Accounting for systematic model error in Data Assimilation

An overview of the IFS and DOP models

Patrick Laloyaux, Massimo Bonavita, Marcin Chrust, Elias Holm, Alban Farci, Mikael Kaandorp, Inna Polichtchouk, Yannick Tremolet, Selime Gurol, Mihai Alexe, Eulalie Boucher, Peter Lean, Ewan Pinnington, Tomas Kral, Simon Lang, Sean Healy, Niels Bormann, Mat Chantry, Tony McNally, and many others at ECMWF

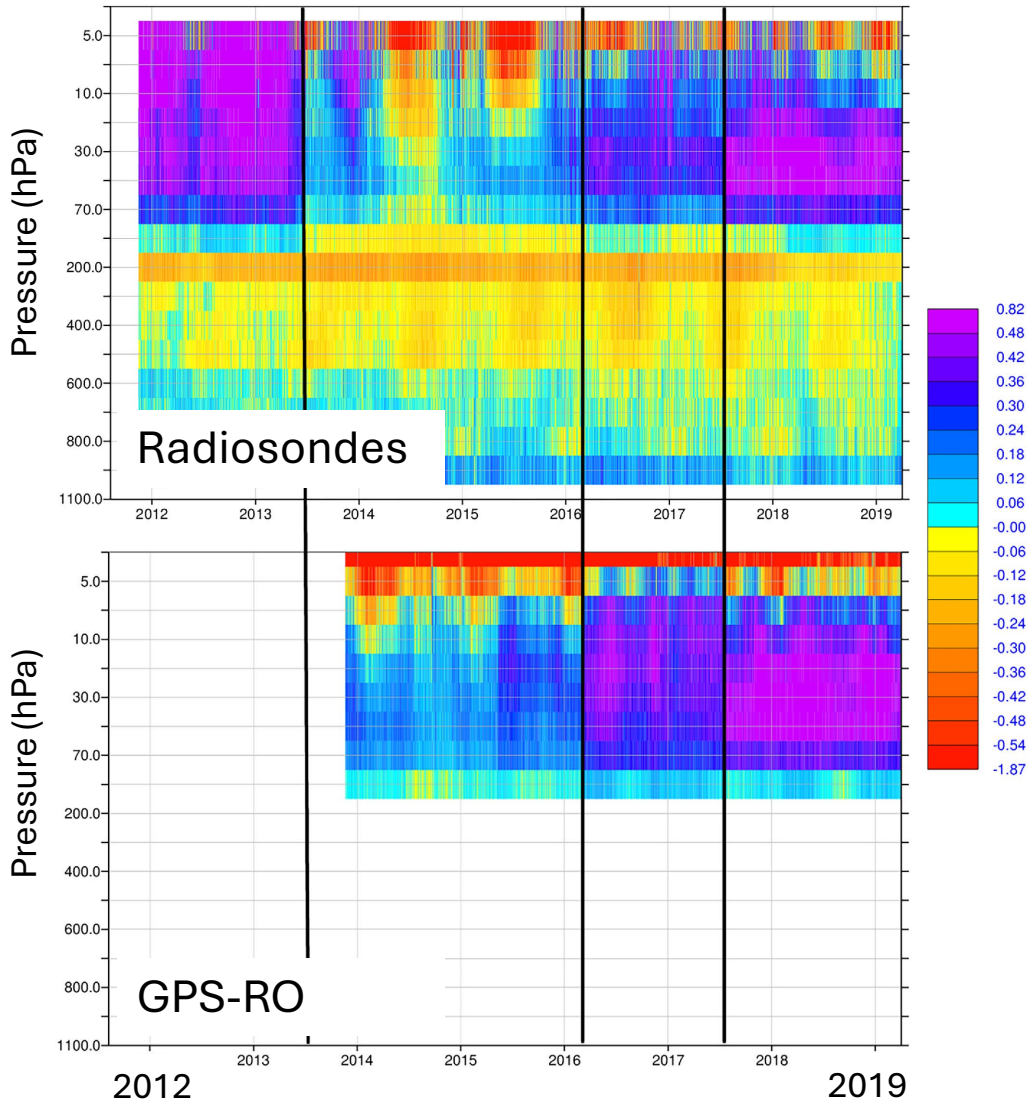


Accounting for systematic model error in Data Assimilation

1. **Integrated Forecasting System (IFS) model**
2. **Direct Observation Prediction (DOP) model**

An overall look at the IFS model error

Difference between obs and IFS trajectory (O-B)



The modelling Section has been working hard to reduce biases in the IFS (challenging with scientific choices and hidden trade-offs)

There is still a residual bias in the stratosphere

➔ cold bias in the lower/mid stratosphere

➔ warm bias in the upper stratosphere

Can Data Assimilation correct model errors?

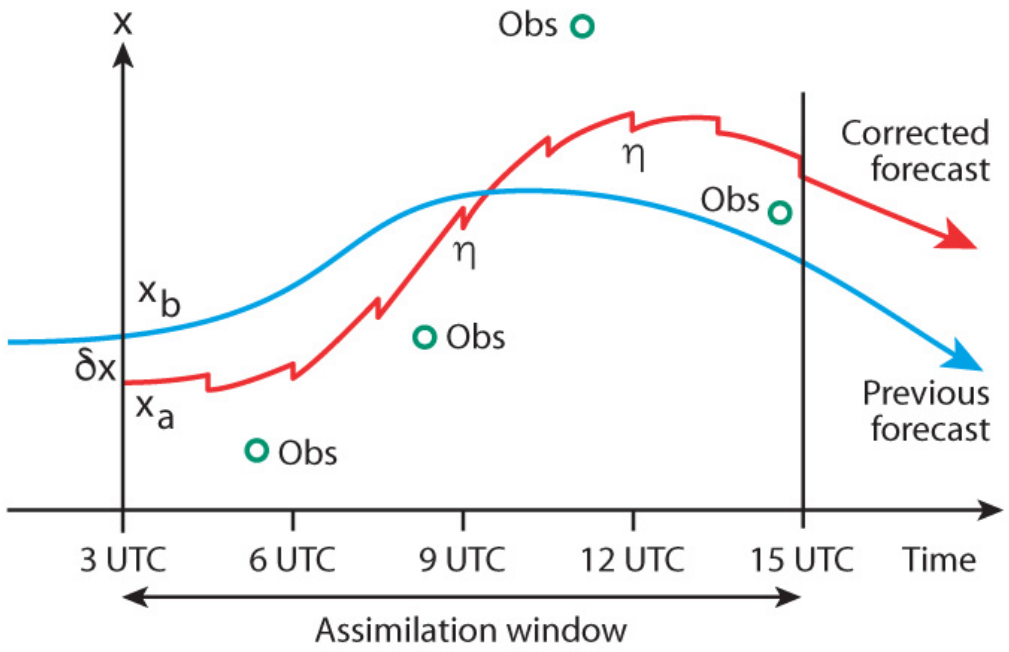
Sasaki (1970)

Derber (1989)

Dee (2005)

Tremolet (2006)

Weak-constraint 4D-Var formulation



$$x_{k+1} = M_{k\to k+1}x_k + \boldsymbol{\eta} \quad \text{where} \quad \eta \sim N(0, Q)$$

$$J(\boldsymbol{x_0}, \boldsymbol{\eta}) = \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\|\eta_0 - \eta_b\|^2_{Q^{-1}} + \frac{1}{2}\sum_{k=0}^{N}\|y_k - \mathcal{H}x_k\|^2_{R^{-1}}$$

Model bias

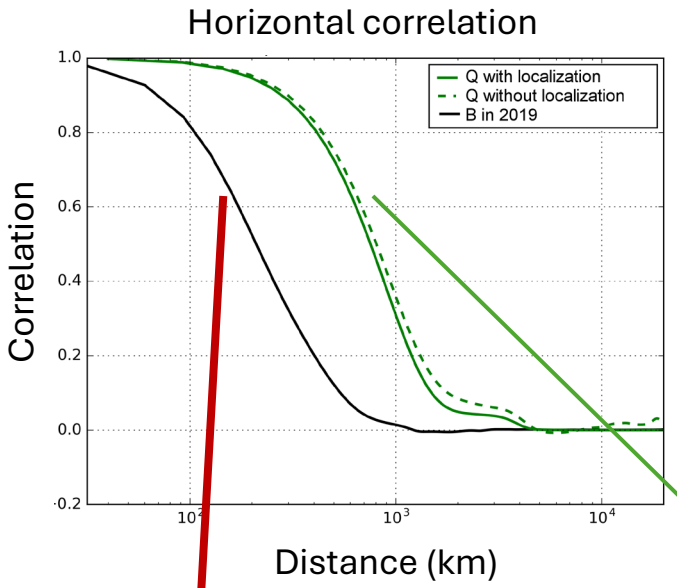
Initial condition

➔ Introduce a model bias correction in the prognostic equations (3D fields: temperature, vorticity and divergence)

➔ 4D-Var estimates simultaneously the optimal initial condition and the model bias correction

➔ Both are estimated by fitting the millions of assimilated observations within 4D-Var

