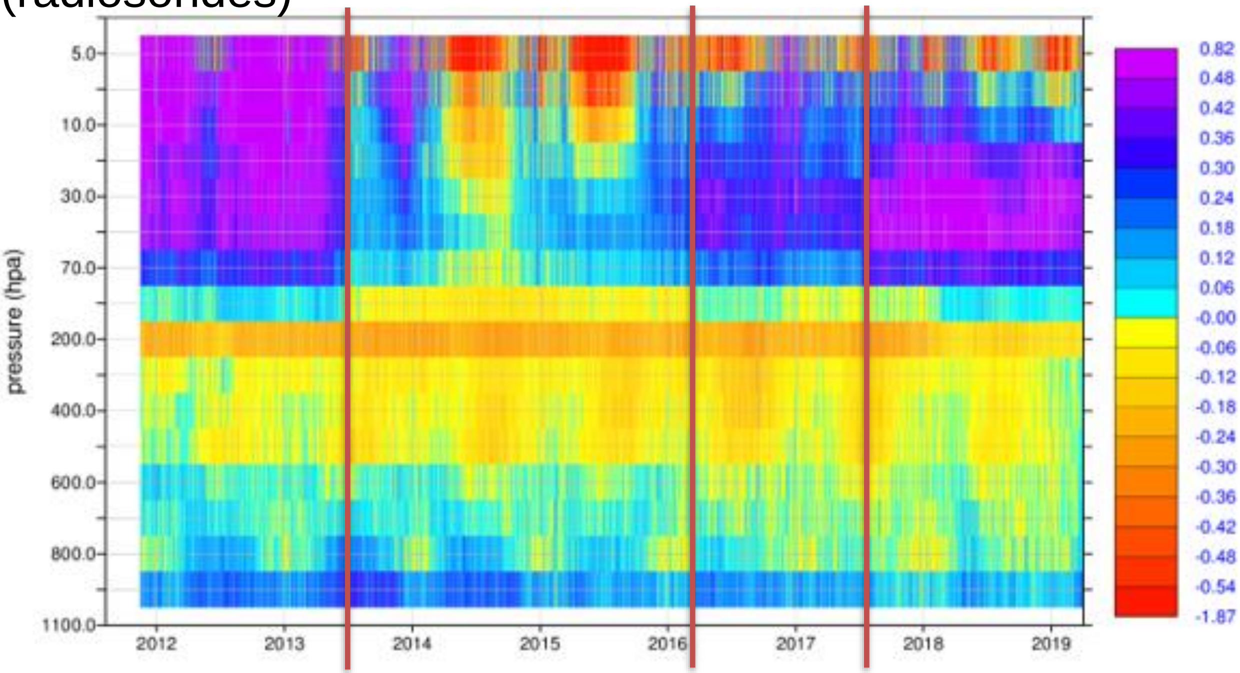
Machine Learning for Model Error in Numerical Weather Prediction

Patrick Laloyaux, Thorsten Kurth, David Hall, Stan Posey, Peter Dueben and Massimo Bonavita

Monitoring the quality of the atmospheric model

Comparison between the 12-hour model trajectory with reference observations (radiosondes)



Systematic error when the atmospheric model is integrated over 12 hours

→ Cold bias in the mid/lower stratosphere (>0.5C)

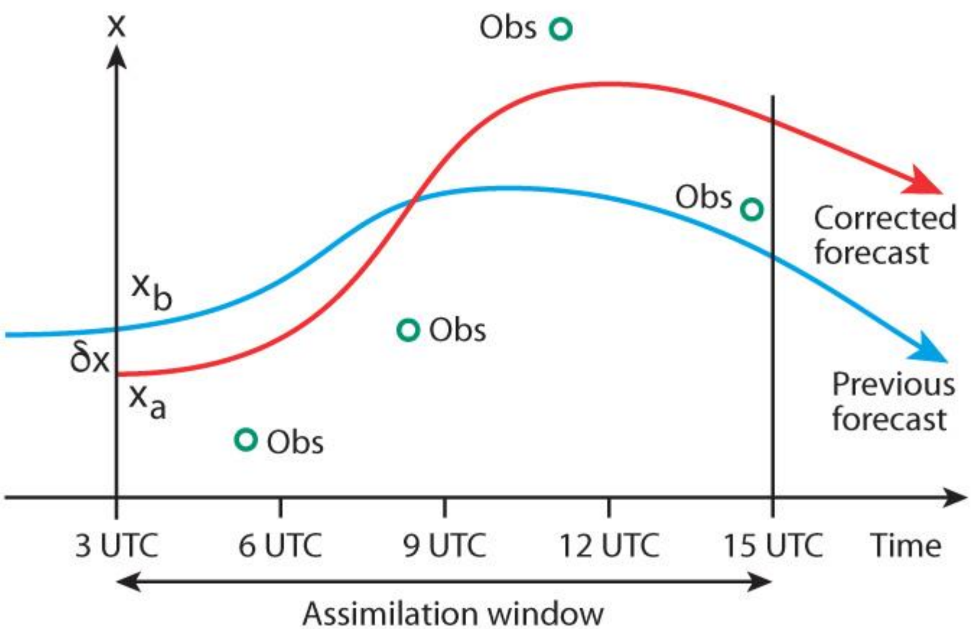
→ Warm bias in the upper stratosphere (>0.5C)

What is the best way to handle model biases? Changing our Data Assimilation system, using a Machine Learning approach or both?

