Bias-aware Data Assimilation and Machine Learning

Massimo Bonavita and Patrick Laloyaux

Research Department, ECMWF

massimo.bonavita@ecmwf.int

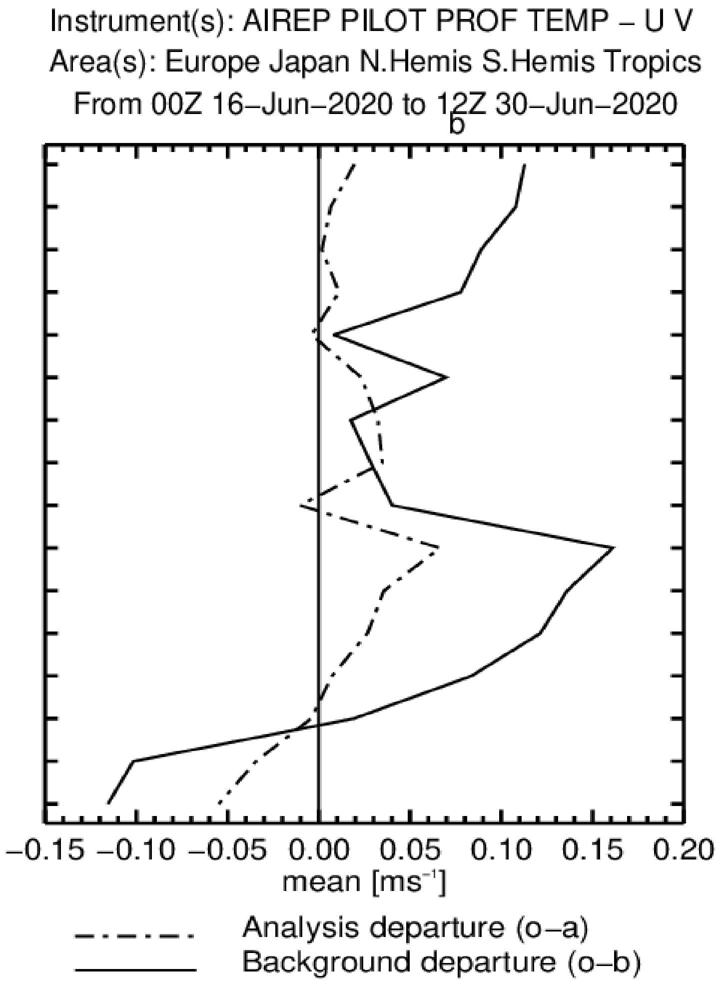
Outline

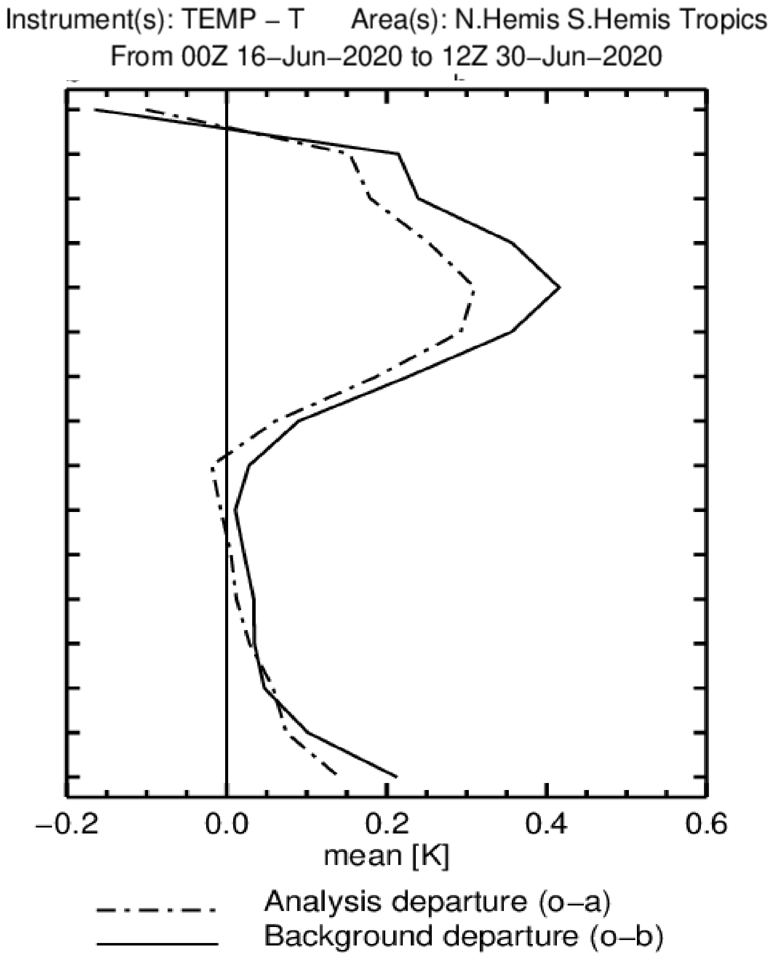
- Systematic errors in NWP and Climate
- Towards bias-aware Data Assimilation: VarBC and Weak Constraint 4D-Var
- What can Machine Learning bring to the table?
- The way forward

Notation and terminology follow closely Dee, 2005: “Bias and Data Assimilation”, where many of the good ideas in the field can be found.

Systematic errors

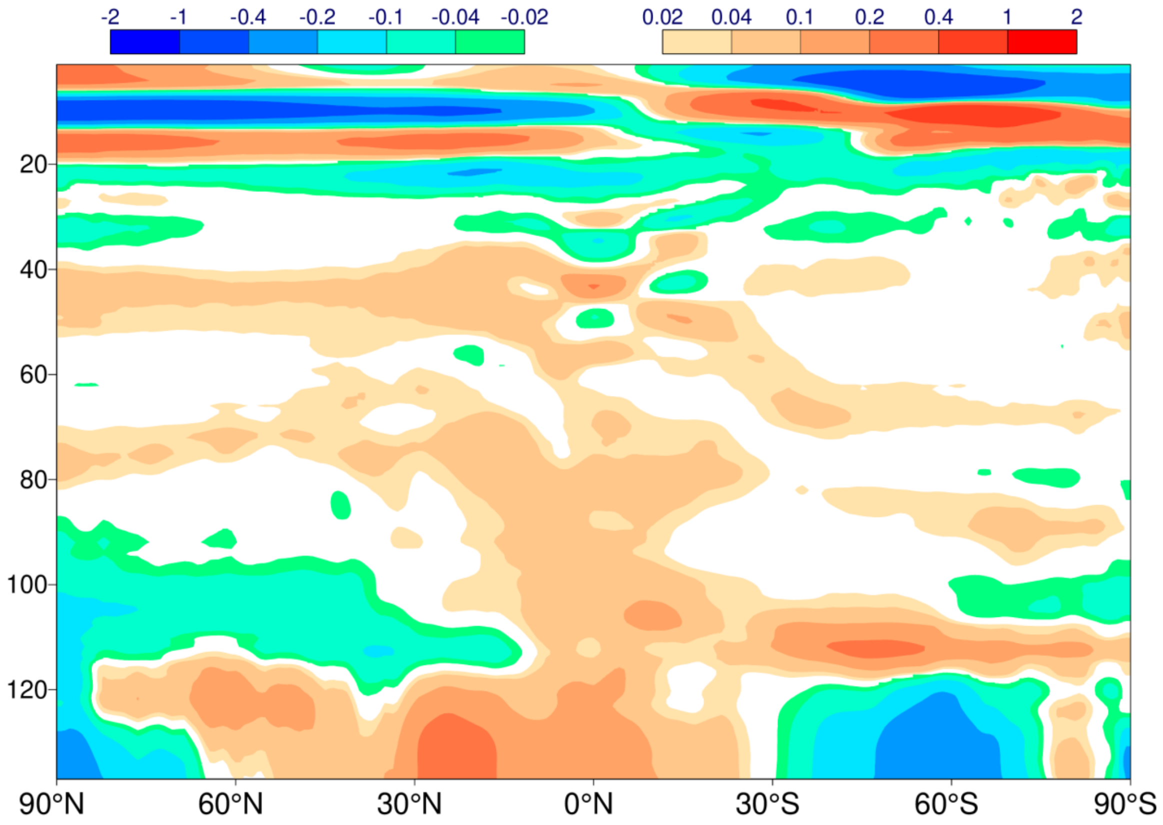
- Systematic errors in both model and observations are a fact of life
- They reveal themselves in innovation departures (OmB) averages:





Systematic errors

- Systematic errors in both model and observations are a fact of life
- They reveal themselves in analysis increments (AmB) mean statistics:



Average Temperature Analysis Increment of the Operational IFS – June 2020

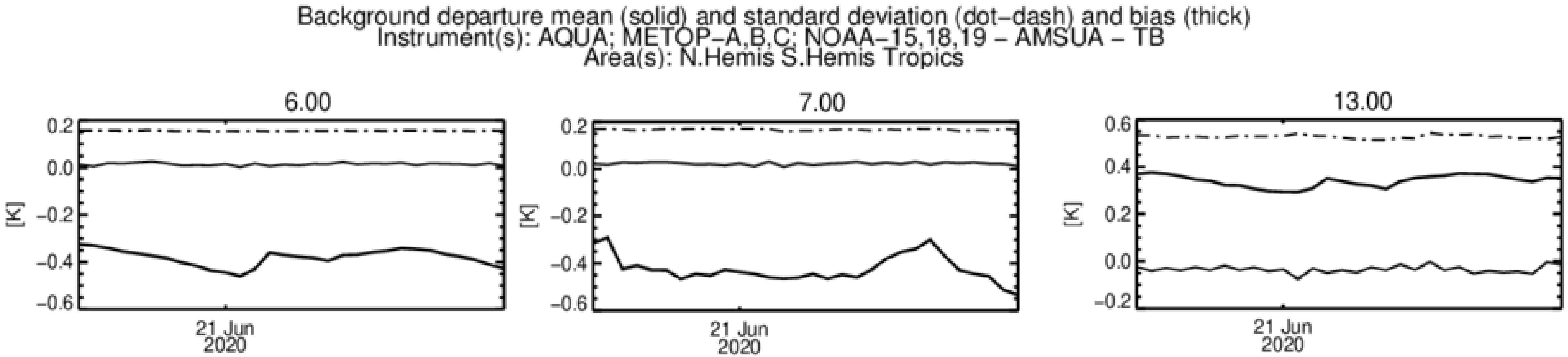
Towards bias-aware Data Assimilation

- Standard "bias-blind" Data Assimilation ignores systematic model and observation errors:

$$J_{4DVar}(\boldsymbol{x}_0) = J_B + J_O = \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^T \mathbf{B}^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) + \frac{1}{2}\sum_{i=0}^{N}(H_i(\boldsymbol{x}_i) - y_i)^T \mathbf{R}_i^{-1}(H_i(\boldsymbol{x}_i) - y_i)\,;$$

$$\boldsymbol{x}_i = M(\boldsymbol{x}_{i-1})$$

- However this is not really an option in the stratosphere (significant systematic model errors) and in general for satellite obs (O-B) innovations:



Towards bias-aware Data Assimilation

- The first step towards a bias-aware DA system is to debias innovations (O-B). This can be done sequentially (debiasing against bg) or inside the Var minimisation, (debiasing against an, Var-BC):

$$J_{4DVar}(x_0, \beta) = J_B + J_O + J_\beta = \frac{1}{2}\left(x_0 - x_0^b\right)^T \mathbf{B}^{-1}\left(x_0 - x_0^b\right) + \frac{1}{2}\sum_{i=0}^{N}\left(H_i(x_i) - \left(y_i - b(x_i, \beta)\right)\right)^T \mathbf{R}_i^{-1}\left(H_i(x_i) - \left(y_i - b(x_i, \beta)\right)\right) + \boxed{\frac{1}{2}\left(\beta - \beta^b\right)^T \mathbf{B}_{\boldsymbol{\beta}}^{-1}\left(\beta - \beta^b\right)};$$

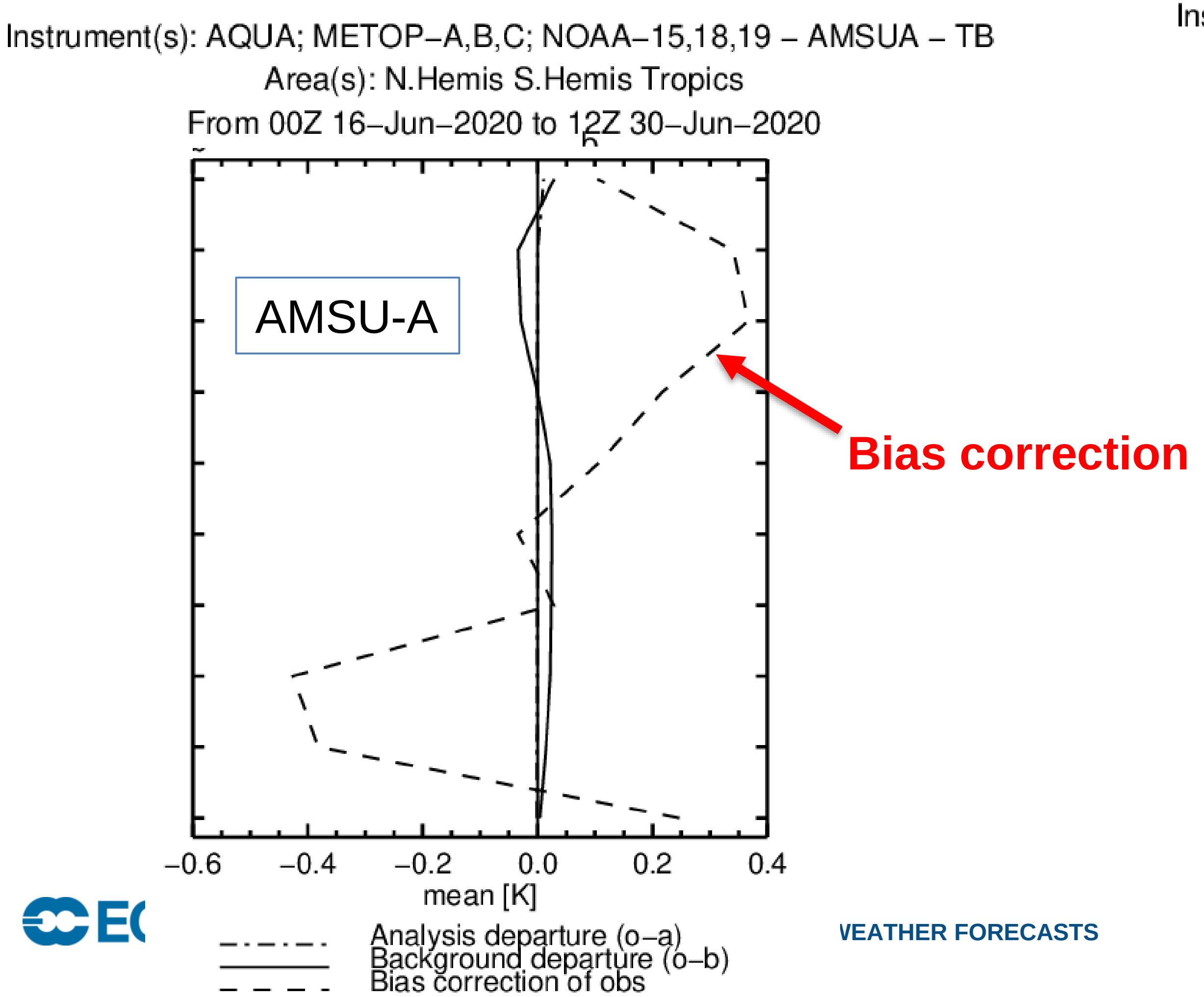
$$x_i = M(x_{i-1})$$

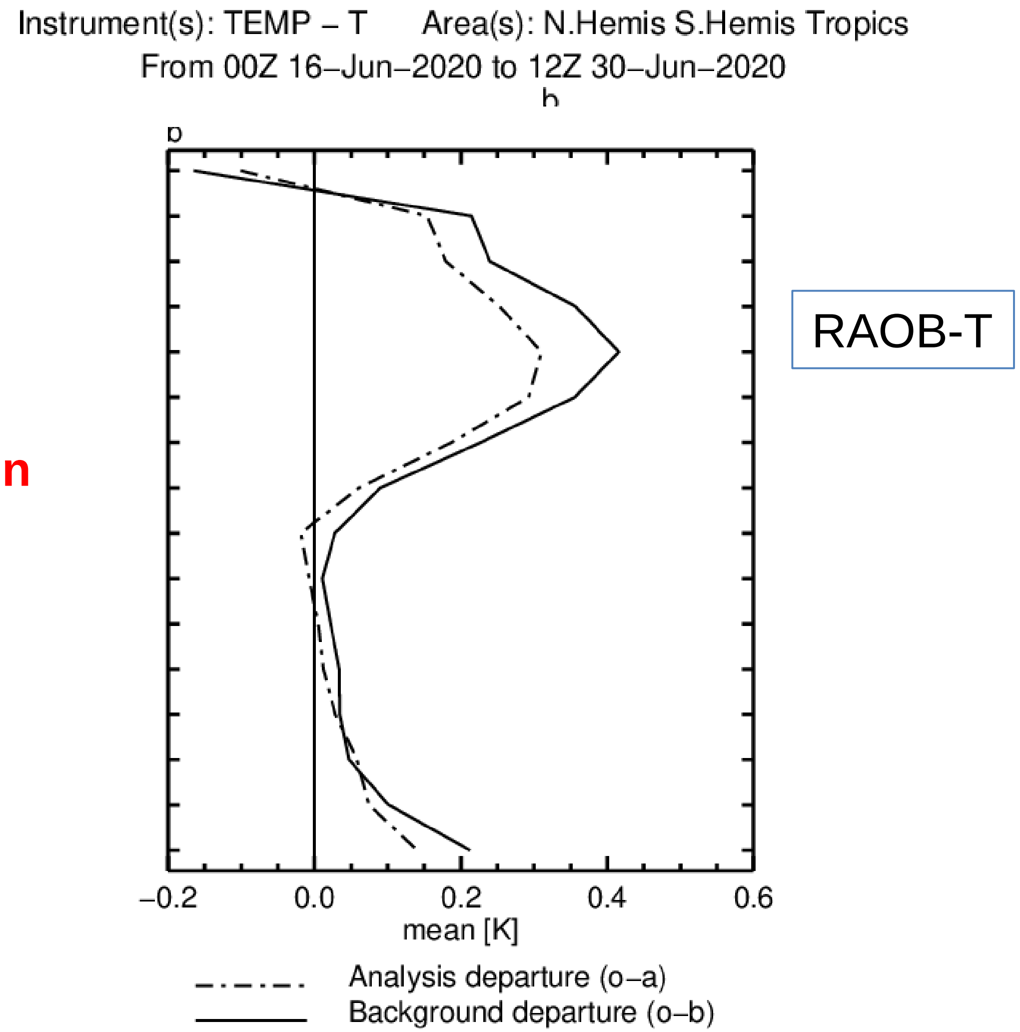
- This is a significant step forward, but it is not the end of the story:

  1. There is an implicit assumption that most of the systematic errors reside with the observations; this leads to discarding useful observational signal
  2. Need to employ “airmass” type of bias predictors points to underlying errors in the models
  3. Model systematic errors are significant: just debiasing the innovations is a post-hoc, sub-optimal solution

Towards bias-aware Data Assimilation

- VarBCis very effective in de-biasing sat. obs, but the model systematic biases are still there: model bias cannot be corrected in obs. space





Towards bias-aware Data Assimilation

- *"Nevertheless, it would be preferable to estimate tendency errors that lead to the bias in the background fields, if this could be used to suppress bias generation during the integration of the model"*. (Dee, 2005)

- This is what weak constraint 4D-Var aims to do:

$$J_{WC4DVar}(x_0, \beta, \eta) = J_B + J_O + J_\beta + J_Q = \frac{1}{2}\left(x_0 - x_0^b\right)^T \mathbf{B}^{-1}\left(x_0 - x_0^b\right) +$$

$$\frac{1}{2}\sum_{i=0}^{N}\left(H_i(x_i) - \left(y_i - b(x_i, \beta)\right)\right)^T \mathbf{R}_i^{-1}\left(H_i(x_i) - \left(y_i - b(x_i, \beta)\right)\right) + \frac{1}{2}\left(\beta - \beta^b\right)^T \mathbf{B}_\beta^{-1}\left(\beta - \beta^b\right) +$$

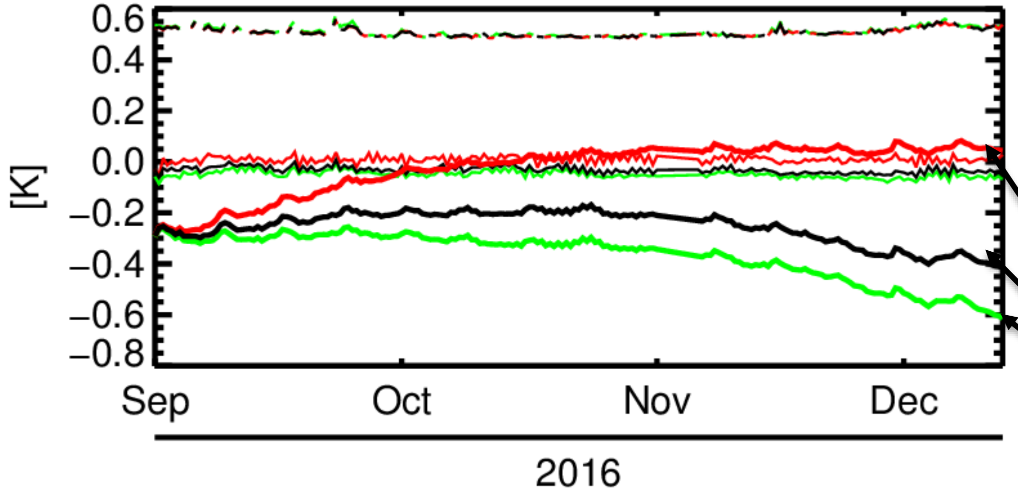
$$\boxed{\frac{1}{2}\left(\eta - \eta^b\right)^T \mathbf{Q}^{-1}\left(\eta - \eta^b\right)},$$

$$x_i = M(x_{i-1}) + \eta$$

Towards bias-aware Data Assimilation

- The new (July 2020, IFS Cycle 47R1) implementation of weak constraint 4DVar is a step change with respect to the old implementation:

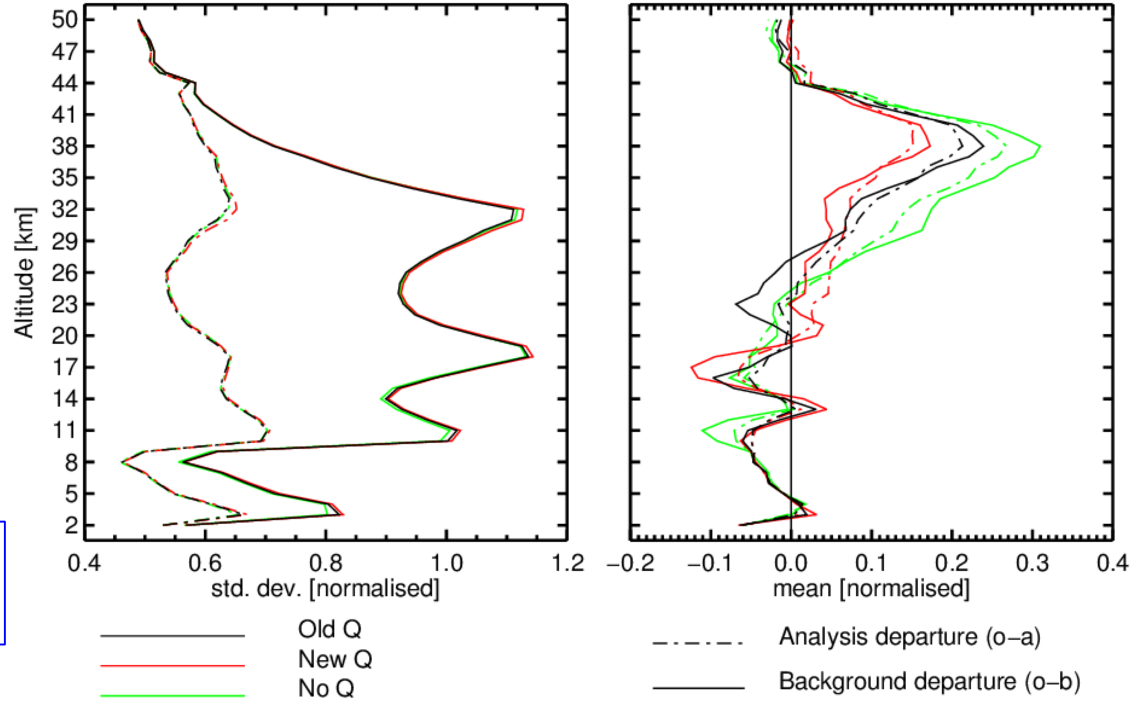
Timeseries of AMSUA ch13
fg departures (fg_departure = obs – fg -VarBC)



VarBC bias correction

------ New WC-4DVar
------ Old WC-4DVar
------ SC-4DVar

Analysis and FG departures - GPS-RO



From P. Laloyaux

Towards bias-aware Data Assimilation

- The new (July 2020, IFS Cycle 47R1) implementation of weak constraint 4DVar is a step change with respect to the old implementation

- The insight there (Laloyaux et al., 2020a,b) was to impose scale separation between the state and the model error corrections (Lscale($\mathbf{B}$)~100km, Lscale($\mathbf{Q}$)~1000km):

  1. Without scale-separation, state and model error corrections alias into one another: There is no mechanism in standard WC-4DVar to separate the two increments (e.g. model error corrections concentrated around airport locations in previous WC-4DVar implementation, Fisher et al., 2011);

  2. We do not have an effective model for the evolution of model systematic errors, thus we use persistence: we target slowly-evolving errors over the timescale of the assimilation window (12 hours)

Towards bias-aware Data Assimilation

- Current version of WC-4DVar is a step change wrto previous versions, but it is not the end of the story (so far):
  - WC-4DVar reduces stratospheric fcst temperature bias by 30-40% against non-bias corrected obs (Radiosondes, GPS-RO): there is a significant residual error
  - Current WC-4DVar has little impact on wind systematic errors
  - Current WC-4DVar is only active in the stratosphere (above ~100hPa). Letting it work on full atmospheric column leads to marginal forecast skill degradation in troposphere (interestingly, it leads to further improvements in the stratosphere)

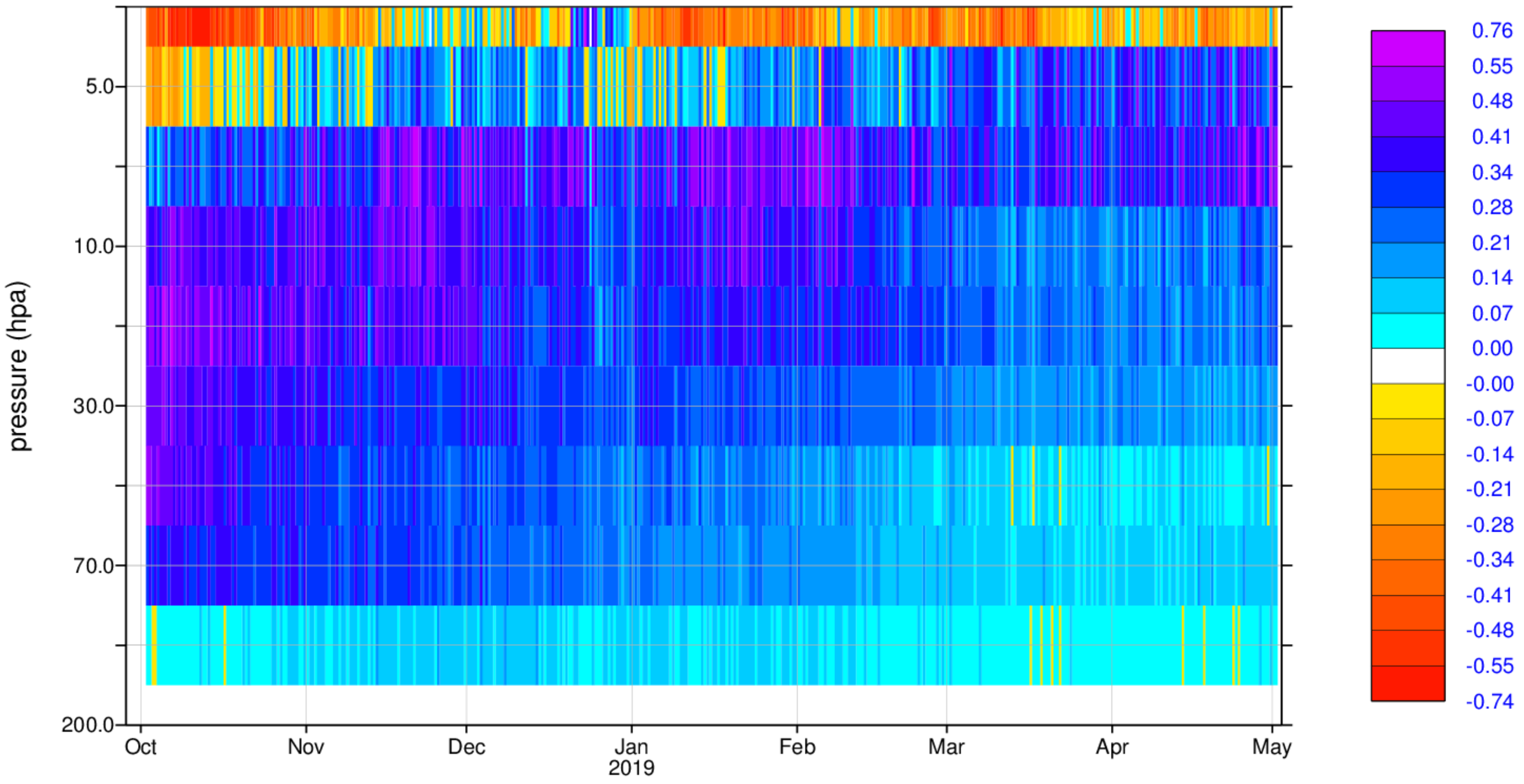
- Can we bring Machine Learning into play?

Towards bias-aware Data Assimilation

- WC-4DVar gradually learns a model error tendency correction and applies it during the assimilation cycle

- WC-4DVar is **an online machine learning** algorithm for model error estimation and correction

Mean first-guess departure with respect to GPS-RO temperature retrievals

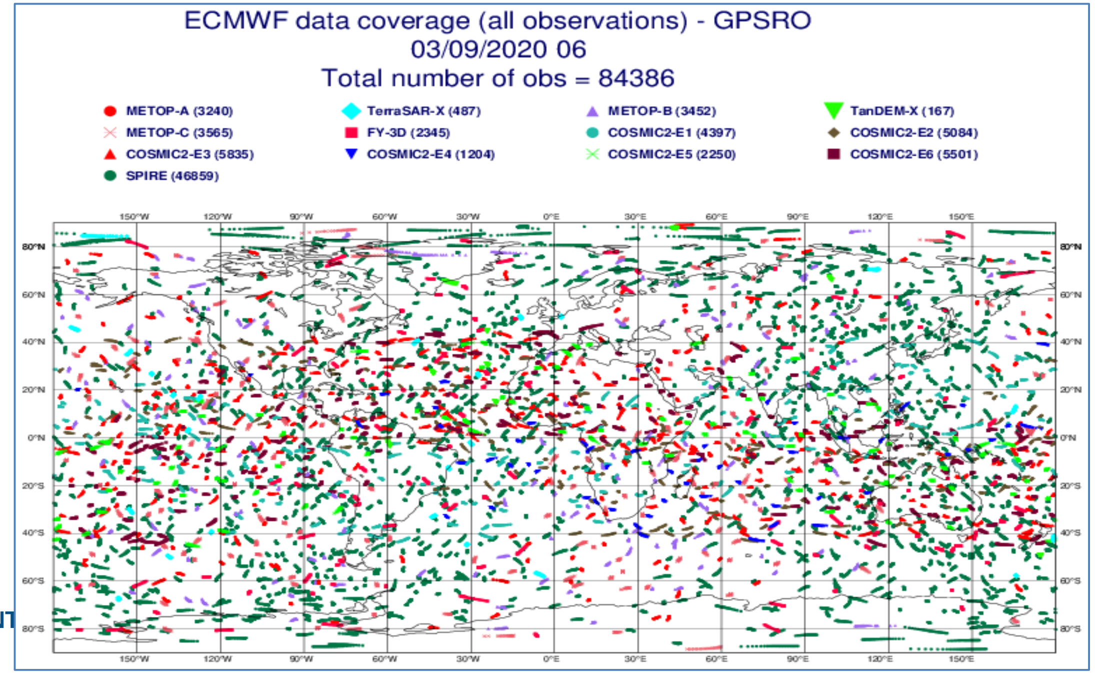


From Laloyaux et al, 2020b

Machine Learning for bias-aware Data Assimilation

What can Machine Learning bring to the problem of model error identification and correction?

