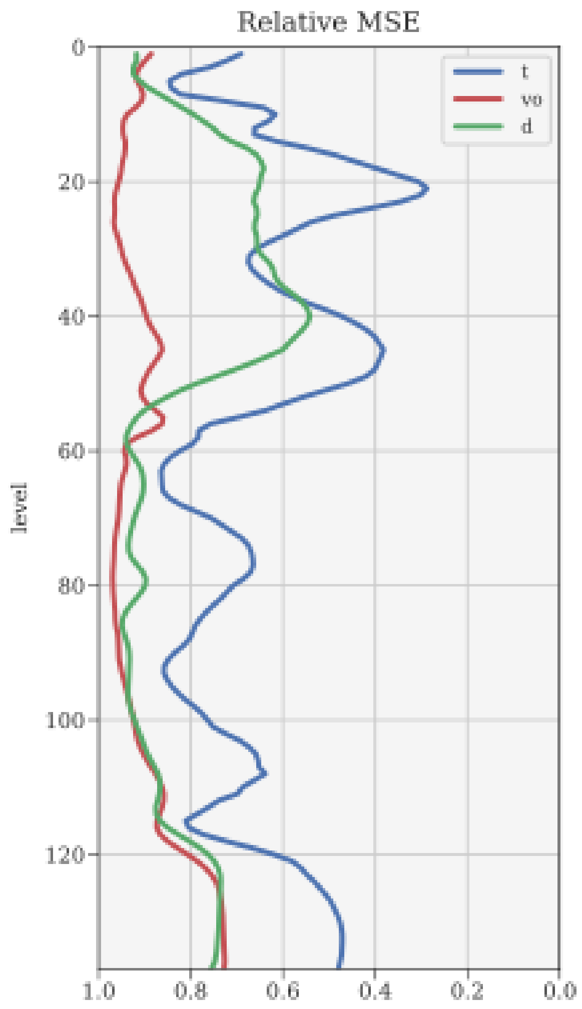

- To focus on large-scale model error, we truncate the data at T15 and interpolated in a 16x32 Gaussian grid
- We then investigate the effect of resolution by progressively increasing the truncation:

  T31, T63, and T95

# Validation scores at T15

Relative MSE on the validation set

| Name | Lnsp | T | Vo | D |
|---|---|---|---|---|
| No correction | 1.000 | 1.000 | 1.000 | 1.000 |
| Mean increment | 1.000 | 0.968 | 1.000 | 1.000 |
| Climatological increment | 0.769 | 0.659 | 0.976 | 0.804 |
| Trained NN | **0.764** | **0.619** | **0.917** | **0.783** |

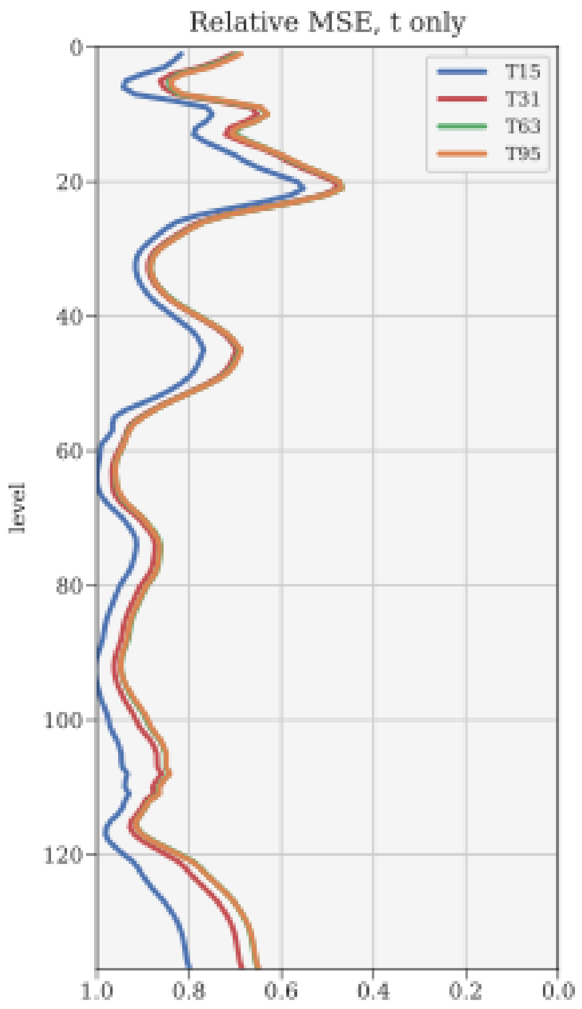



- Overall, the NN can predict 10 to 40% of the increments
- The increments for t, lnsp, and d are more predictable than for vo
- The predictions are in general most accurate close to the surface and in the stratosphere (where WC is active)

# Validation scores at higher resolution

Relative MSE on the validation set **at T127**

| Training truncation | Lnsp | T | Vo | D |
|---|---|---|---|---|
| T15 | 0.969 | 0.835 | 1.000 | 0.997 |
| T31 | 0.900 | 0.757 | 0.996 | 0.993 |
| T63 | 0.880 | 0.740 | 0.996 | 0.994 |
| T95 | 0.882 | 0.739 | 0.996 | 0.994 |

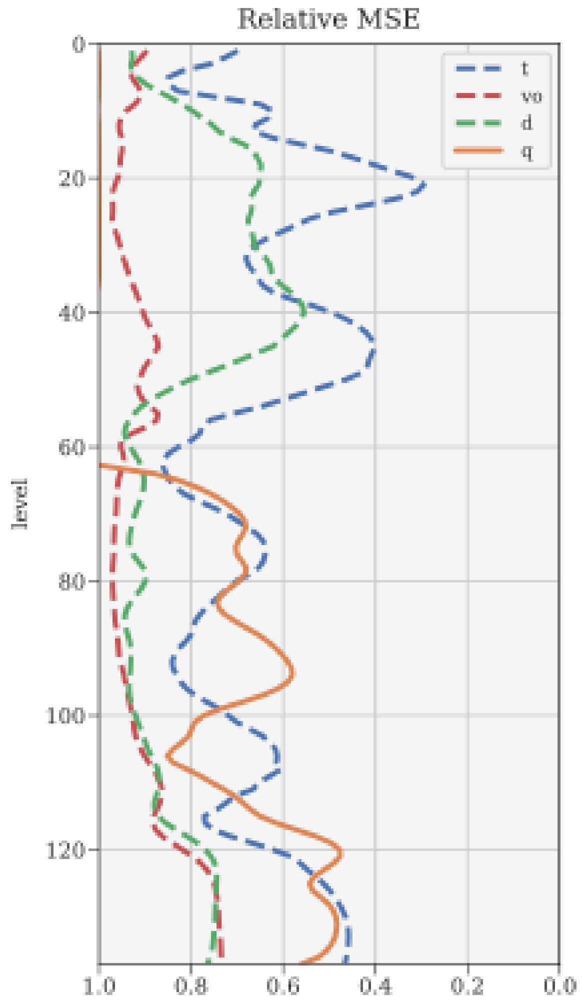



- To ensure fairness, all NNs are evaluated at the same truncation, T127
- At higher resolution, vo and d are almost unpredictable for this NN architecture
- For t and lnsp, training at higher resolution is beneficial, until ~T63