Data Assimilation Approach

Data assimilation and standard 4D-Var formulation

4D-Var is a popular algorithm to find the optimal initial state by minimising the discrepancies with the prior estimate and the observations



Model's equation

$$x_k = \mathcal{M}_k(x_{k-1})$$

4D-Var cost function

$$\begin{aligned} J(x_0) &= \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) \\ &+ \frac{1}{2}\sum_{k=0}^{K}[y_k - \mathcal{H}(x_k)]^T \mathbf{R}_k^{-1}[y_k - \mathcal{H}(x_k)] \end{aligned}$$

- ➔ 4D-Var fills the blanks between observations using the model's equations and the regularisation term in the cost function
- ➔ The cost function (loss function) is minimised using adjoint integrations (backpropagation)
- ➔ Equivalence with the standard Kalman Filter analysis update (linear operators)

4D-Var formulation with biased model (weak-constraint)

Because the model is biased, we add an error term η in the model equation

$$x_k = \mathcal{M}_k(x_{k-1}) + \eta \quad \text{for } k = 1, 2, \cdots, K$$

The model bias correction η contains 3 physical fields

- temperature
- vorticity
- divergence

$$\begin{aligned} J(x_0, \beta, \eta) &= \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) \\ &+ \frac{1}{2}\sum_{k=0}^{K}[y_k - \mathcal{H}(x_k) - b(x_k, \beta)]^T \mathbf{R}_k^{-1}[y_k - \mathcal{H}(x_k) - b(x_k, \beta)] \\ &+ \frac{1}{2}(\beta - \beta_b)^T \mathbf{B}_\beta{}^{-1}(\beta - \beta_b) \\ &+ \frac{1}{2}(\eta - \eta_b)^T \mathbf{Q}^{-1}(\eta - \eta_b) \end{aligned}$$

Model state

Observation bias correction

Model bias correction

→ Introduce additional controls to fit background and observations

→ The model error covariance matrix Q constrains the model error field

→ Constant model error forcing over the assimilation window to correct the model bias

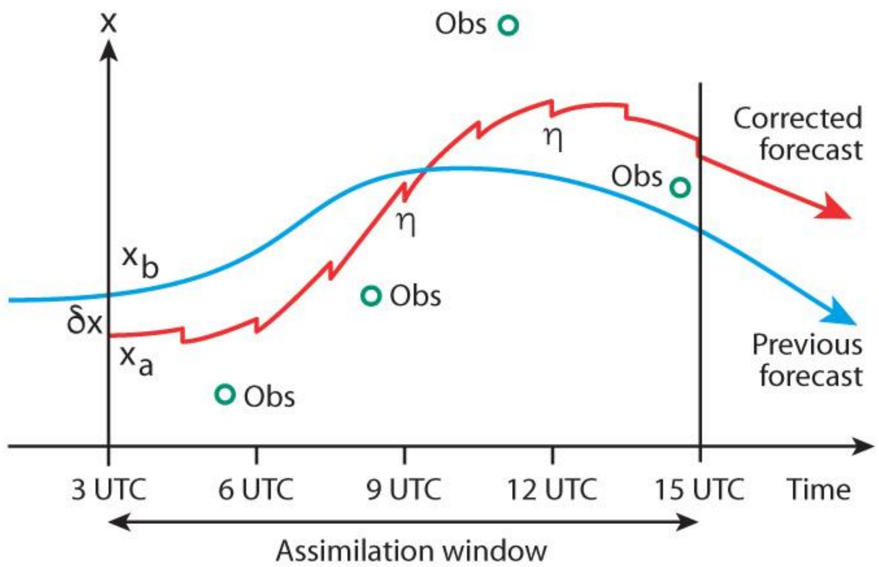
4D-Var formulation with biased model (weak-constraint)

We assume that the model is not perfect, adding an error term η in the model equation

$$x_k = \mathcal{M}_k(x_{k-1}) + \eta \quad \text{for } k = 1, 2, \cdots, K$$

The model error estimate η contains 3 physical fields

- temperature
- vorticity
- divergence



→ Introduce additional controls to fit background and observations

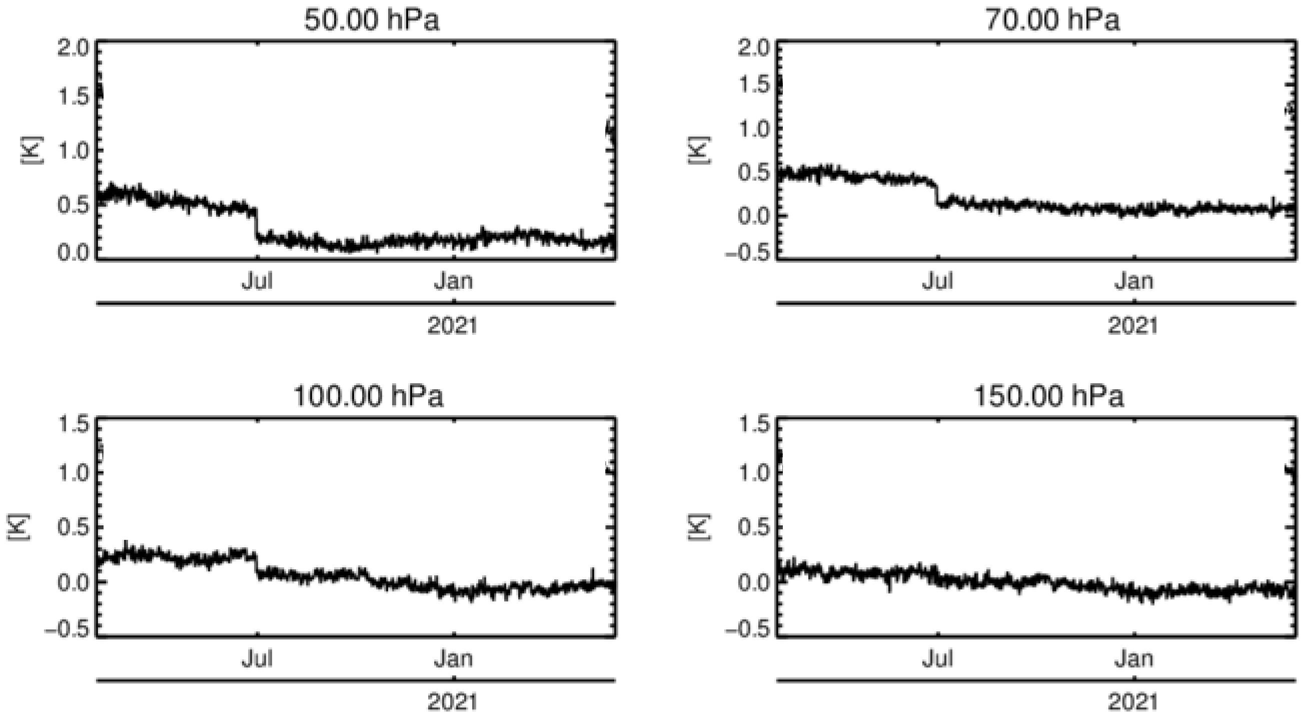
→ The model error covariance matrix Q constrains the model error field