Weak-constraint 4D-Var formulation (scale separation in B and Q)



$$x_{k+1} = M_{k\to k+1}x_k + \eta \quad \text{where} \quad \eta \sim N(0, Q)$$

$$J(x_0, \eta) = \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\|\eta_0 - \eta_b\|^2_{Q^{-1}} + \frac{1}{2}\sum_{k=0}^{N}\|y_k - \mathcal{H}x_k\|^2_{R^{-1}}$$

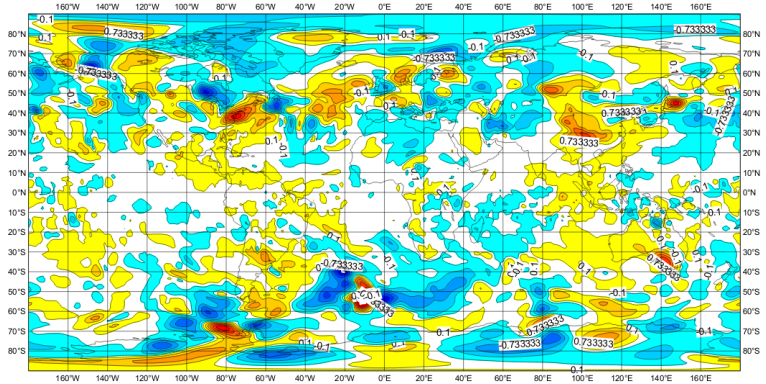
Model bias

Initial condition

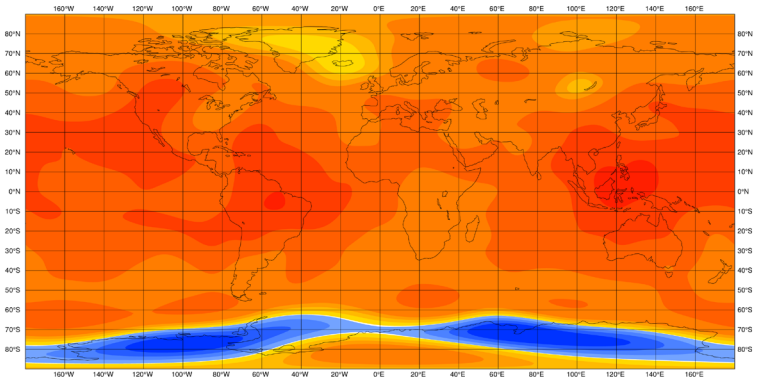
B contains smaller horizontal correlations

Q contains larger horizontal correlations

Initial condition correction



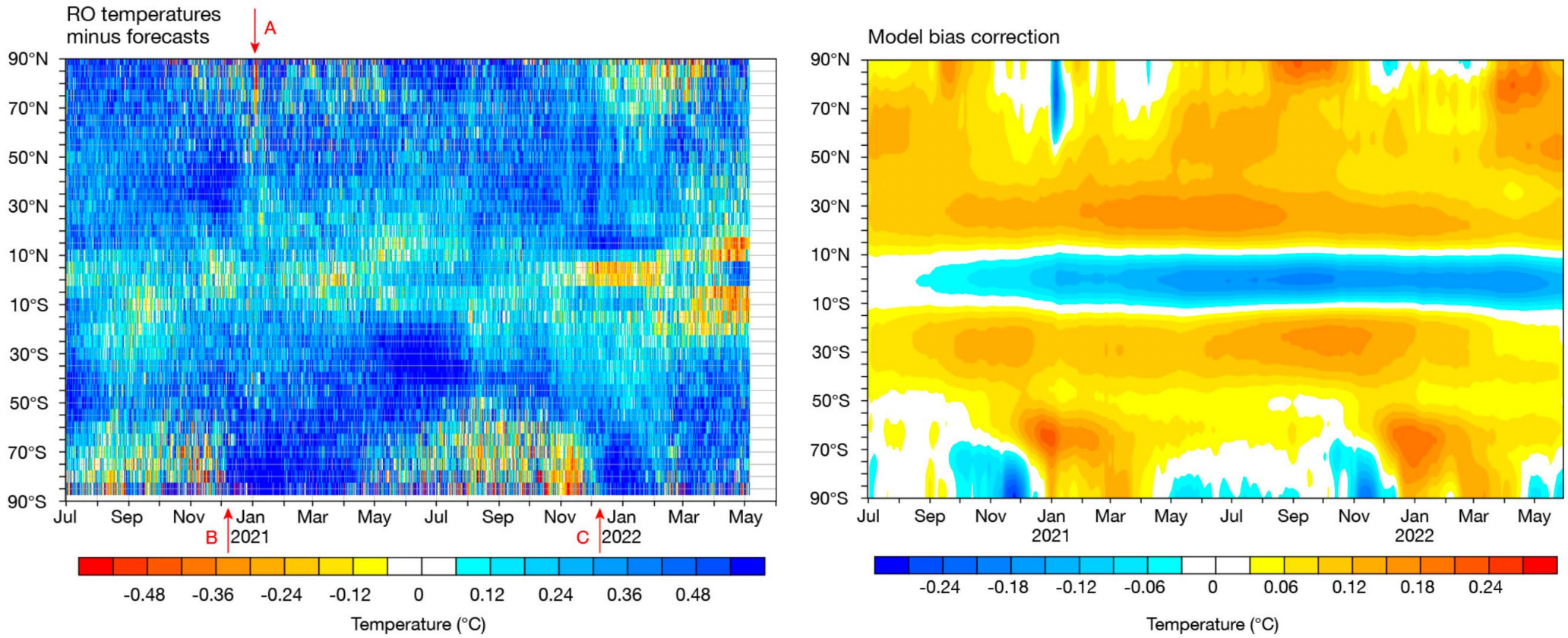
Model bias correction



Without scale separation, the assimilation algorithm cannot distinguish whether a mismatch with observations is caused by an incorrect initial condition or model error correction

See Laloyaux et al, 2020a

Weak-constraint 4D-Var in operations in the stratosphere since 47r1



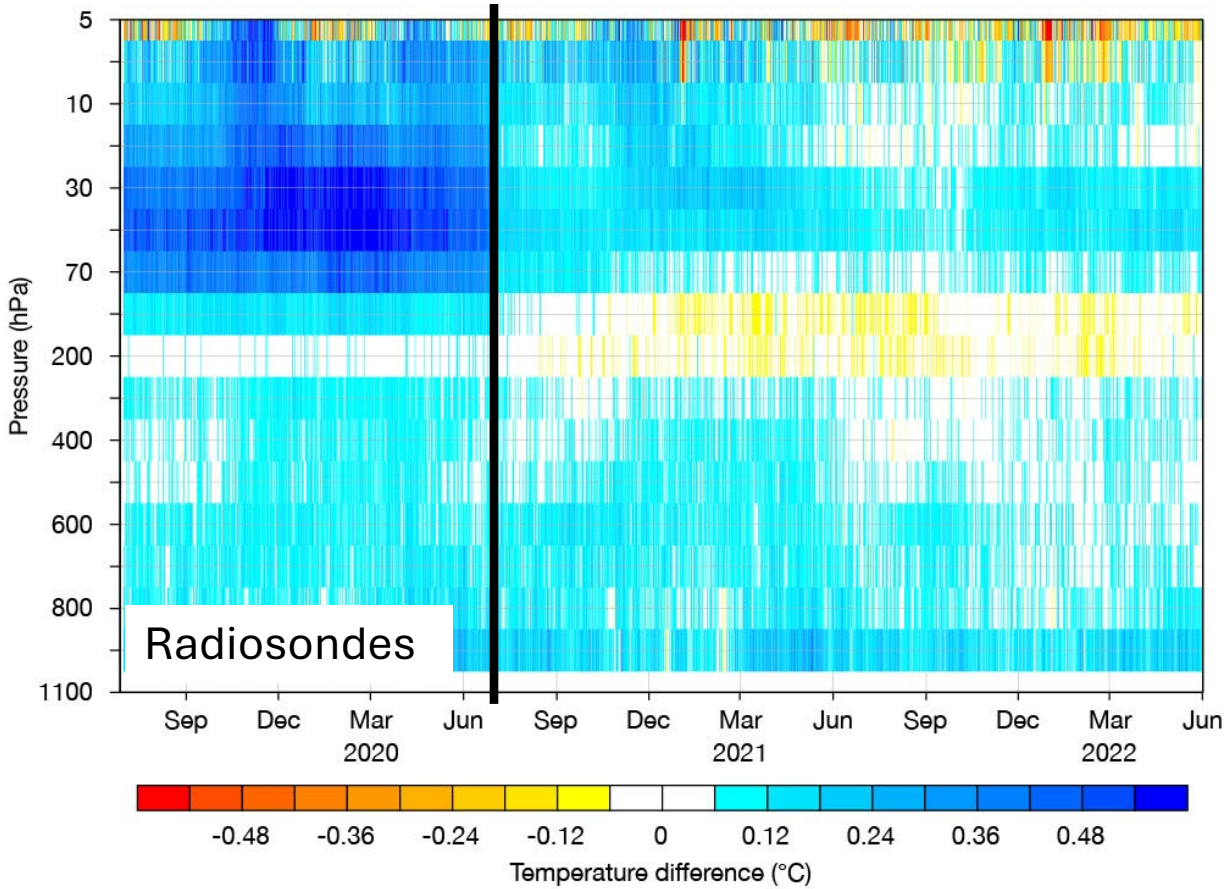
A) On 31 December 2020, a Sudden Stratospheric Warming (SSW) event started over the northern hemisphere