# Extension 1: include q

Relative MSE on the validation set

| Name | Lnsp | T | Vo | D | Q |
|---|---|---|---|---|---|
| No correction | 1.000 | 1.000 | 1.000 | 1.000 | 1.00 |
| Mean increment | 1.000 | 0.968 | 1.000 | 1.000 | 0.956 |
| Climatological increment | 0.769 | 0.659 | 0.976 | 0.804 | 0.620 |
| Trained NN wo q | **0.764** | **0.619** | **0.917** | **0.783** | |
| Trained NN w. q | **0.766** | **0.619** | **0.924** | **0.788** | **0.570** |

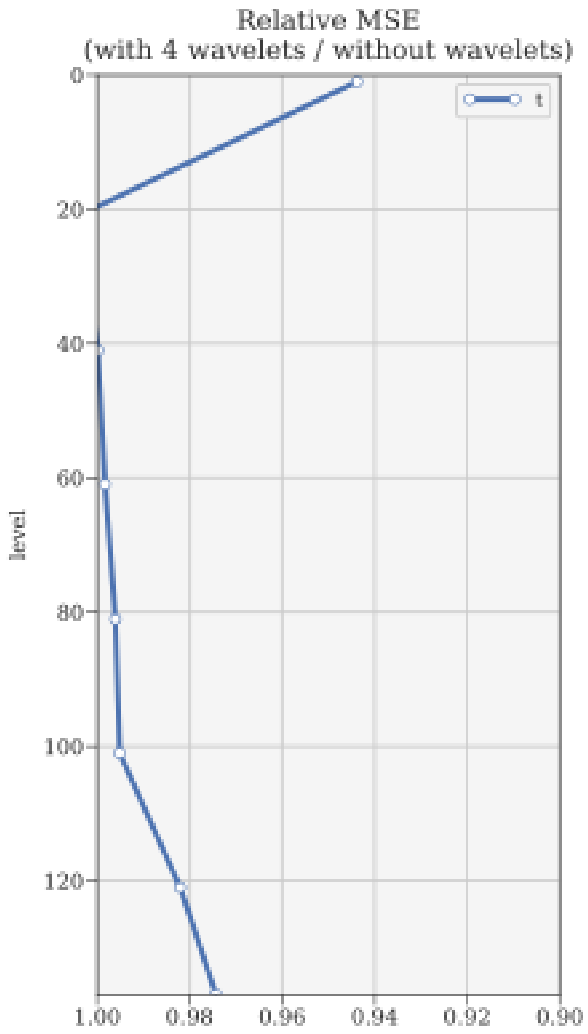



- Originally, our choice of 4 atmospheric variables (t lnsp vo d) is based on the results from Bonavita and Laloyaux (2020)
- Later results (Chen et al 2022) indicate that it could be relevant to include specific humidity in the corrected variables
- In our experiments, the increments for q are highly predictable

# Extension 2: include horizontal information

- The choice of a column architecture is driven by computational constraints: in the online system, each atmospheric column is stored in the same processor unit
- Even within the column framework, it is possible to include horizontal information, beyond the latitude and longitude extra predictors:
    - Use the horizontal gradients as extra predictors of the NN (Kochkov et al. 2023)
    - Use a wavelet decomposition of the predictors
- Challenge: how to deal with the increased input dimension?

- Results from a toy experiment with 14 years of ERA-5 data using a selection of model levels at T63 resolution indicate that using a wavelet decomposition of the predictors is beneficial.



# Extension 3: increase the frequency of the NN correction

- In the offline training dataset, we have samples every 12h, at 09:00 and 21:00
- In the online experiments, the NN correction is used as a tendency correction, but for consistency, the correction is updated only every 12h
- We believe that updating the correction more frequently should be beneficial, because it would allow us to better represent the daily variability of model error

- Challenges:
  - Is it necessary to update the offline training step? Most probably yes.
  - What offline training data can be used? Analysis and first-guess within the window?

# Outline

- Methodology
- Offline pre-training
- Online forecast and training experiments

# Motivation for online training

- Offline Training
  - Uses historical data (reanalysis – same model, long dataset record, 0001 – short record or changing model)
  - Static correction model with fixed parameters
  - Can become outdated as model and model error evolve (numerical model updates, slowly evolving model climate)
  - Requires periodic retraining
- Why train model error online?
  - Continuous adaptation to changing model error
  - Model error diagnosed using latest observations
  - Model error correction constrained by and consistent with TLM dynamics
  - Mitigated model error in data assimilation