→ Constant model error forcing over the assimilation window to correct the model bias

4D-Var formulation with biased model (weak-constraint)

This technique is used operationally since 30 June 2020 to correct the stratospheric biases

Mean first-guess departure with respect to temperature measurements from radiosondes

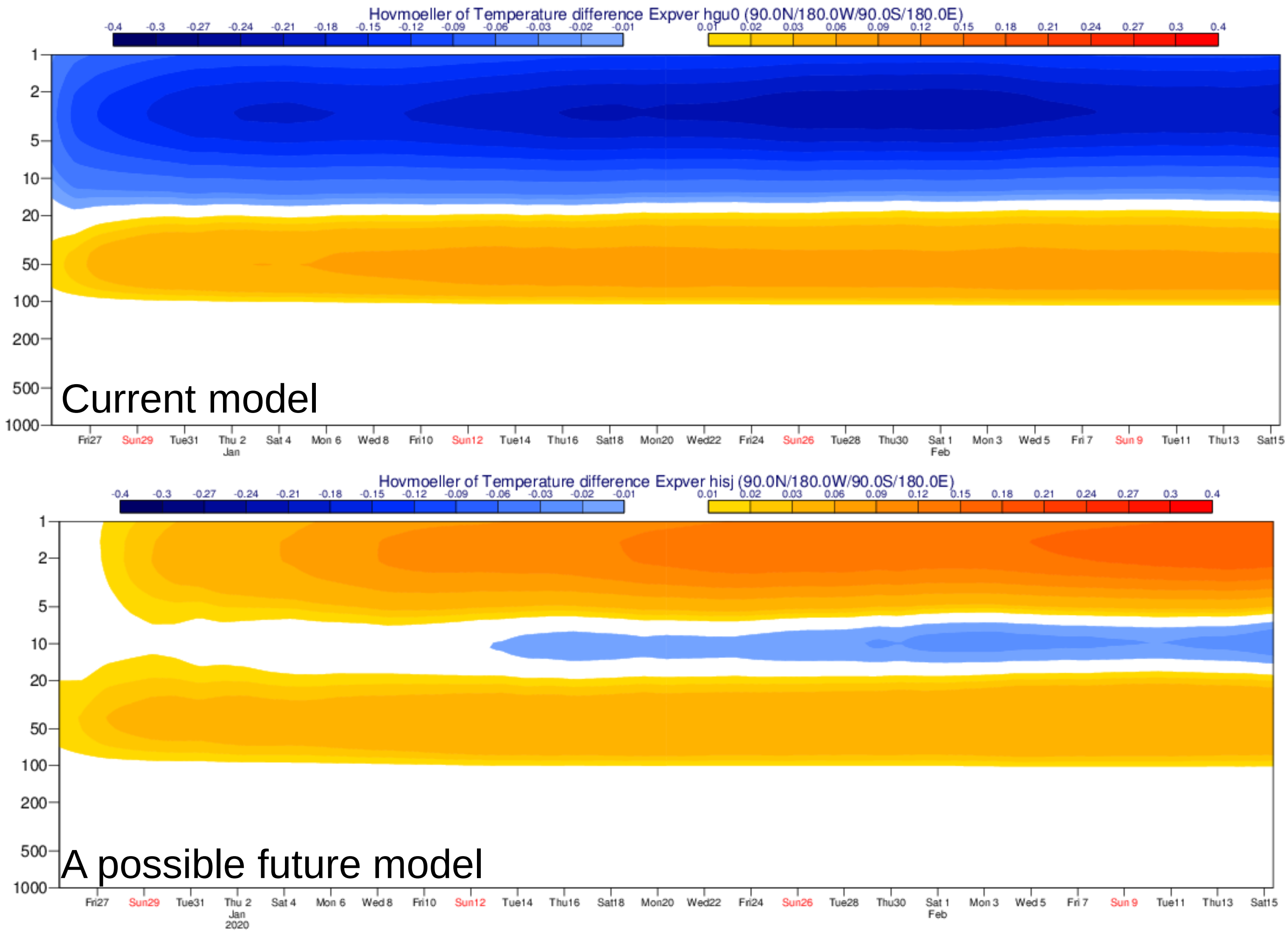


Model upgrade at ECMWF

Implementation date	Summary of changes	Resolution	Full IFS documentation
11-May-2021	Cycle 47r2	ENS (Vertical)	CY47r2
30-Jun-2020	Cycle 47r1	Unchanged	CY47R1
11-Jun-2019	Cycle 46r1	Unchanged	CY46R1
05-Jun-2018	Cycle 45r1	Unchanged	CY45R1
05-Nov-2017	Implementation of Seasonal Forecast SEAS5	Unchanged	Documentation
11-Jul-17	Cycle 43r3	Unchanged	CY43R3
22-Nov-16	Cycle 43r1	Ocean (Horizontal & vertical)	CY43R1

The ECMWF model is upgraded every year. The bias of the new model is different and need to be estimated

Learning rate of weak-constraint 4D-Var



Weak-constraint 4D-Var learns model biases rather quickly (~4 weeks)

Machine Learning Approach

‘Science and Technology’ strategic actions

The ‘Science and Technology’ strategic actions are linked to enhancements in the exploitation of observations, data assimilation, modelling and exploitation of new technologies, computational science and operational processes.

Strengthen leadership in Earth system data assimilation

ECMWF will strengthen its leadership position in data assimilation by progressing in coupled assimilation, algorithmic development and integration of approaches. This will include the incorporation of machine learning, with 4D-Var data assimilation being uniquely positioned to benefit from integrating machine learning technologies because the two fields share a common theoretical foundation and use similar computational tools.

