Biases in the model

Patrick Laloyaux

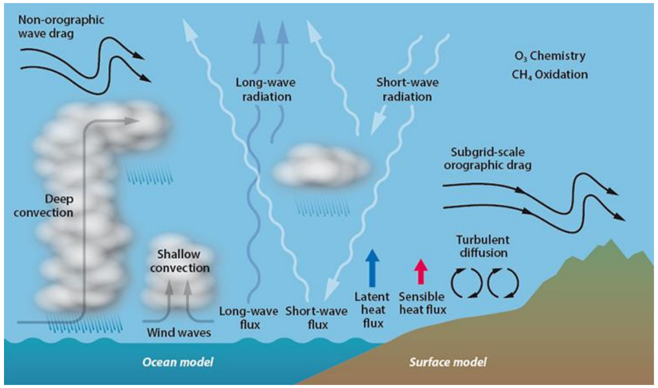
Identify biases in models

Learn how to develop bias correction methods

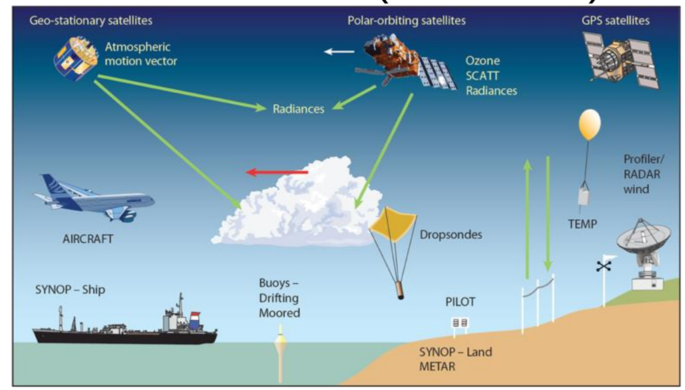
Prospect for future developments

What you have seen so far on data assimilation (1/2)

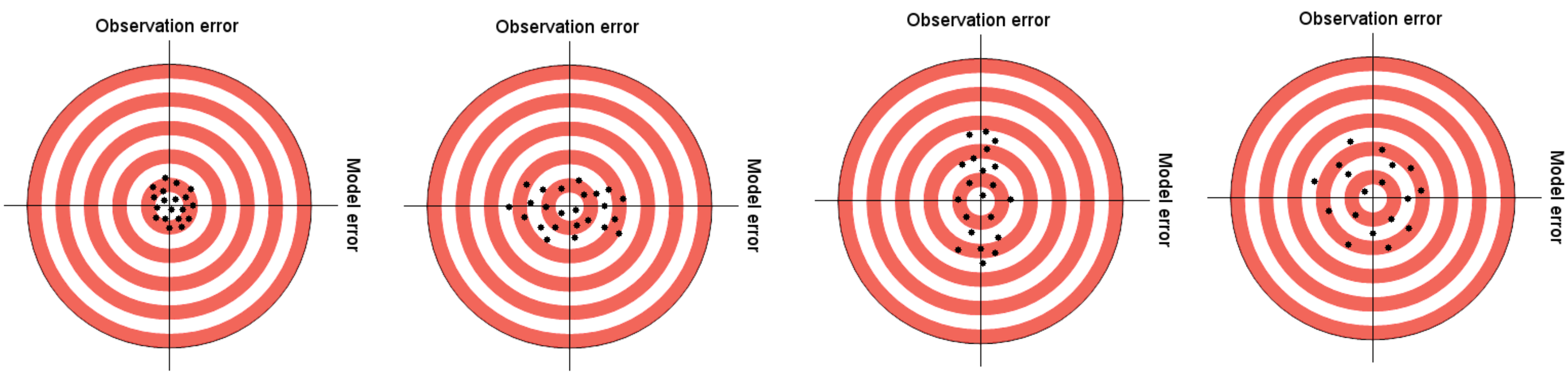
Model (with errors)



Observations (with errors)

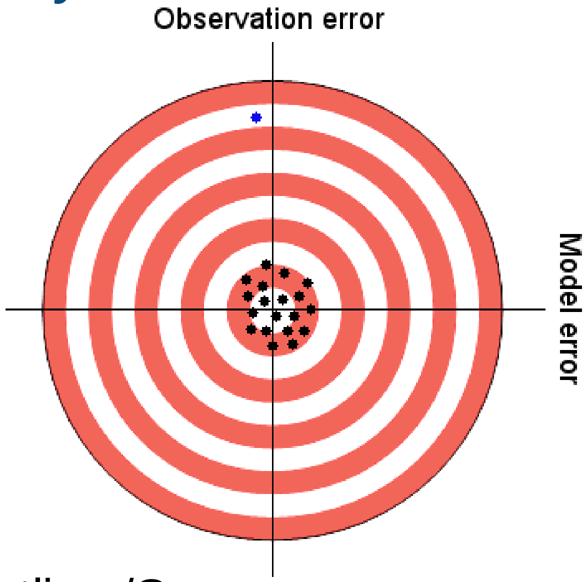


4D-Var has been initially designed to deal with random errors



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

What you have seen so far on data assimilation (2/2)

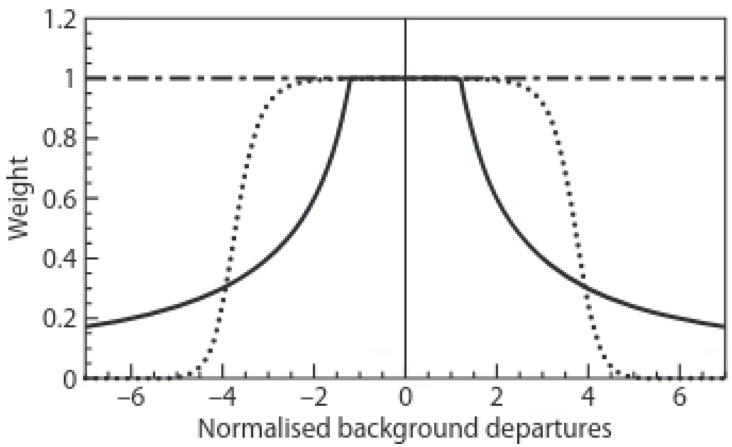


➔ Outliers/Gross error

➔ Variational Quality Control (VarQC)

➔ Systematic observation errors

➔ Variational Bias Control (VarBC)

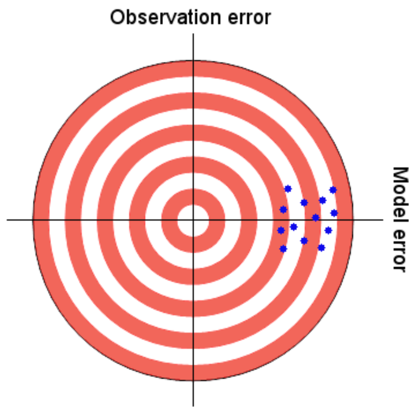


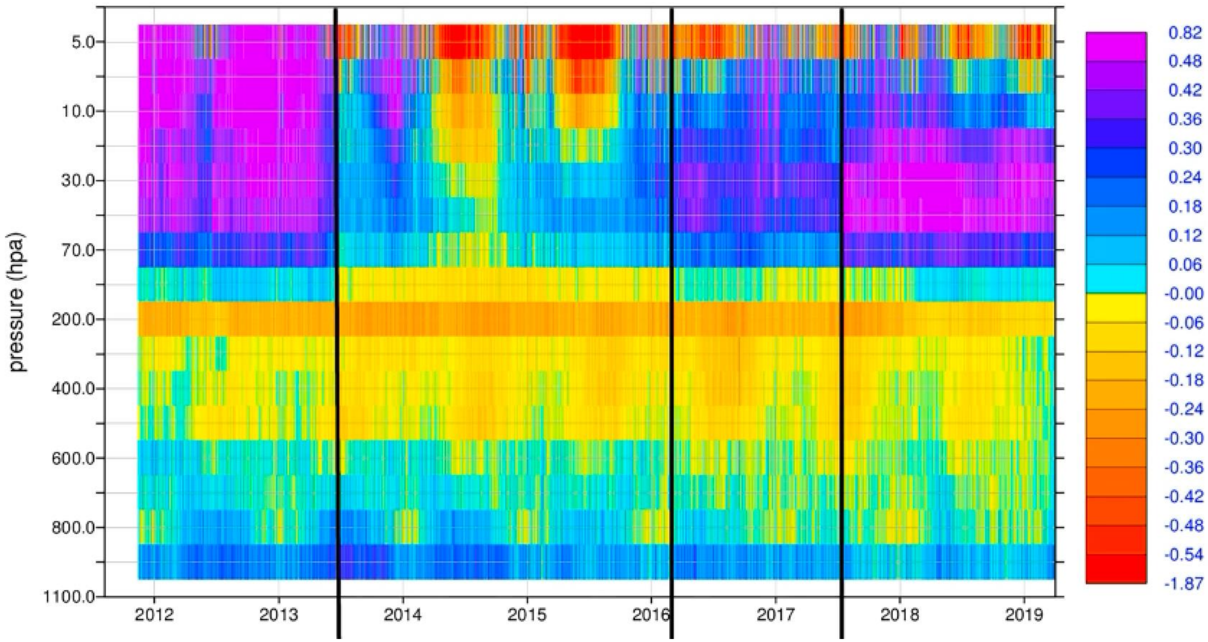
$$\begin{aligned} J(x_0,\beta) &= \frac{1}{2}(x_0-x_b)^T\mathbf{B}^{-1}(x_0-x_b) \\ &+ \frac{1}{2}(\beta-\beta_b)^T{\mathbf{B}_\beta}^{-1}(\beta-\beta_b) \\ &+ \frac{1}{2}\sum_{k=0}^{\text{Radiosonde}}[y_k-\mathcal{H}(x_k)]^T\mathbf{R}_k^{-1}[y_k-\mathcal{H}(x_k)] \\ &+ \frac{1}{2}\sum_{k=0}^{\text{GPSRO}}[y_k-\mathcal{H}(x_k)]^T\mathbf{R}_k^{-1}[y_k-\mathcal{H}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