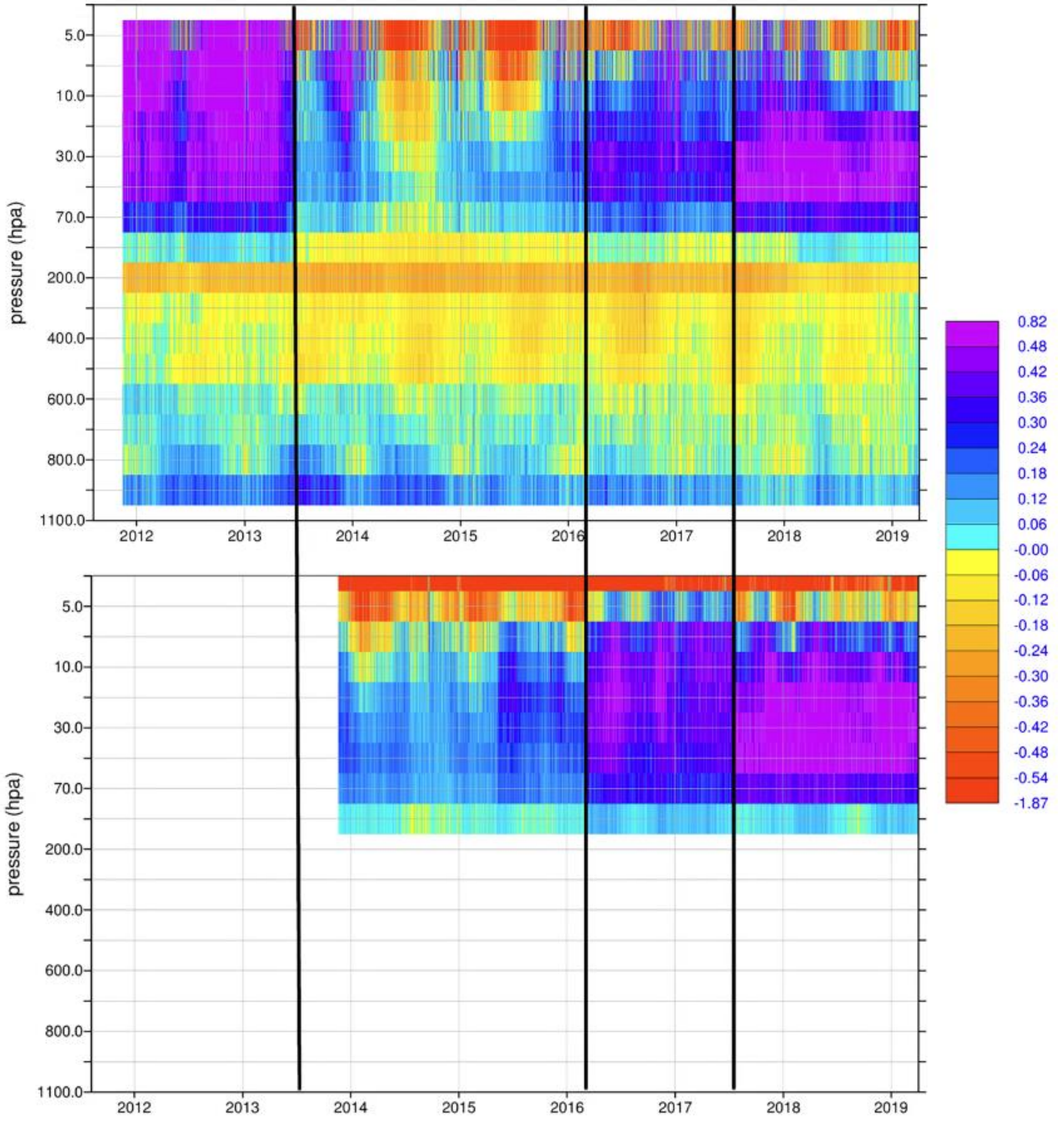



**Fig.** Time series of the difference between observations and temperature first-guess trajectory from the ECMWF operational analysis cycle for radiosonde (top) and RO (bottom). From Laloyaux et al., 2020.

ECMWF

# Performance of the model error trained offline in the forecast experiments (HRES)

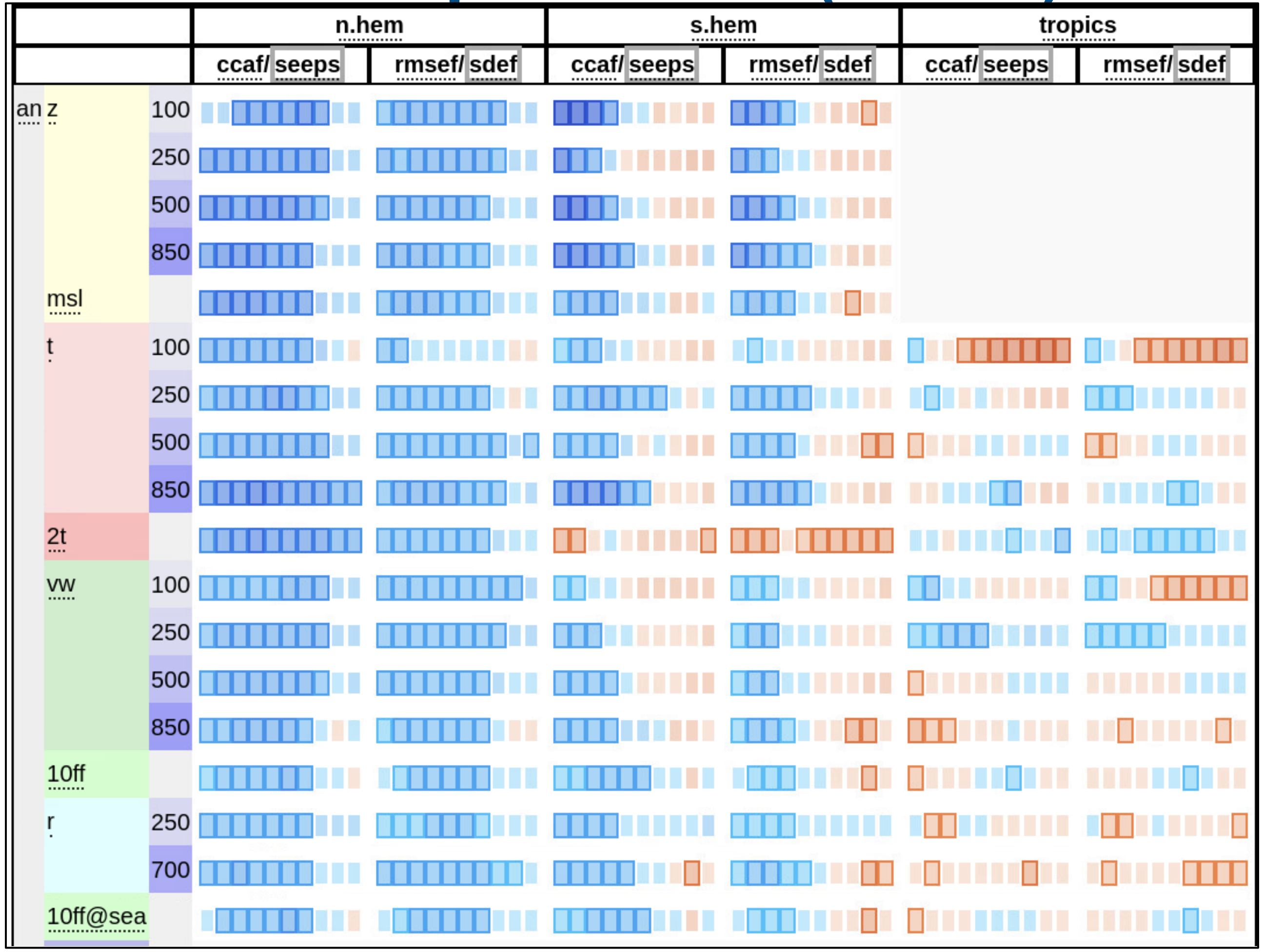



**Fig.** NN trained offline. Scorecard showing the change in forecast anomaly correlation and RMSE as a function of lead time. 20240101-20240228.