Cost / loss function equivalence of ML and variational DA

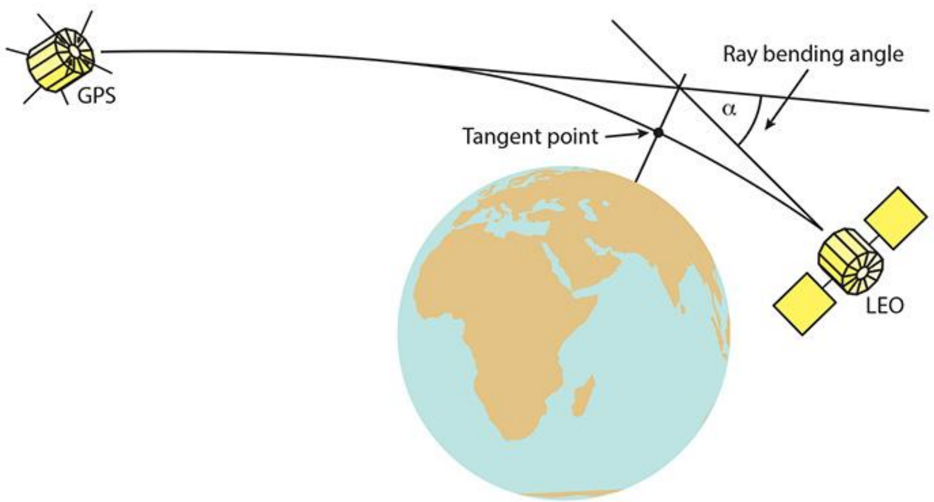
A. Geer (2021) Learning earth system models from observations: machine learning or data assimilation?

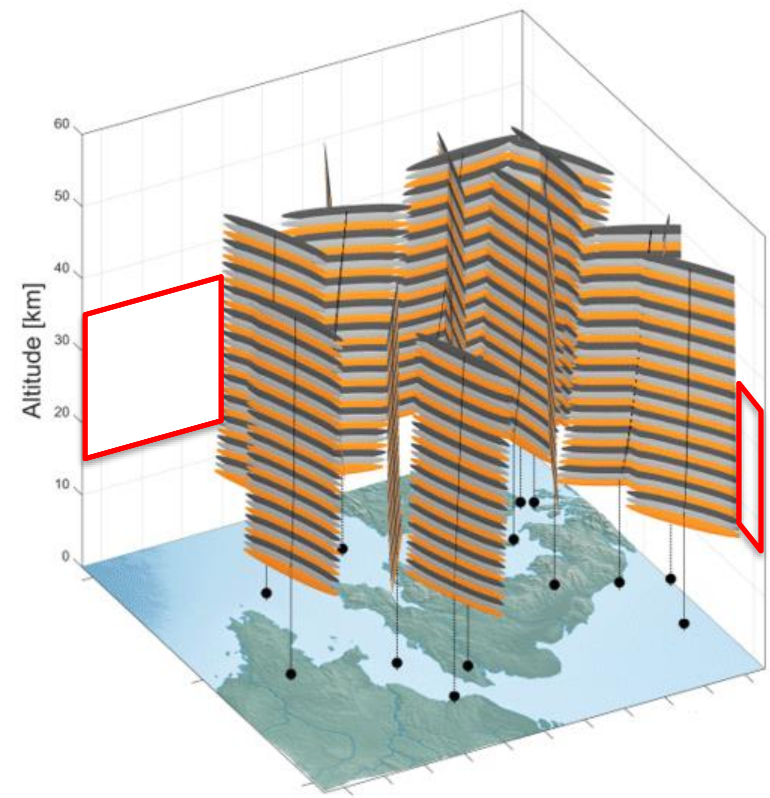
ML	Loss function	Basic loss function	Feature error?	Weights regularisation
DA	Cost function	Observation term	Prior knowledge of state	Prior knowledge of model

$$J(x,w) = \underbrace{\frac{(y-h(x,w))^2}{(\cancel{\sigma^y})^2}}_{J^y} + \underbrace{\cancel{\frac{(x^b-x)^2}{(\sigma^x)^2}}}_{J^x} + \underbrace{\frac{(\cancel{w^b}-w)^2}{(\cancel{\sigma^w})^2}}_{J^w}$$

How to correct biases using Machine Learning?

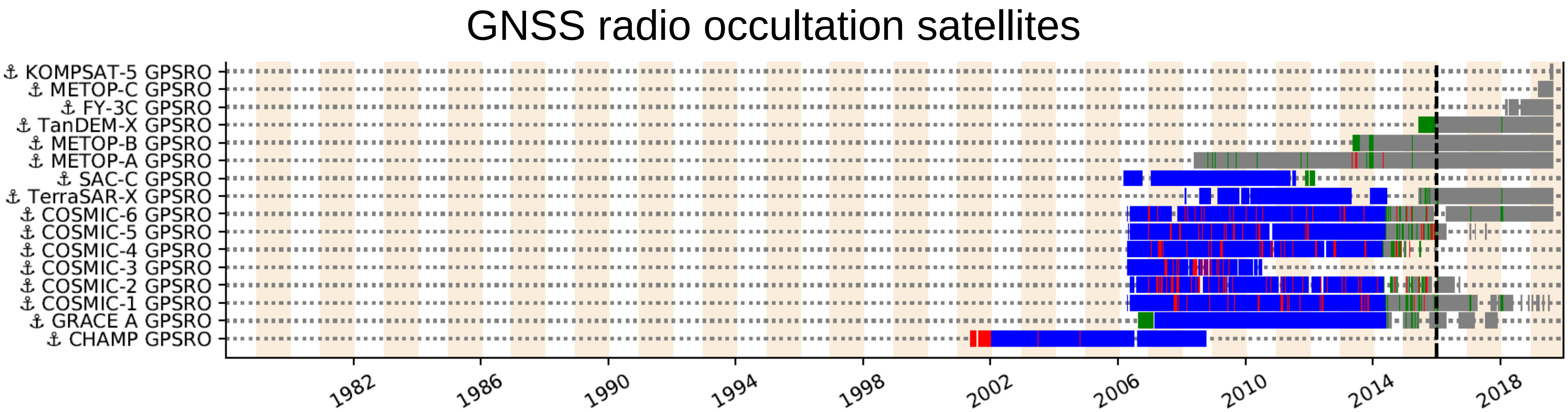
Radio occultation technique: Temperature profiles can be derived from the bending caused by the atmosphere along paths between a GNSS satellite and a LEO satellite

