Observation and Model Bias Correction

Patrick Laloyaux

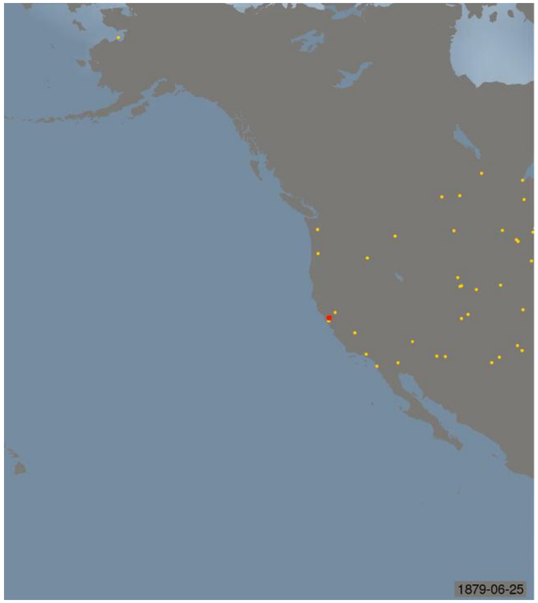
To question how much we should trust observations and models

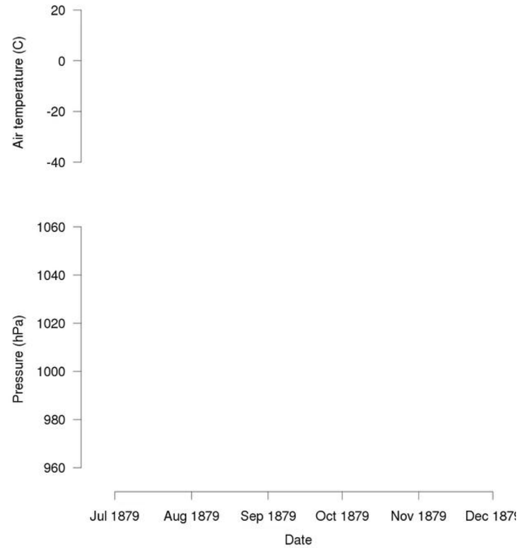
To understand how 4D-Var needs to be modified to take biases into account

To identify the challenges with this new 4D-Var formulation

Examples of biases in observations (1/3)

The USS Jeannette (1879, Artic, 33 crew members)





SST measurements from standard buckets have a cold bias (~0.4C)

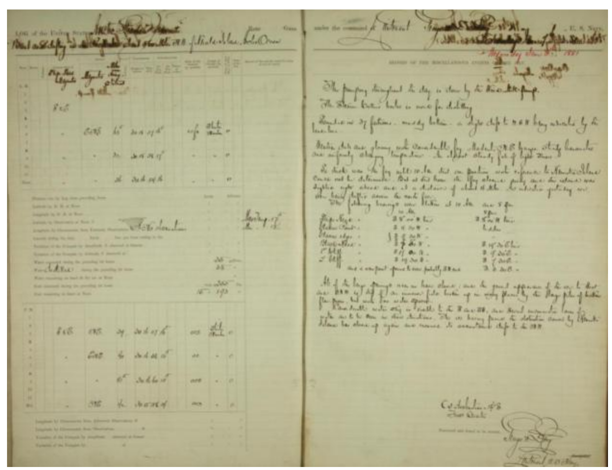
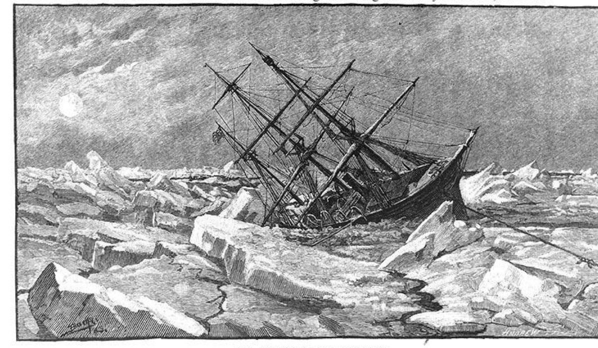


Photo # NH 52000 Steamer Jeannette sinking after being crushed by Arctic ice, June 1881



THE SINKING OF THE JEANNETTE.

Photo # NH 52002 Jeannette's crewmen drag their boats over the Arctic ice, June-August 1881

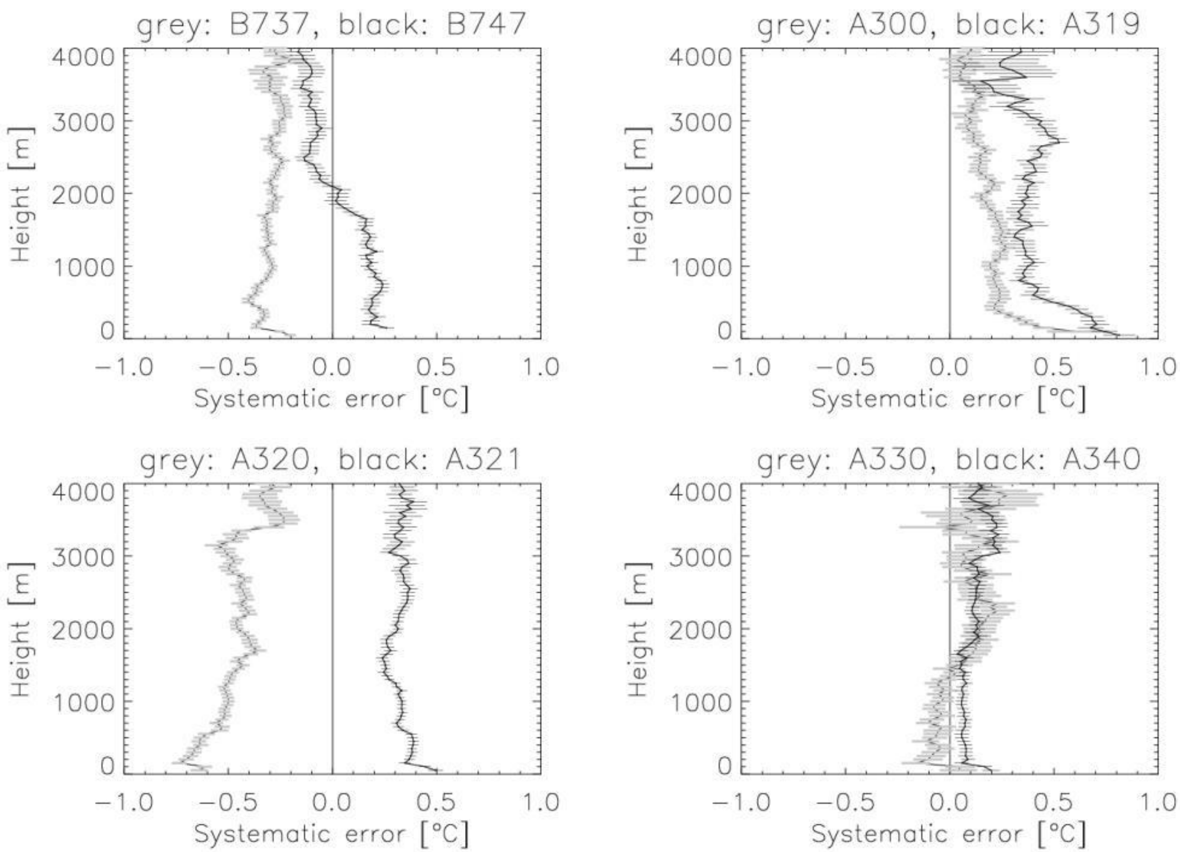
DRAGGING THE BOATS OVER THE ICE.

Photo # NH 92142 LCdr. DeLong and his party wading ashore on the Lena Delta, Siberia, 17 Sept. 1881

WADING ASHORE.

Examples of biases in observations (2/3)

One year of measurements from aircrafts landing at Frankfurt



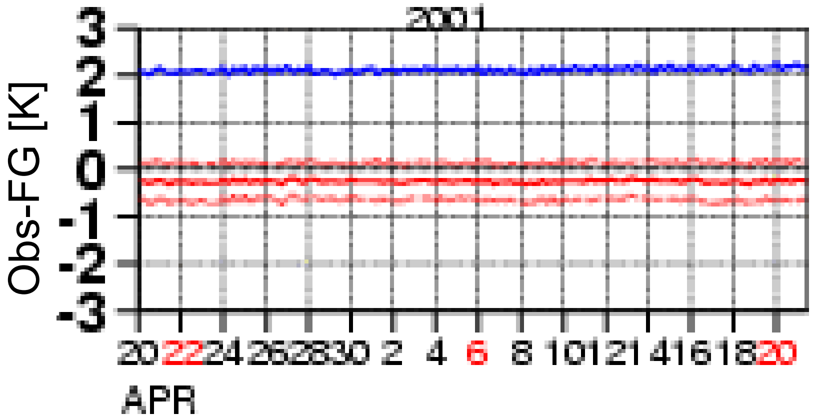
Estimation of observation biases done by inter-comparison between instruments

➔ Involve experts knowing the instruments

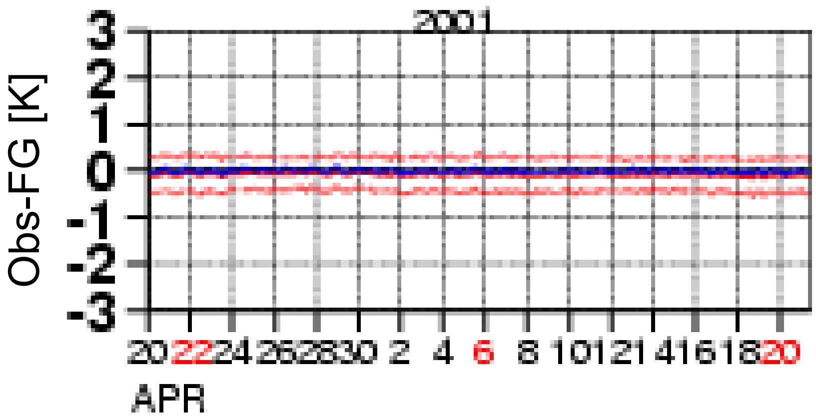
➔ observation bias is estimated using the hourly mean of all measured profiles

Examples of biases in observations (3/3)

High-Resolution Infrared Radiation Sounder (HIRS) channel 5 is sensitive to to atmospheric temperature in the middle troposphere



HIRS channel 5 on NOAA-14 satellite has +2.0K radiance bias against FG (blue line)



HIRS channel 5 on NOAA-16 satellite has no significant bias against FG (blue line)

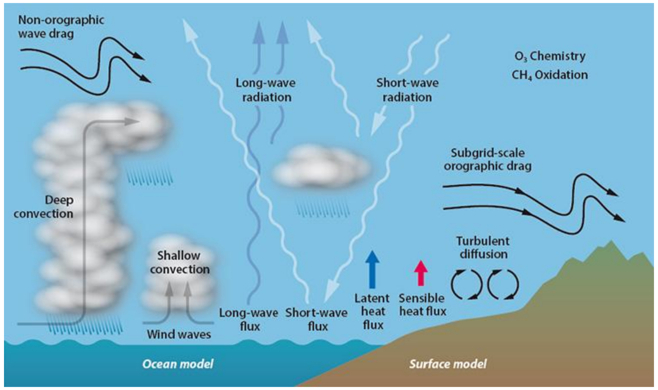
Estimation of observation biases done by inter-comparison between instruments