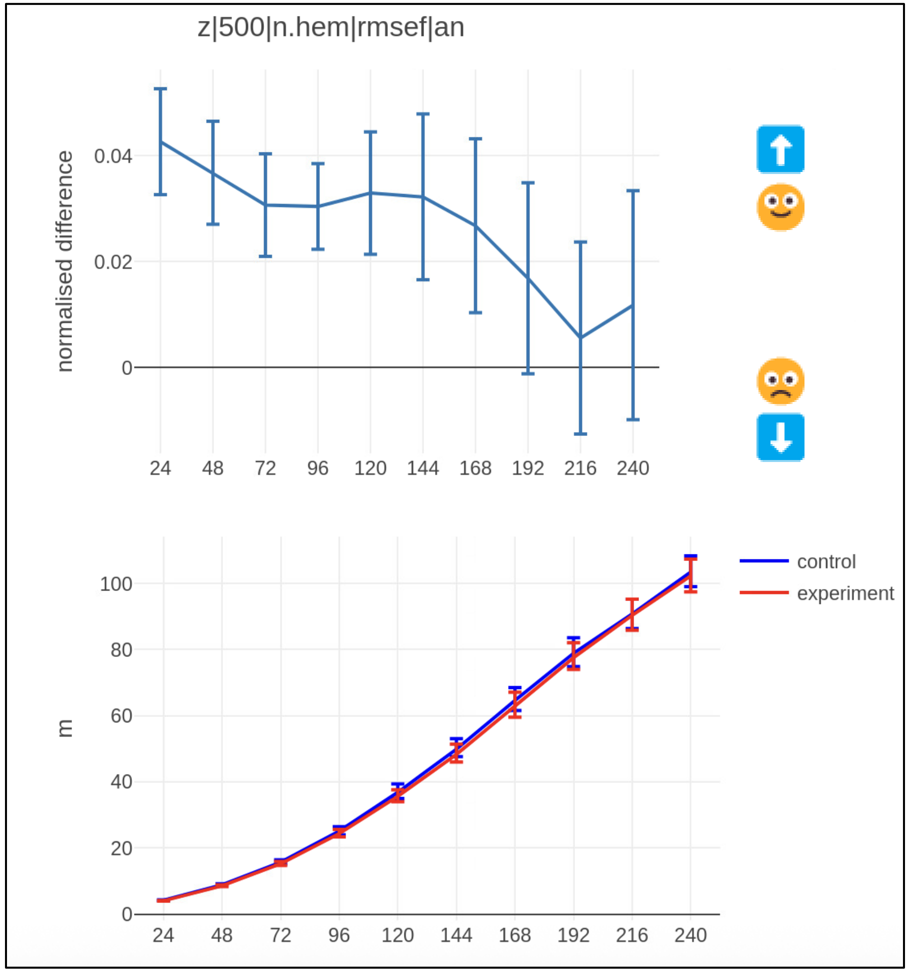
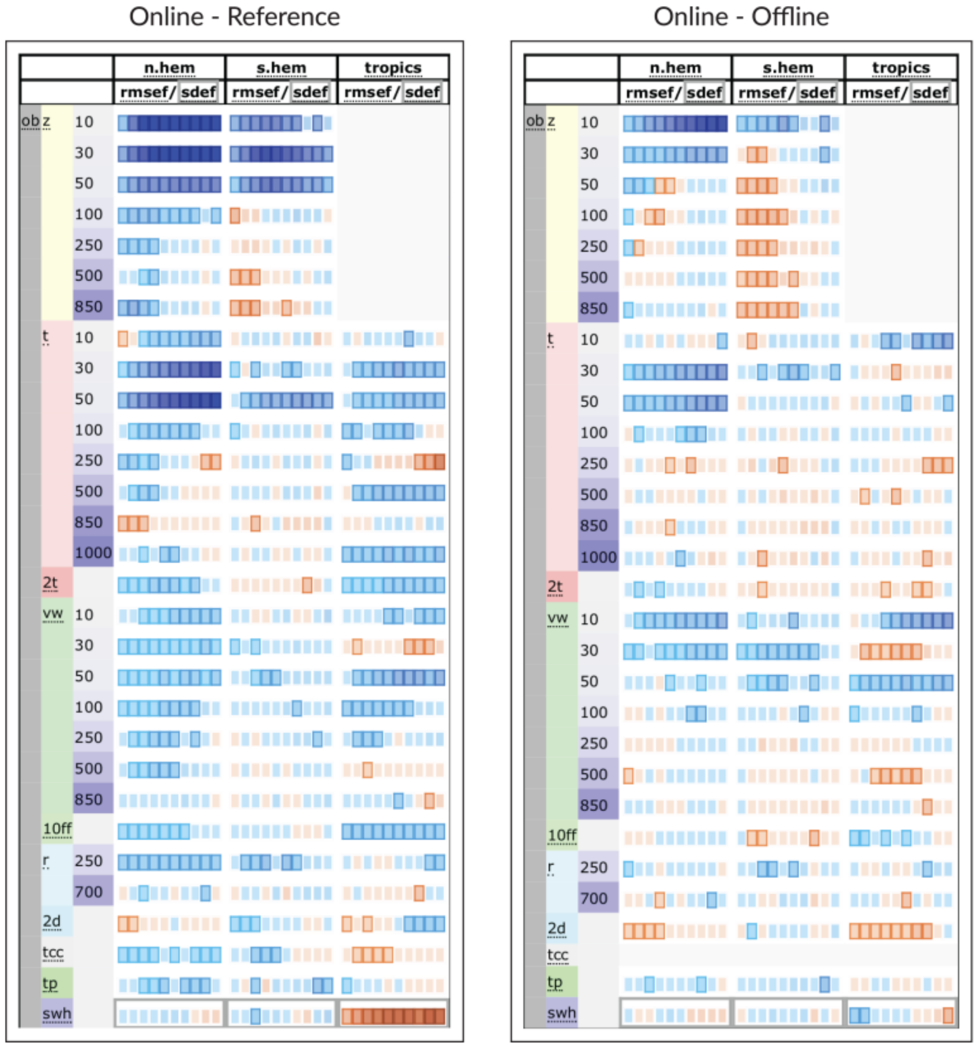