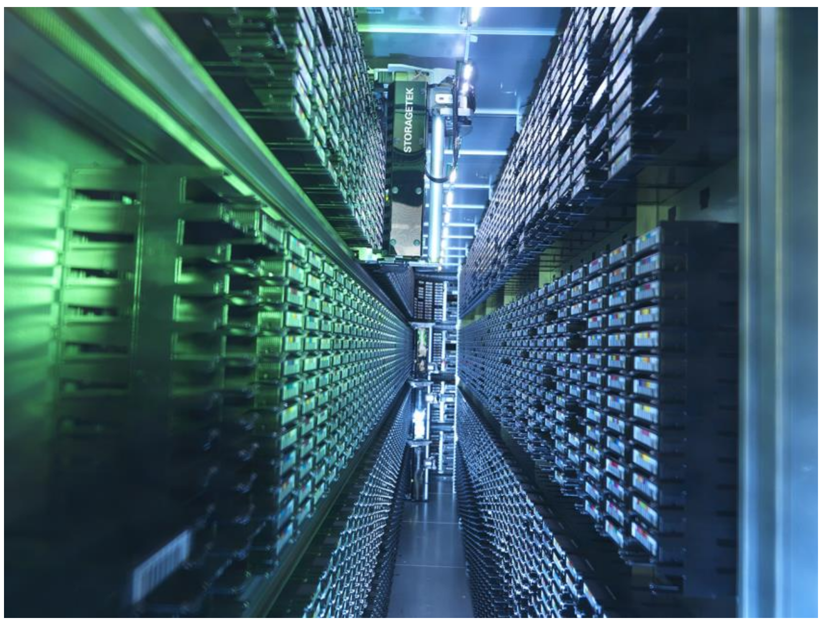


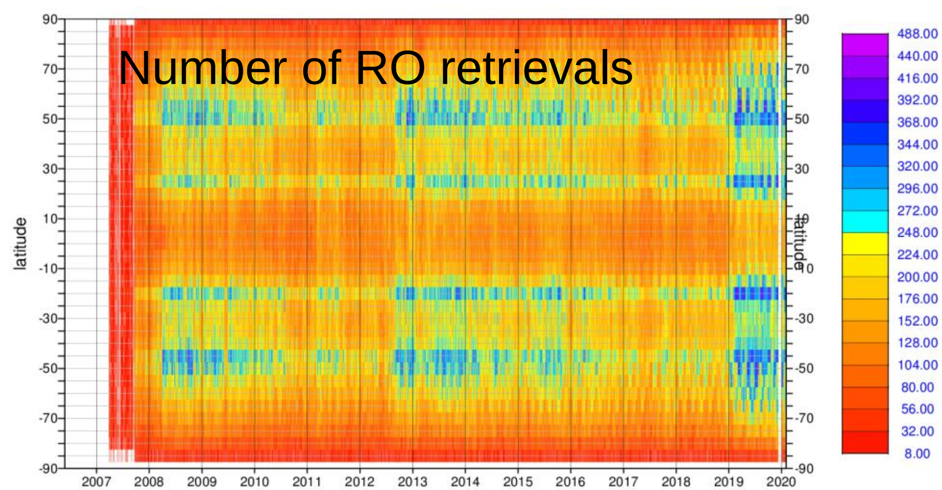


Temperature profiles are very accurate in the stratosphere (between 10-50 km). They are good for highlighting errors/biases

How to correct biases using Machine Learning?



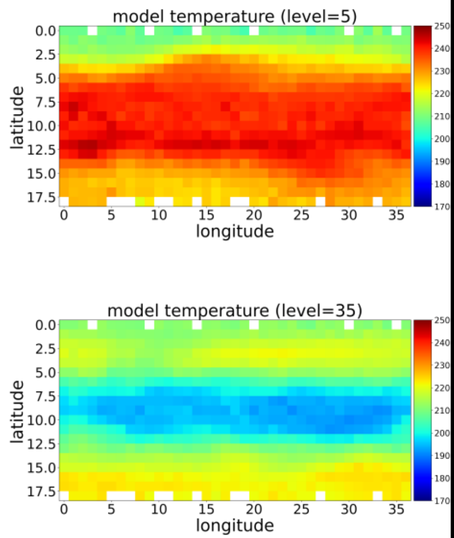




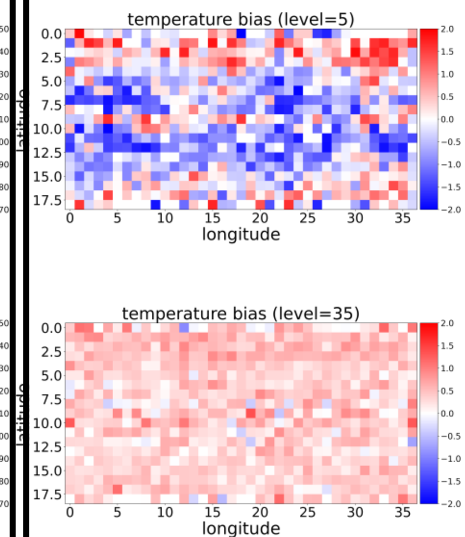
250 millions of observations and their model equivalent (ERA5) have been extracted to create a dataset (2008-2021)

Building a dataset for Machine learning

Predictors (input)
Temperature estimate from the model



Predictand (output)
Difference between the temperature from the model and the retrieval



The atmospheric state is never fully observed in NWP. Average measurements on a 10-degree grid every 10 days. Interpolate to fill the gaps