4D-Var has multiple ways to fit the observations

How to estimate model biases (1/4)

The first-guess trajectory of the model can be compared to accurate observations





Difference between radiosonde temperature observations and the IFS first-guess trajectory (O-B)

Errors in models are often systematic rather than random, zero-mean

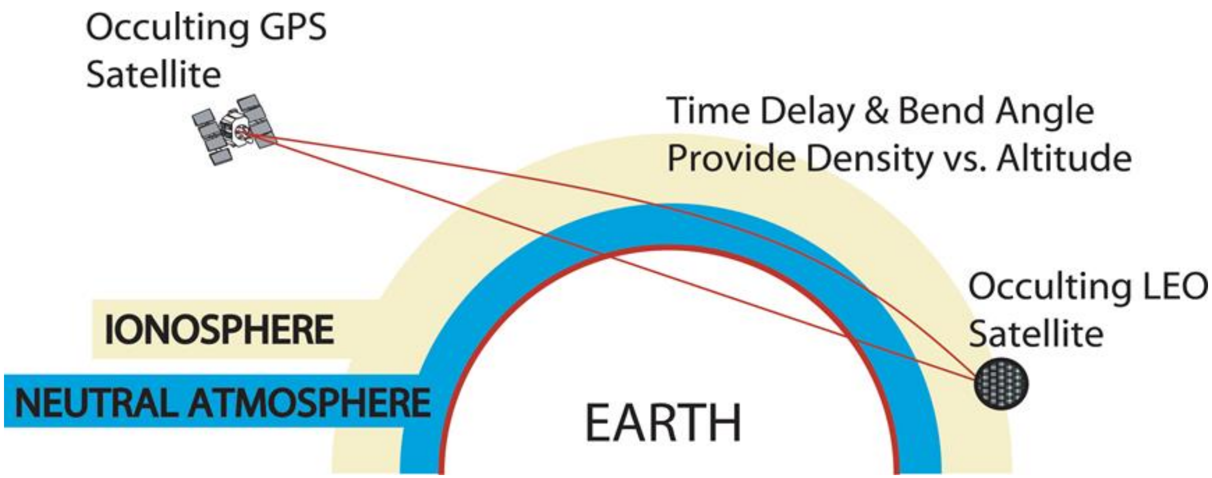
→ Largest bias in the stratosphere

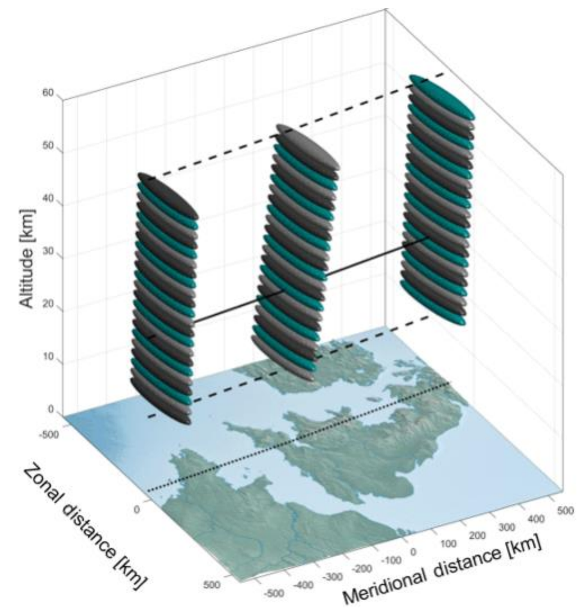
→ Model has a temperature cold bias in the lower/mid stratosphere

→ Model has a warm bias in the upper stratosphere

How to estimate model biases (2/4)

GPS-RO (Radio Occultation) analyses the bending caused by the atmosphere along paths between a GPS satellite and a receiver placed on a low-earth-orbiting satellite.





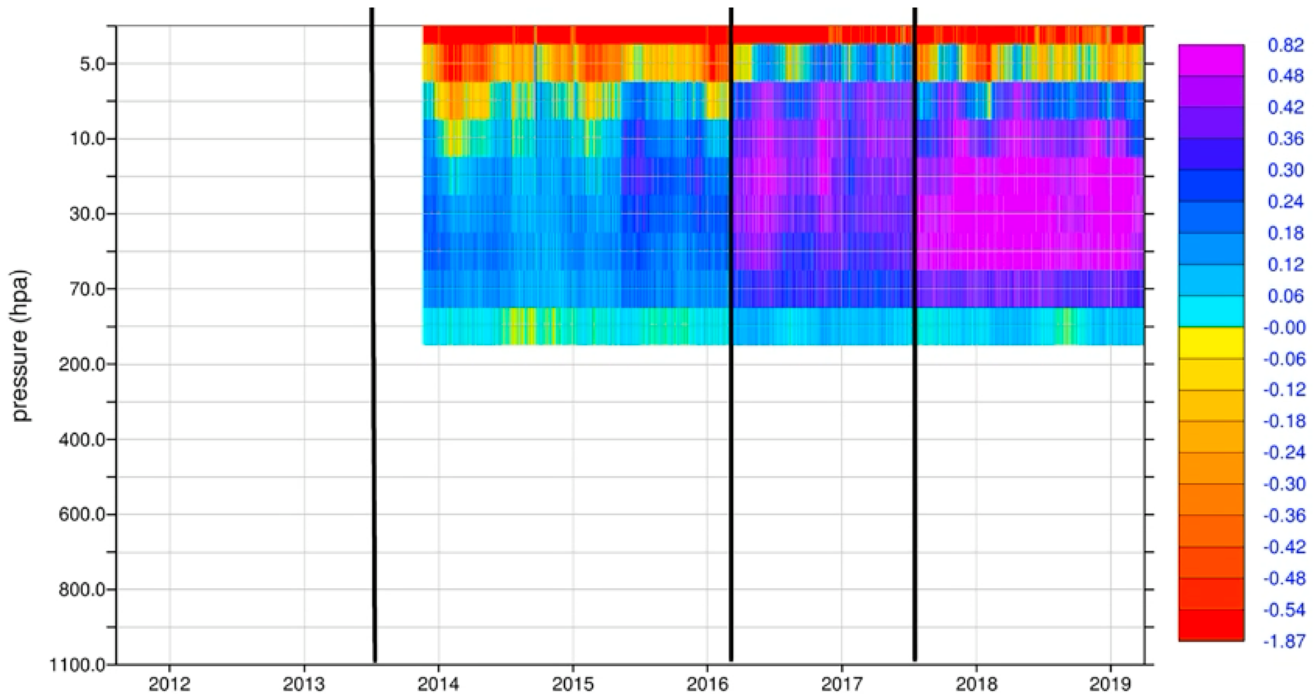
→ As the LEO moves behind the earth, we obtain a profile of bending angles

→ Temperature profiles can then be derived

→ GPS-RO can be assimilated without bias correction. They are good for highlighting errors/biases

How to estimate model biases (3/4)

The first-guess trajectory of the model can be compared to accurate observations



Difference between GPS-RO temperature retrievals and the IFS first-guess trajectory (O-B)

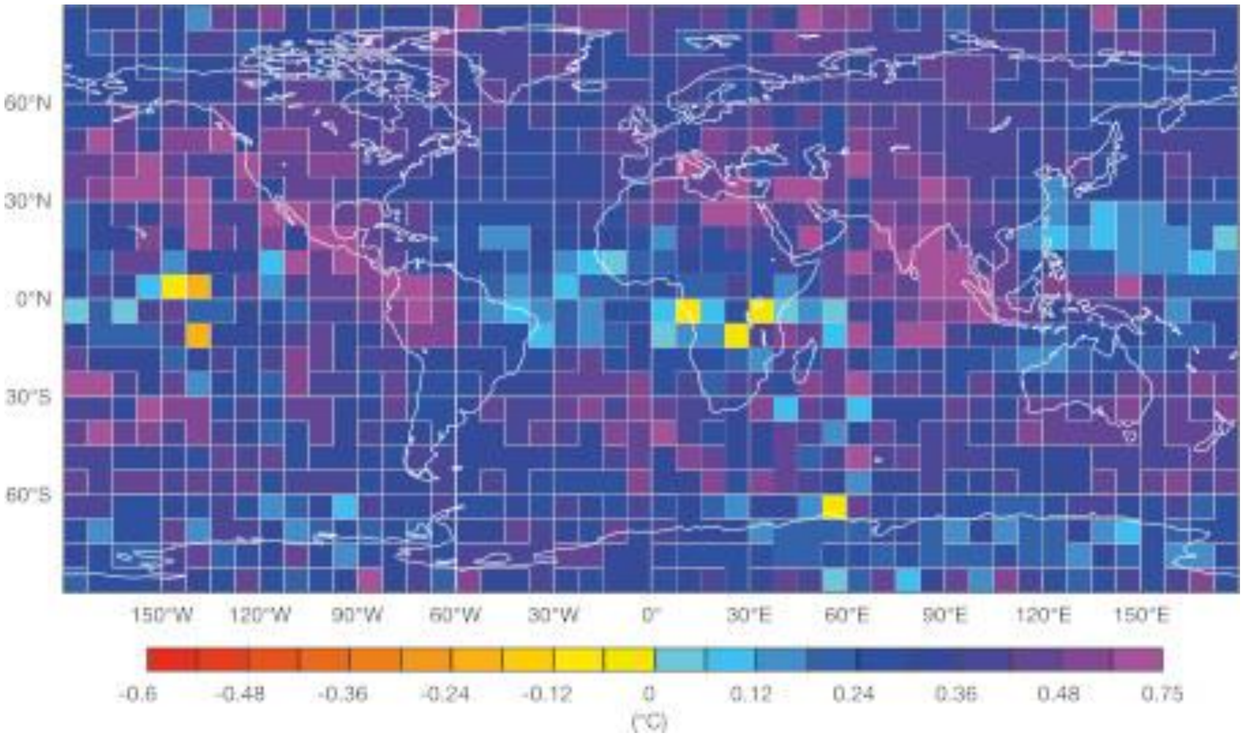
Errors in models are often systematic rather than random, zero-mean