➔ Involve experts knowing the instruments

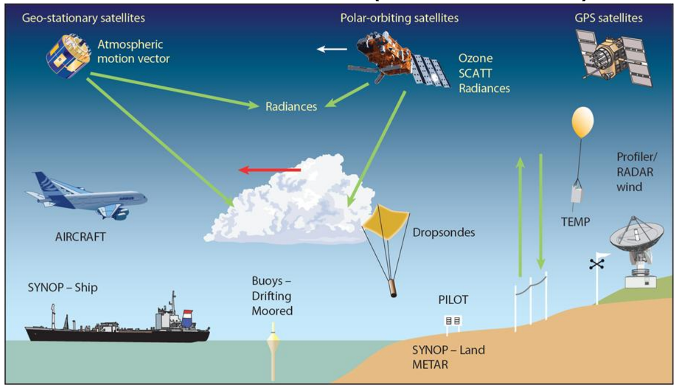
➔ observation bias is estimated comparing obs with the model (time/space average)

What you have seen so far on data assimilation

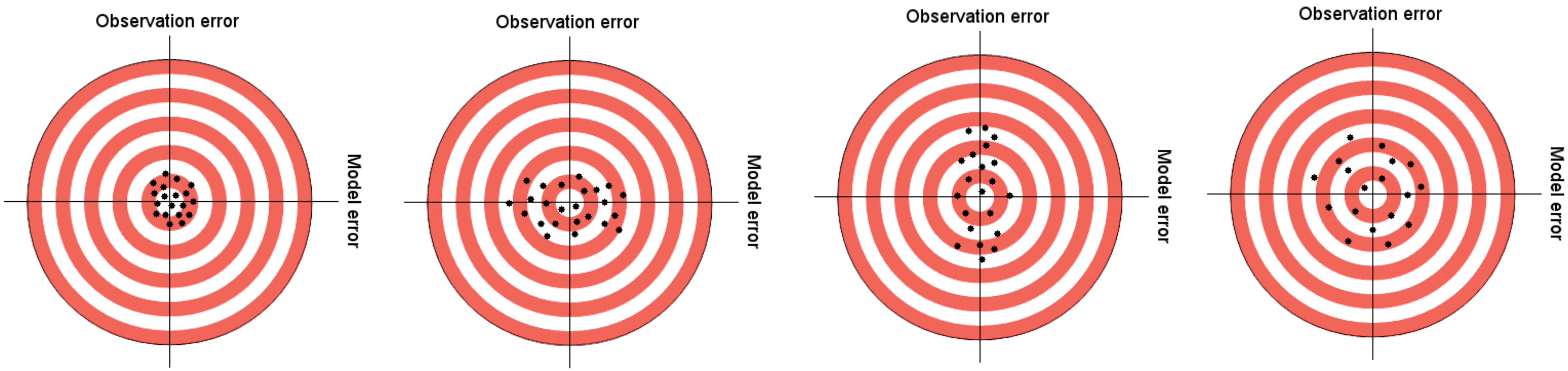
Model (with errors)



Observations (with errors)



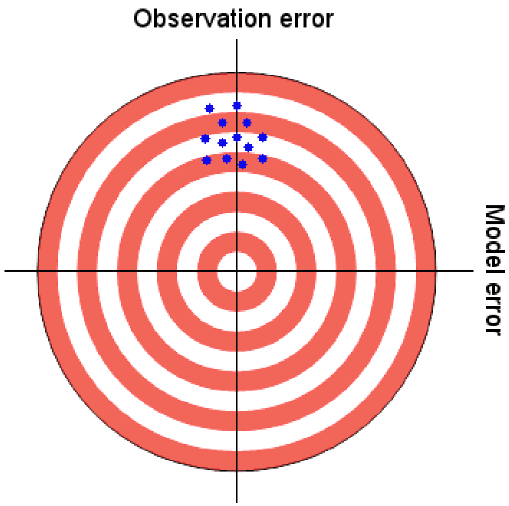
If you are lucky, model and observations are **accurate** (no biases, mean error is zero)



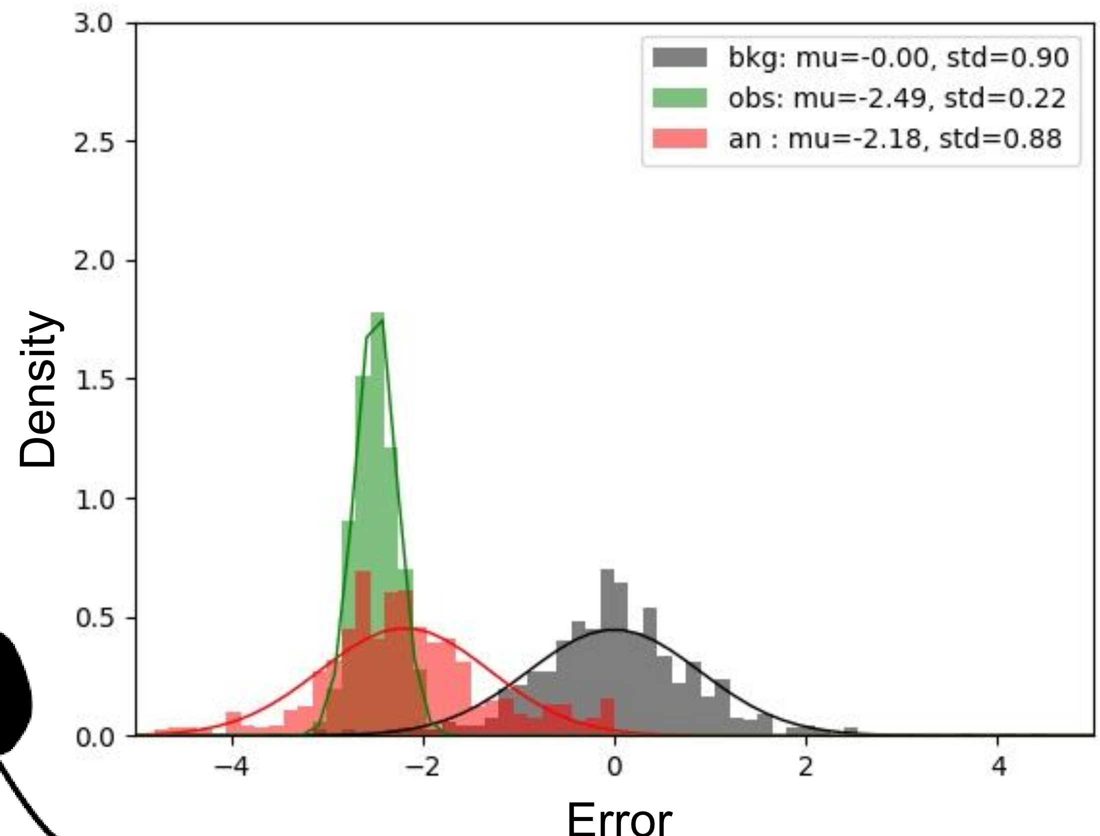
$$J(x_0) = \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) + \frac{1}{2}\sum_{k=0}^{K}[y_k - \mathcal{H}(x_k)]^T \mathbf{R}_k^{-1}[y_k - \mathcal{H}(x_k)]$$

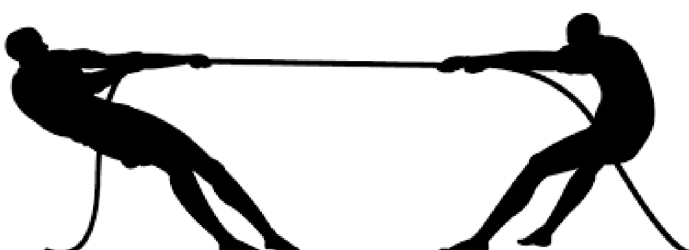
Most of the time, we are unlucky!

Observation biases matter



- If standard 4D-Var is used to assimilate biased observations (systematic errors), the resulting analysis will be biased.
- In this case the background is more accurate than the analysis!





Changing the 4D-Var formulation (introducing VarBC)

$$
\begin{aligned}
J(x_0, \beta) &= \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) \\
&+ \frac{1}{2}(\beta - \beta_b)^T \mathbf{B}_\beta{}^{-1}(\beta - \beta_b) \\
&+ \frac{1}{2}\sum_{k=0}^{\text{Radiosonde}} [y_k - \mathcal{H}(x_k)]^T \mathbf{R}_k^{-1}[y_k - \mathcal{H}(x_k)] \\
&+ \frac{1}{2}\sum_{k=0}^{\text{NOAA-16}} [y_k - \mathcal{H}(x_k)]^T \mathbf{R}_k^{-1}[y_k - \mathcal{H}(x_k)] \\
&+ \frac{1}{2}\sum_{k=0}^{\text{NOAA-14}} [y_k - \beta - \mathcal{H}(x_k)]^T \mathbf{R}_k^{-1}[y_k - \beta - \mathcal{H}(x_k)]
\end{aligned}
$$

Model state

Observation bias parameters

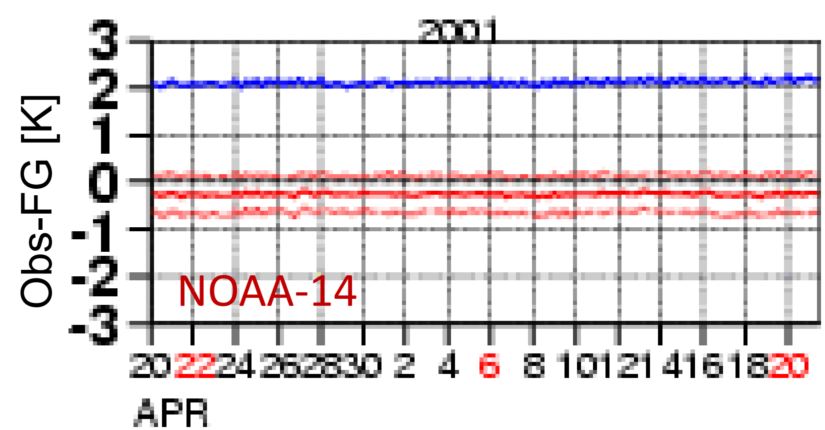
Unbiased observations (anchor)

Biased observations

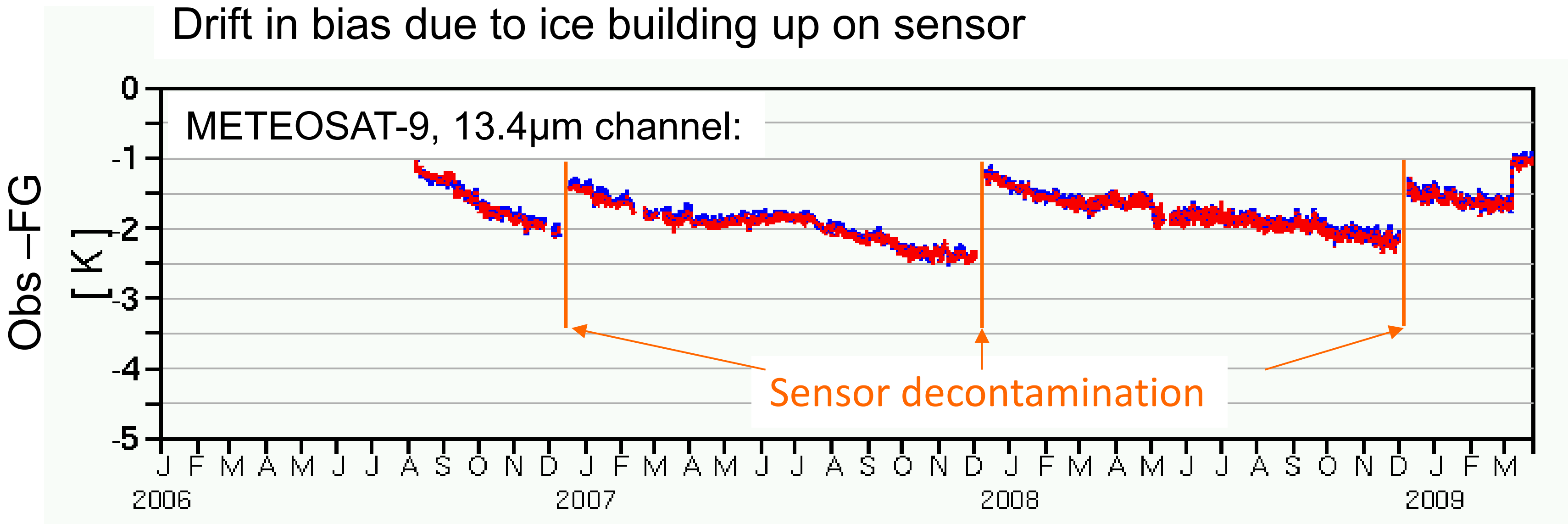
Bias model

Variational Bias Correction (VarBC)