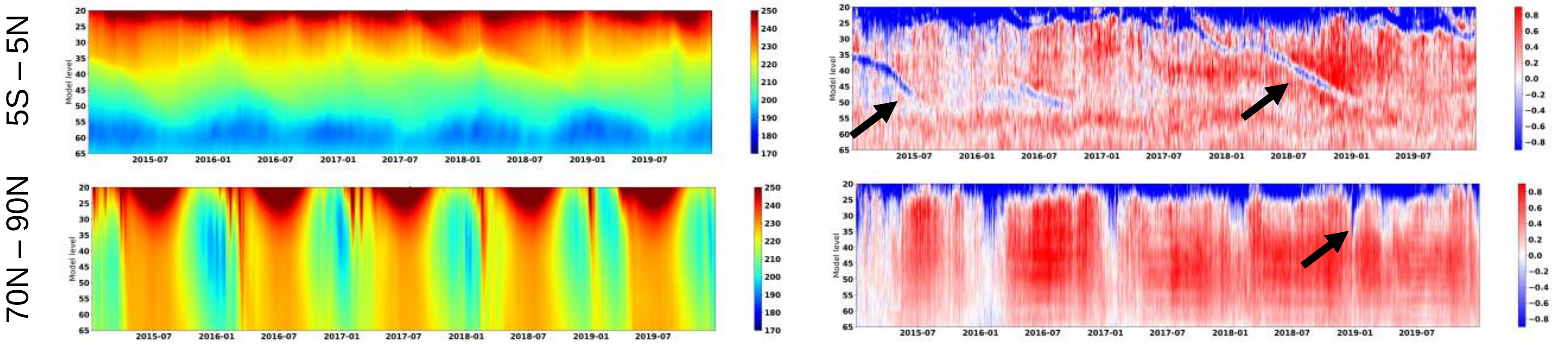
Dataset size

- input, output: 19x37x45 (31635)
- training set: 2008-2018 (2300 samples)
- validation set: 2019
- test set: 2020 and 2021

Building a dataset for Machine learning

NN input: temperature field from IFS

NN target: temperature bias correction (RO minus model)



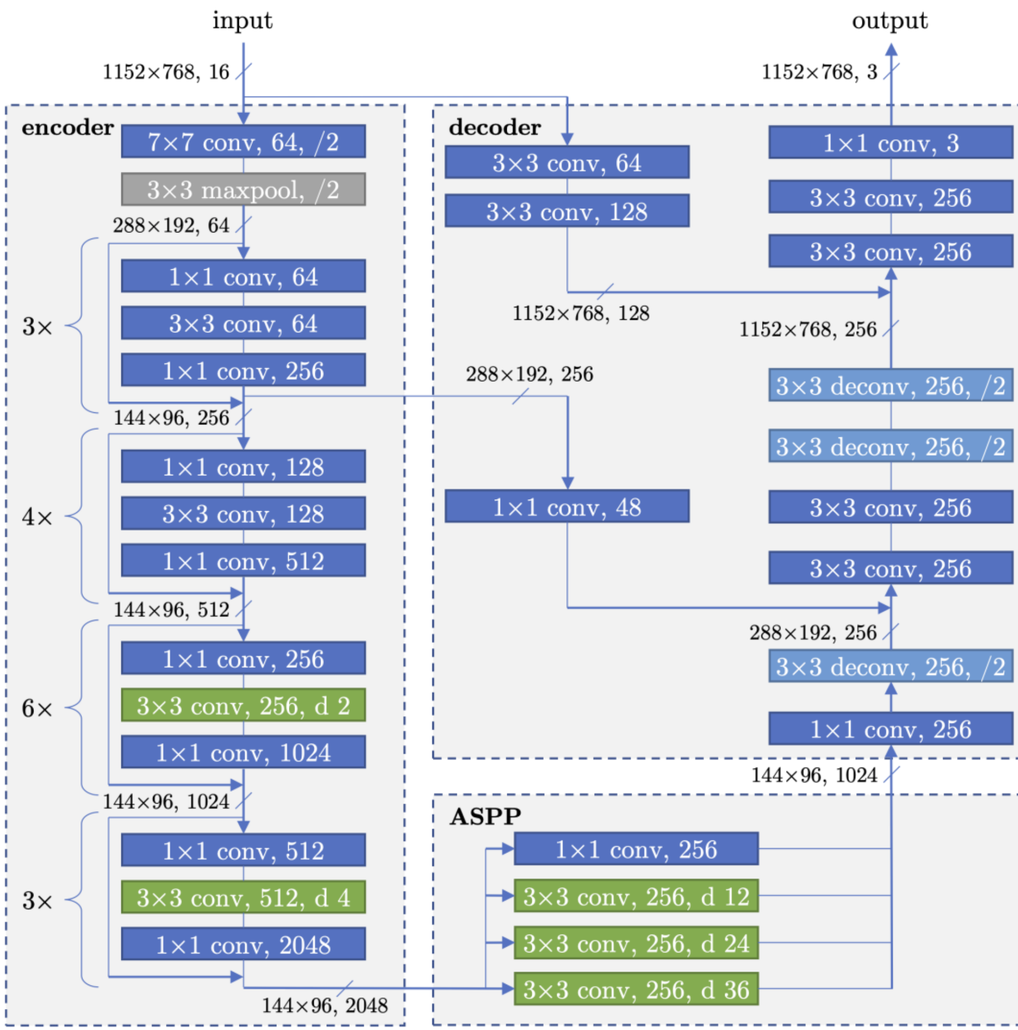
Features that the NN should learn

➔Positive correction below 30, negative correction above 30

➔Blue diagonal stripes linked to the Quasi-Biennial Oscillation (QBO)