→ Model has a temperature cold bias in the lower/mid stratosphere

→ Model has a warm bias in the upper stratosphere

How to estimate model biases (4/4)

The first-guess trajectory of the model can be compared to accurate observations between 70 hPa and 100 hPa over the period 31 August 2018 to 31 January 2019



Difference between GPS-RO temperature retrievals and the IFS first-guess trajectory (O-B)

Errors in models are often systematic rather than random, zero-mean

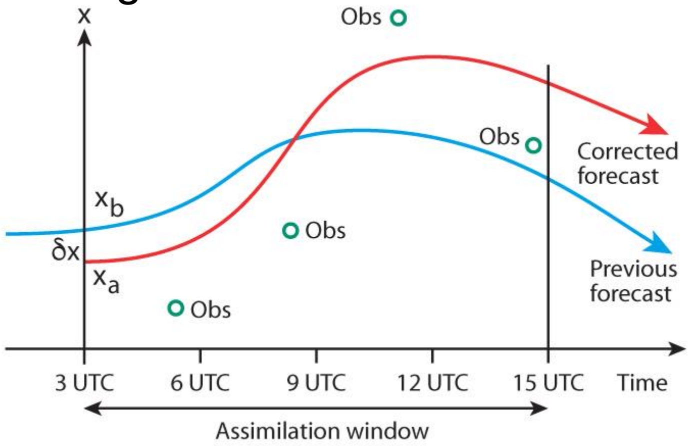
→ Model has a temperature cold bias in the lower/mid stratosphere

→ Model has a warm bias in the upper stratosphere

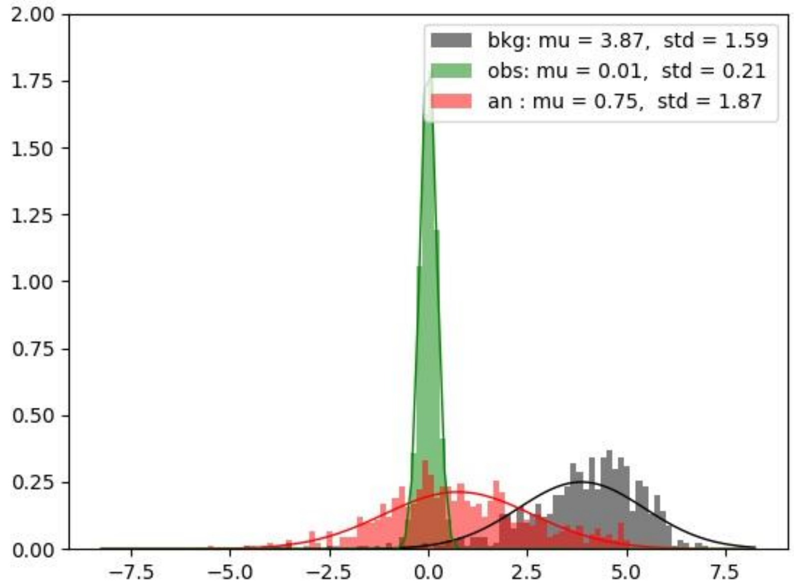
→ **Model biases have large-scale structure**

Model bias correction: weak-constraint 4D-Var (1/3)

Strong constraint 4D-Var

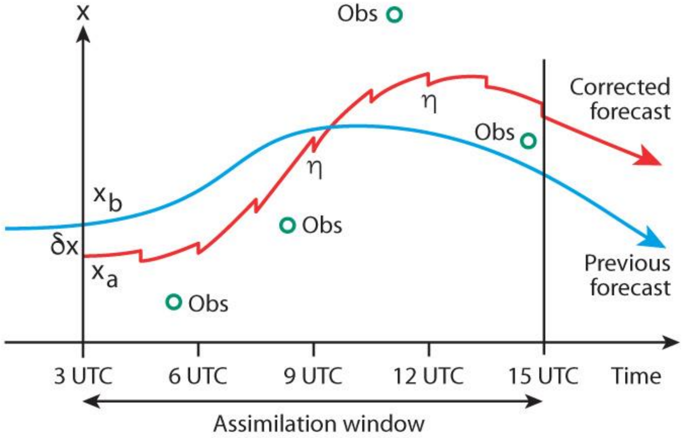


$$x_k = \mathcal{M}_k(x_{k-1})$$

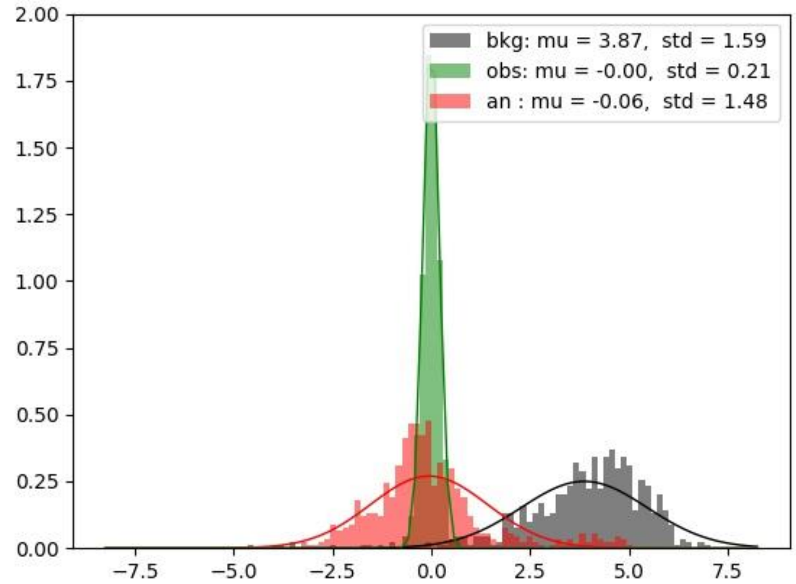


→ Large bias and standard deviation in the analysis

Weak constraint 4D-Var



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



→ Bias in the analysis has been reduced, standard deviation as well

Model bias correction: weak-constraint 4D-Var (2/3)

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 3D fields

- temperature
- vorticity
- divergence

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

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

Model state (x_0)
Observation bias (β)
Model bias (η)

→ Introduce additional controls to target an unbiased analysis

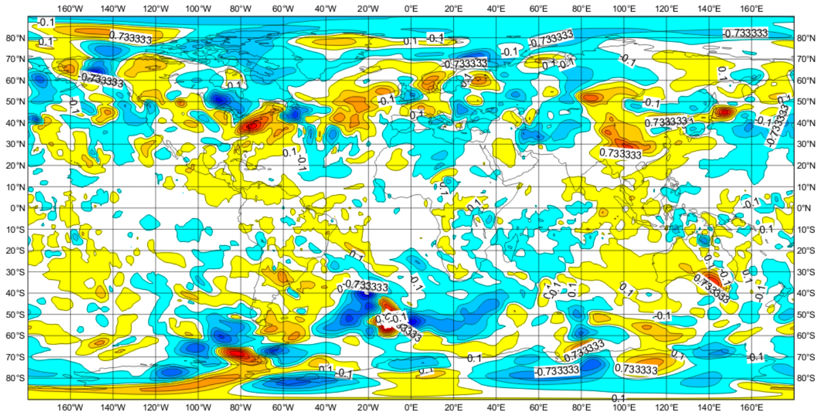
→ This looks very much like VarBC with a constant predictor, but in the model space!