- We choose which observations we want to correct and which observations we trust
- We choose the bias model b(β) = β
- 4D-Var minimization estimates the value of the VarBC parameters



Changing the 4D-Var formulation (introducing VarBC)



VarBC needs to correct for observation bias changing over time

- Bias model = b(β) = β
- β is evolving over time depending how much ice is building up

Changing the 4D-Var formulation (introducing VarBC)

$$
\begin{aligned}
J(x_0,\beta) = {} & \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) \\
& + \frac{1}{2}(\beta - \beta_b)^T \mathbf{B}_\beta^{-1}(\beta - \beta_b) \\
& + \frac{1}{2}\sum_{k=0}^{\text{Radiosonde}} [y_k - \mathcal{H}(x_k)]^T \mathbf{R}_k^{-1}[y_k - \mathcal{H}(x_k)] \\
& + \frac{1}{2}\sum_{k=0}^{\text{NOAA-16}} [y_k - \mathcal{H}(x_k)]^T \mathbf{R}_k^{-1}[y_k - \mathcal{H}(x_k)] \\
& + \frac{1}{2}\sum_{k=0}^{\text{NOAA-14}} [y_k - \beta - \mathcal{H}(x_k)]^T \mathbf{R}_k^{-1}[y_k - \beta - \mathcal{H}(x_k)]
\end{aligned}
$$

Model state

Observation bias parameters

Parameter estimates from previous analysis

Background covariance matrix for VarBC parameters

Variational Bias Correction (VarBC)

- A cycling scheme for updating the bias parameter estimates
- Specification of the background covariance matrix $\mathbf{B}_\beta$ (large value ➔ fast adaptation, small value ➔ slow adaptation)

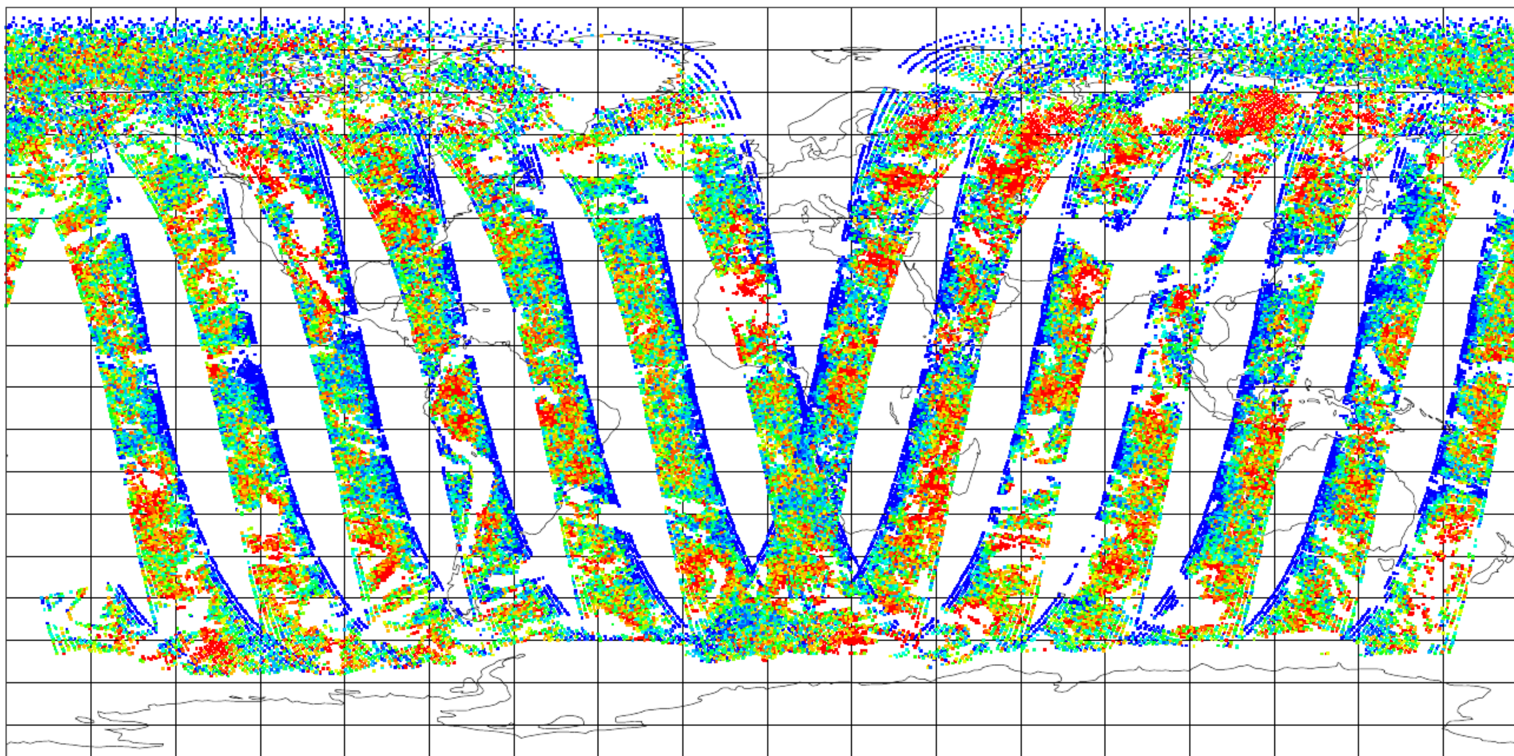
$$
\mathbf{B}_{\boldsymbol{\beta}} = \begin{bmatrix} \mathbf{B}_{\boldsymbol{\beta}}^{(1)} & & 0 \\ & \ddots & \\ 0 & & \mathbf{B}_{\boldsymbol{\beta}}^{(J)} \end{bmatrix}
$$

Building models of observation biases (a more complex case)

ECMWF is assimilating polar-orbiting Metop-C satellite (launched on 7 November 2018)

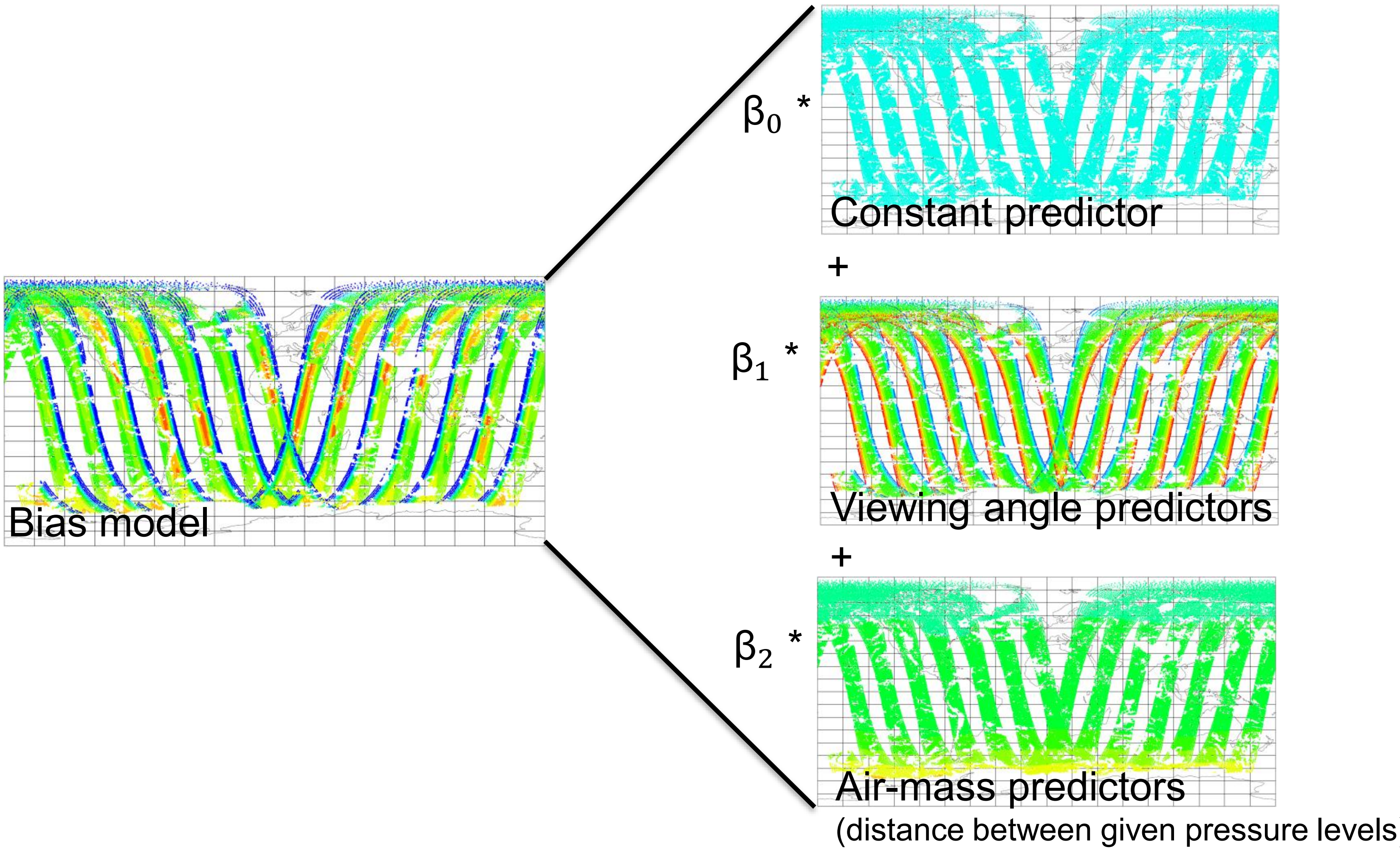
Observation bias is estimated inside 4D-Var
➔ comparing measurements with model
➔ specifying the structure of the model bias

Metop-C AMSUA-A Channel 5 (obs-model)

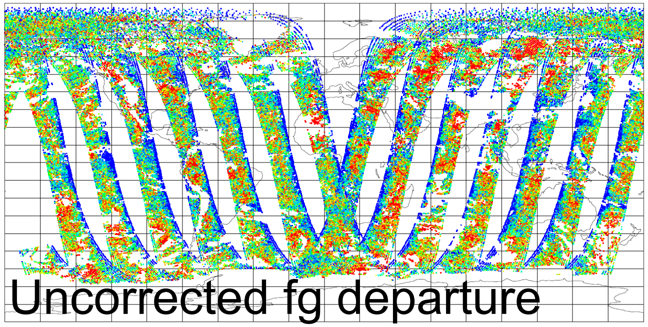


Building models of observation biases (a more complex case)