B & C) Clear seasonal cycle in the model bias over the southern hemisphere with a sharp transition in early December 2020 and 2021

See Laloyaux et al, 2020b

Weak-constraint 4D-Var in operations in the stratosphere since 47r1

Difference between obs and the (corrected) IFS trajectory (O-B)



With stratospheric biases significantly reduced (with quintic vertical interpolation), developments of bias correction in the boundary layer and at the surface started

More challenging because

➔ Involving faster processes

➔ Involving diurnal cycle

➔ Biases coming from different earth system components

See Laloyaux et al, 2020b

Weak-constraint 4D-Var in the boundary layer

From a constant correction over the assimilation window

$$x_{k+1} = M_{k\to k+1}x_k + \boldsymbol{\eta} \quad \text{where} \quad \eta \sim N(0, Q)$$

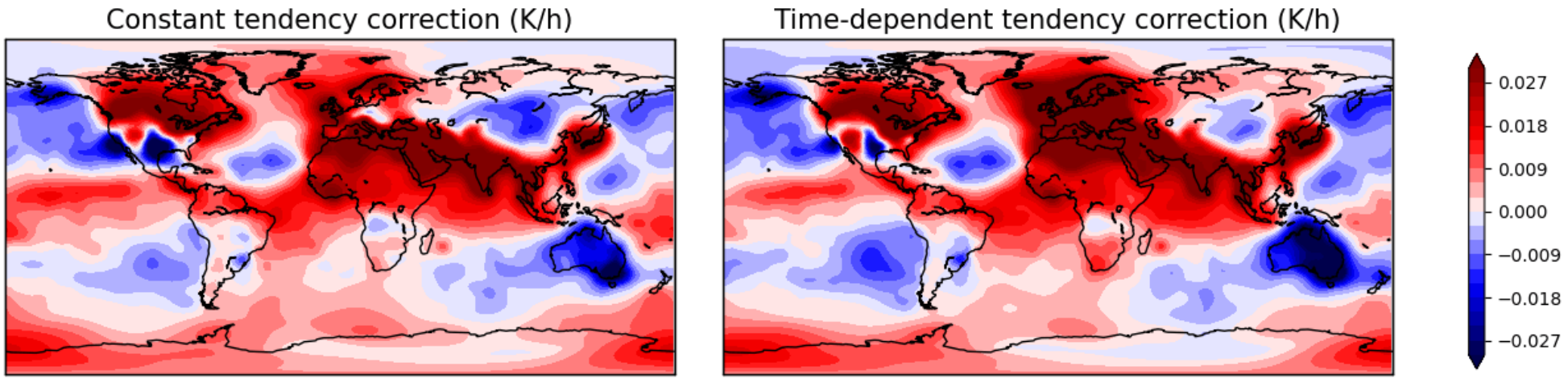
To a time-varying model error

$$x_{k+1} = M_{k\to k+1}x_k + \boldsymbol{\eta_k} \quad \text{where} \quad \eta_k \sim N(0, Q)$$

$$\boldsymbol{\eta_k} = \gamma_0 + \gamma_1 \sin\left(2\pi\frac{k\Delta t}{24*3600}\right) + \gamma_2 \cos\left(2\pi\frac{k\Delta t}{24*3600}\right)$$

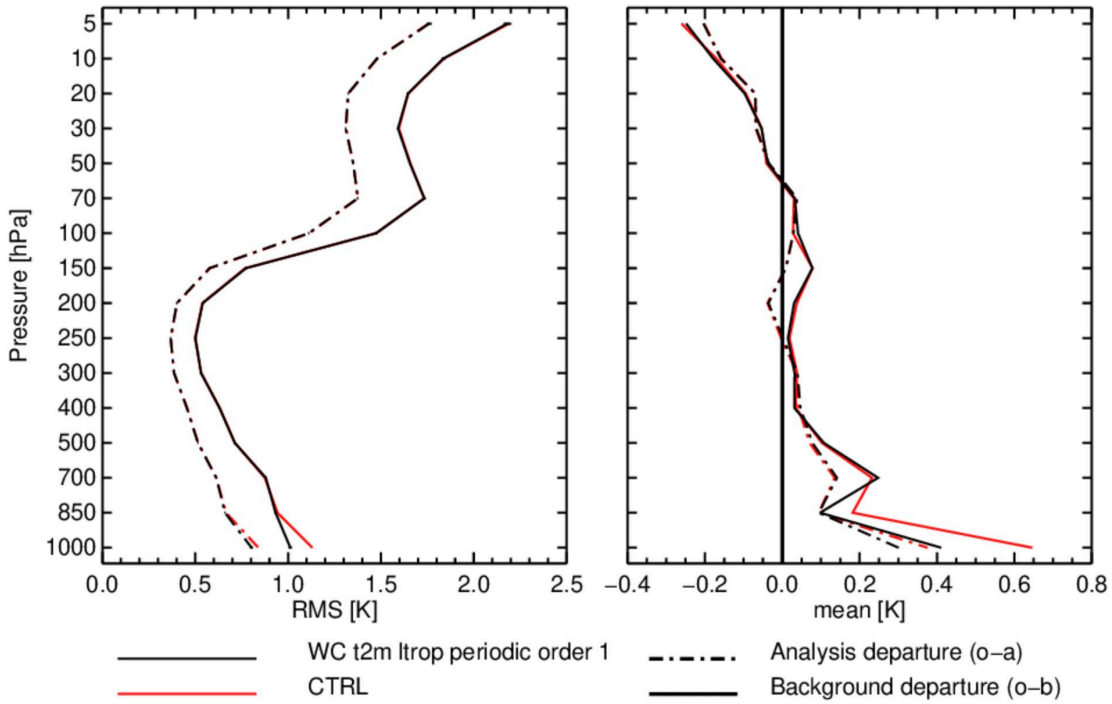
➔ Time-varying within the assimilation window (first-order Fourier decomposition)
➔ Designed to capture a diurnal cycle
➔ Used to correct temperature, vorticity, and divergence, as well as surface pressure and soil temperature

20220720 09:00

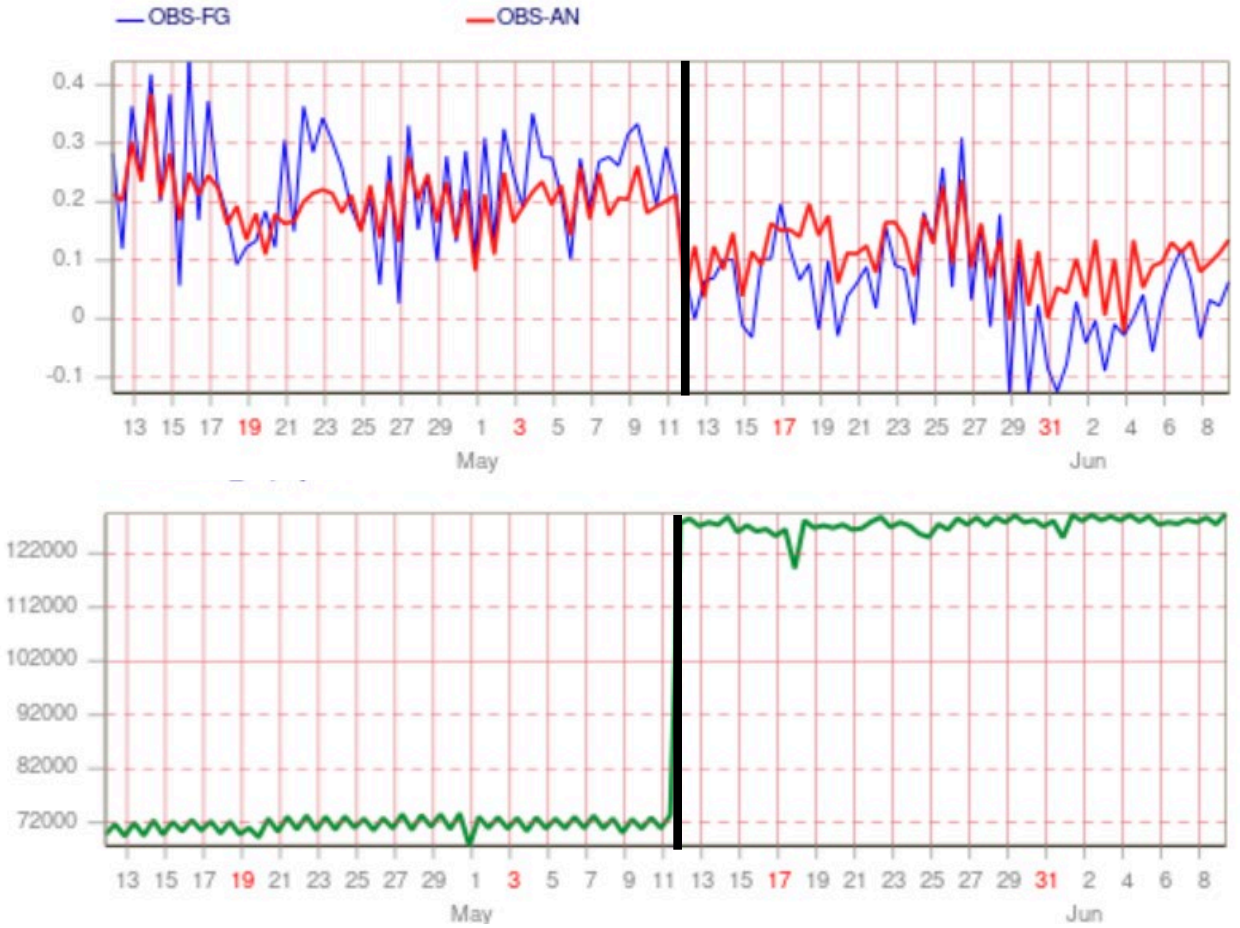


Weak-constraint 4D-Var in the boundary layer in 50R1 (May 2026)