→ 4D-Var has multiple ways to fit the observations

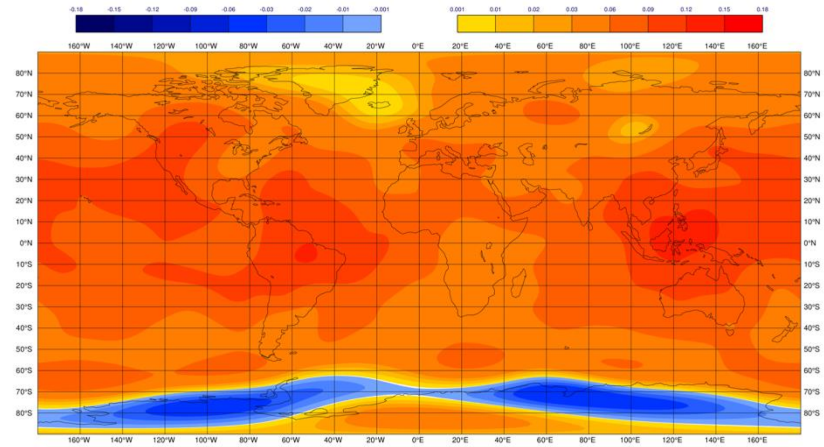
Model bias correction: weak-constraint 4D-Var (3/3)

How can we separate background and model error?

background correction (increment)

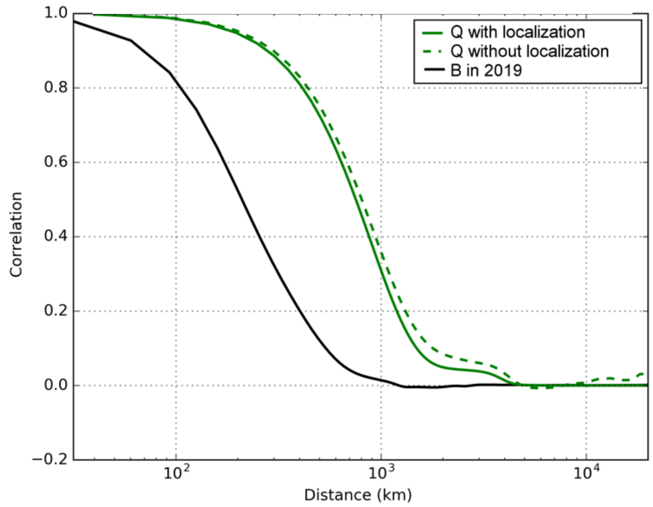


Model correction



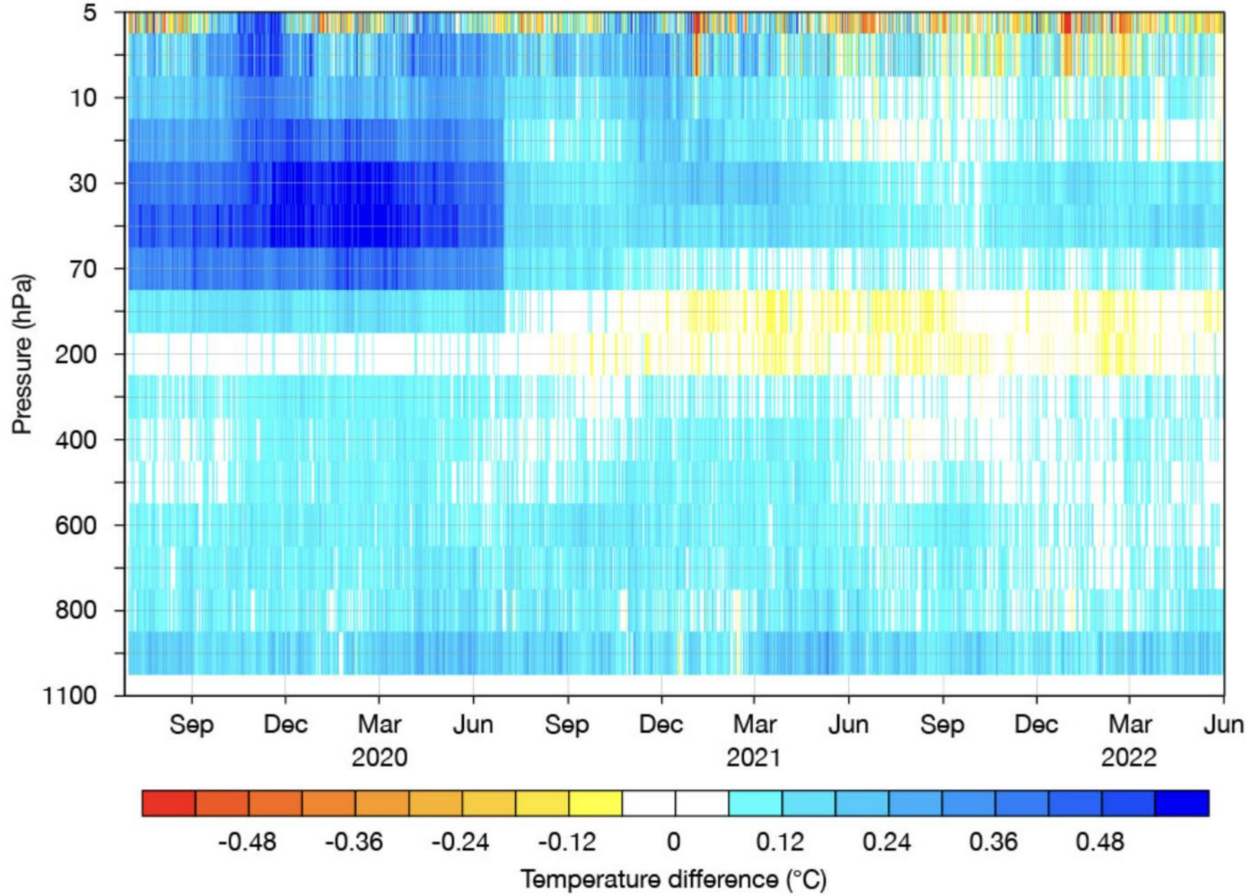
Background errors tend to be small scales while model errors tend to be large scale

Horizontal correlation

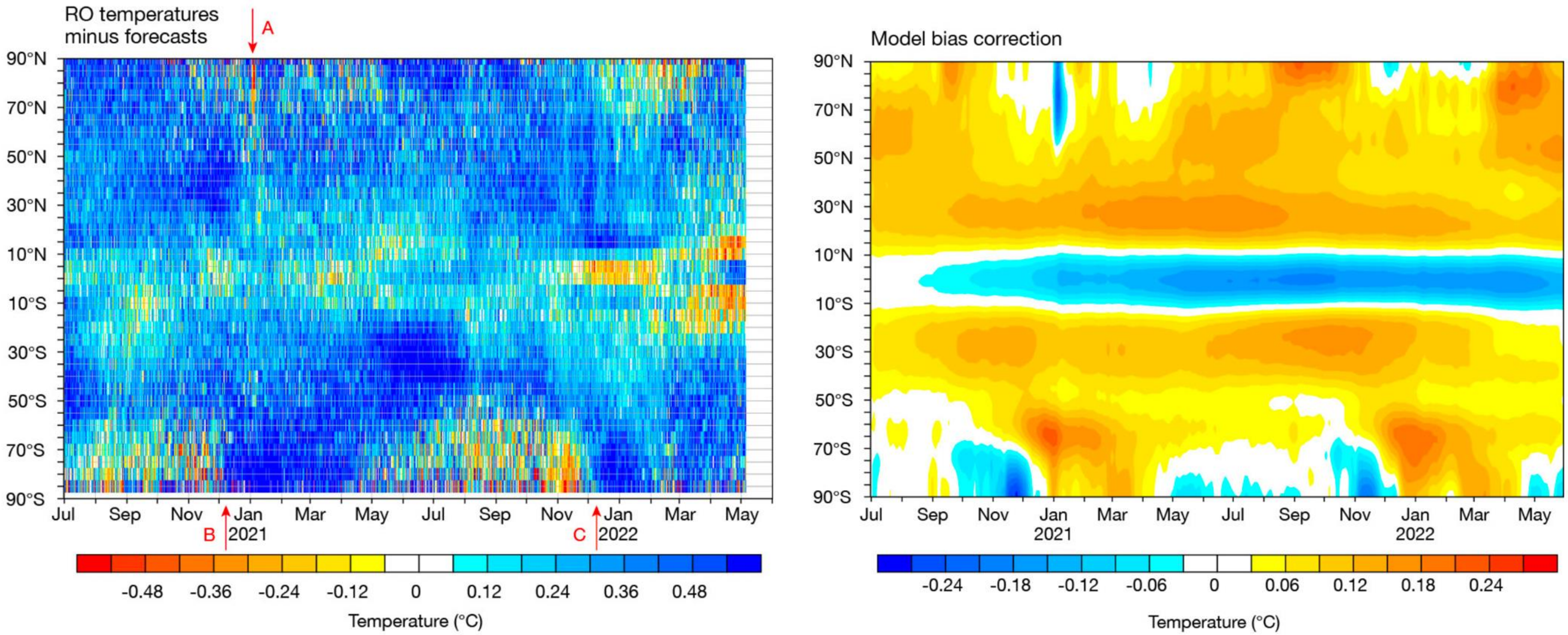


Weak-constraint 4D-Var in operations for the stratosphere (1/2)

Time series of the difference between radiosonde temperature observations and model first-guess (47r1 implemented on 30 June 2020)



Weak-constraint 4D-Var in operations for the stratosphere (2/2)



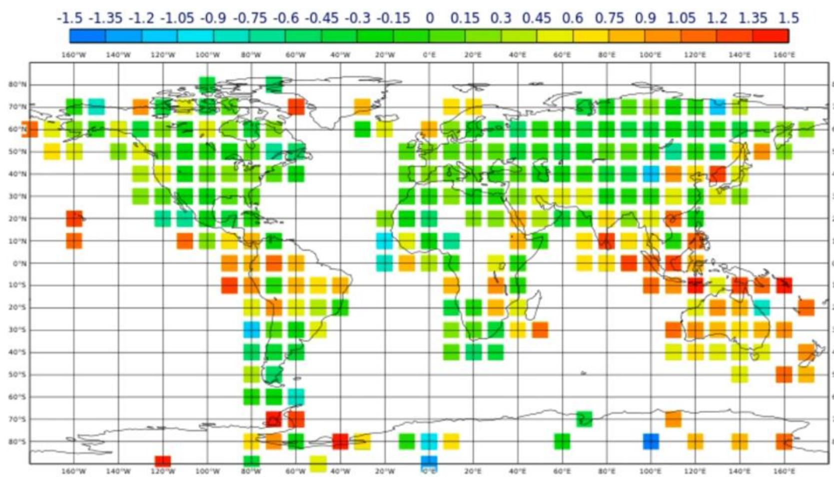
A) On 31 December 2020, a Sudden Stratospheric Warming (SSW) event started over the northern hemisphere