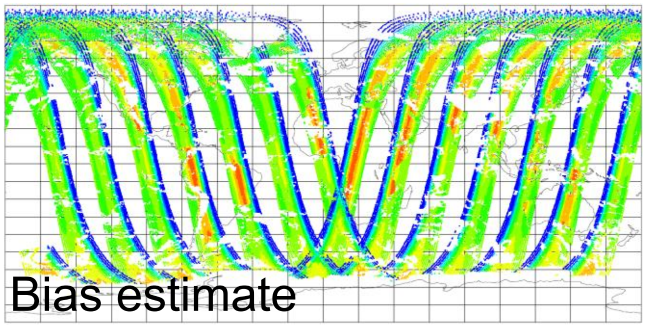
Bias model = $b(\beta) = b(\beta_0 , \beta_1 , \beta_2) = \beta_0 + \beta_1$ * viewing angle + β_2 * air-mass



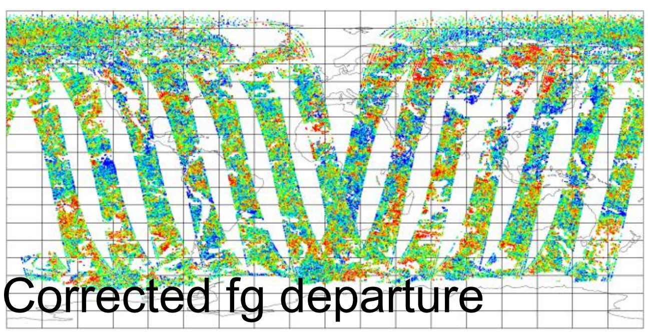
Building models of observation biases (a more complex case)



-



=



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

Do not include too many predictors in the bias correction models

➔ to avoid correcting for other sources of errors (background errors/model error)

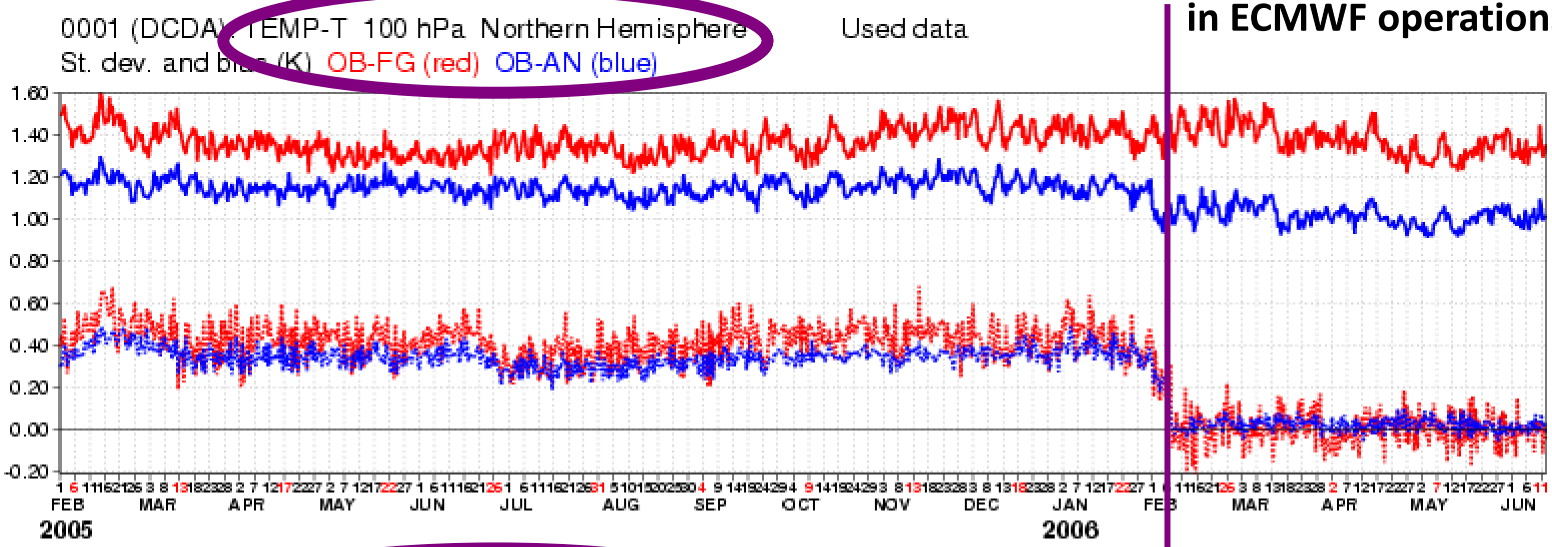
➔ corrected fg departure should still contain some information to constrain x_0

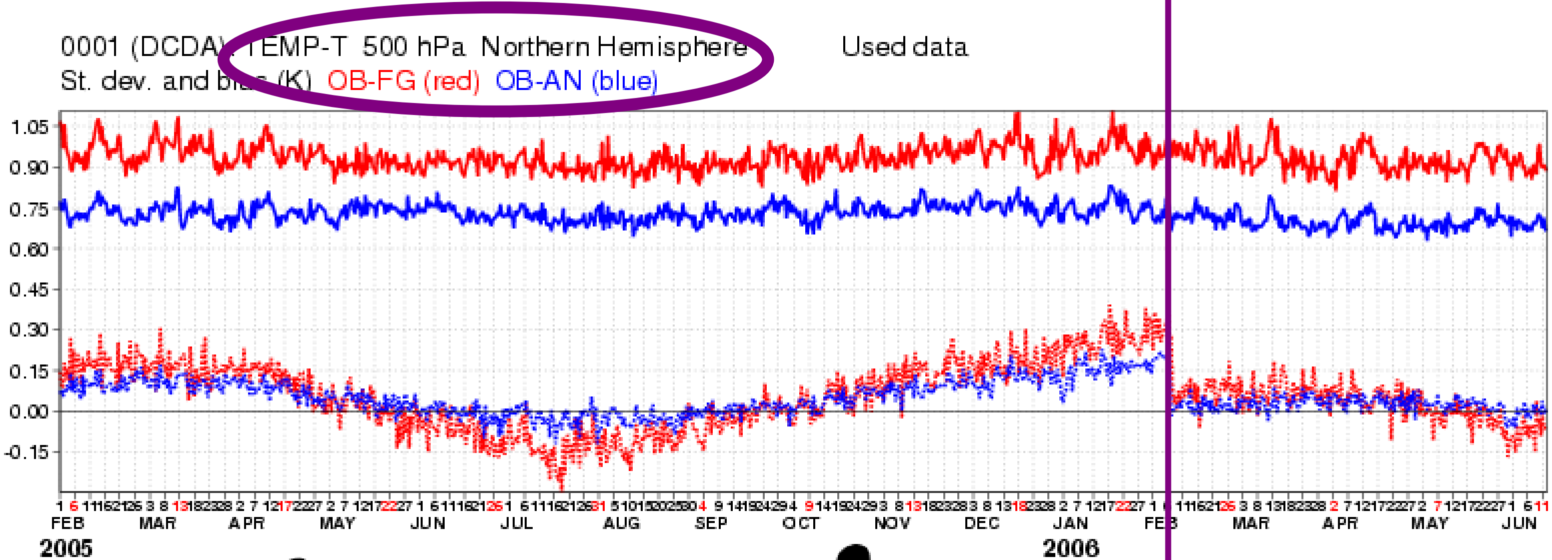
Generic VarBC formulation

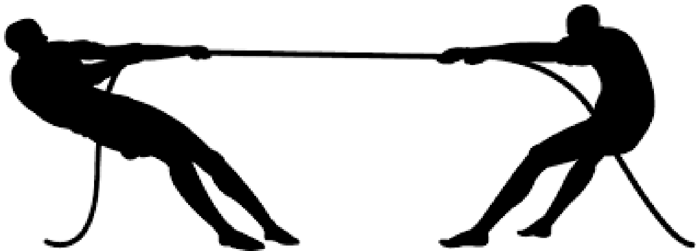
$$b(\beta,x_k)=\beta_0+\sum_{i=0}^{N}\beta_i p_i(x_k)$$

VarBC introduced in operations at ECMWF

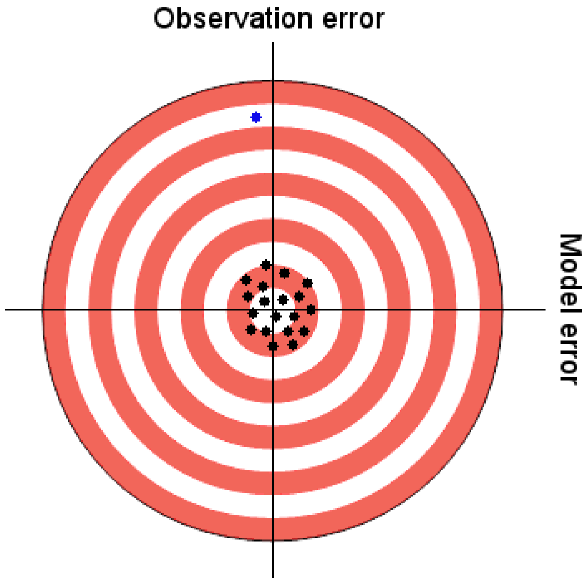
Introduction of VarBC in ECMWF operations





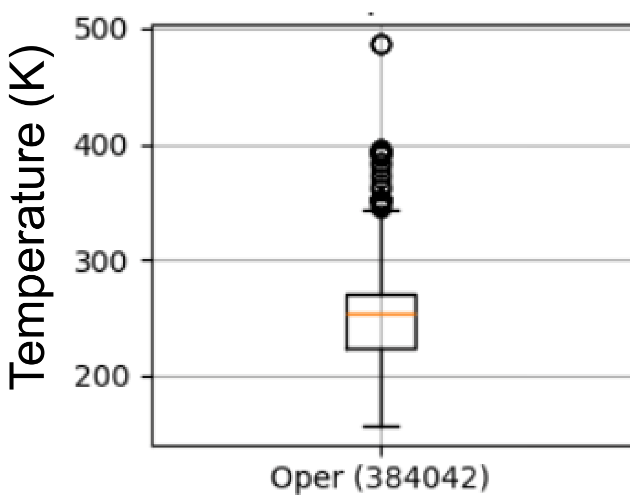


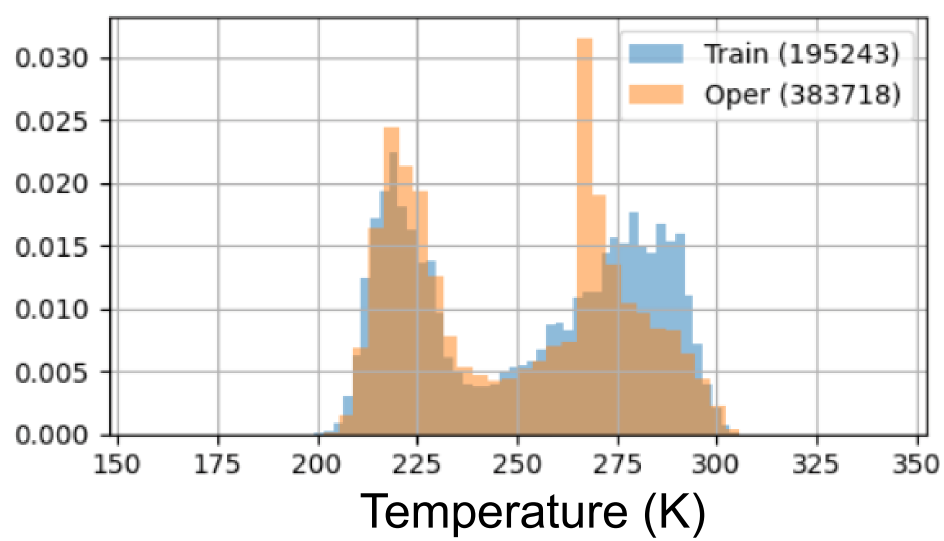
Not the job of VarBC



Gross (obvious) errors ➔ Quality control is required

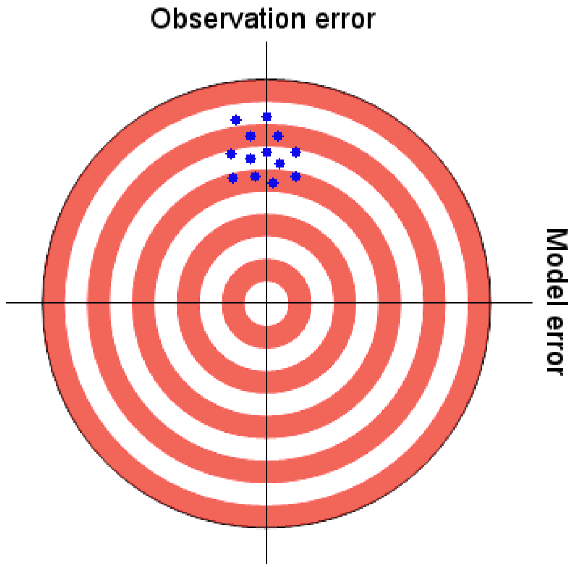
Temperature observations received last Saturday from aircrafts