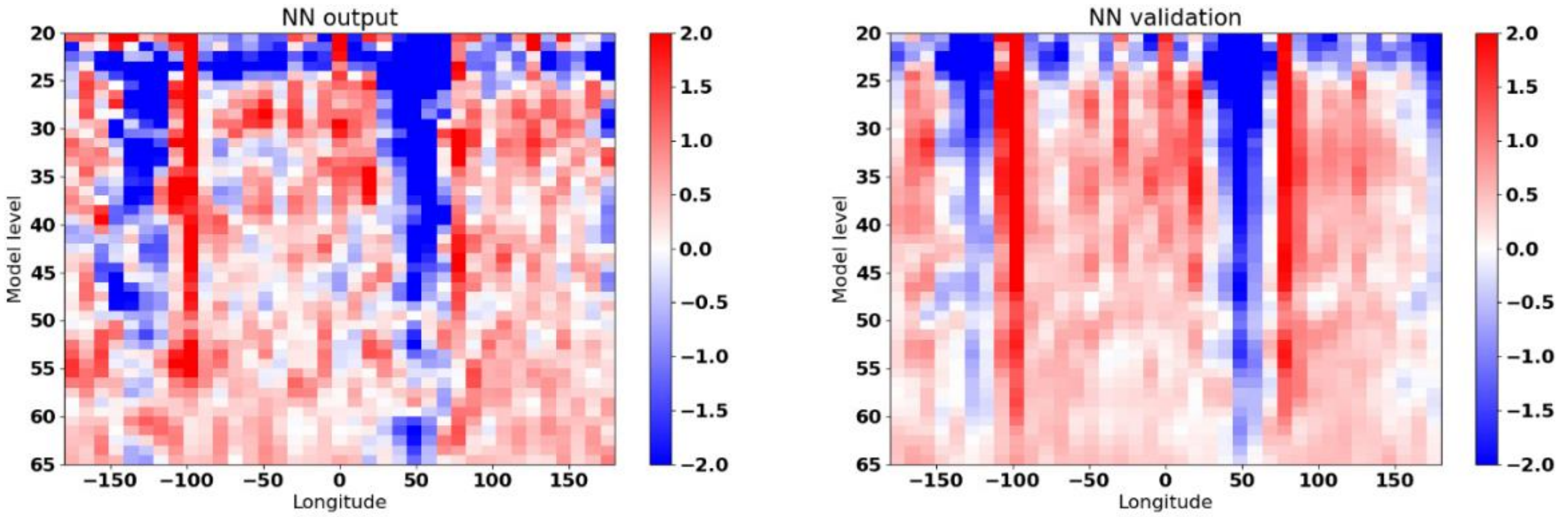
- Basic idea: Train an Artificial Neural Network (ANN) to “learn” the statistically predictable part of model error
- The training of the ANN can be realised in at least two ways
- Most direct approach is to train the ANN on a database of OmB departures, choosing an approx. homogeneous, dense and unbiased observing system, eg GPS-RO:



Machine Learning for bias-aware Data Assimilation

- Pluses: Effectively by-passing the assimilation cycle, which avoids the systematic errors present in the analyses
- Minuses: Lack of homogeneous coverage, need of long timeseries (=> stable model system, eg re-analysis) , how to extrapolate the corrections to unobserved locations and unobserved state variables (ie, how to make an ANN produce an analysis?)

OmB on 1-5-2019 averaged between 10S and 20N (left) and the prediction of the Convolutional Neural Network (right)



From P. Laloyaux and T. Kurth

Machine Learning for bias-aware Data Assimilation

- The other possibility is to train the ML model on a database (fixed or rolling) of analysis increments, again under the assumption that the learnable part of the increments is due to systematic model deficiencies

- This idea is not new in Data Assimilation, e.g. Dee, 2005, proposed an online version of this idea:

"*In the presence of bias, therefore, certain components of the increments are systematic and therefore predictable. … Provided the predictable part of the increment can be attributed to model errors, the algorithm*

$$\mathbf{dx}_k^p = \mathbf{f}_k(\mathbf{dx}_{k-L}, \dots, \mathbf{dx}_{k-1}) \qquad (43)$$

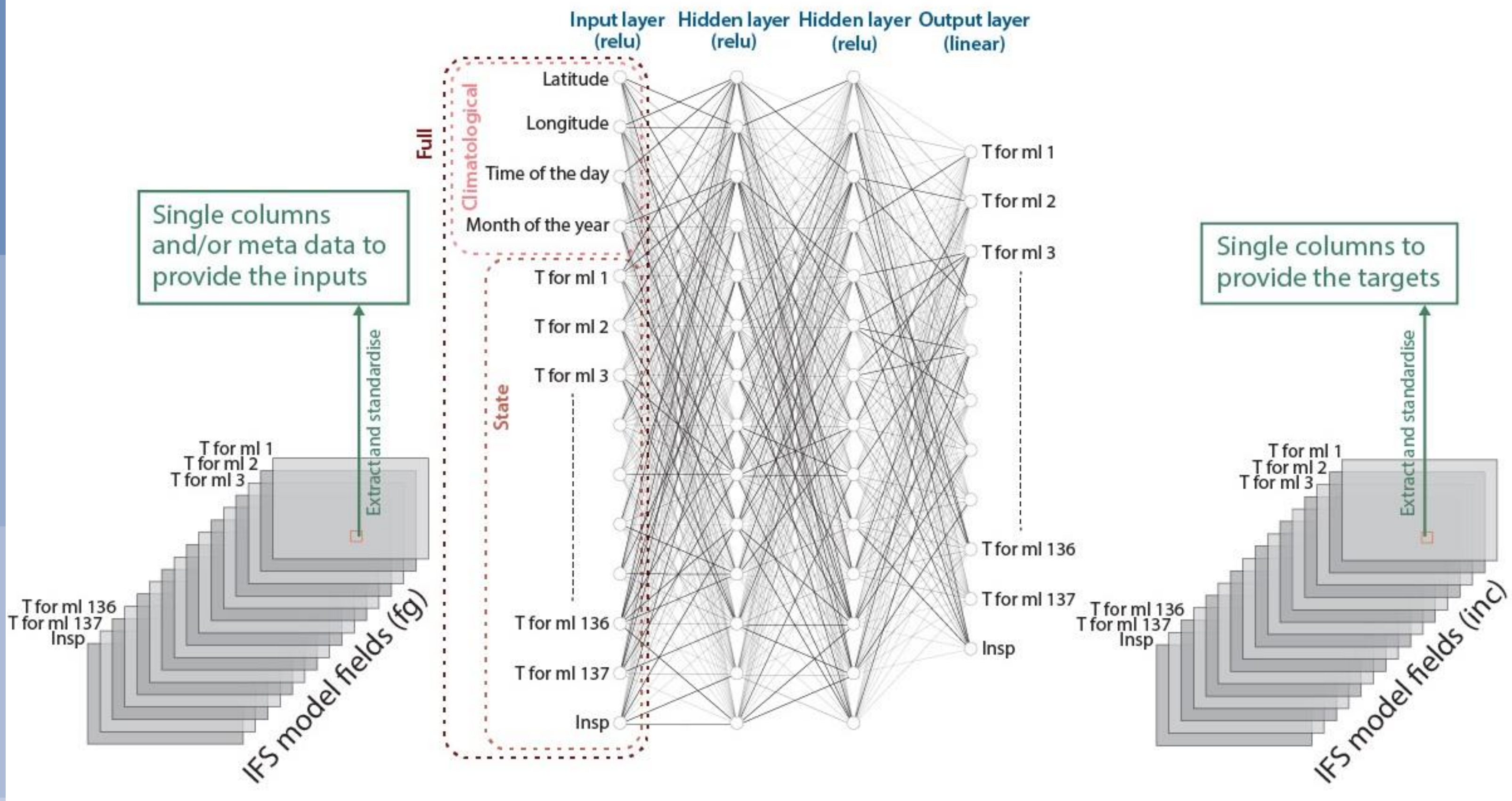
$$\mathbf{dx}_k = \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{h}\left(\mathbf{x}_k^b - \mathbf{dx}_k^p\right)\right) \qquad (44)$$

will correct the model background and produce unbiased analyses."

Machine Learning for bias-aware Data Assimilation

- We are testing this approach by training an ANN to learn the analysis increments of the operational version of the ECMWF IFS (more results in Bonavita and Laloyaux, 2020)
- The size of the state vector is $\mathcal{O}(10^{10})$. Even assuming a much lower effective dimension for the model error vector, a primary design consideration for the ANN has been to reduce its size
- This led us to define a set of predictors made up of the concatenation of climatological predictors (time of day, month, lat, lon) and the vertical columns (137 levels) of the model first guess msin prognostic variables of the model (t, lnsp, vo, div, q).
- This choice to split the 3D regression problem into a 1D x 2D problem is similar to having a separable representation of a 3D covariance matrix and can be justified by two considerations:
    1. *We can consider the atmospheric flow to be subject to homogeneous dynamics and heterogeneous forcings*;
    2. *Physical parameterisations are computed and applied over model columns*.