Impact in the mean state against radiosondes



By improving the model used inside 4D-Var, t2m is now more realistic and t2m obs can be assimilated over the whole window (only first 6 hours before)



Weak-constraint 4D-Var in ERA6 to correct stratospheric biases

START

Start with 2007 onward

Run WC4D-Var in the stratosphere on modern observing system

TRAIN

Learn model error

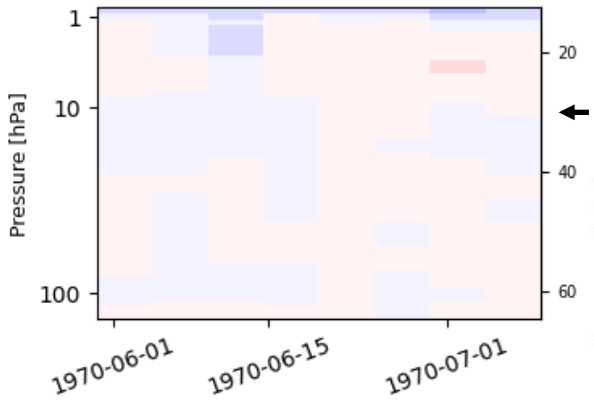
Use well-observed years to emulate model error corrections

EXTEND

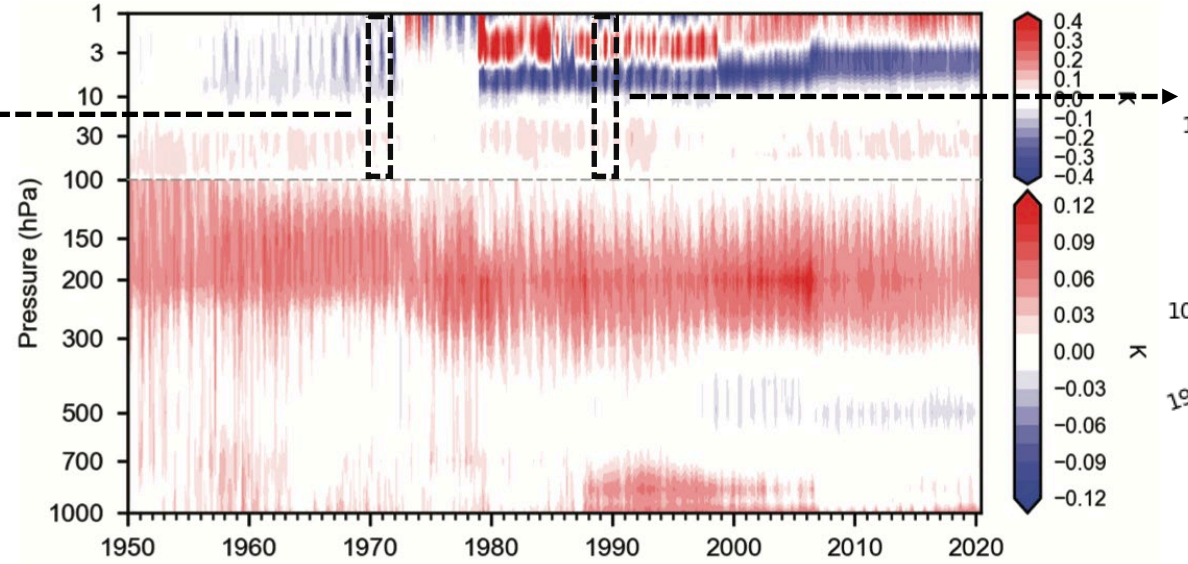
Move to pre-2007 streams

Apply the learned correction where anchor observations are sparse

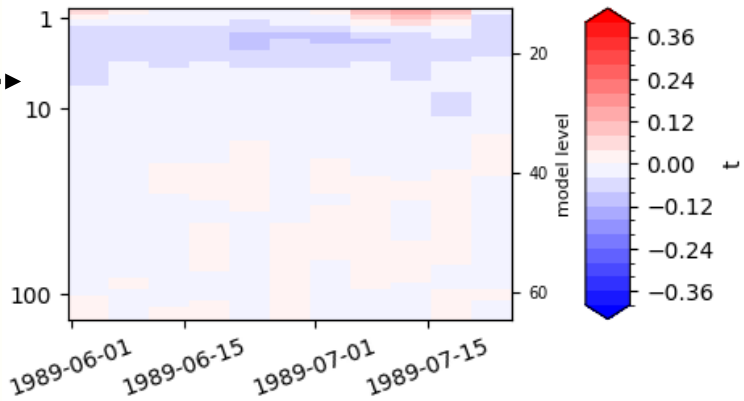
ERA6 prototype experiment



Mean ERA5 temperature increments



ERA6 prototype experiment



Future of weak-constraint 4D-Var and model bias correction in IFS

Model biases should be corrected at the source if possible

➔ ECMWF keeps improving the model: parametrized physics (e.g. moist processes, clouds, ozone…)

and dynamical core (e.g. conservation aspects, equation formulations…)

➔ ECMWF is moving away from (impractical) manual tuning to exploring new optimization methods (e.g. Bayesian optimisation of parameters, adaptive parameter tuning)

Data assimilation is a great framework to improve further the model (compares the model to the observations

➔ Weak-constraint 4D-Var is an efficient approach with a hidden crux (i.e. specification of B and Q)

➔ Can ML alleviate this challenge for model bias correction? (Bonavita et al. 2020)

Thursday 14:15 – 15:00	Building a robust hybrid model framework for the IFS	Alban Farchi (ECMWF) Marcin Chrust (ECMWF)

Accounting for systematic model error in Data Assimilation

1. **Integrated Forecasting System (IFS) model**
2. **Direct Observation Prediction (DOP) model**

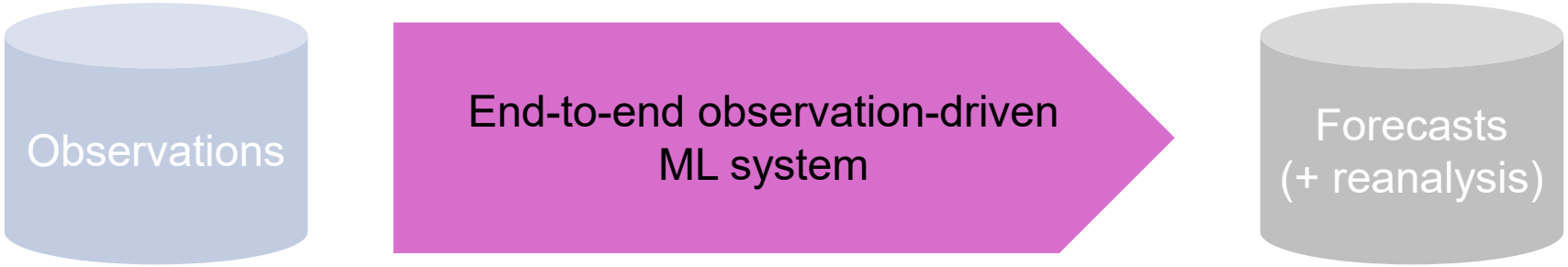
Two parallel but complementary forecast streams

1) Physics/Hybrid
(analysis-based)