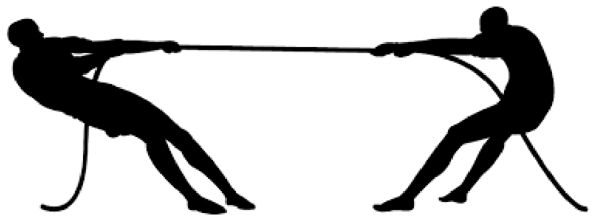




Take-away messages (1/3)

To question how much we should trust observations and models
To understand how 4D-Var needs to be modified to take biases into account
To identify the challenges of this approach





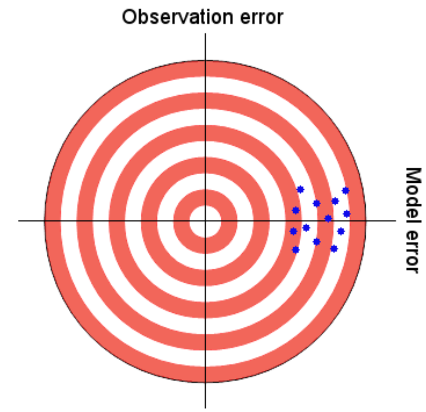
➔ we only have information about differences $\mathbf{y} - \mathbf{h}(\mathbf{x}_b)$

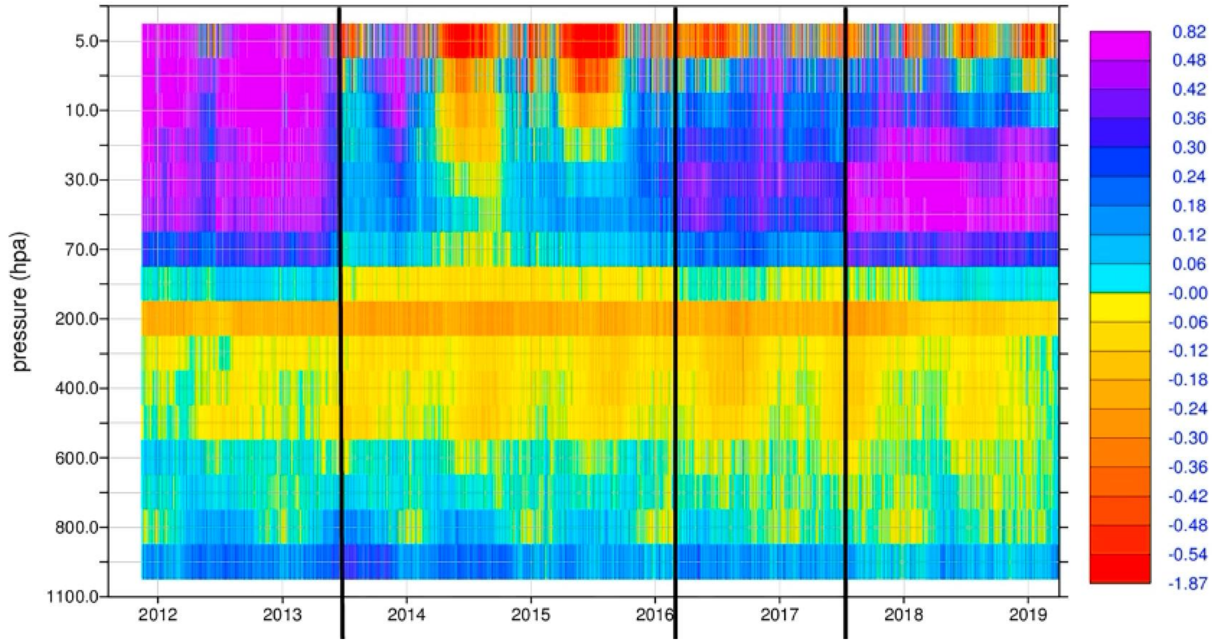
➔ there is no true reference in the real world!

➔ the success of VarBC relies on ***anchoring*** and ***redundancy***

How to estimate model biases (1/2)

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

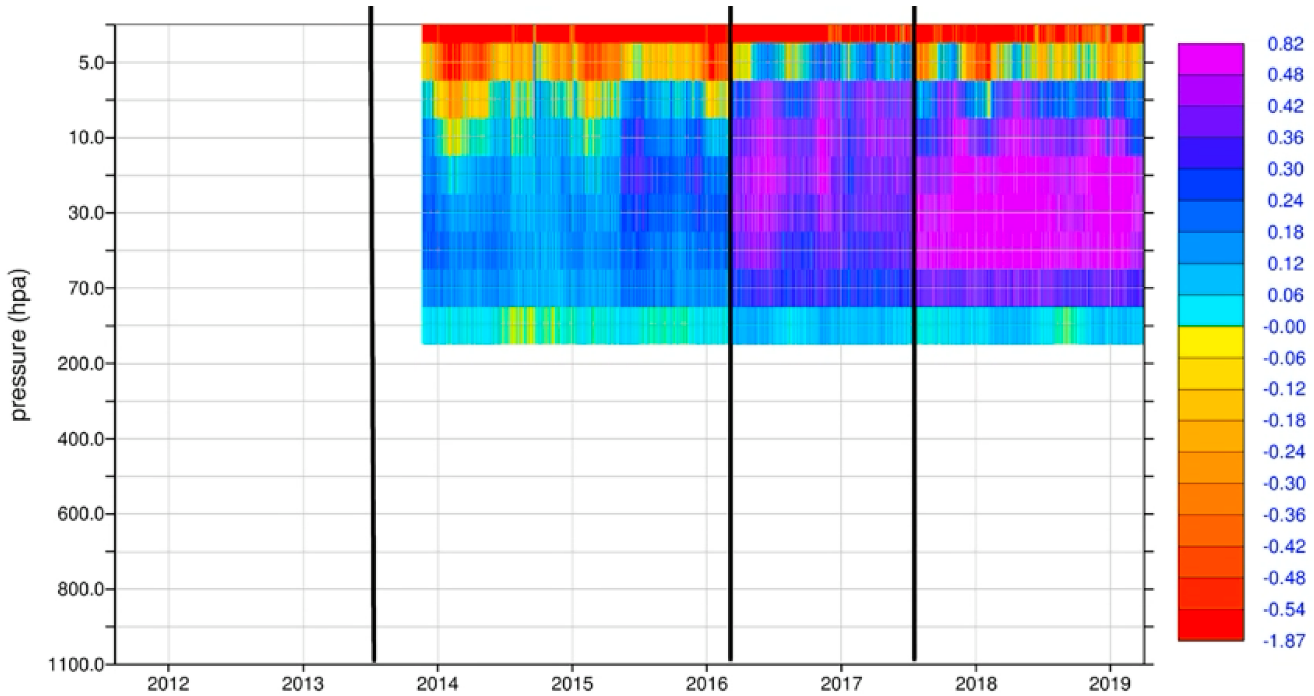
→ Largest bias in the stratosphere

→ 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 (2/2)

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)

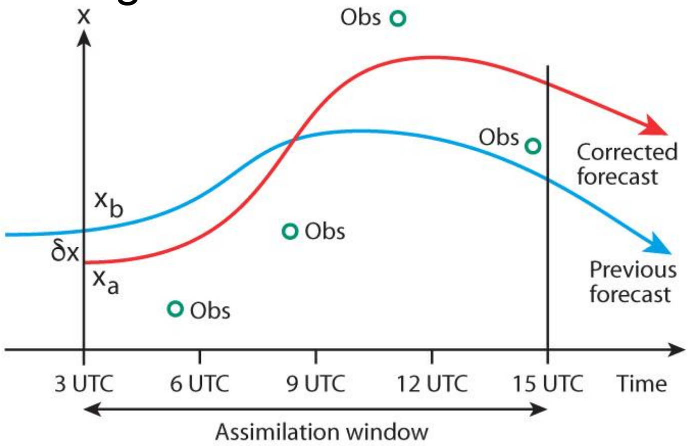
Errors in models are often systematic rather than random, zero-mean

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

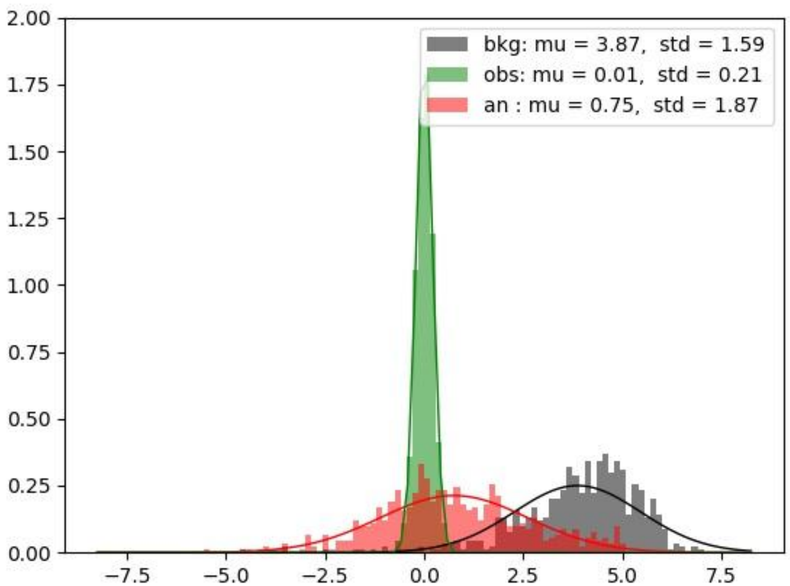
→ Model has a warm bias in the upper stratosphere