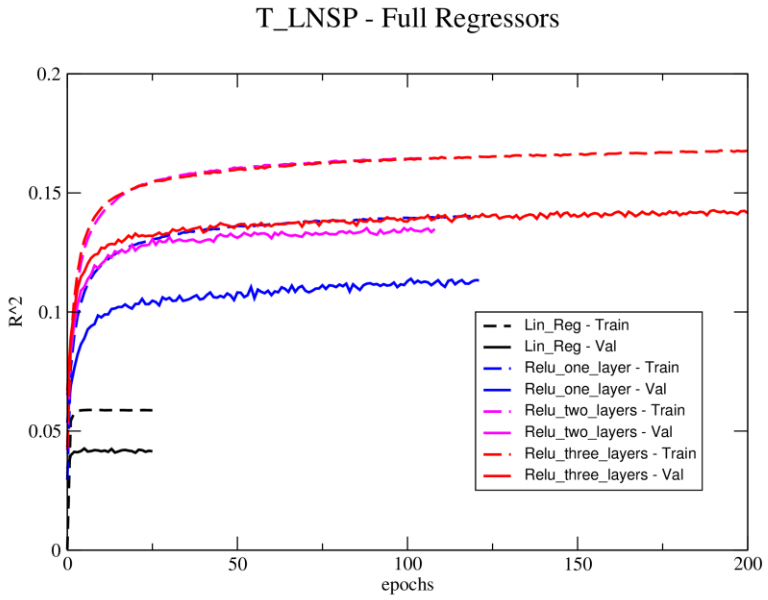
Machine Learning for bias-aware Data Assimilation

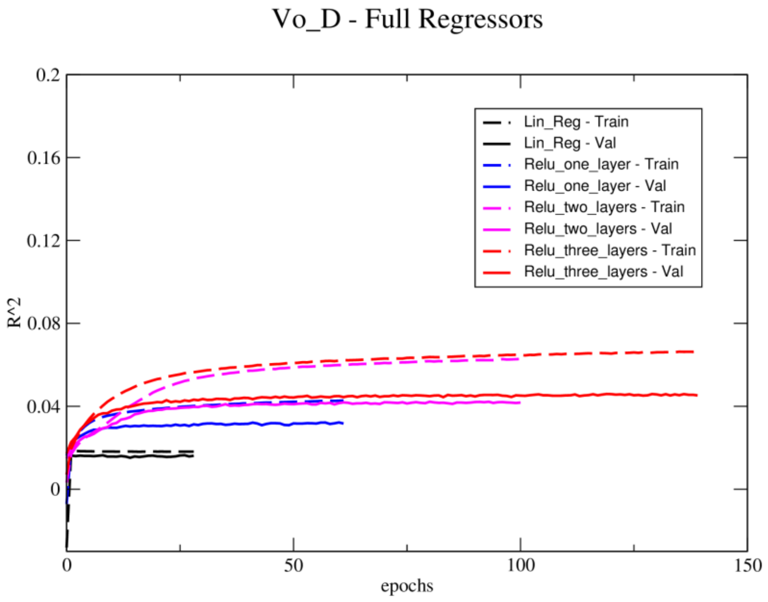


- Dense Neural Network with Relu activation
- Three layers with nonlinear activations give best results: problem with only moderate nonlinearities
- Dropout layers used to control overfitting, input/outputs pre-normalised for training, Adam minimiser
- Number of trainable parameters ~$6*10^4$, size of training dataset ~10^6

From Bonavita & Laloyaux, 2020

Machine Learning for bias-aware Data Assimilation





- Training/Testing curves are shown in terms of explained variance (R^2)
- Saturation of explained variance is used as stopping criterion in the training
- Mass (T, lnsp) errors can be better predicted (~14-15% explained variance) than wind (~4-5% explained variance) and humidity (~0%) errors.
- State-dependent predictors (first guess values) have more predictive power than climatological predictors: in forecast mode it is important to have an online model error correction.

From Bonavita & Laloyaux, 2020

ML for bias-aware Data Assimilation: Experiments

- So what happens when we use this trained Neural Network to correct model errors inside the data assimilation system of the operational IFS?

- The initial investigations focussed on two different configurations:

  a) Use the NN model errors instead of the Weak Constraint 4D-Var model errors (NN_SC in the following);

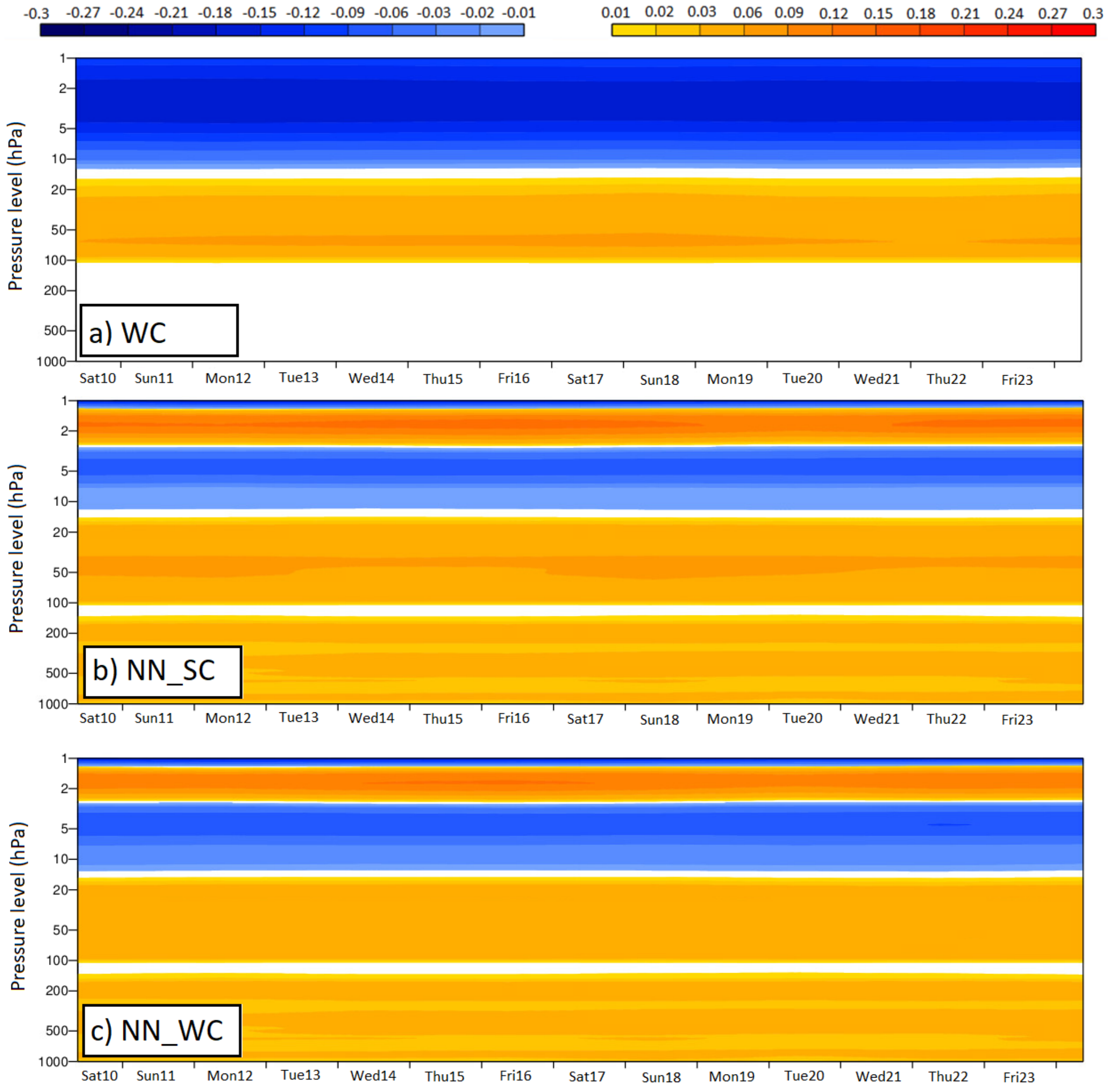
  b) Use the NN model errors as a first-guess for the Weak Constraint 4DVar (NN_WC)

- Baseline configuration was the currently operational version of Weak Constraint 4D-Var at full resolution (T1279, ~9km grid spacing) (WC) and the previously operational Strong Constraint 4D-Var (SC)

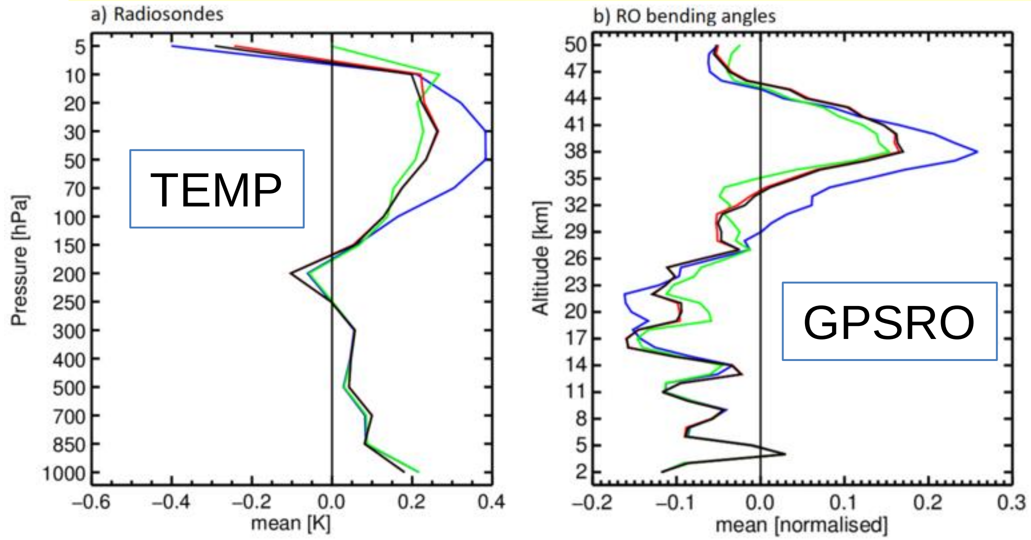
- Note: In the NN-informed experiments we let the model error corrections work on the full atmospheric column

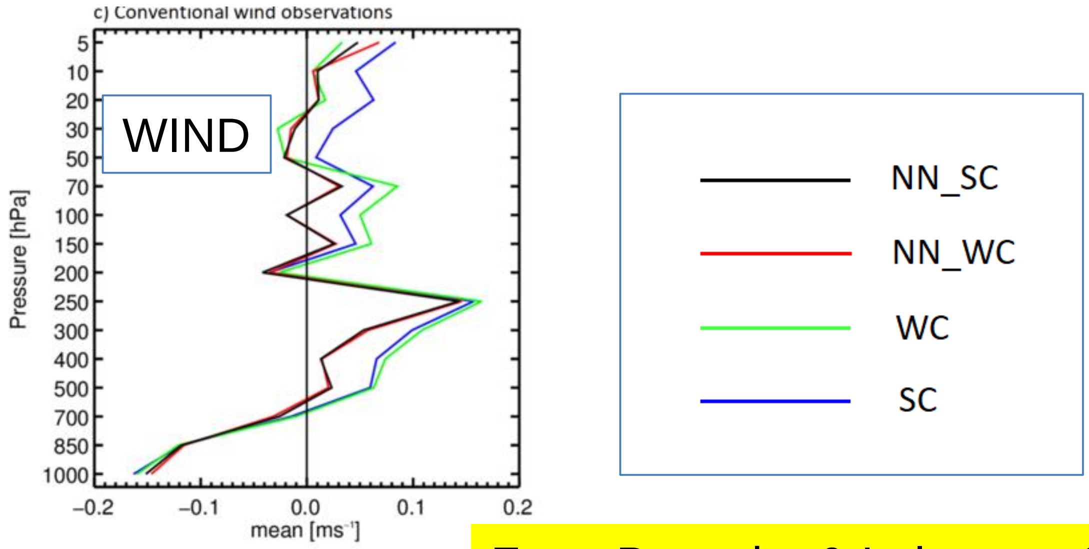
ML for bias-aware Data Assimilation: Experiments