**Fig.** NN trained offline. Z500hPa NH RMSE. 20240101-20240228.

ECMWF

# Training online with the standard DA setup (12h window)




- The neural network is trained online at a resolution that is lower or equal to the inner loop resolution
- The neural network correction is applied at the resolution consistent with online training
- We obtained the best impact on forecast skill when training online/applying the correction online at the **T31** resolution
- The largest impact is visible in the stratosphere

**Fig.** Score cards showing the change in forecast RMSE verified against independent observations. Left panel: the impact of online training (Online) with respect to the standard weak-constraint configuration (Reference). Right panel: the impact of online training (Online) with respect to offline training (Offline). From Farchi et al. (2024).

# Training online with long DA window

- Why would we want to train the model of model error with a long DA window?
  - Improved constraint on model error – more observational information able to better constrain evolving model errors
  - Stronger coupling between observations and model dynamics – enhance the ability to identify systematic model errors
  - Helps capture model errors that develop over longer time-scale
- What are the obstacles?
  - Cannot run in the critical path unless past observations are used
  - The quality of the initial conditions may significantly deteriorate for the very long window (observation usage, QC)
  - Computational cost

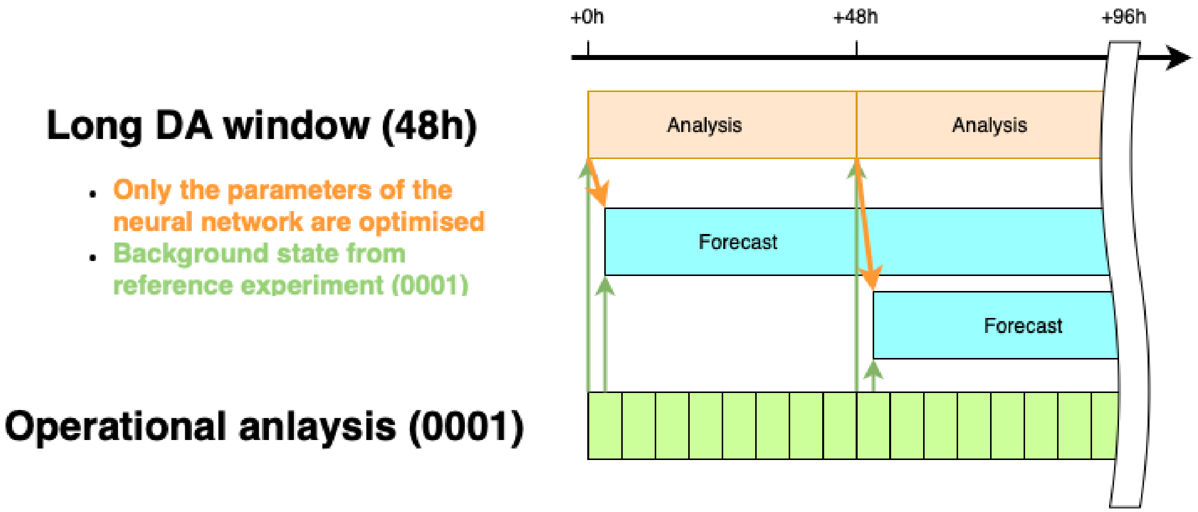



**Fig**. Online training setup with a long DA window (48h).

$$\mathcal{J}^{nn}(\theta) = \frac{1}{2}\left\|\theta - \theta^{b}\right\|^{2}_{P^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\left\|y_k - \mathcal{H}_k \circ \mathcal{M}^{\theta}_{k:0}(x_0)\right\|^{2}_{R_k^{-1}}$$

**Fig**. Modified cost function for the long DA window (48h).

Long DA window training setup:

- Only the parameters of the NN are optimized
- The atmospheric background state is not cycled and in each cycle is replaced with the operational analysis (0001).