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

Strong constraint 4D-Var

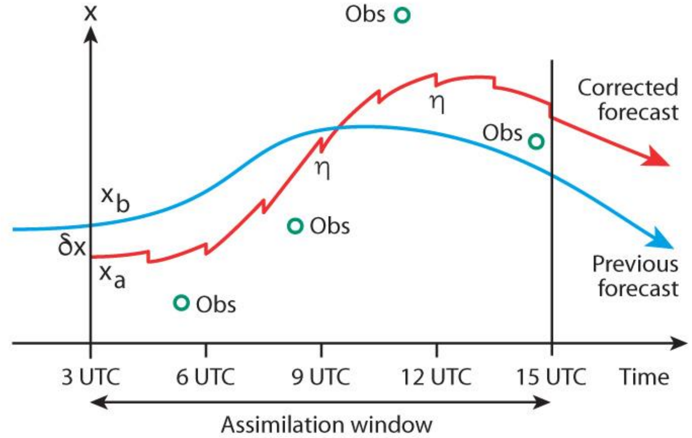


$$x_k = \mathcal{M}_k(x_{k-1})$$

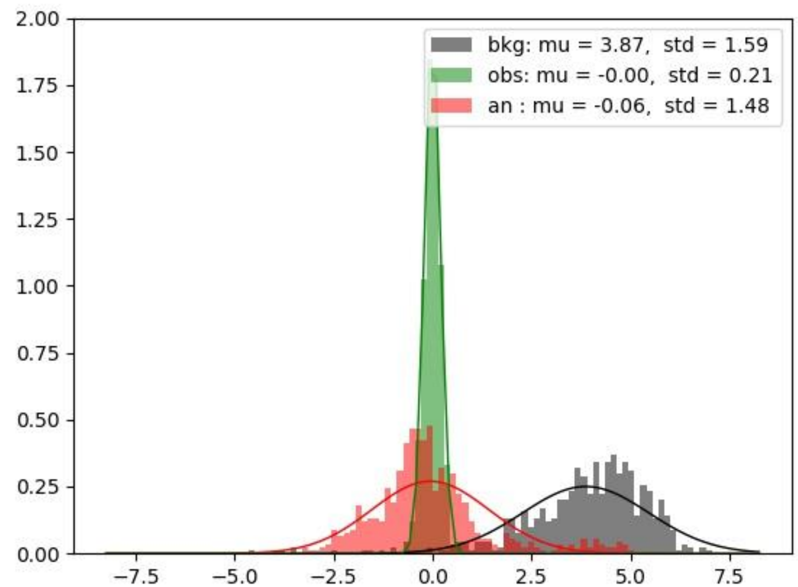


→ Large bias and standard deviation in the analysis

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

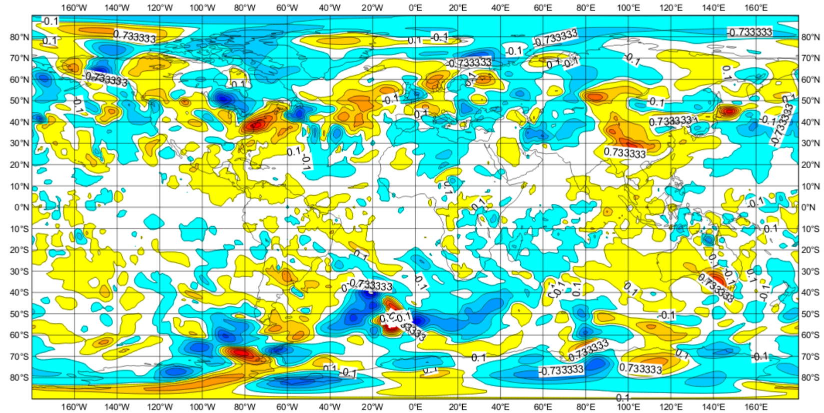
→ This looks very much like VarBC with a constant predictor, but in the model space!

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

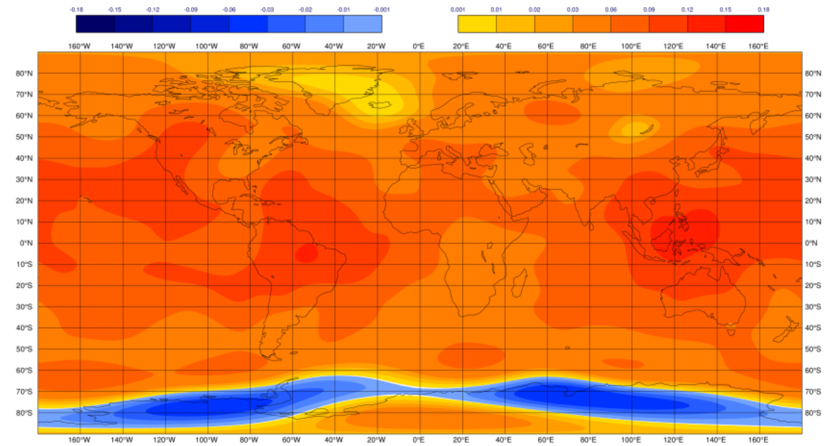
Model bias correction: weak-constraint 4D-Var (3/3)

How can we separate background and model error?

background correction (analysis increment)

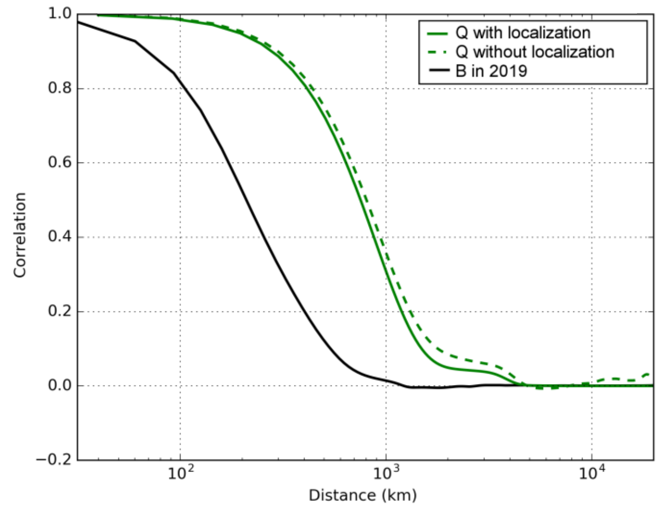


Model correction

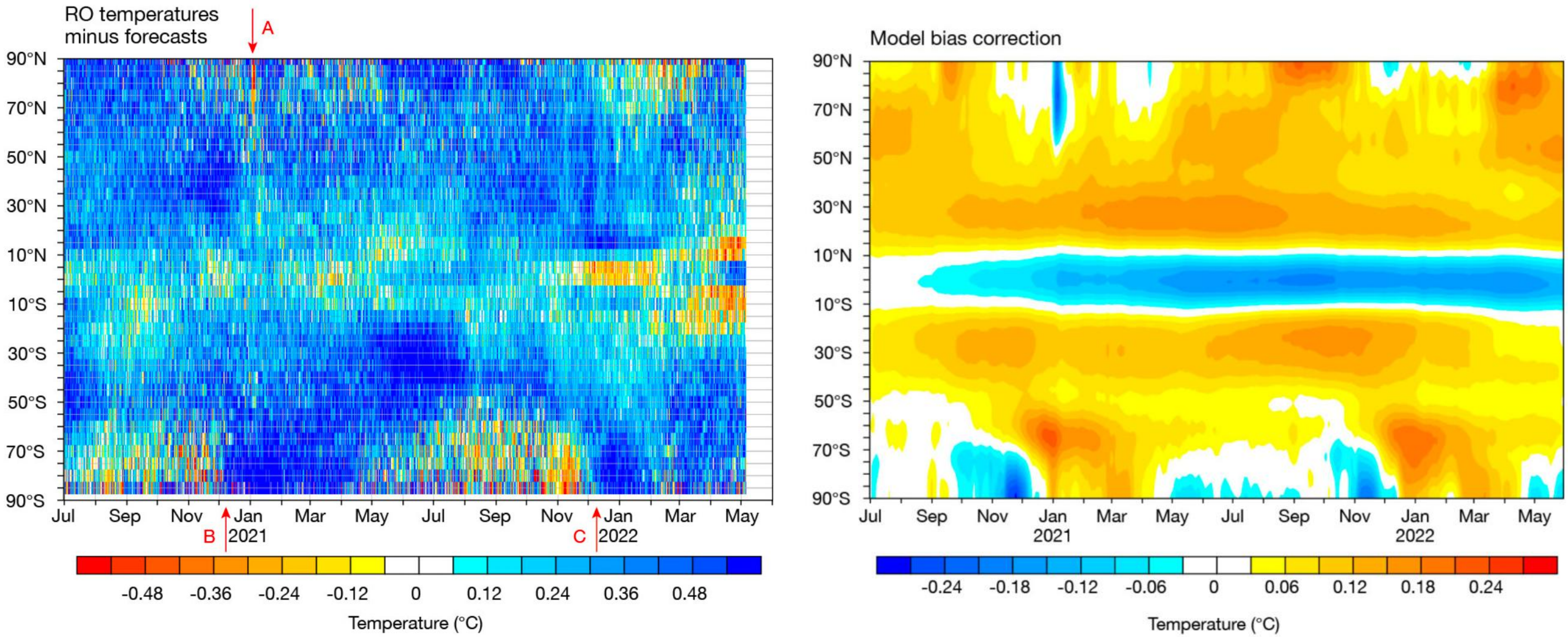


Background errors tend to be small scales while model errors tend to be large scale

Horizontal correlation



Weak-constraint 4D-Var in operations



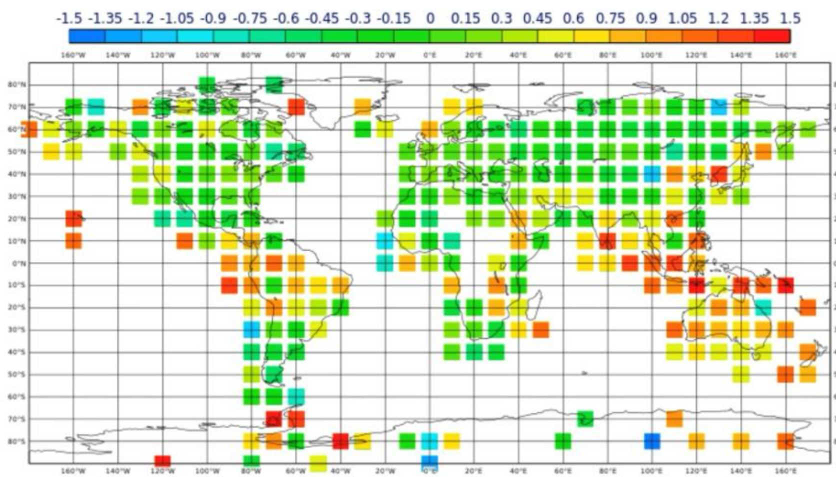
A) On 31 December 2020, a Sudden Stratospheric Warming (SSW) event started over the northern hemisphere

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

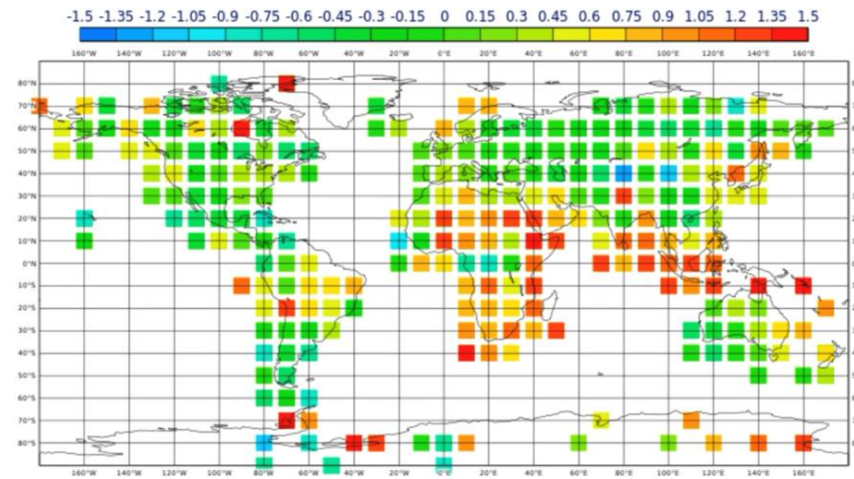
Model biases in the boundary layer (1/3)