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

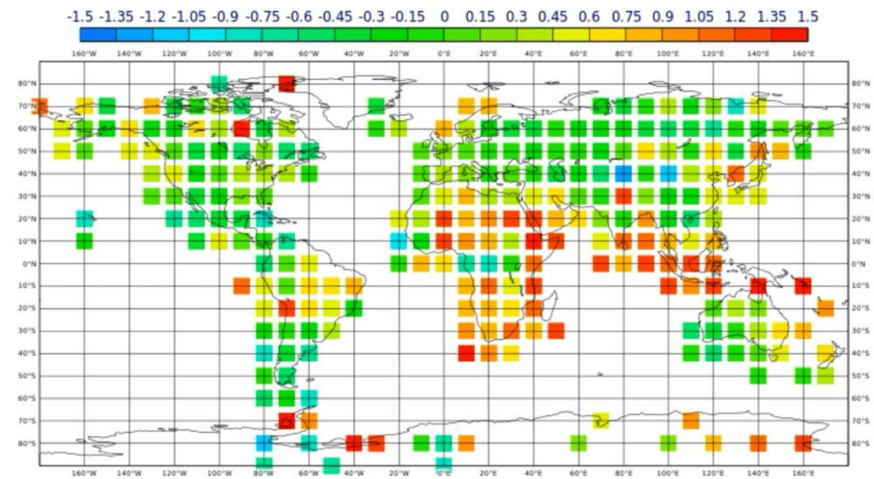
Model biases in the boundary layer (1/3)

Several diagnostics shows that the structure of model biases is time-correlated

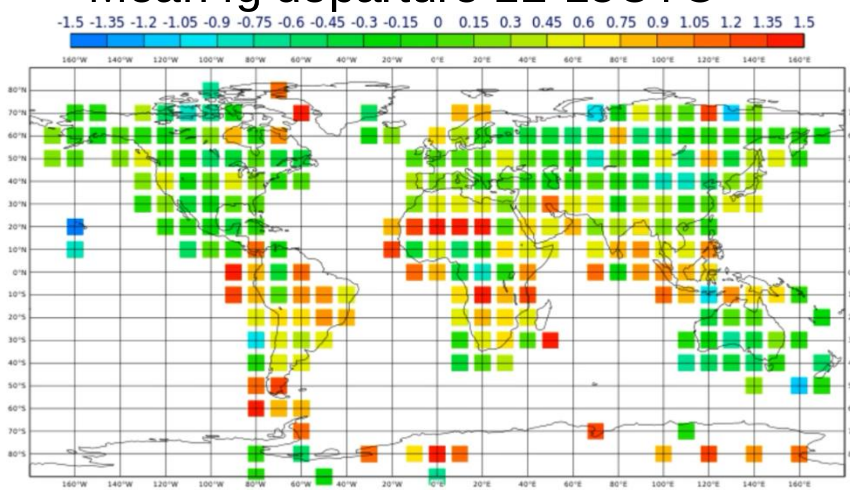
Mean fg departure 00-03UTC



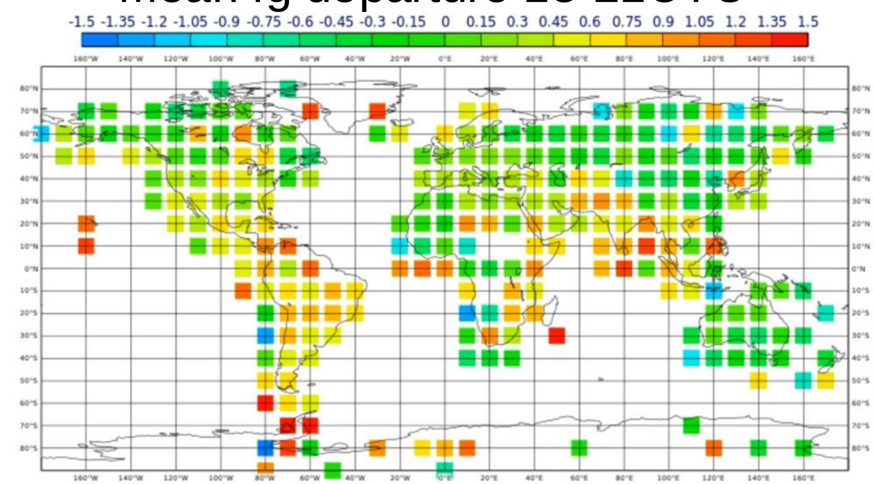
Mean fg departure 06-09UTC



Mean fg departure 12-15UTC

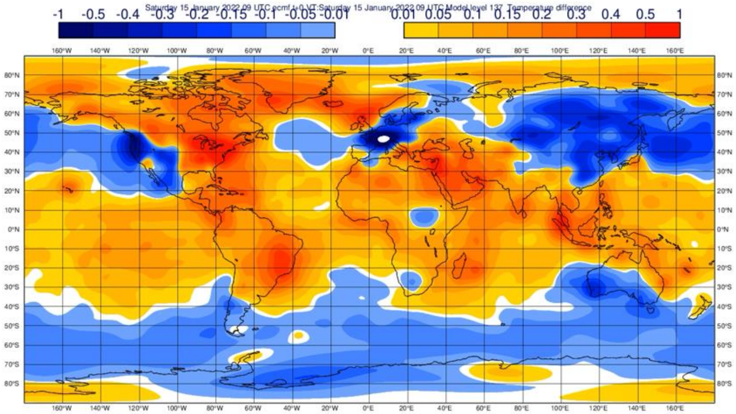


Mean fg departure 18-21UTC

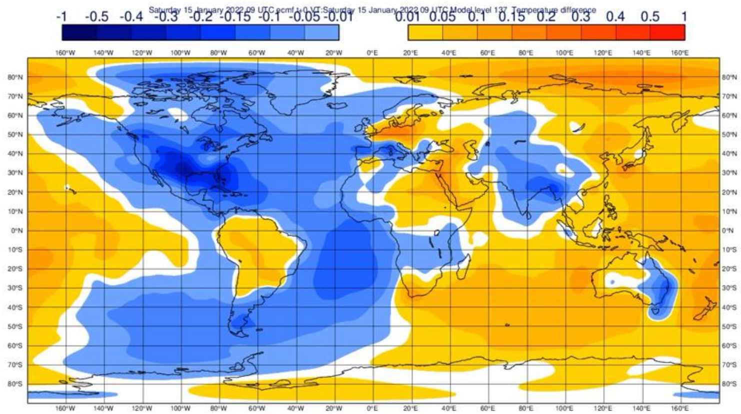


Model biases in the boundary layer (2/3)

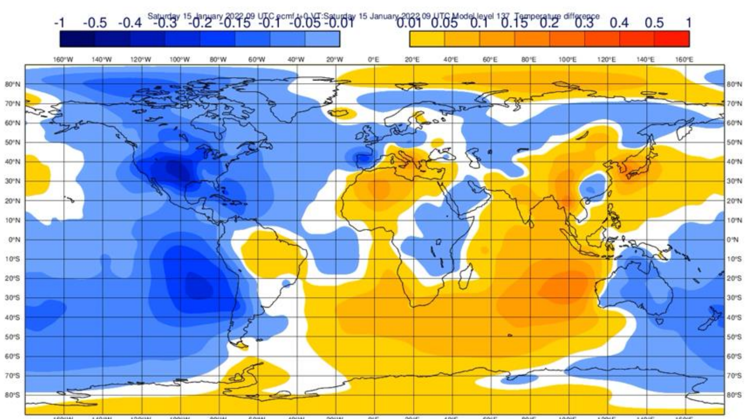
η_0



η_1



η_2



New representation of the model bias:

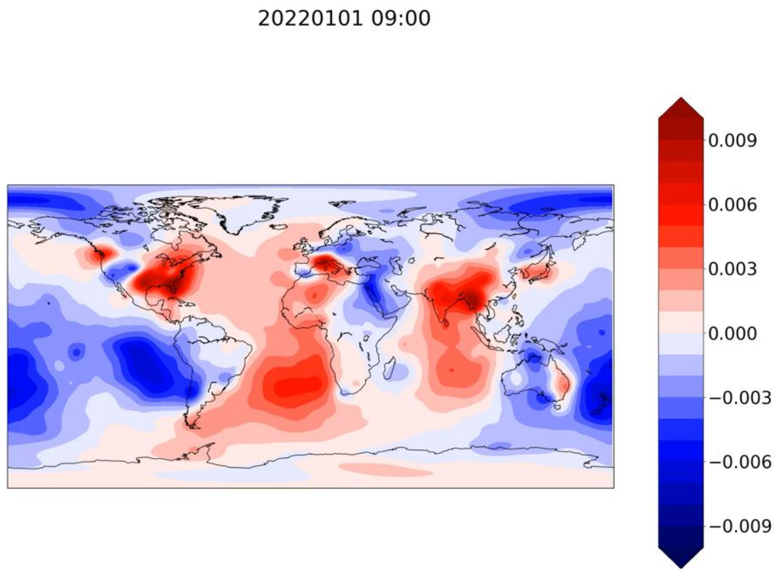
$$\eta_0 + \eta_1 \sin(2\pi \frac{t}{24}) + \eta_2 \cos(2\pi \frac{t}{24})$$

- ➔ Time-varying within the assimilation window
- ➔ Designed to capture a diurnal cycle