# Training online with long window

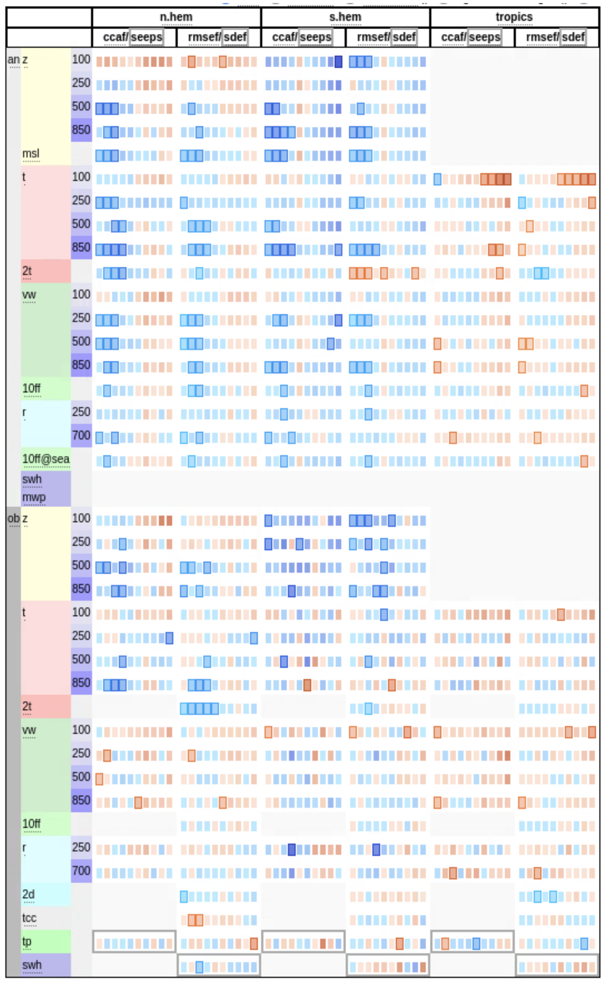


**Fig.** NN trained offline. Scorecard.

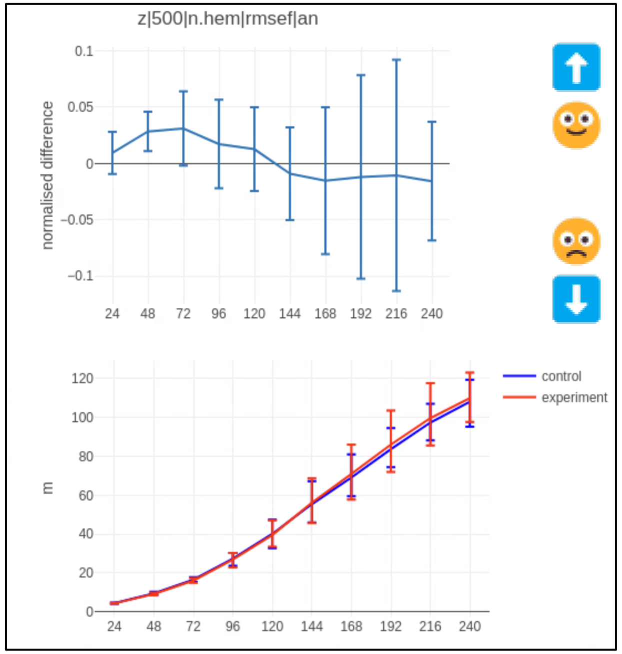



**Fig.** NN trained offline, Z500, NH, RMSE.

**Fig.** NN trained online. Scorecard.

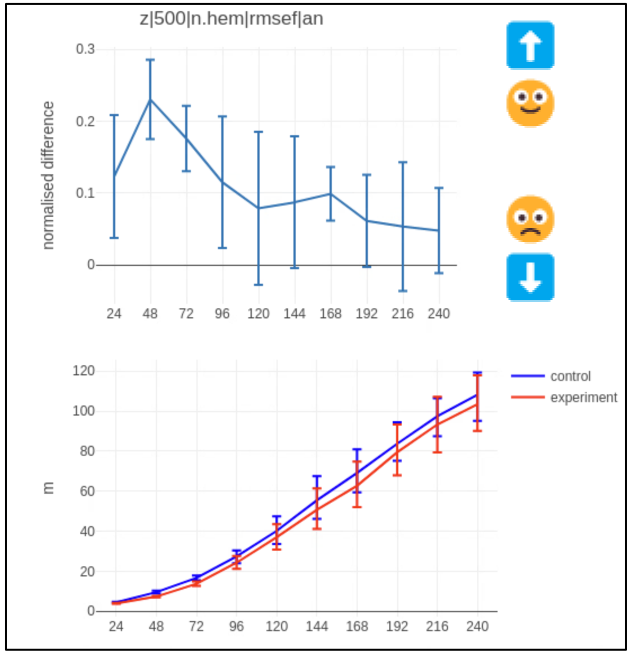



**Fig.** NN trained online, Z500, NH, RMSE.

- Training/Evaluation period: 20240101 – 20240331
- Window length: 48h

# Impact on forecast activity in Northern Hemisphere

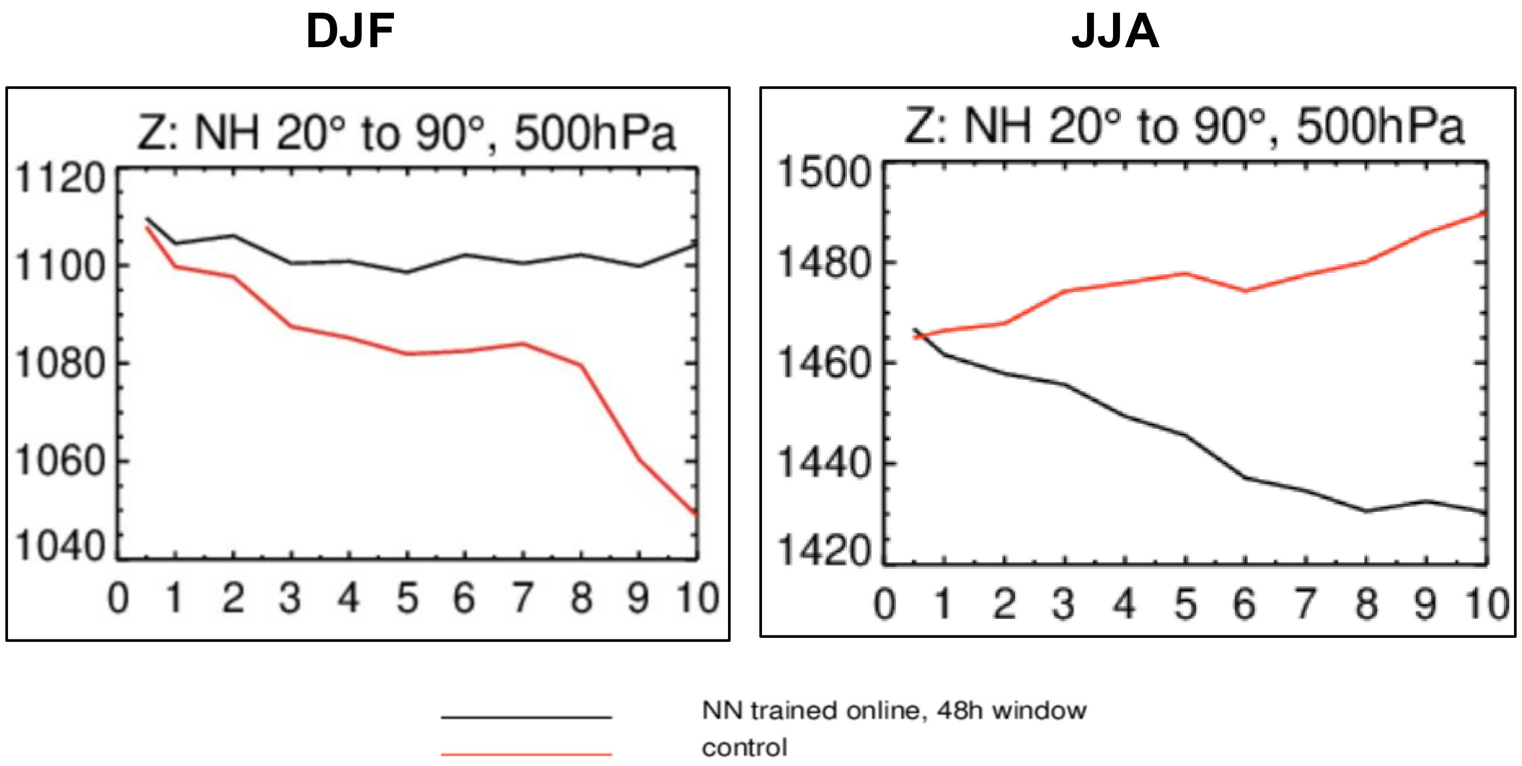
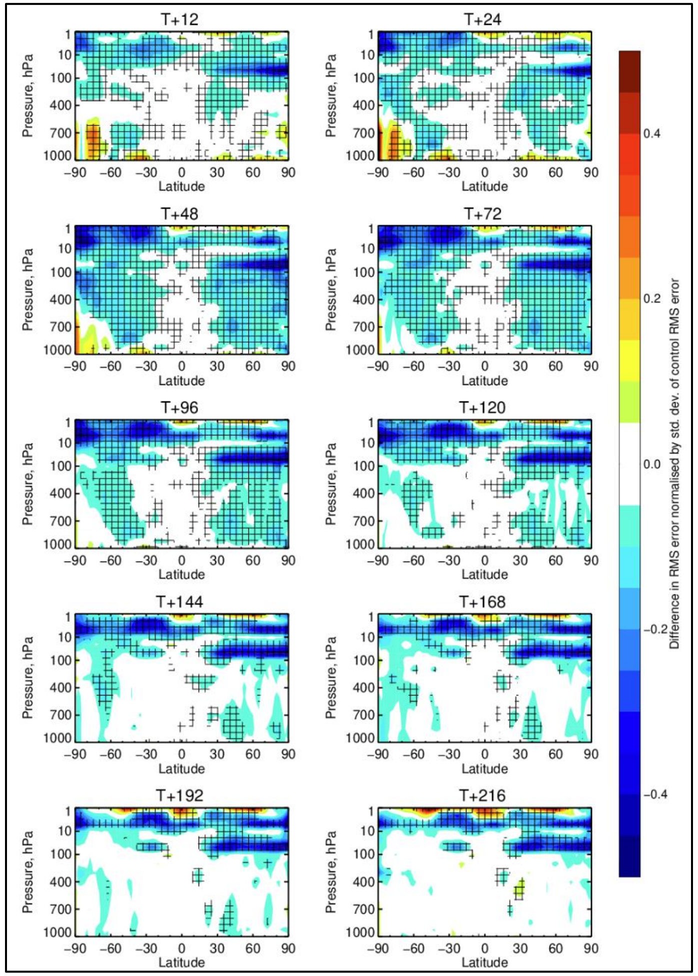