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

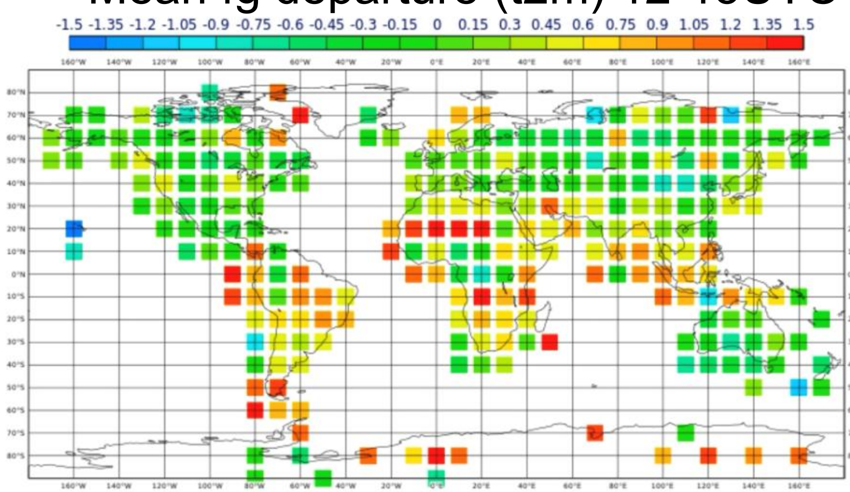
Mean fg departure (t2m) 00-03UTC



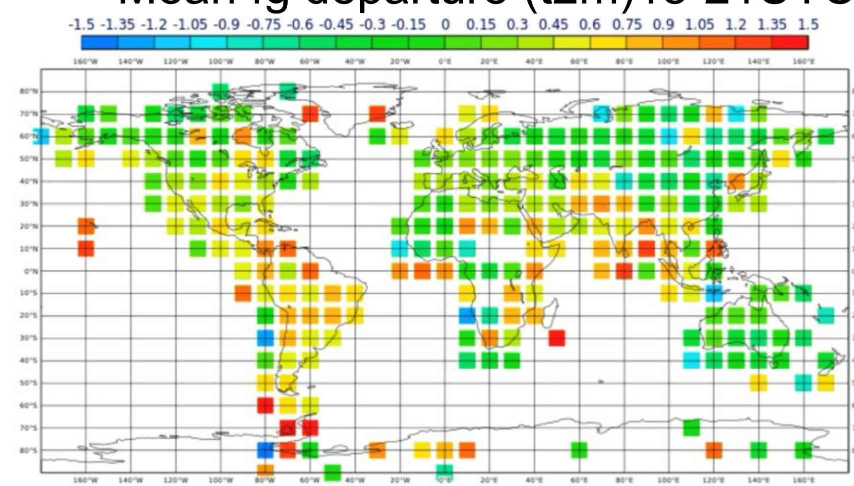
Mean fg departure (t2m) 06-09UTC



Mean fg departure (t2m) 12-15UTC

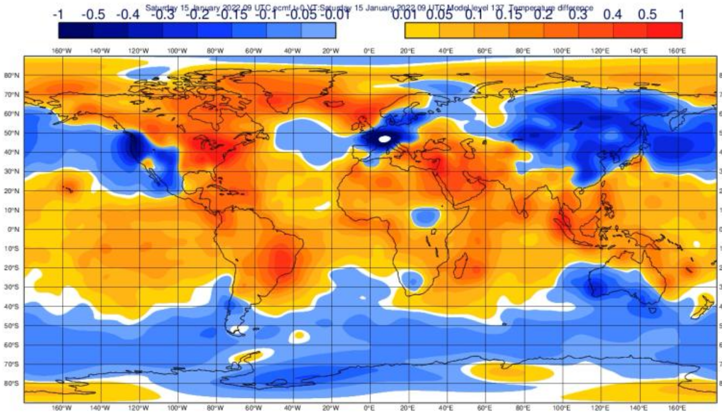


Mean fg departure (t2m)18-21UTC

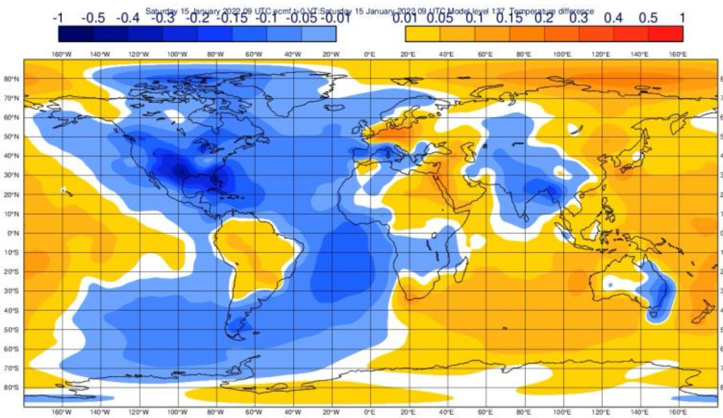


Model biases in the boundary layer (2/3)

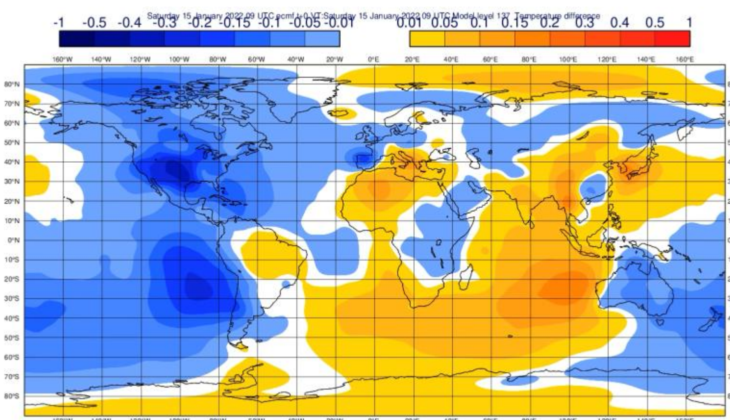
η_0



η_1



η_2



So far:

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

New representation of the model bias:

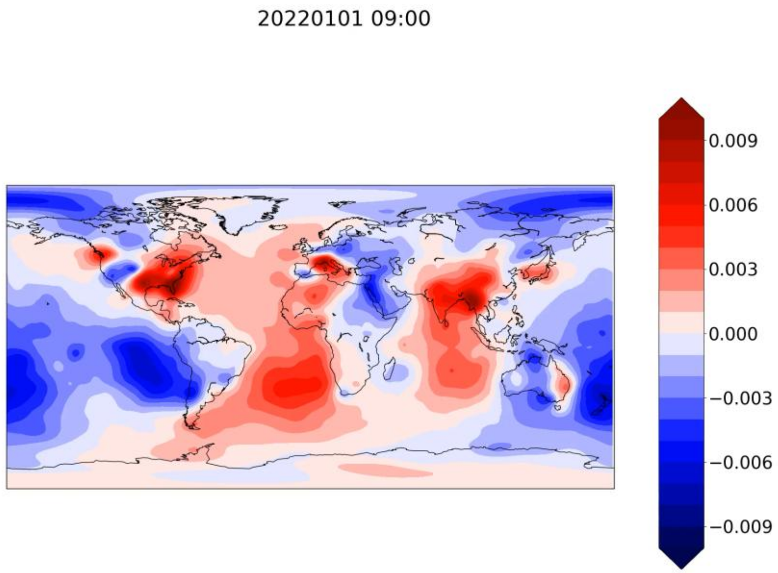
$$\eta_0 + \eta_1 \sin(2\pi\frac{t}{24}) + \eta_2 \cos(2\pi\frac{t}{24})$$

➔ Time-varying within the assimilation window

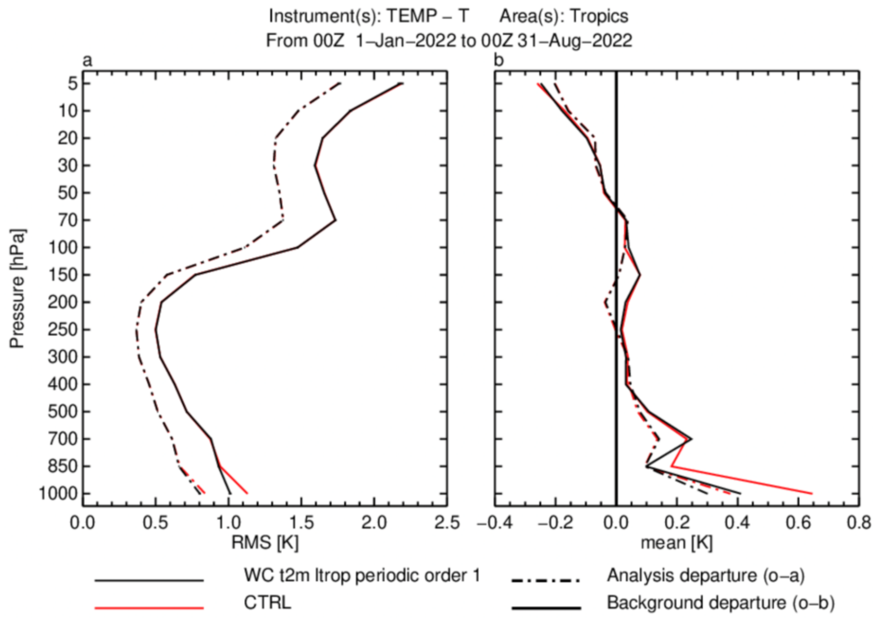
➔ Designed to capture a diurnal cycle

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 May 2026 (50R1)