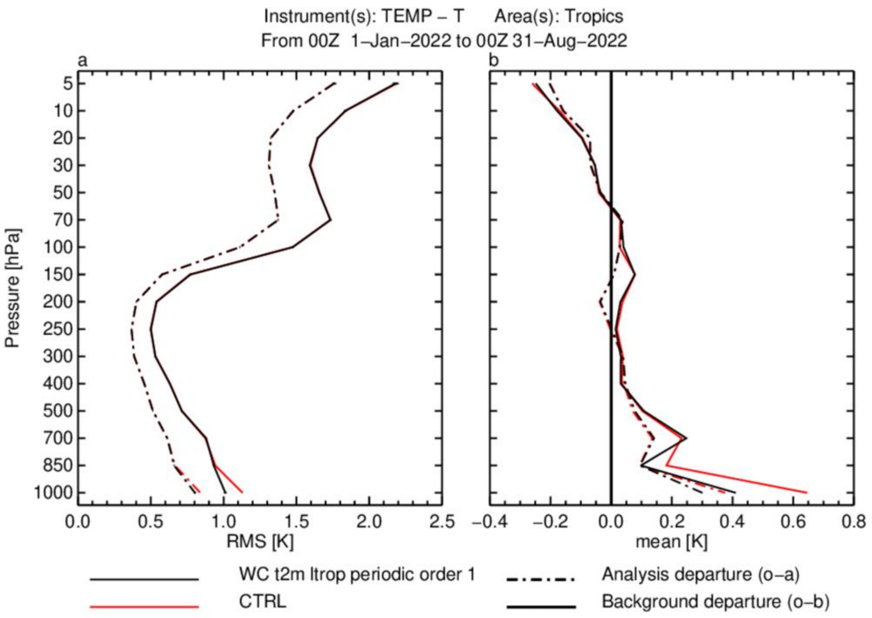
$$\begin{aligned} J(x_0, \beta, \eta) &= \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) \\ &+ \frac{1}{2}\sum_{k=0}^{K}[y_k - \mathcal{H}(x_k) - b(x_k, \beta)]^T \mathbf{R}_k^{-1}[y_k - \mathcal{H}(x_k) - b(x_k, \beta)] \\ &+ \frac{1}{2}(\beta - \beta_b)^T \mathbf{B}_\beta{}^{-1}(\beta - \beta_b)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

Model biases in the boundary layer (3/3)

Model bias correction (level 137)

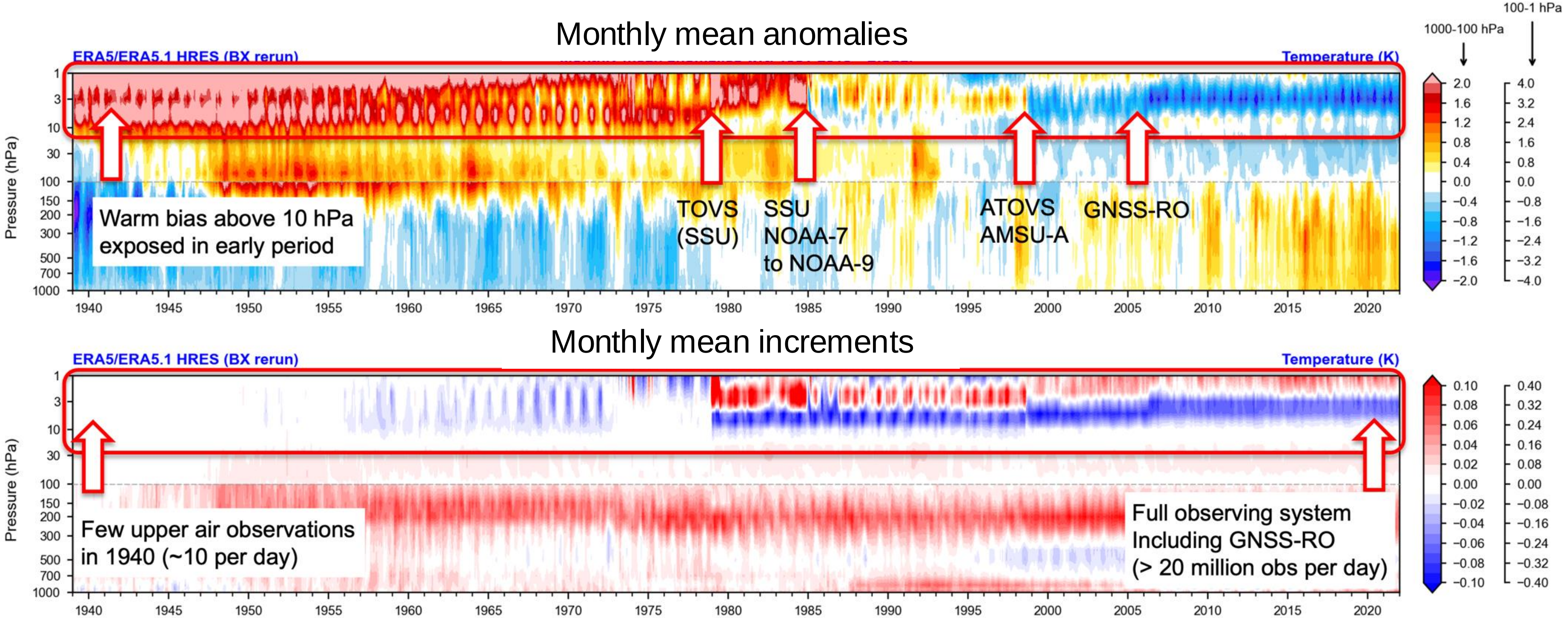


Impact in the mean state against radiosondes



Going in the next operational upgrade in few months (50R1)

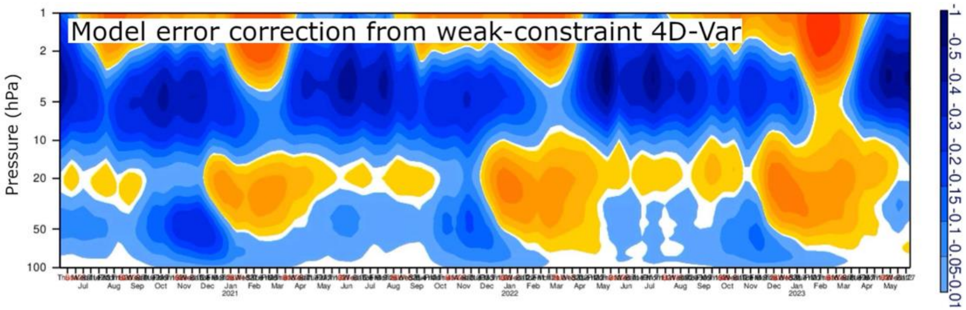
Using WC-4DVar for reanalysis (1/3)



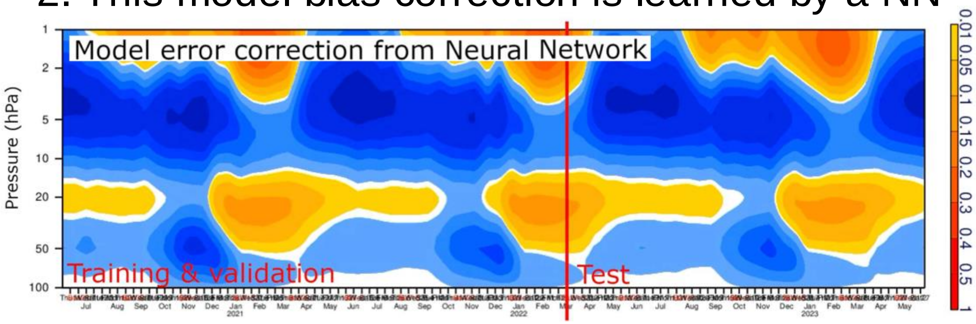
Challenge: reduce artefacts in the stratosphere coming from model biases while preserving climate trends. Amplitude of current spurious signal can be large (>1K)

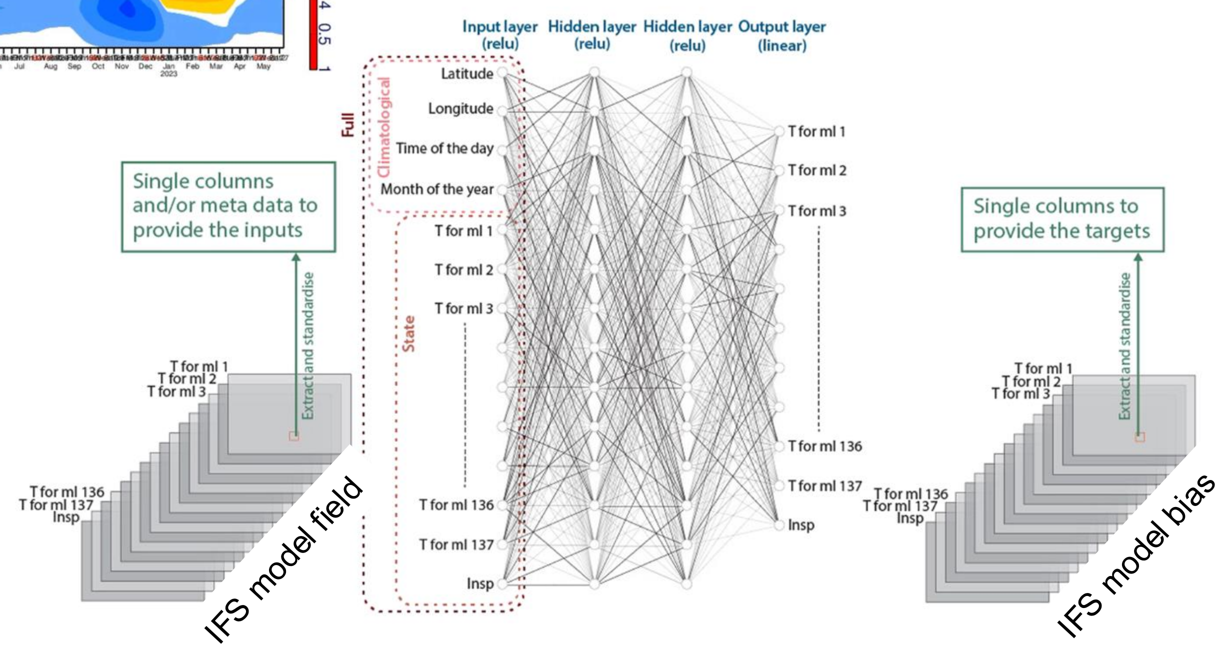
Using WC-4DVar for reanalysis (2/3)