➔Sharp vertical blue lines linked to Sudden Stratospheric Warming (SSW) event

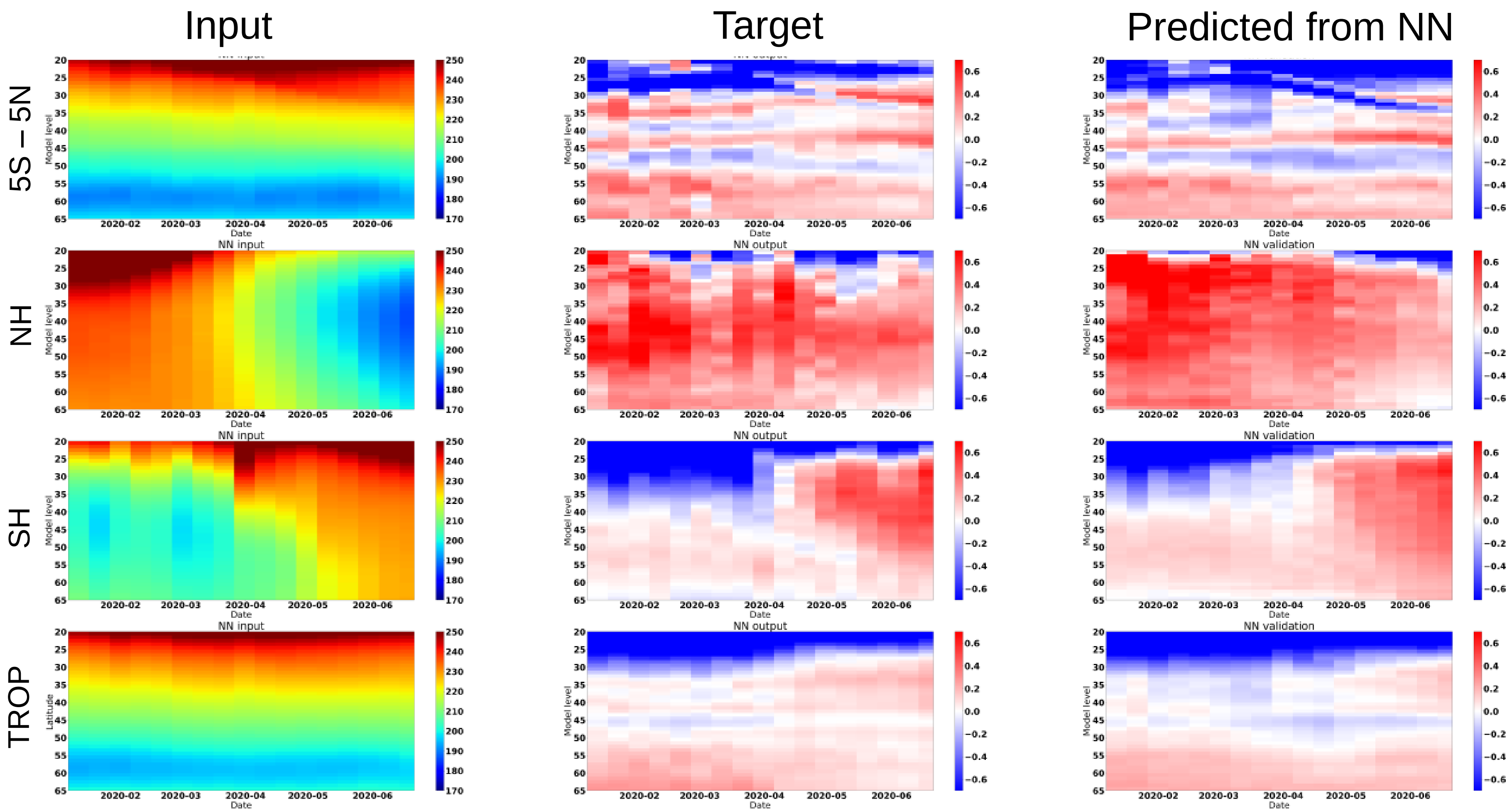
Training 3D Convolutional Neural Network



Convolutional neural network (CNN) are the best to learn computer-vision task. The usual 3 channels (RGB) have been replaced by 45 channels (vertical levels in the stratosphere)

Kurth et al., doi:10.5555/3291656.3291724

Results from the NN



NN results are really good but it required a large dataset for training (2008-2018)

Retraining of the NN

Implementation date	Summary of changes	Resolution	Full IFS documentation
11-May-2021	Cycle 47r2	ENS (Vertical)	CY47r2
30-Jun-2020	Cycle 47r1	Unchanged	CY47R1
11-Jun-2019	Cycle 46r1	Unchanged	CY46R1
05-Jun-2018	Cycle 45r1	Unchanged	CY45R1
05-Nov-2017	Implementation of Seasonal Forecast SEAS5	Unchanged	Documentation
11-Jul-17	Cycle 43r3	Unchanged	CY43R3
22-Nov-16	Cycle 43r1	Ocean (Horizontal & vertical)	CY43R1