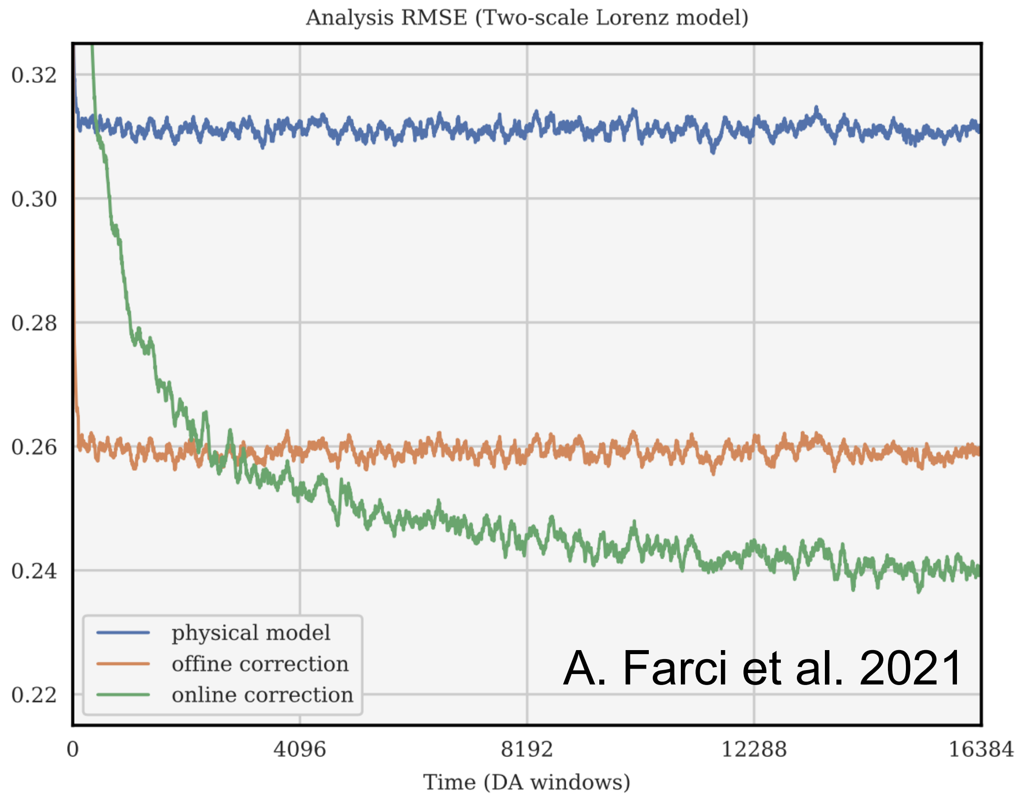
A ML approach to correct model biases

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

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

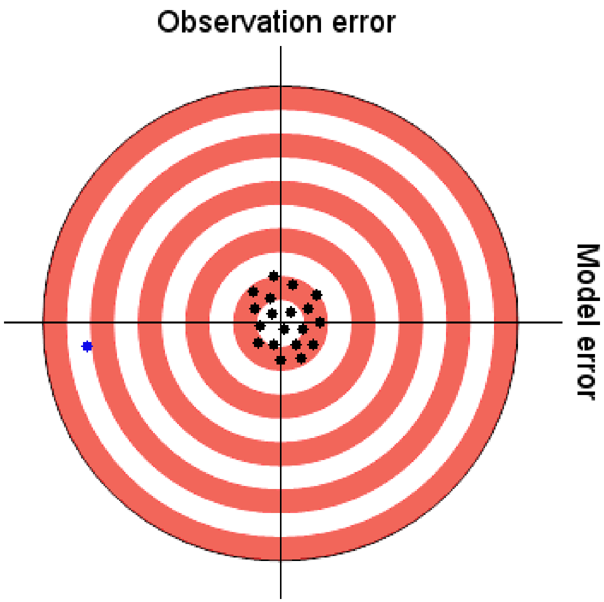
NN online loss function

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

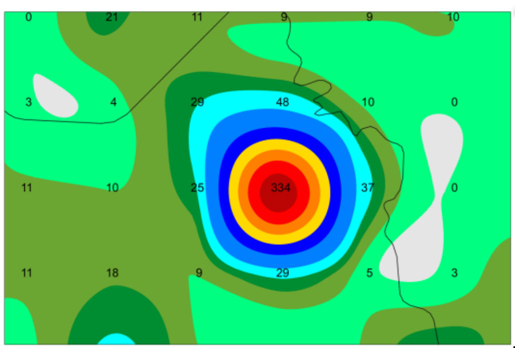


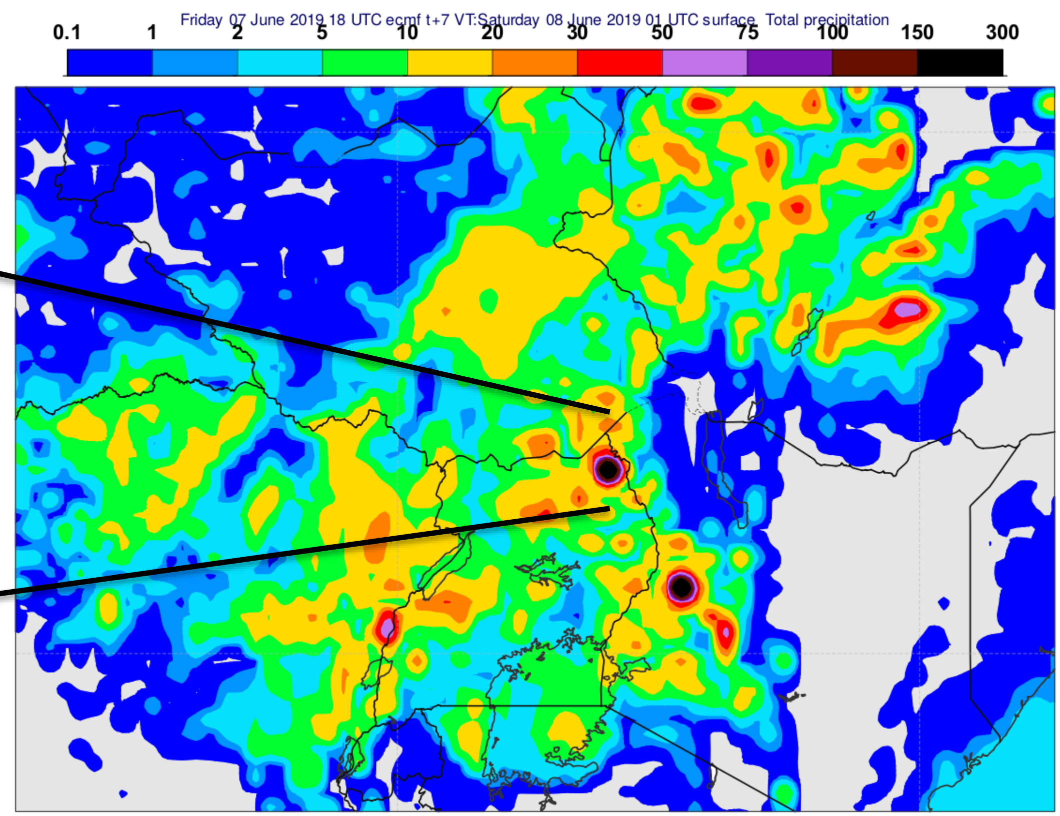
- ➔ learn both model state and NN parameters from observations
- ➔ the online correction steadily improves the model, learning from observations

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



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



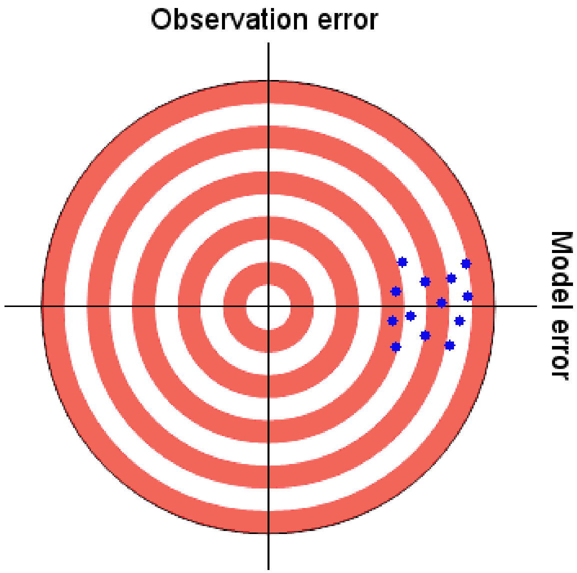


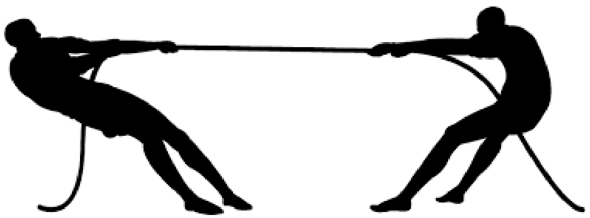
→ Continuous monitoring

→ Keep improving the model

Take-away messages (2/3)

To question how much we should trust observations and models
To understand how 4D-Var needs to be modified to take biases into account
To identify the challenges of this approach





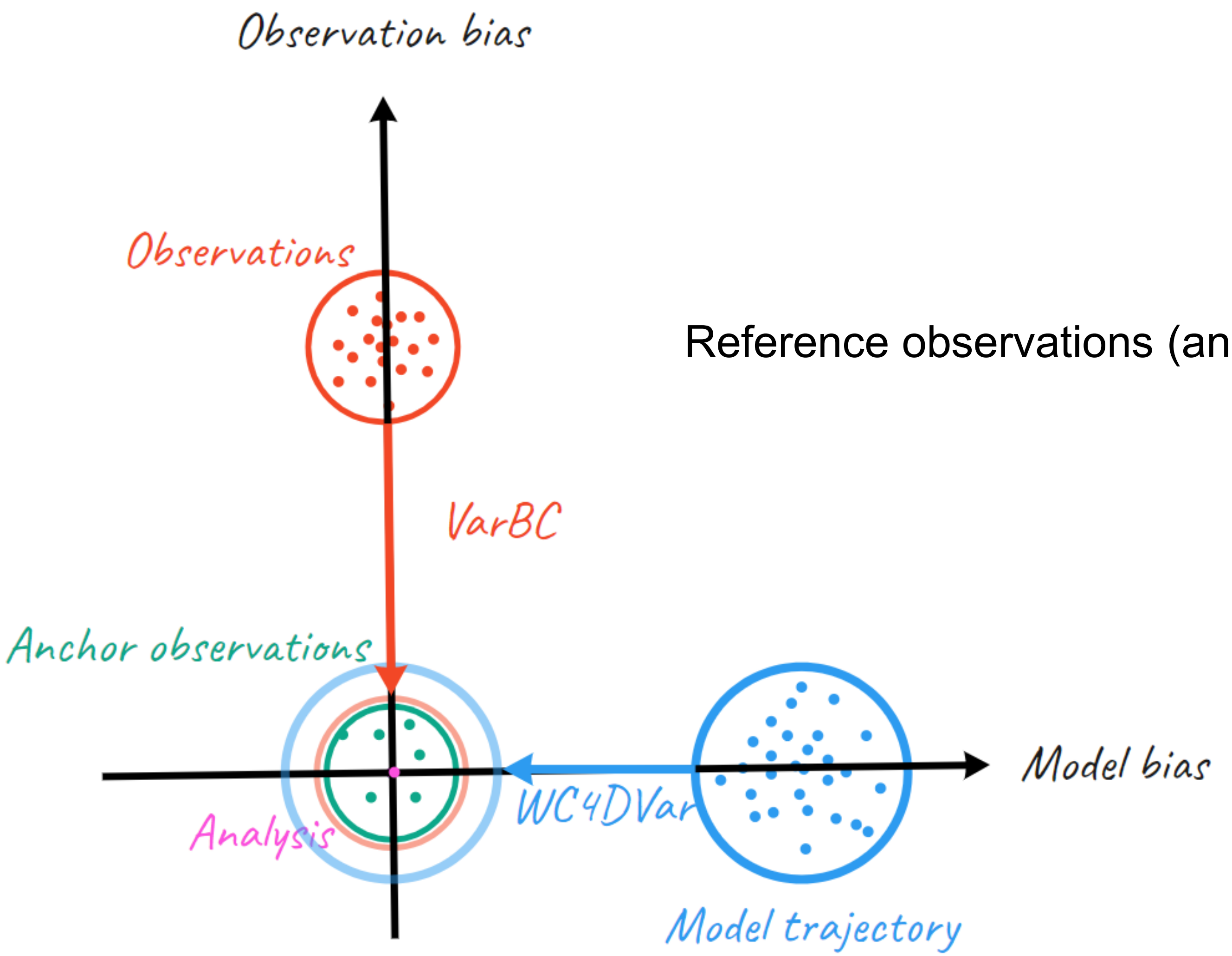
➔ we only have information about differences $\mathbf{y} - \mathbf{h}(\mathbf{x}_b)$

➔ there is no true reference in the real world!

➔ Disentangle observation and model biases is difficult and is still an open question

Take-away messages (3/3)

From bias-blind to bias-aware data assimilation



Reference observations (anchors) are critical!

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