**Fig.** Impact of applying a NN model error correction on Z500 NH forecast activity. **Left**: DJF period and **Right**: JJA period.

- Applying NN-based model error correction impacts forecast activity:
  - Forecast activity is maintained in Northern Hemisphere (NH) winter
  - Forecast activity is suppressed in NH summer
- What are the reasons behind affecting the forecast activity in NH:
  - Only large-scale tendencies are corrected, while summer variability is driven by smaller-scale processes
  - The same correction is applied uniformly throughout 12h which may affect diurnal variability

# Effect of the temporal resolution

## 12h update frequency




## 3h update frequency

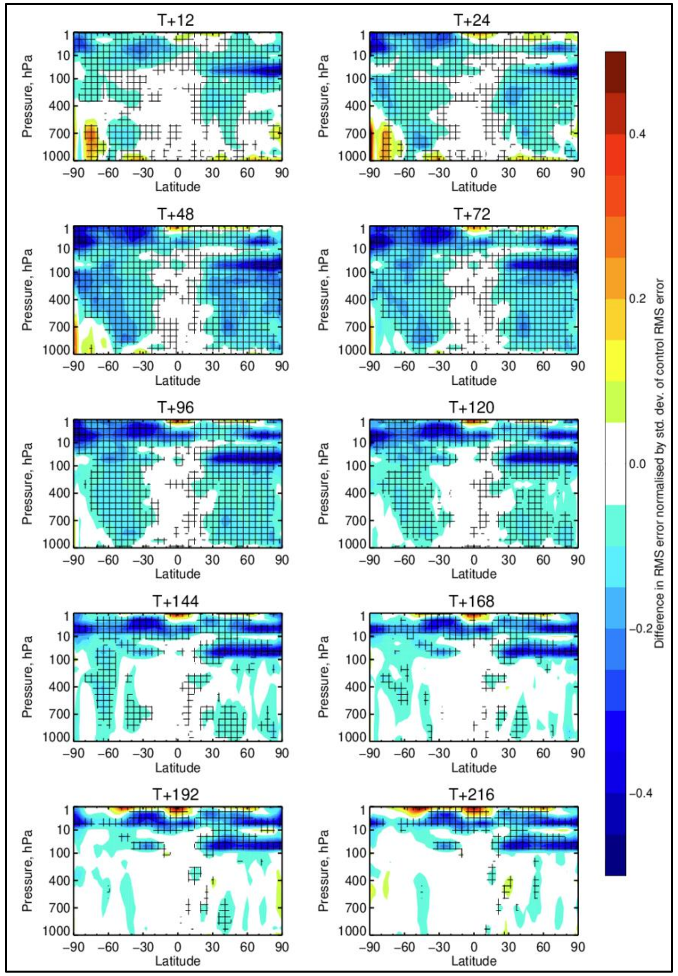
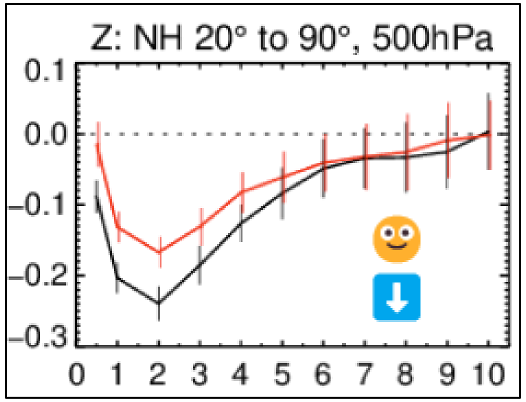



**Fig.** Change in RMS error in T. Online training with 48h DA window. **Left**: 12h correction update frequency. **Right**: 3h correction update frequency.