Globally-averaged Mean Temperrature Correction



Globally-averaged Obs-Fg Mean Difference

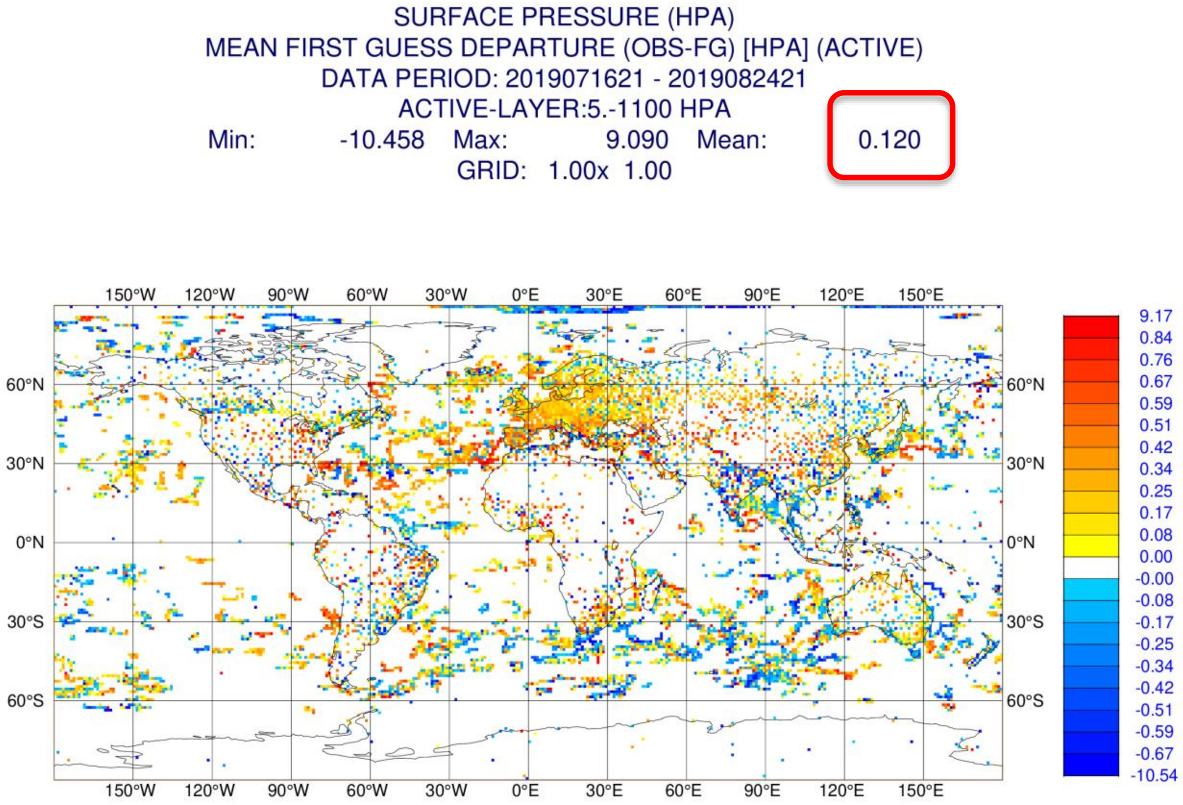




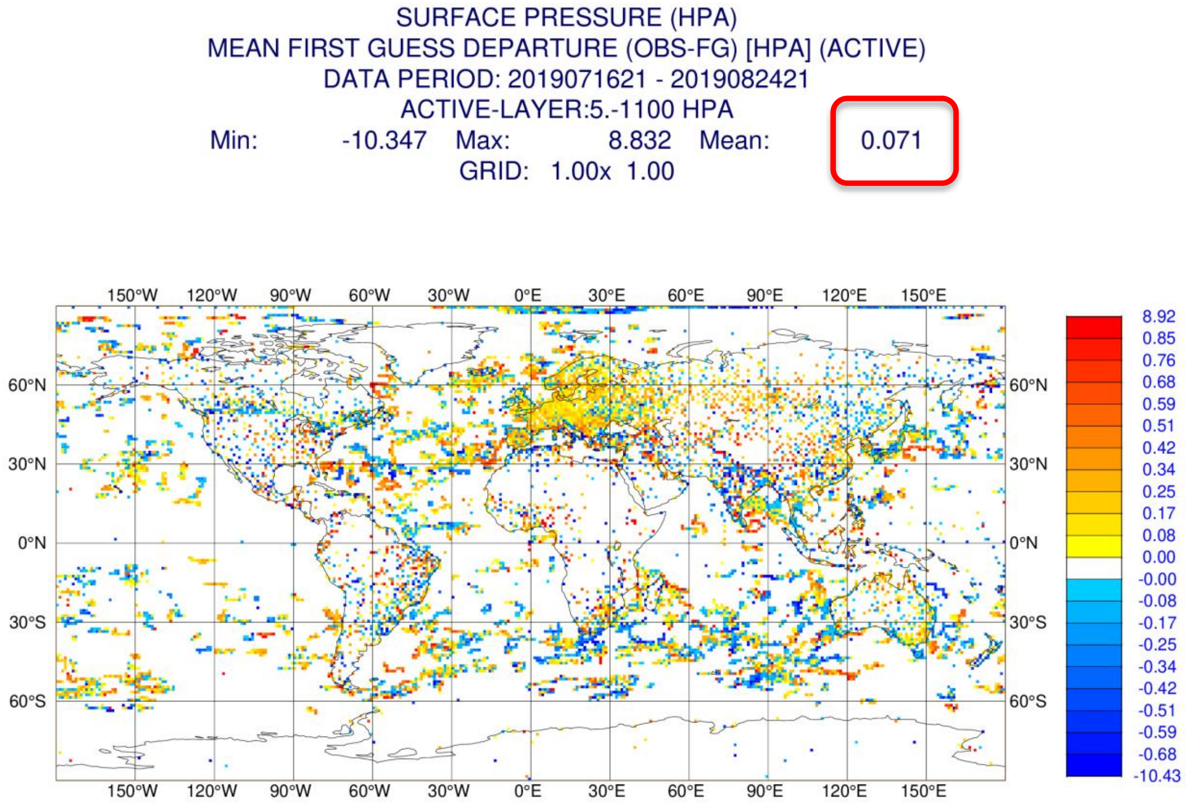
From Bonavita & Laloyaux, 2020

ML for bias-aware Data Assimilation: Mean errors

O-B Surf. Press. Obs: Standard WC_4DVar

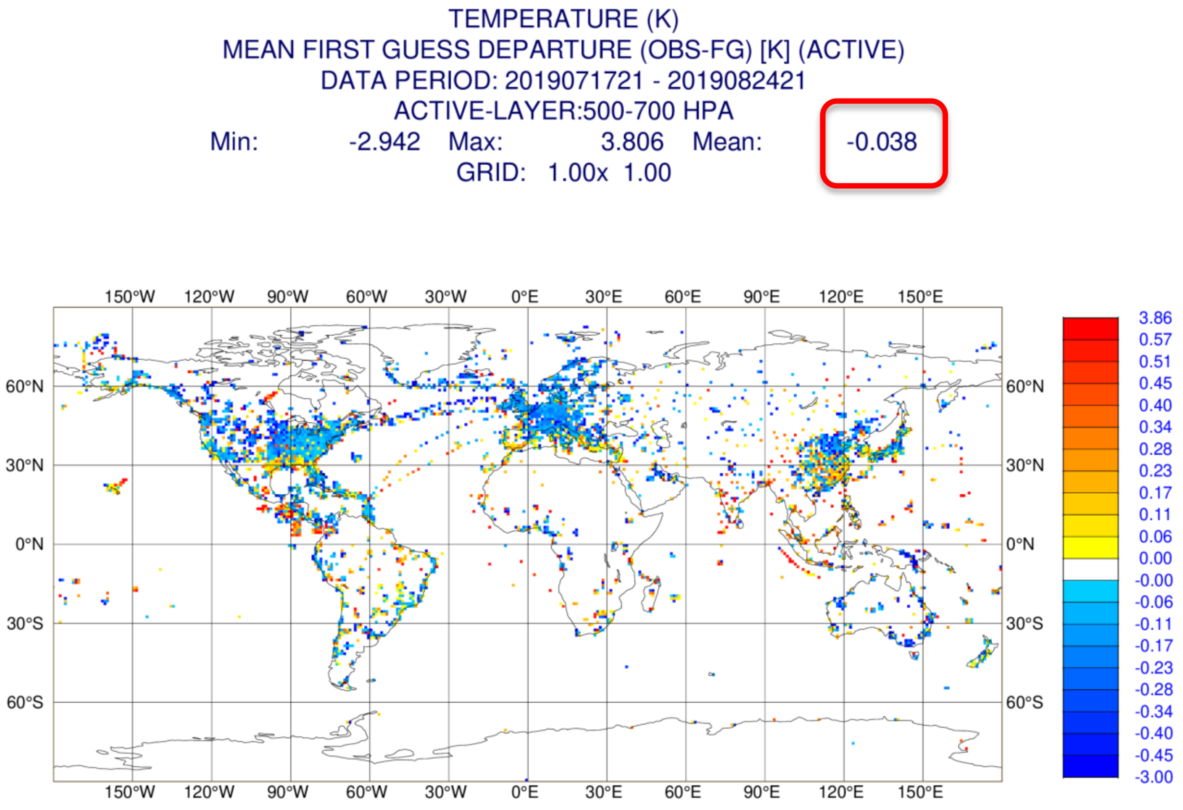


O-B Surf. Press. Obs: Hybrid NN-WC_4DVar

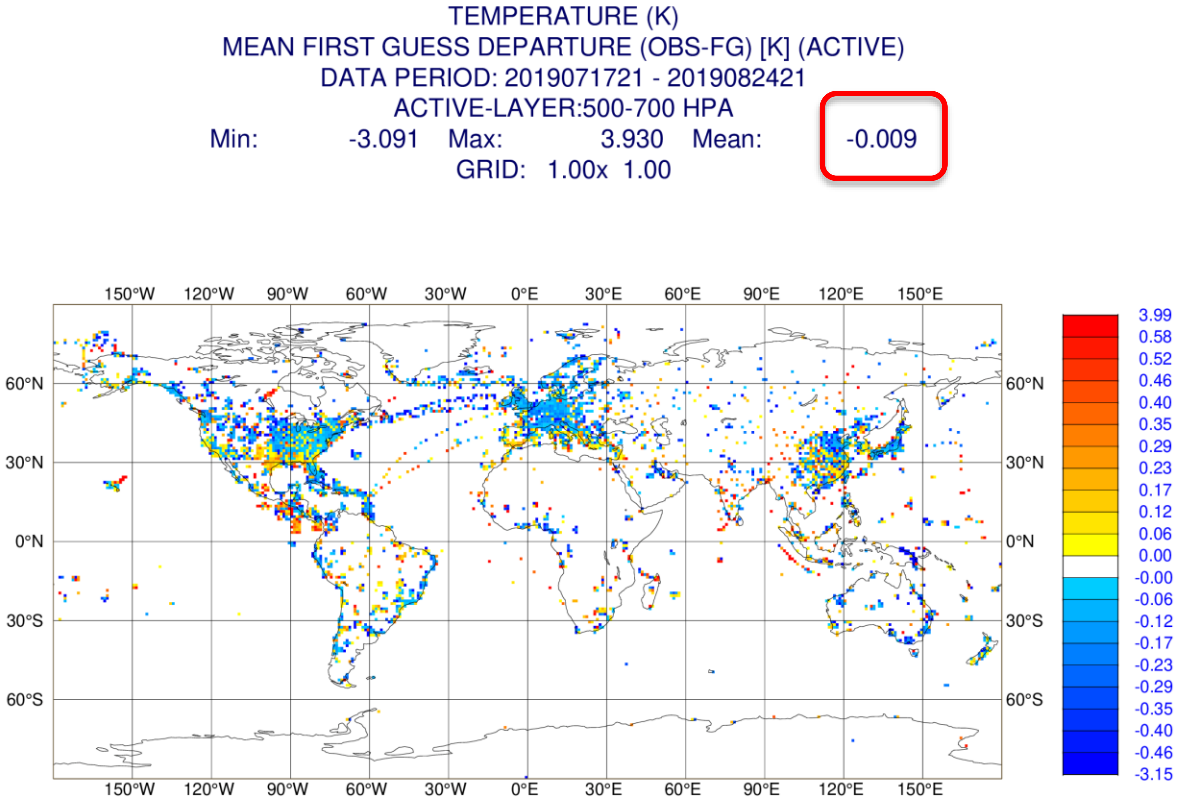


ML for bias-aware Data Assimilation: Mean errors

O-B Temp. 500-700 hPa: Standard WC_4DVar

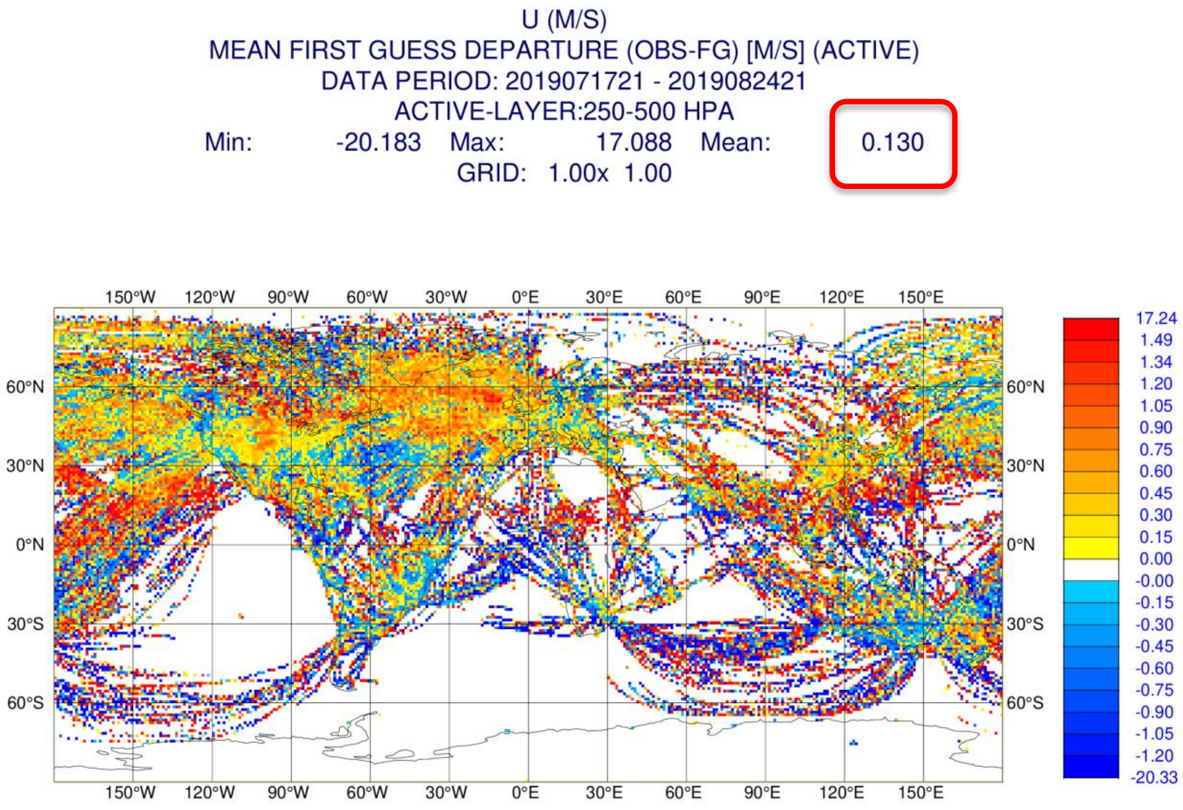


O-B 500-700 hPa : Hybrid NN-WC_4DVar

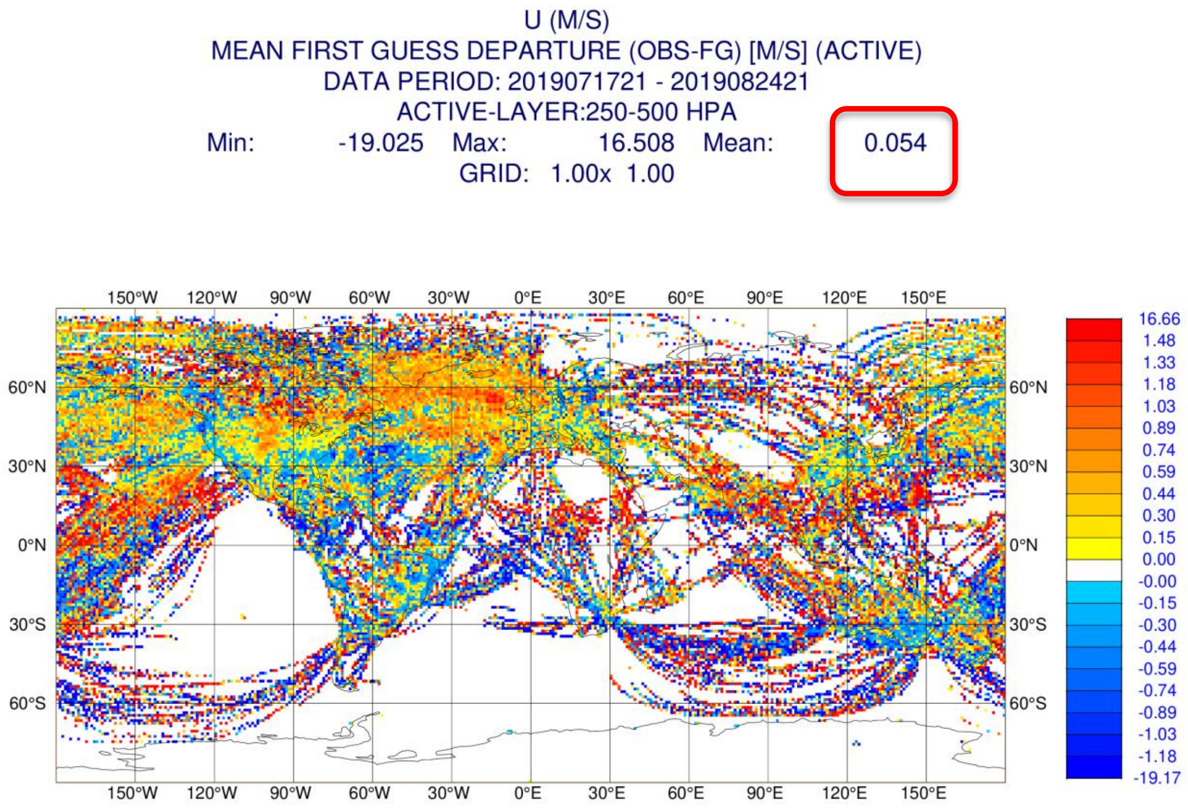


ML for bias-aware Data Assimilation: Mean errors

O-B U-wind 250-500 hPa: Standard WC_4DVar

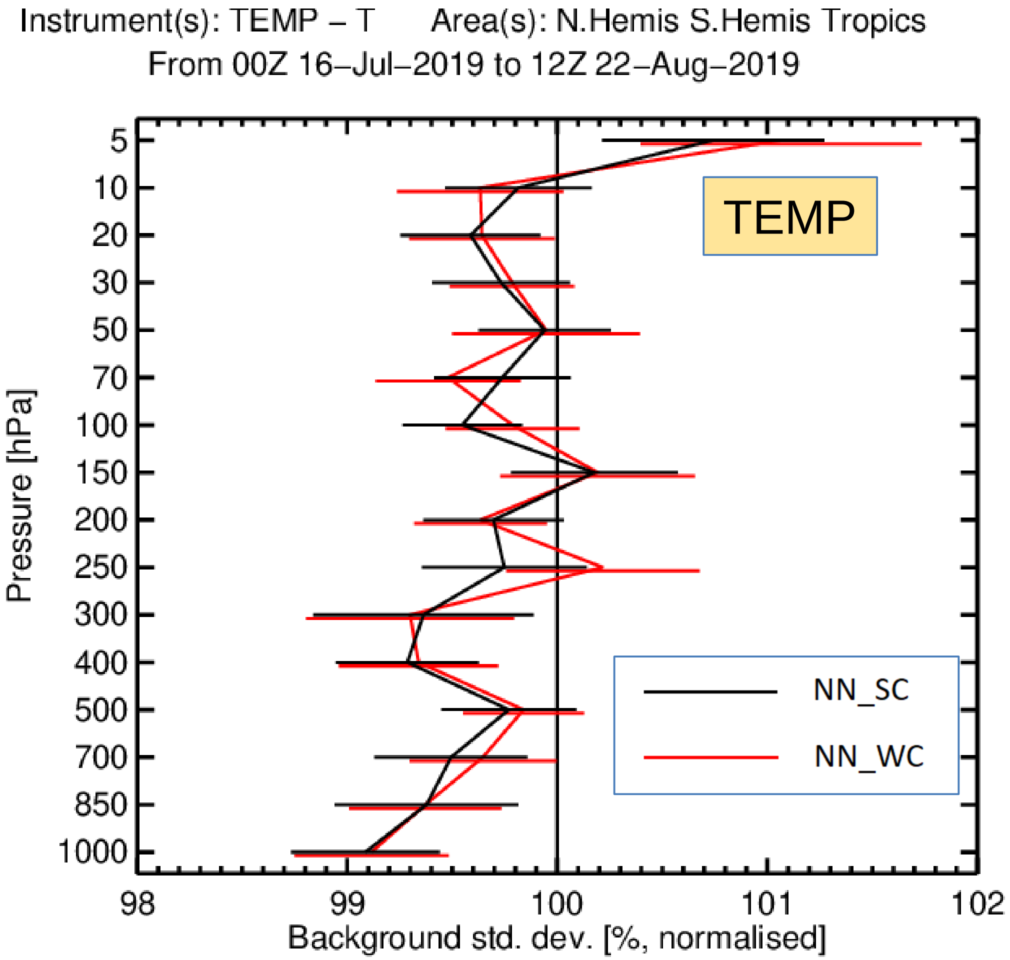


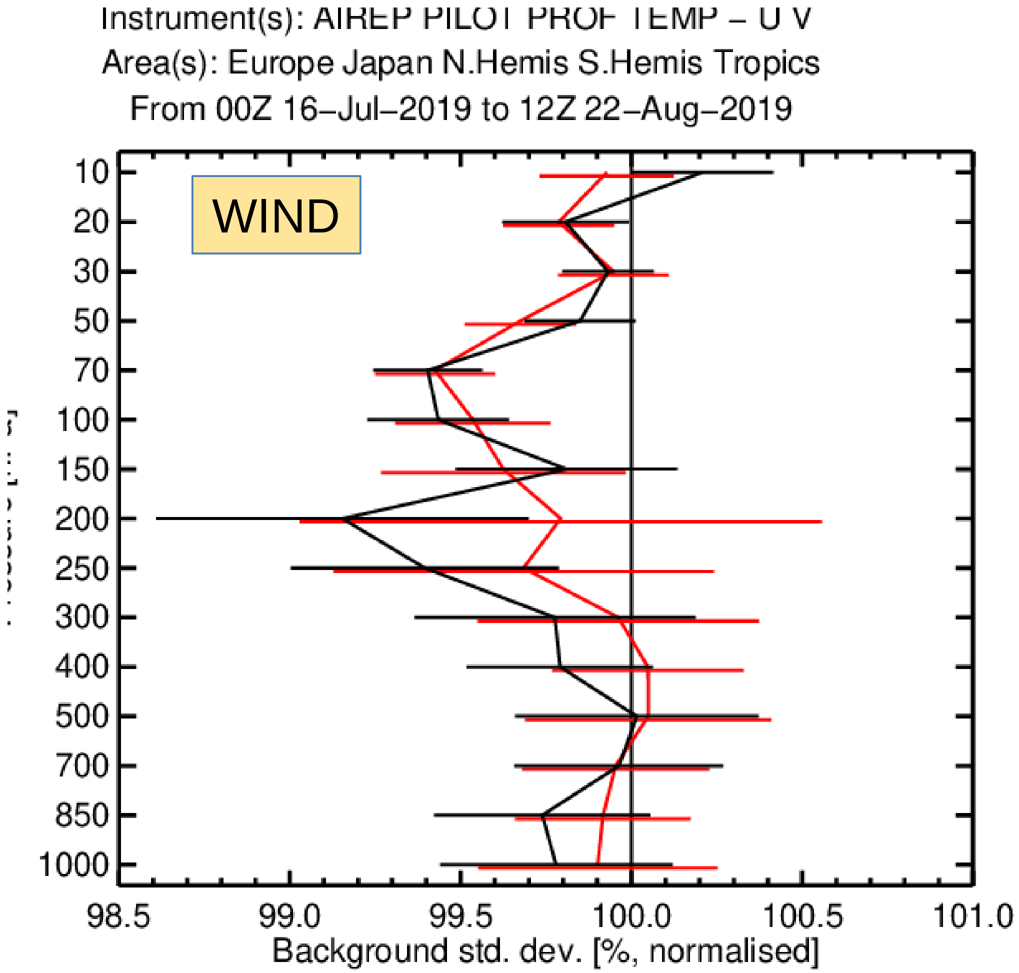
O-B U-wind 250-500 hPa: Hybrid NN-WC_4DVar



ML for bias-aware Data Assimilation: Random errors

Globally-averaged Observation-First Guess StDev norm. diff.





From Bonavita & Laloyaux, 2020

ECMWF

Note: 100% Baseline is current operational Weak Constraint 4D-Var

ML for bias-aware Data Assimilation: Discussion

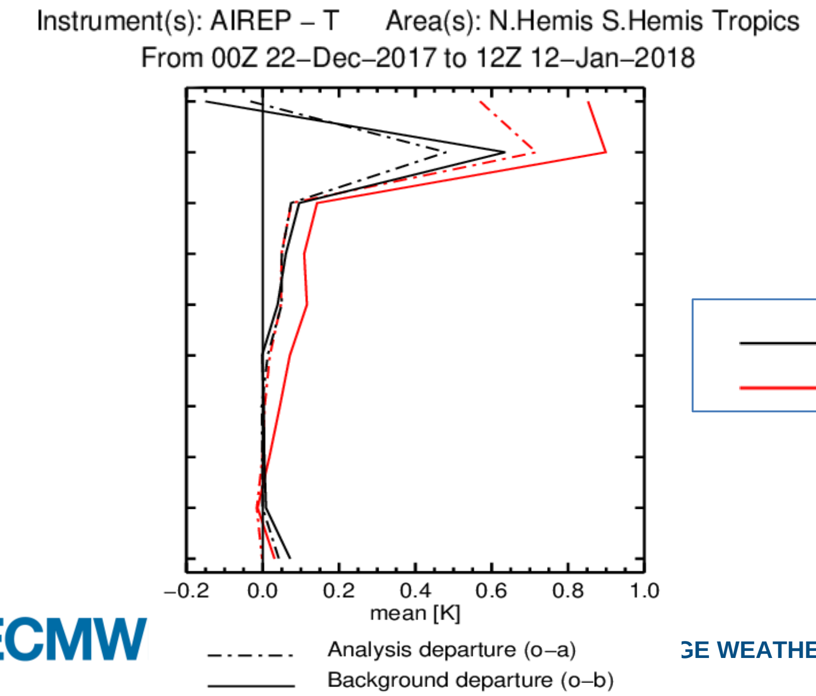
- A hybrid NN-WC4DVar appears to be better than a pure WC4DVar (at least in its current configuration)
- In particular it allows to extend WC4DVar to the whole atmospheric column without analysis/forecast skill degradation (in fact, with noticeable improvements in both)
- Mean (and StDev of) first guess errors are generally reduced in the troposphere, while performance in the stratosphere is comparable