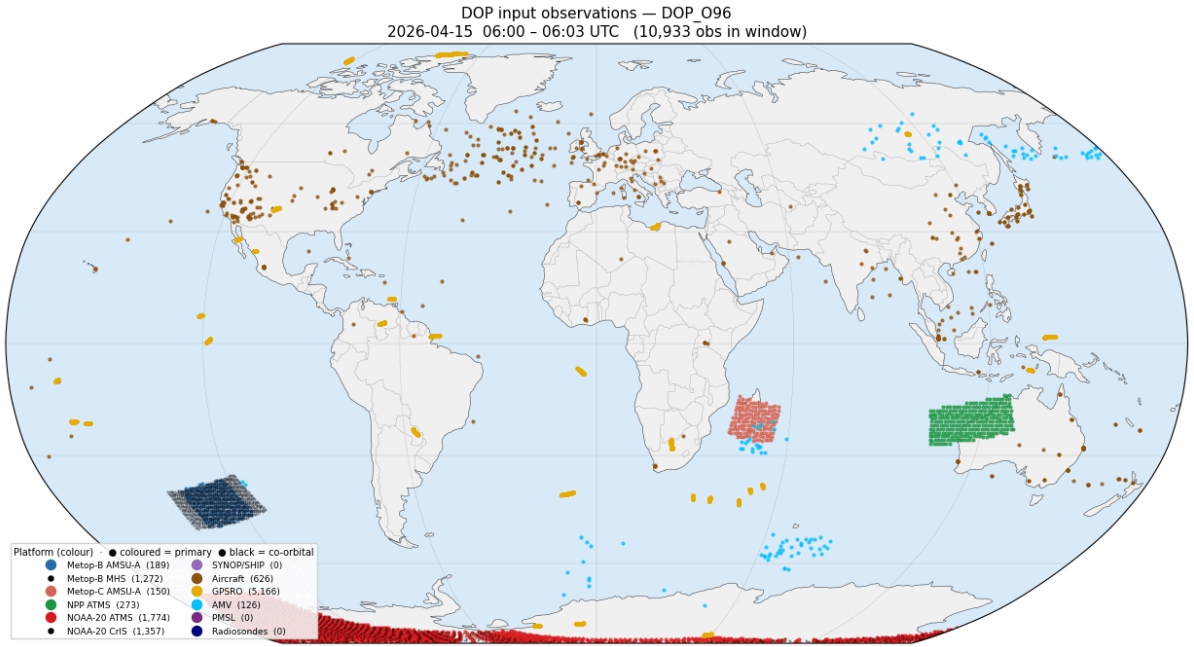
→ These forecasting systems rely on 4D-Var (NEMOVAR / LDAS) for initialisation/training

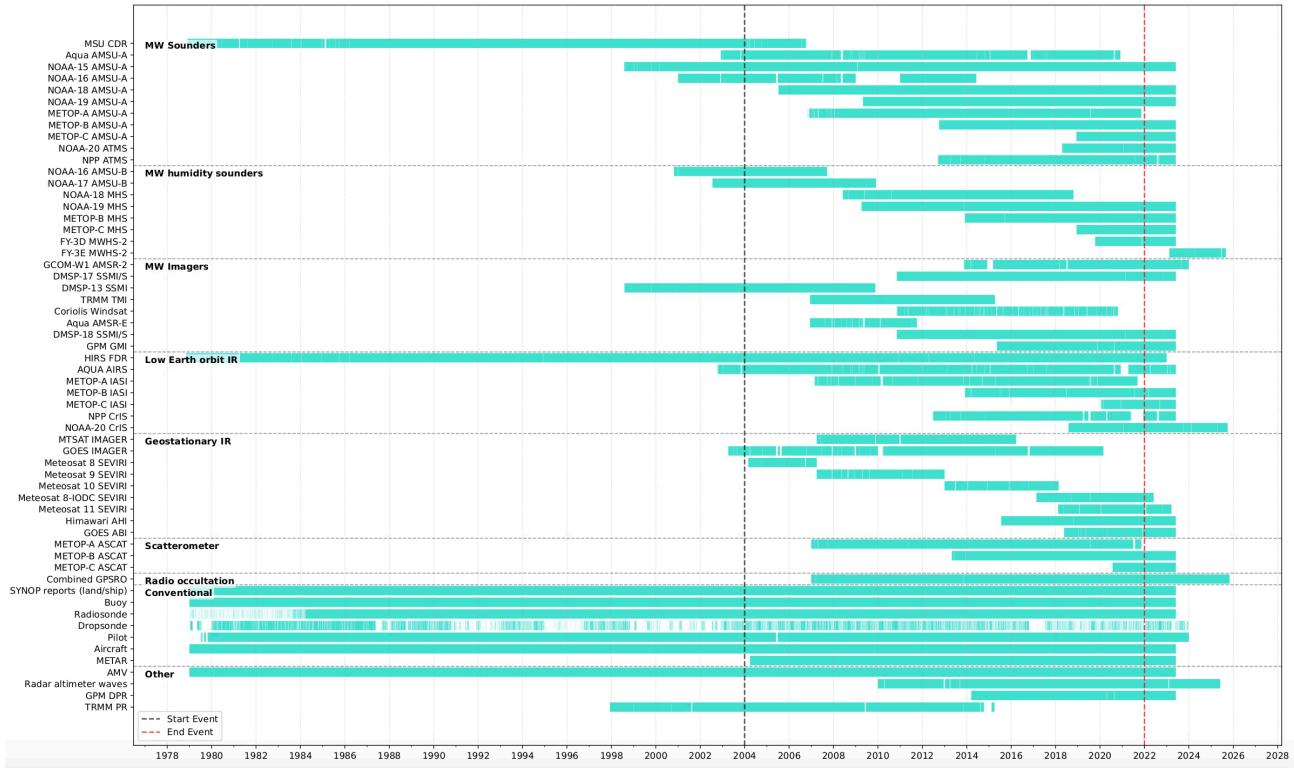
2) Observation-driven



→ What that system will eventually look like is an open research question (DOP + AIFS + WeatherGenerator + …)

DOP approach is based on large observation datasets





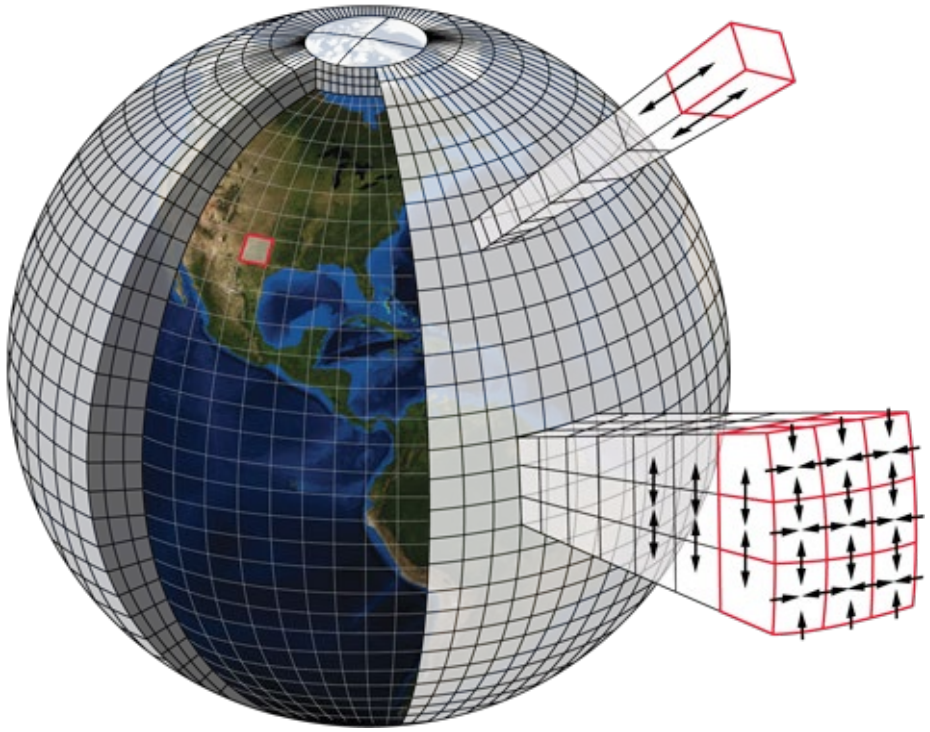
Large observation datasets are used (no analysis or reanalysis) to train a model. The challenge is to learn spatial and temporal correlations between all these measurements

How does data assimilation happen in DOP?

How are observation and model errors accounted for in DOP?

DOP approach is based on reduced Gaussian grids

IFS and DOP models based on reduced Gaussian grid (e.g. O96, N320, …)