1. Weak-constraint 4D-Var estimates model biases effectively over recent periods (2021/2023)



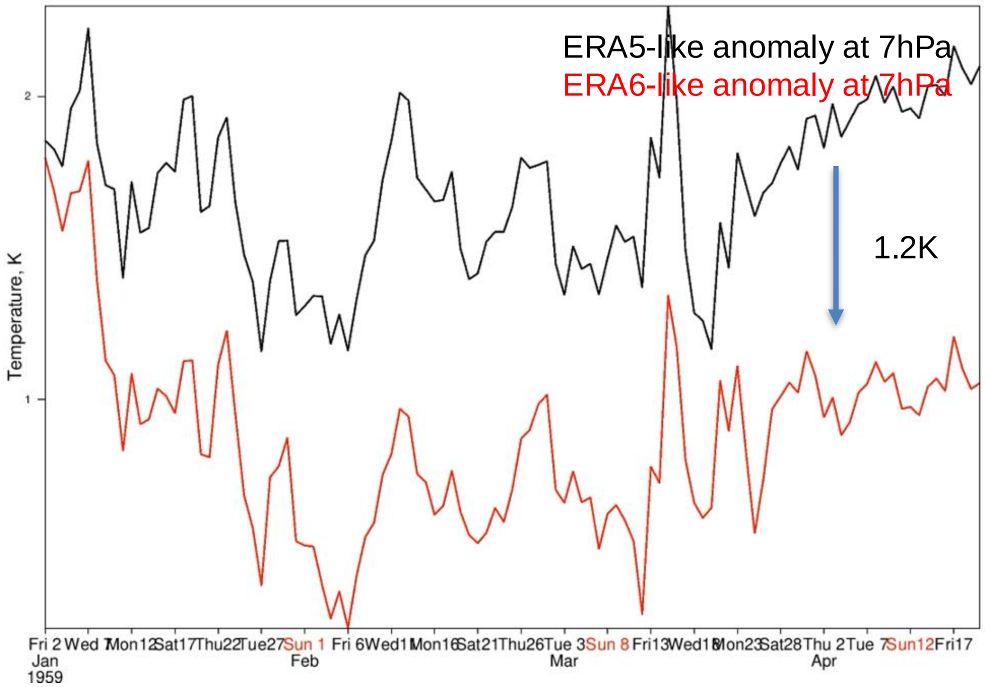
2. This model bias correction is learned by a NN





Using WC-4DVar for reanalysis (3/3)

3. The ML correction can be applied over any reanalysis period (e.g. Jan 1959 to May 1959)



4. The NN cools down the upper stratosphere to account for the warm bias

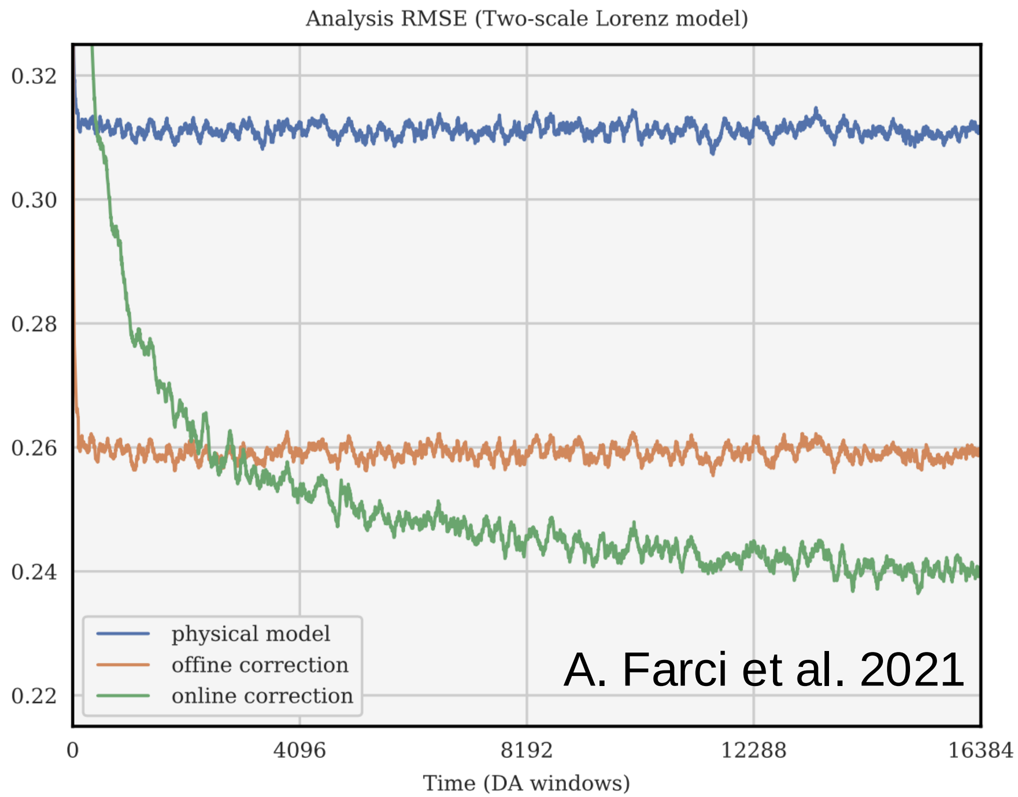
A ML approach to correct model biases (1/2)

The hybrid model (physical model + NN correction) is estimated inside 4D-Var

$$\mathbf{x}_{k+1} = \mathcal{M}^{\mathrm{nn}}_{k+1:k}(\mathbf{p}, \mathbf{x}_k) = \mathcal{M}_{k+1:k}(\mathbf{x}_k) + \mathcal{F}(\mathbf{p}, \mathbf{x}_k)$$

NN online loss function

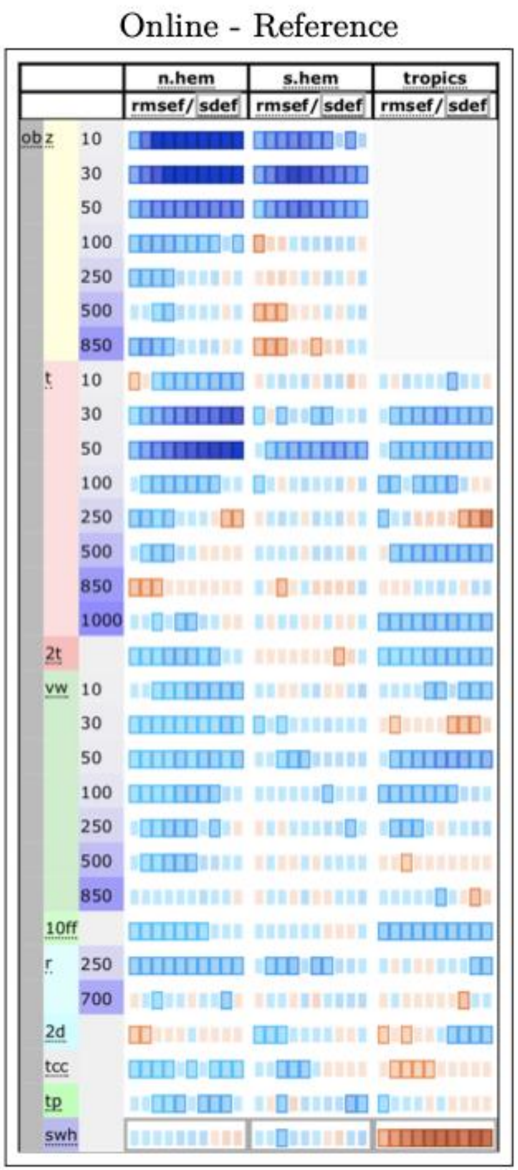
$$\mathcal{J}^{\mathrm{nn}}(\mathbf{p}, \mathbf{x}_0) = \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\left\|\mathbf{p} - \mathbf{p}^{\mathrm{b}}\right\|^2_{\mathbf{P}^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\left\|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}^{\mathrm{nn}}_{k:0}(\mathbf{p},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



➔ learn both model state and NN parameters from observations

➔ the online correction steadily improves the model, learning from observations

A ML approach to correct model biases (1/2)