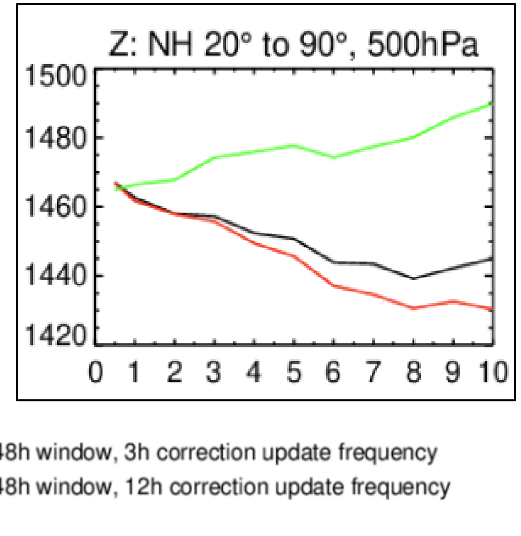



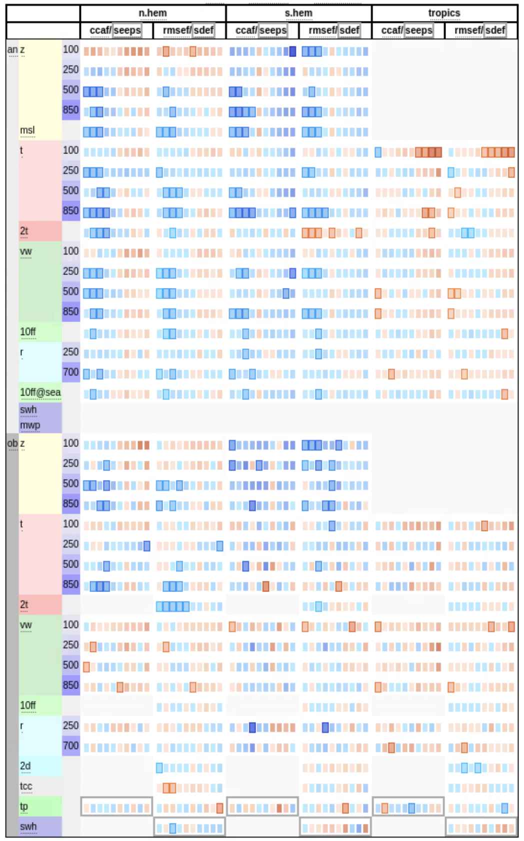

**Fig.** Impact of the correction update frequency on **Left**: Z500hPa NH RMSE, **Right**: forecast activity Z500hPa NH RMSE.

Increasing the frequency of the correction update:

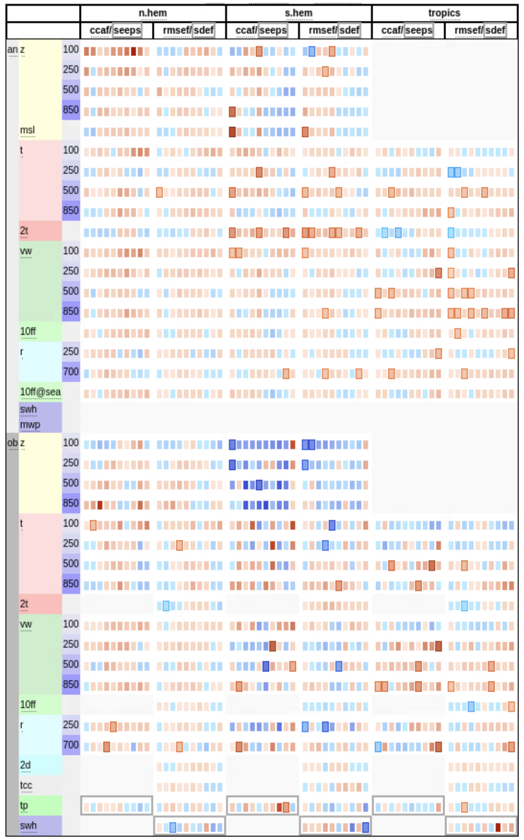
- Improves forecast scores
- Is positive for the forecast activity

Training/Evaluation period: 20220601 – 20220930

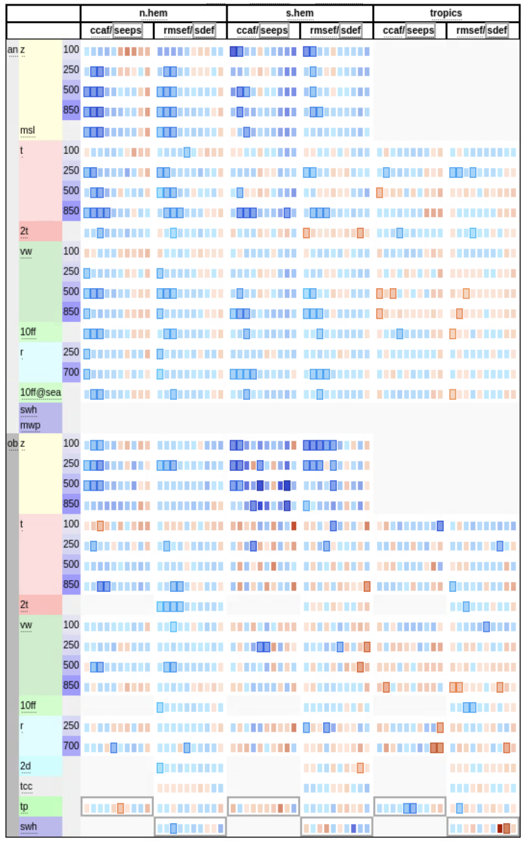
# Challenges: performance in out-of-training dates




NN trained offline




NN trained online in **lagged mode** with a large learning rate




NN trained online in **lagged mode** with a moderate learning rate

- A long DA window (48h) is not feasible in operations due to unavailable observations and timeliness constraints
- The online training would require years of training for the NN to generalise well beyond the training dataset
  - Significant changes to the ECMWF suites would be required to pursue this avenue
- A lagged system using previous 48h observations could be implemented

ECMWF

Training/Evaluation period: 20240101 – 20240331

# Conclusions

- Developed a physically consistent hybrid model that enhances the purely physics-based model
- Two step training of the neural network correction: offline followed by online fine-tuning within the framework of 4D-Var owing to the differentiable model and observation operators
- The neural network is directly constrained with observations in the online training step

- Online training with a long DA window:
  - Significant forecast improvements across the board
  - Weak generalisation of the NN to out of training samples
  - Need to train on multi-year datasets rather than 3 months

- Lagged DA window:
  - Feasible to implement in an operational environment
  - NN constantly updated with the latest observations
  - Improved performance with respect to the NN trained offline

- **Next steps**:
  - Offline pre-training at a 3h frequency
  - Integration of specific humidity online
  - Integration of the wavelet transformation online
  - Development of non-sequential online learning?