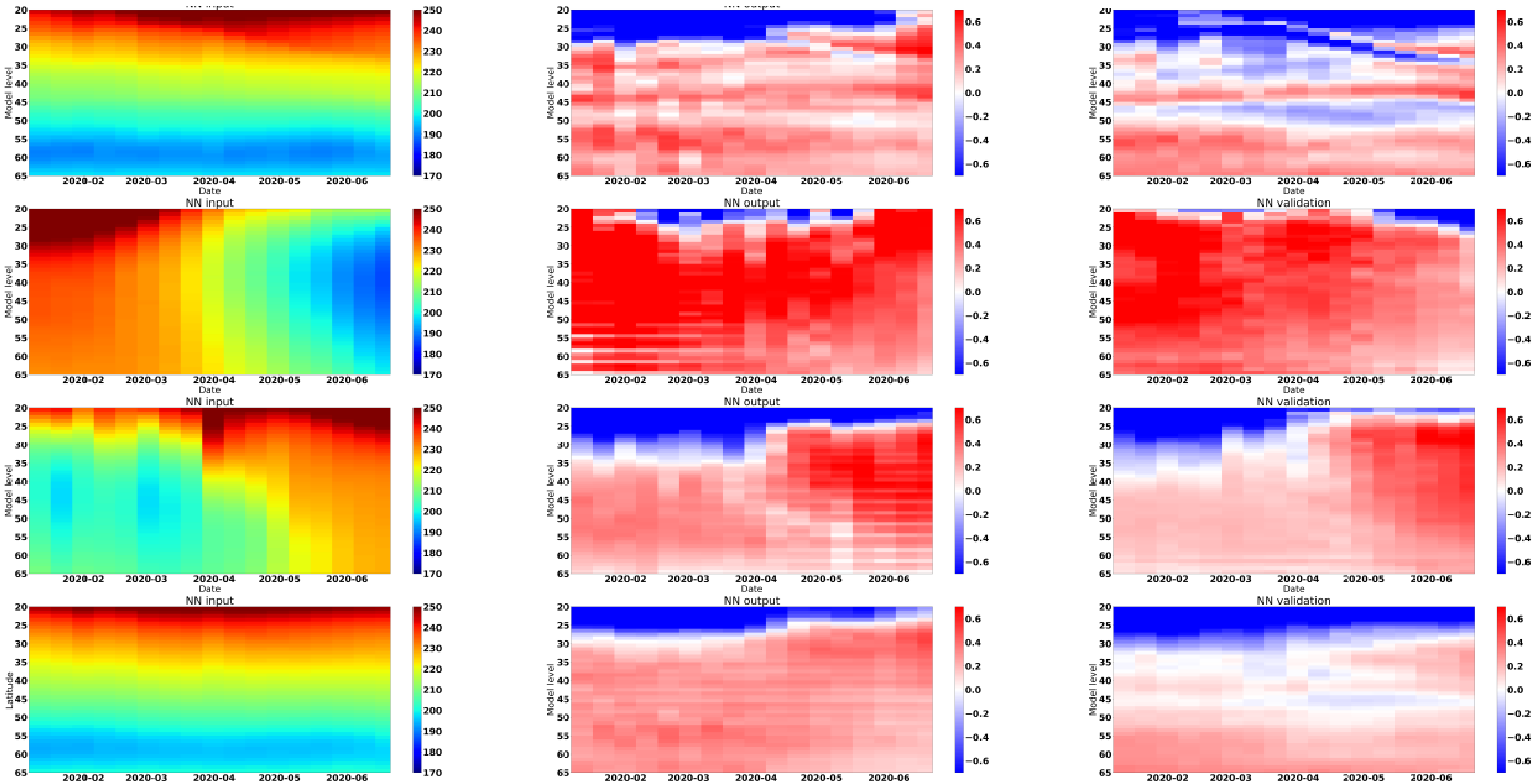
The ECMWF model is upgraded every year of so which means that the model bias is changing and need to be retrained

How can we retrain the NN to estimate the bias from the new model?

Retraining of the NN

We can only compute a one year dataset with the new model
Training (last 6 months in 2019, 20 samples)
Test (first 6 months in 2020, 20 samples)

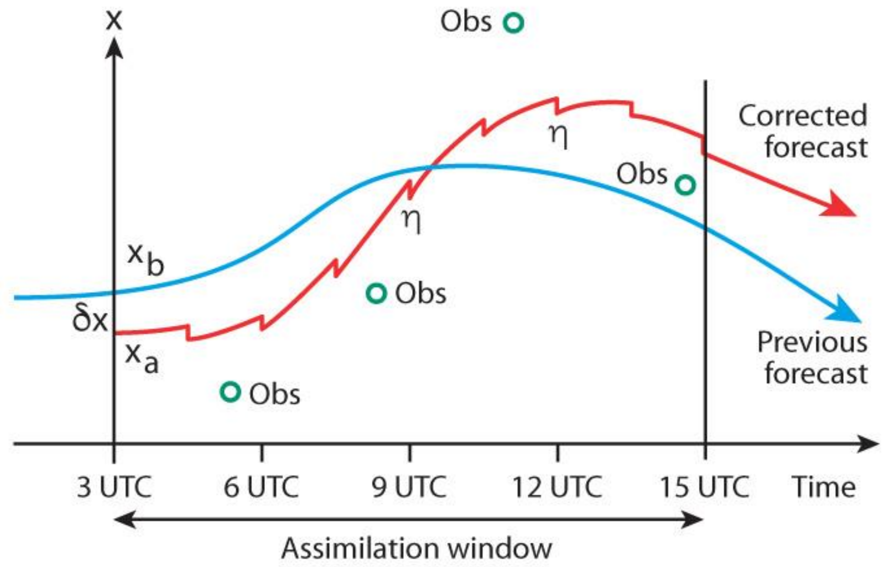
NEW Input | New Target | Predicted from the NN

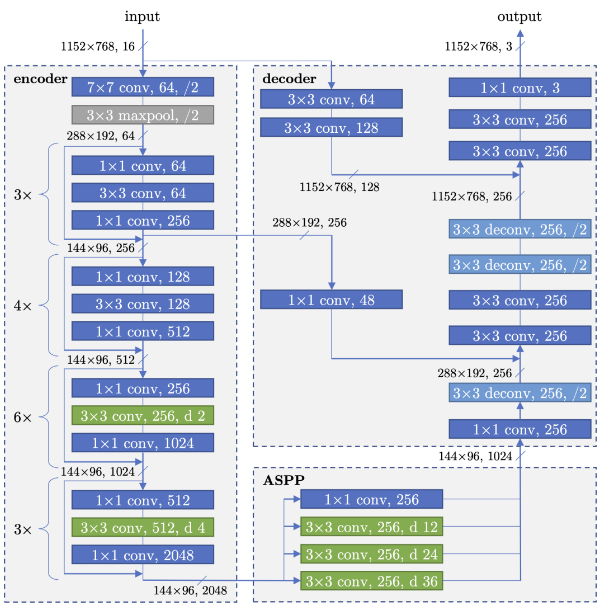


NN results are not as good, mainly due to the small number of samples

Conclusion and future work

ECMWF has implemented a weak-constraint 4D-Var in operations that learns and correct model biases in the stratosphere





- NN technique is able to learn the model bias but requires large datasets.
- Retraining is challenging as limited availability of samples (~3-6 months of data)
- Creating the best dataset is difficult as it is a tradeoff between resolution, noise and sparsity.

Next steps: integrate the NN into the full DA pipeline and compare this approach with weak-constraint 4D-Var