- Best results were obtained using the ANN model error estimate as first-guess/background for a full column WC-4DVar analysis: this is similar to what we do in our hybrid background error modelling, i.e. we hybridise a climat. B with an online B from the latest EDA
- So we hypothesize that the hybrid ANN-WC_4DVar system outperform standard WC_4DVar for similar reasons, i.e ability to capture systematic, slow-evolving part of model error
- Note however that ANN is also a state-dependent ME estimator, not only a climatological estimator: it can be applied on extended-range forecasts

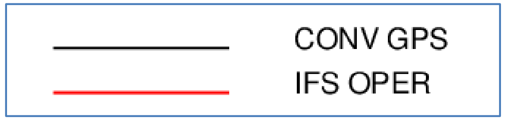
ML for bias-aware Data Assimilation: Perspectives

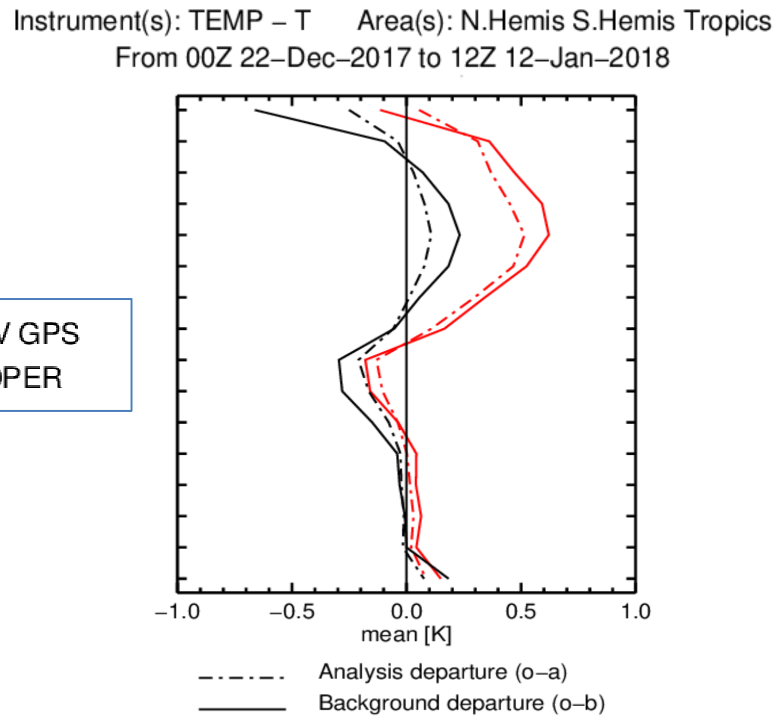
- The hybrid NN-WC4DVar setup presented today is just a first attempt to leverage ML technologies in the DA system
- There is room for improvement, e.g. selection of predictors and structure of the ANN
- Also the resolution of the training dataset could be increased (currently T21~1000 km), as model errors should be more heterogenous in the troposphere due to heterogeneity of surface forcings
- Are there better approaches?

ML for bias-aware Data Assimilation: Perspectives

- The ability of the ANN to de-bias the model forecasts is fundamentally limited by the residual systematic errors in the operational analyses
- Training the ANN on OmB increments requires very long timeseries and effectively asks the NN to perform an analysis of model error, which is not built to do
- A possible way forward is to train the ANN on a database of analyses with a reduced set of anchoring observations (conv. + gpsro + anchor sat. obs), whose analyses show reduced systematic errors in the DA cycle (Bonavita, 2014)







Thanks for your attention!