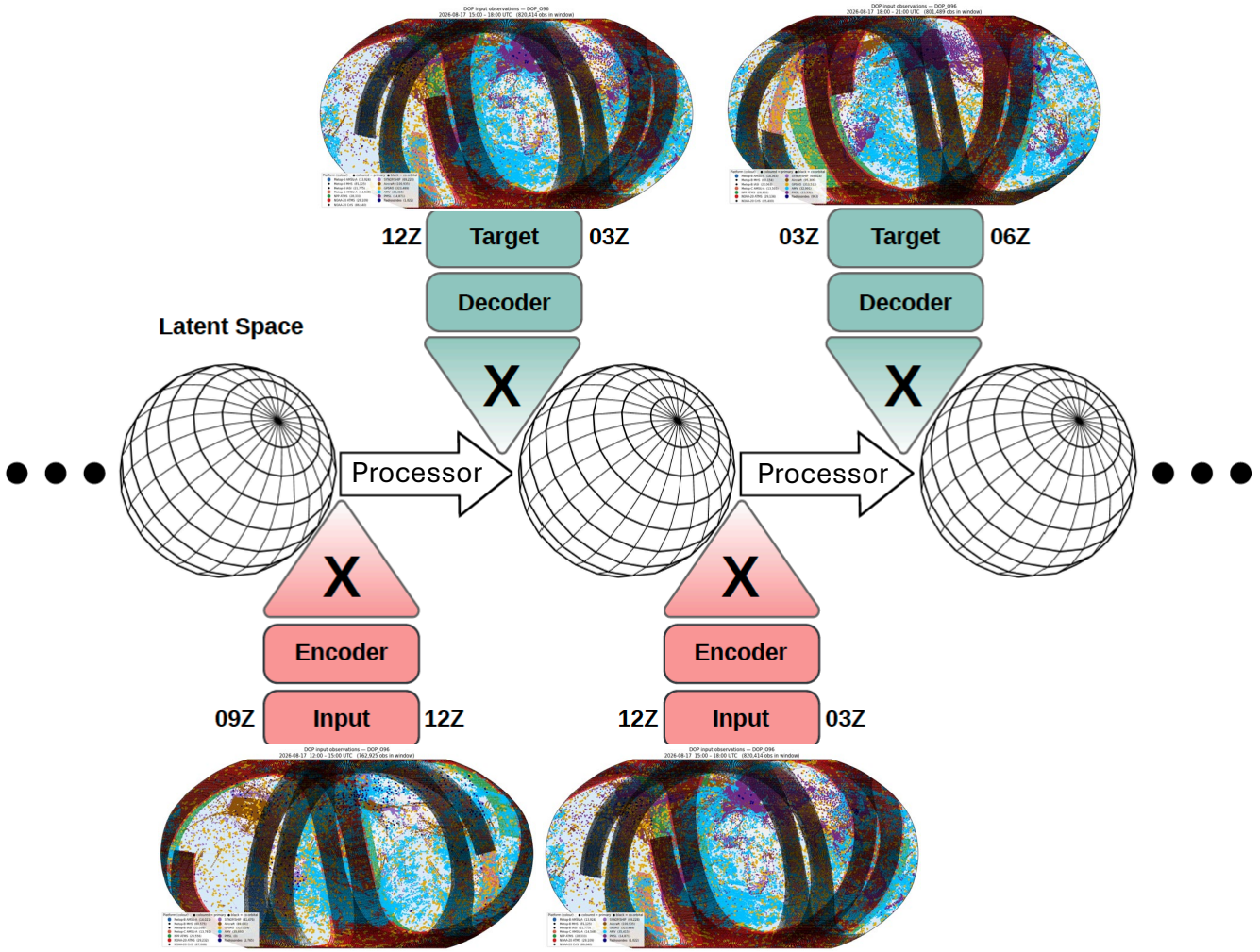


➔ A column of the IFS model represents the atmosphere (temperature, vorticity, divergence, humidity, …) on 137 vertical levels

➔ A column of the DOP model represents the state of the atmosphere summarized in an **abstract/latent** space (e.g. 1024 values).

➔ To produce a forecast, both model are integrated in time (recursive loop, involving vertical and horizontal exchanges between columns)

DOP learns the encoder-processor-decoder blocks



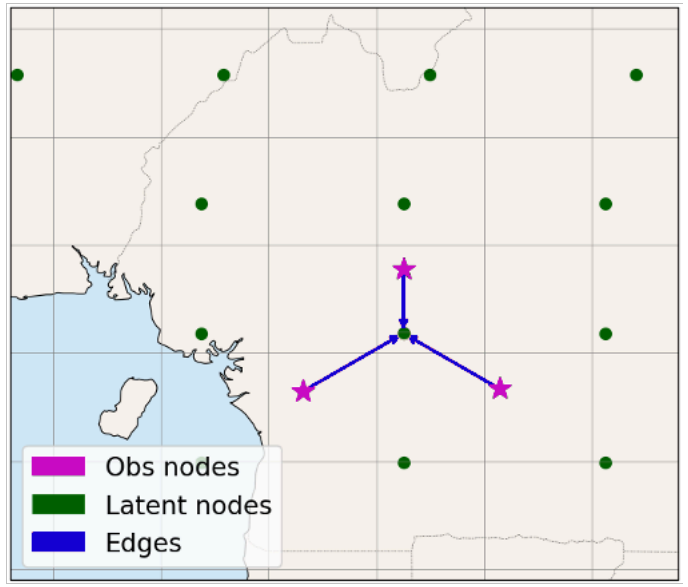
Learned observation operators (decoder) which map from the model latent state to observation physical space

A **learned forecast model (processor)** which evolves the latent state forward in time

A **learned Data Assimilation function (encoder)** updates (e.g. every 3 hours) the current background state in the latent space using a window of observations (cross-attention)

Learned Data Assimilation function (embedding 1/3)

Zoomed in around a specific grid point



Latent background

z_i^b | 1 | 2 | 3 | ... | ... | ... | ... | 1023 | 1024 |

$z_i = LN(z_i)$ LN=Layer Normalization

Raw observation

$\hat{y}_j$ | | | | | | |

Metadata Measurements

Embedded observation (same dimension)

y_j | 1 | 2 | 3 | ... | ... | ... | ... | 1023 | 1024 |

$y_j = LN(\boldsymbol{W_Y}\hat{y}_j)$

Raw edge

$\widehat{e_{i,j}}$ | | | | |

Distance and direction

Embedded edge (same dimension)

$e_{i,j}$ | 1 | 2 | 3 | ... | ... | ... | ... | 1023 | 1024 |

$e_{i,j} = \boldsymbol{W_G}\widehat{e_{i,j}}$

All $\boldsymbol{W}$ are learned projections into the 1024-dimensional embedding space

Learned Data Assimilation function (cross-attention 2/3)

1. Compute the ***query*** *from the latent background*

$$q_i = QKnorm(W_Q z_i^b)$$

$\mathbf{q_i}$ what the grid point needs to check, according to what it believes

2. Compute the ***keys*** and ***values*** *from observations and edges*

$$k_{i,j} = QKnorm(W_K y_j) + W_E e_{i,j}$$

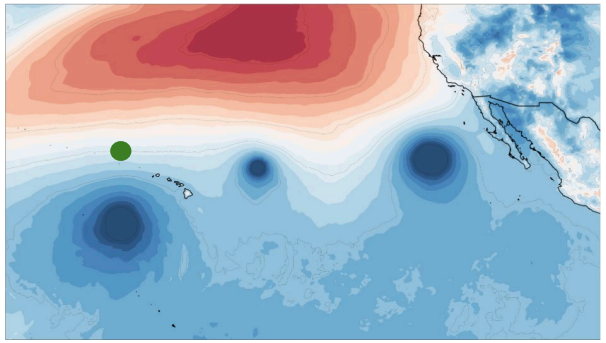
$\mathbf{k_{i,j}}$ the kind of information that observation j knows about grid point i

$$v_{i,j} = W_V y_j + W_E e_{i,j}$$

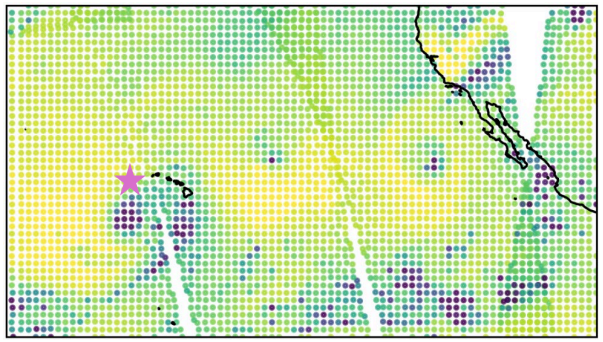
$\mathbf{v_{i,j}}$ the actual information that observation j knows about grid point i

3. Compute the cross-attention between grid point ***i*** and observation ***j***

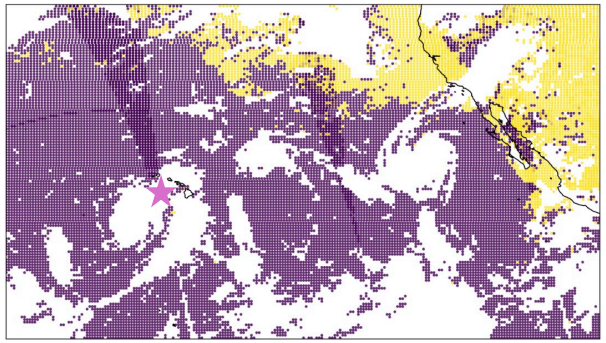
$$\alpha_{i,j} = softmax\left(\frac{q_i^T k_j}{\sqrt{C}}\right)$$



Background believes that a storm is coming
➔ q_i needs to check humidity channels and pressure to confirm or infirm



This observation contains MW humidity information ($k_{i,j}$)
➔ strong match with q_i , $\alpha_{i,j} \gg 0$



This observation contains NO2 concentration ($k_{i,j}$)
➔ orthogonal to q_i , $\alpha_{i,j} = 0$

Learned Data Assimilation function (latent space update 3/3)

Latent space update

$$\boldsymbol{\alpha_{i,j}} = softmax\left(\frac{q_i^T k_j}{\sqrt{C}}\right)$$

$$\boldsymbol{z_a} = \underbrace{(\boldsymbol{I} + (\boldsymbol{W_O W_S} LN))\boldsymbol{z_b}}_{\text{green}} + \underbrace{\boldsymbol{W_O} \sum_j \boldsymbol{\alpha_{i,j}} \boldsymbol{v_{i,j}}}_{\text{pink}}$$

Kalman filter update

$$\mathbf{K} = BH^T(HBH^T + R)^{-1}$$

$$\mathbf{x^a} = \underbrace{(\boldsymbol{I} - \boldsymbol{K}H)\boldsymbol{x^b}}_{\text{green}} + \underbrace{\mathbf{K}\, y}_{\text{pink}}$$