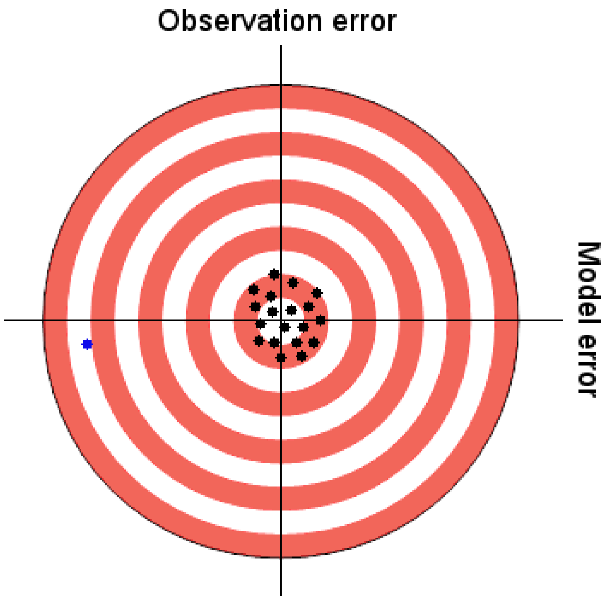


A. Farci et al. 2024

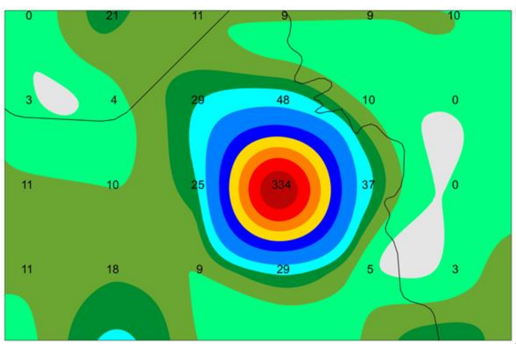
The hybrid model (physical model + NN correction) is applied in the 10-day forecast

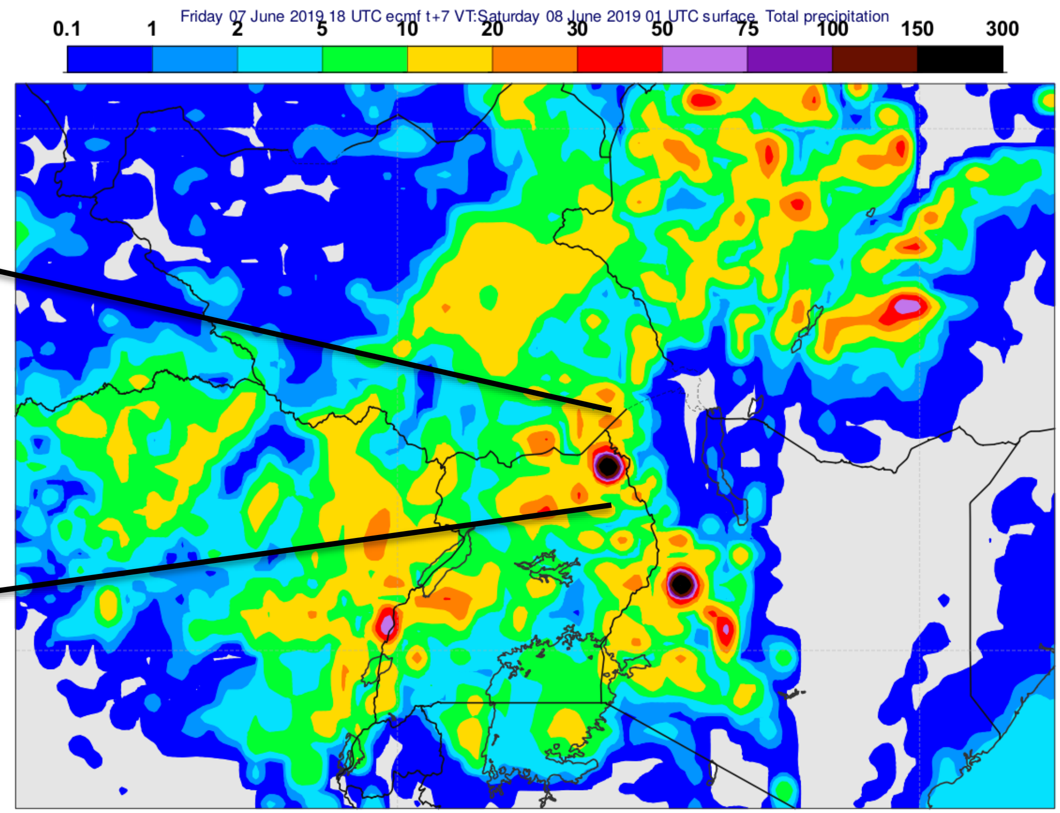
$$\mathbf{x}_{k+1} = \boldsymbol{\mathcal{M}}^{\mathrm{nn}}_{k+1:k}\left(\mathbf{p}, \mathbf{x}_k\right) = \boldsymbol{\mathcal{M}}_{k+1:k}\left(\mathbf{x}_k\right) + \boldsymbol{\mathcal{F}}\left(\mathbf{p}, \mathbf{x}_k\right)$$

Not the job of weak-constraint 4D-Var: Model gross errors



Total precipitation on 07 June 2019
(accumulated over 6 hours)

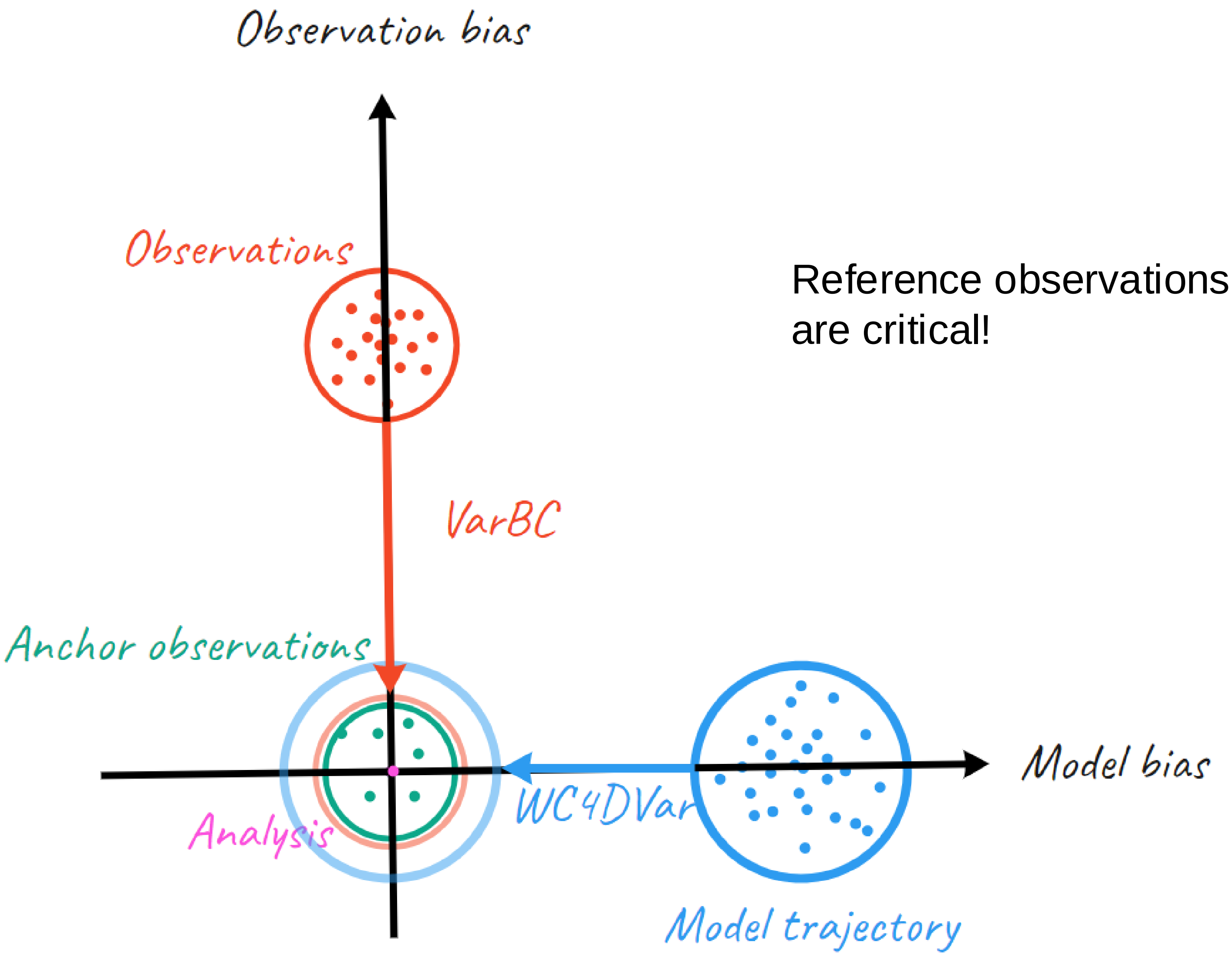




→ Continuous monitoring

→ Keep improving the model

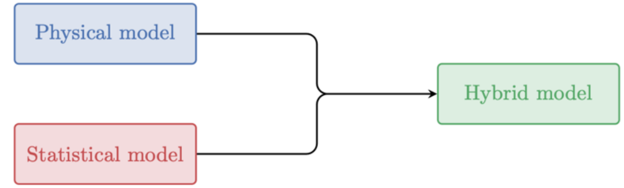
Summary 1/1



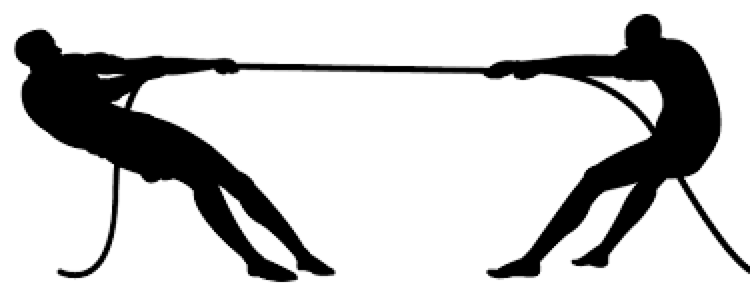
Reference observations (anchors) are critical!

Summary 2/2

From bias-blind to bias-aware data assimilation



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



Any questions? Feel free to contact me patrick.laloyaux@ecmwf.int