Drawing some tentative parallels, both updates split into similar terms

➔ $\boldsymbol{v_{i,j}}$ contains the information of the observation j at grid point i, like applying BH^T to an observation

➔ $\sum_j \boldsymbol{\alpha_{i,j}}$ cross-attention values tell the importance of each observation, like $(HBH^T + R)^{-1}$

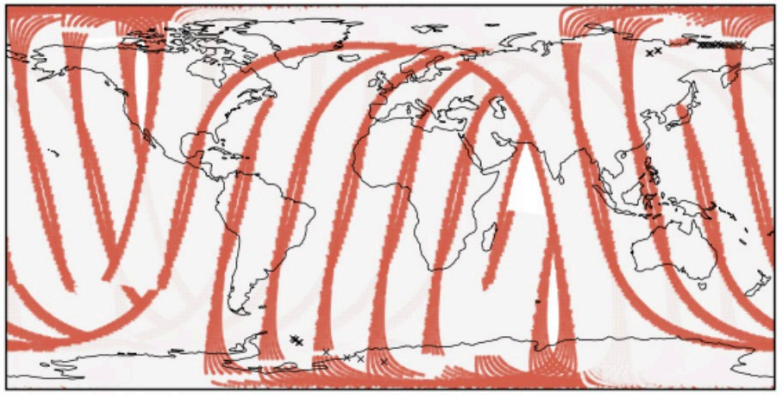
➔ Does $\boldsymbol{W_O}$ and $\boldsymbol{W_S}$ play the same role as $-\boldsymbol{K}H$ (weighting the background)?

The cross-attention $\boldsymbol{\alpha_{i,j}}$ depends on the key $\boldsymbol{k_{i,j}}$ which is computed from the observation $\boldsymbol{y_j}$. If the observation value is changing, the cross-attention will be different (online quality control)?

Note: In practice multi-heads and localization are used but not described here

How does DOP deal with observation errors?

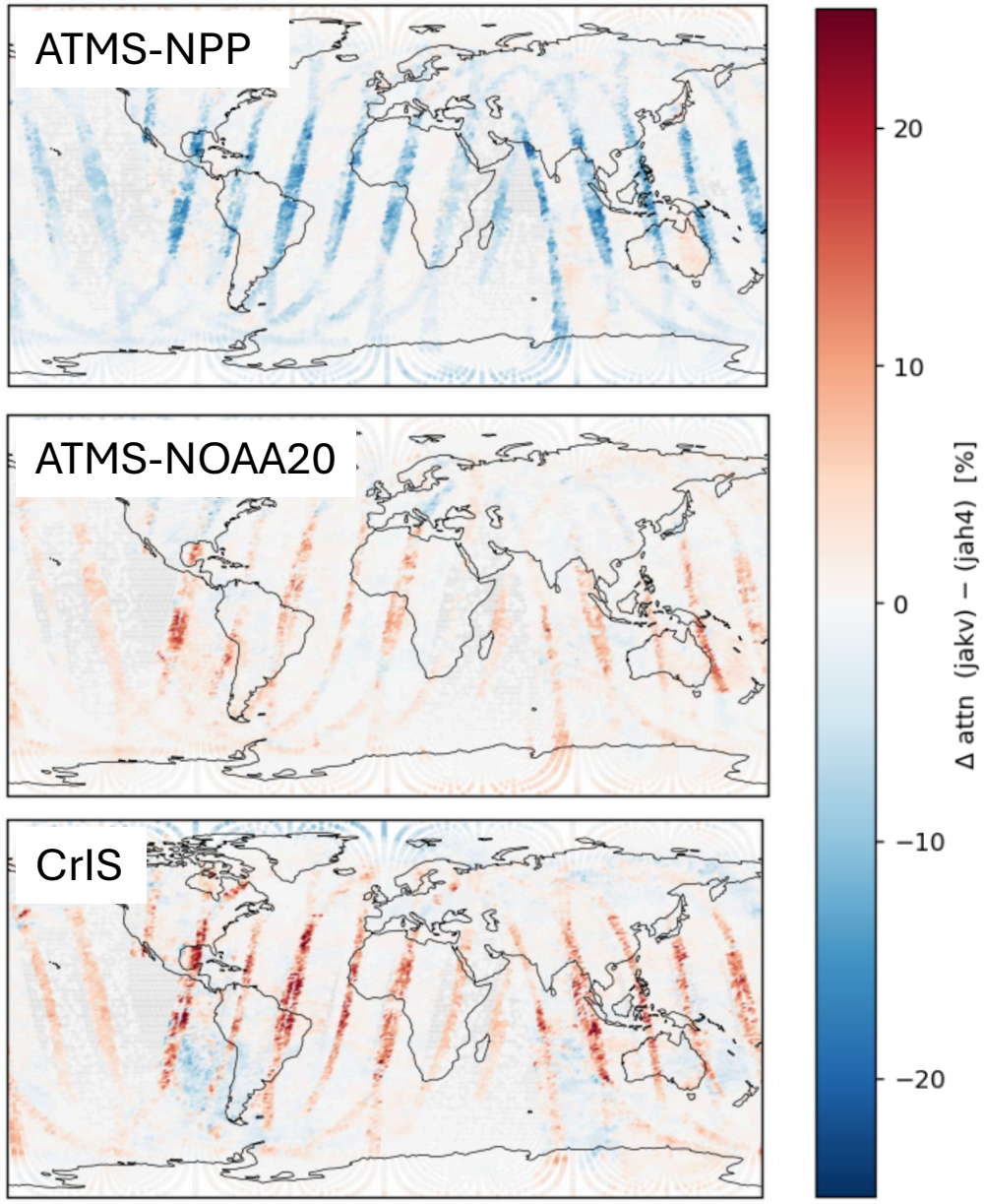
A 3K bias is introduced in ATMS-NPP channels (increasing the limb effect)



$$\boldsymbol{\alpha_{i,j}} = softmax\left(\frac{\boldsymbol{q_i^T k_j}}{\sqrt{C}}\right)$$

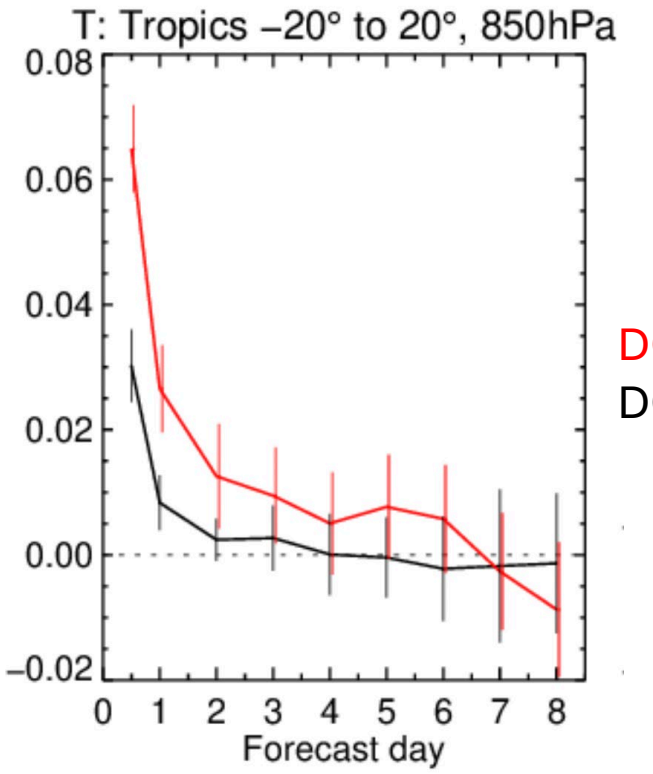
➔ The attention to these biased observations is decreased at inference (blue in ATMS-NPP) as key and query mismatch

➔ The attention for other instruments is bumped up (red in ATMS-NOAA20 and CrIS)



How does DOP deal with observation errors?

Online quality control from cross-attention is not enough. A change in observation biases (or noisy observations) still increase the forecast error in DOP

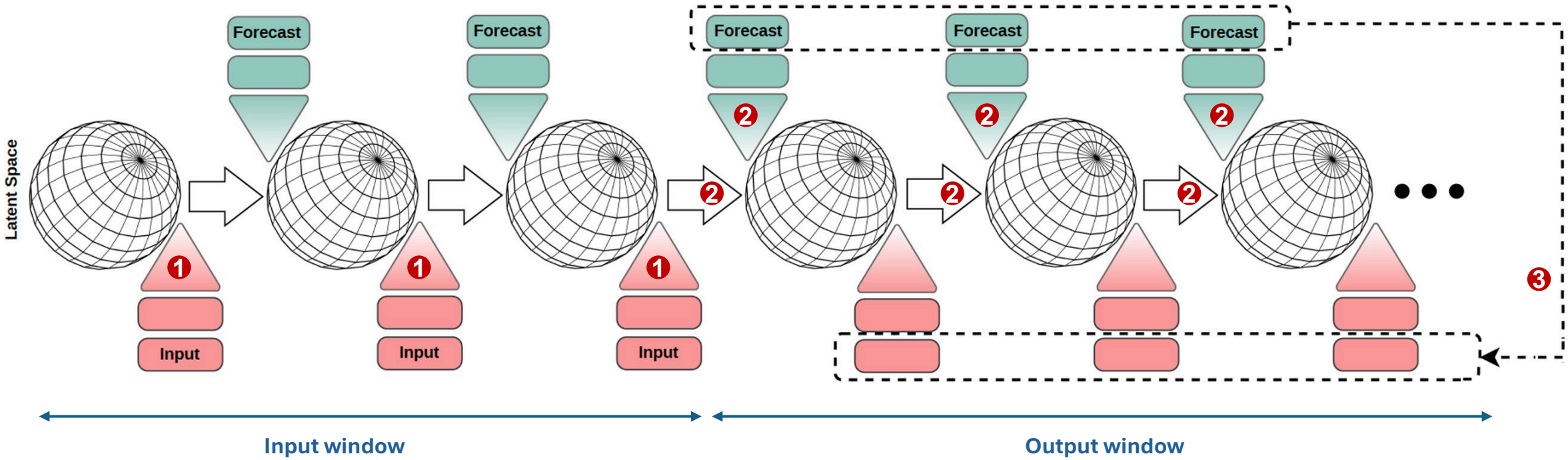


DOP with systematic error in ATMS-NPP (3K)

DOP with random error in ATMS-NPP (3K)

More than ever, expert knowledge is critical to understand the quality of the different instruments and develop methods to perform QC in training and inference

How does DOP deal with model errors?



1. **warm up:** observations available in the input window are encoded in the latent space
2. **latent rollout:** Processor carries the information forward in time and decoder computes predicted observations in the output window
3. **observation rollout:** Predicted observations are fed back into the model to ensure a self-consistent latent space

Dual rollout strategy can be seen as assimilating predicted observations in the forecast model

How does DOP deal with model errors?

DOP processor carries the latent space forward in time

$$o_i = z_i + attention\ (z_i, z_j\ \epsilon\ window(i))$$

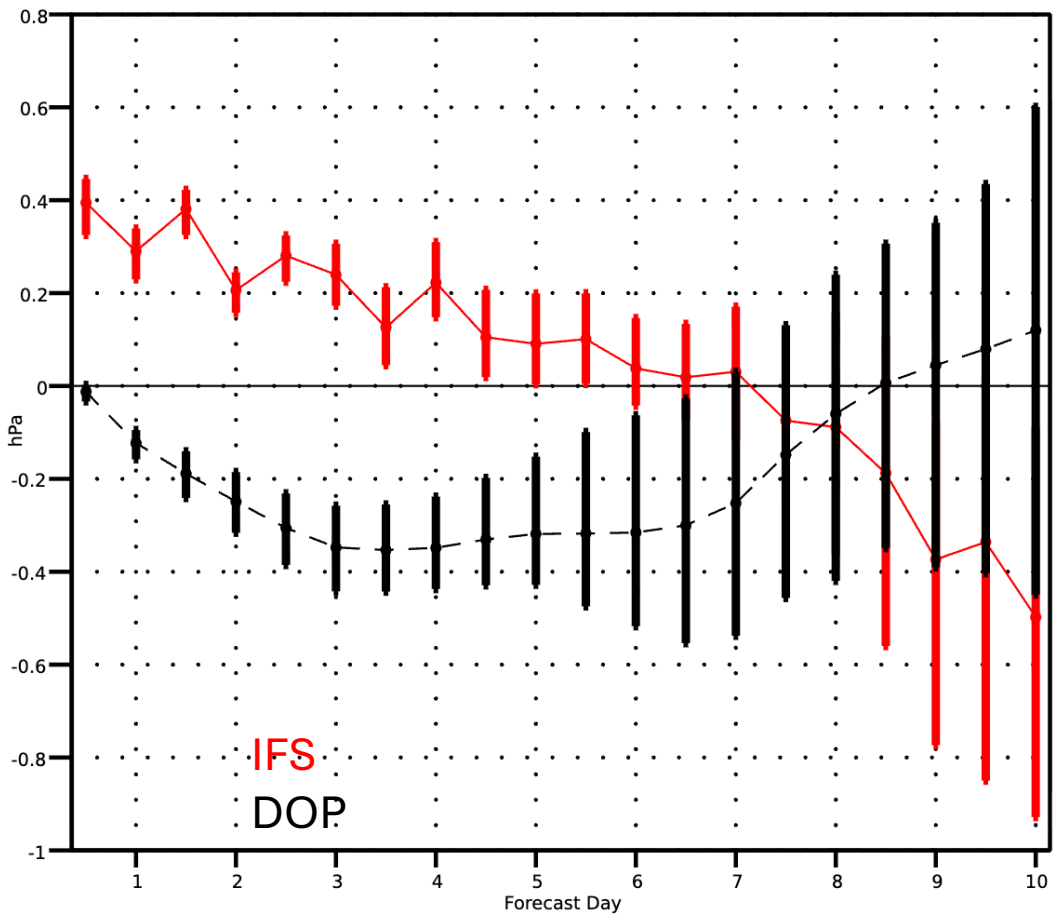
$$z_i^{new} = o_i + MLP(o_i)$$

Weak-constraint 4D-Var formulation

$$x_{k+1} = M_{k \to k+1} x_k + \eta_k \quad \text{where} \quad \eta_k \sim N(O, Q_k)$$

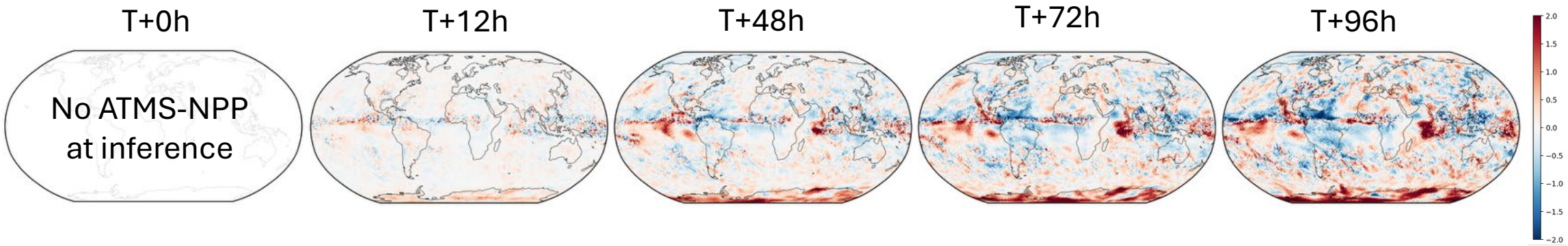
➔ In DOP, the MLP (dense, fully connected NN) represents a non-linear correction to represent processes missed in the attention mechanism

Mean forecast error (against mslp obs, NH)

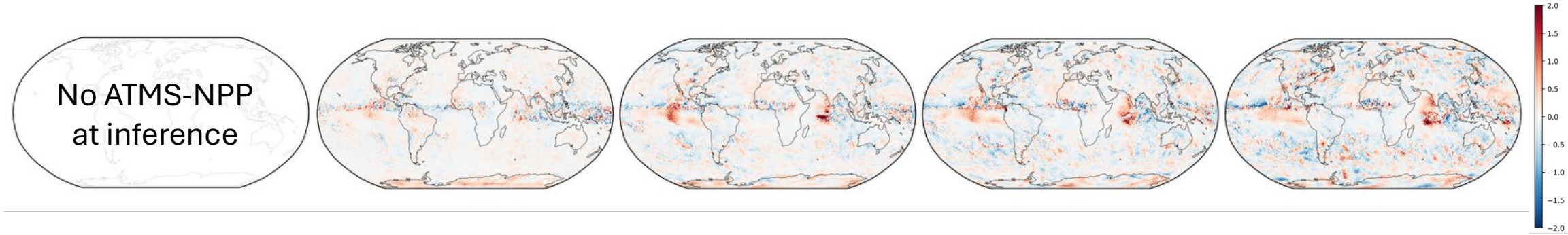


How does DOP deal with model biases (when an instrument is missing)?

Large model biases in the forecast from the loss of NPP (DOP has never been trained without the instrument)



Model biases reduced significantly by feeding back ATMS-NPP predicted observations (observation rollout)



Dual rollout (latent+obs) is more robust to a satellite loss, by keeping the model inside the conditions it was trained for

Final thoughts on IFS and DOP models

In **IFS**, DA is a powerful tool to further improve the model (looking at the discrepancies with observations)

➔ Weak-constraint 4D-Var is a very successful story with 47r1 in 2020 and 50r1 in 2026

➔ The challenge is the specification of ***B*** and ***Q***

DOP forecasts the weather directly from observations

➔ It learns a DA function in latent space (via cross-attention mechanism)

➔ Error statistics (***B***, ***R***, ***Q***) are somehow learned within the encoder-processor-decoder blocks

➔ More diagnostics are required to ensure that DOP has captured correctly these error statistics

Thursday 16:15 – 17:00	End-to-end real-time weather forecasting with Machine Learning	Eulalie Boucher (ECMWF) Mihai Alexe (ECMWF)
Friday 09:00 – 09:45	The prospects for future Machine Learning climate reanalysis directly from observations	Ewan Pinnington (ECMWF) Peter Lean (ECMWF)

Thank you!