Bonavita, M. (2014), On some aspects of the impact of GPSRO observations in global numerical weather prediction. Q.J.R. Meteorol. Soc., 140: 2546-2562. doi:10.1002/qj.2320

Bonavita, M. and P. Laloyaux, 2020: Machine Learning for Model Error Inference and Correction, JAMES, accepted. https://doi.org/10.1002/essoar.10503695.1

Dee, D.P. (2005), Bias and data assimilation. Q.J.R. Meteorol. Soc., 131: 3323-3343. doi:10.1256/qj.05.137

Laloyaux, P., Bonavita, M., Chrust, M., Gürol, S., 2020a:. Exploring the potential and limitations of weak-constraint 4D-Var. Q J R Meteorol Soc. 1– 16. https://doi.org/10.1002/qj.3891

Laloyaux, P, Bonavita, M, Dahoui, M, et al. 2020b: Towards an unbiased stratospheric analysis. Q J R Meteorol Soc. 146: 2392–2409. https://doi.org/10.1002/qj.3798

Model Error Estimation and Correction in the IFS: Forecast Skill

- ANN in combination with Weak Constraint 4DVar improves the fit of observations to the model, both in the mean and in the random component.
- What can the ANN bring to forecast skill?

From Bonavita et al., in preparation

Data Assimilation and Machine Learning

- Ideally we would like to estimate the state and the model consistently and simultaneously, i.e. to solve the full Bayesian estimation problem (Bocquet et al., 2020):

$$p(\boldsymbol{x}_{0:K}, \mathbf{A} | \boldsymbol{y}_{0:K}) = \frac{p(\boldsymbol{y}_{0:K} | \mathbf{A}, \boldsymbol{x}_{0:K}) p(\boldsymbol{x}_{0:K} | \mathbf{A}) p(\mathbf{A})}{p(\boldsymbol{y}_{0:K})}$$

- In low-order geophysical systems it has been shown to be possible to solve this problem of state and model estimation (e.g. Brajard et al., 2020, Bocquet et al., 2020, Bocquet et al., 2019) using a coordinate descent approach:

FARCHI ET AL. 5

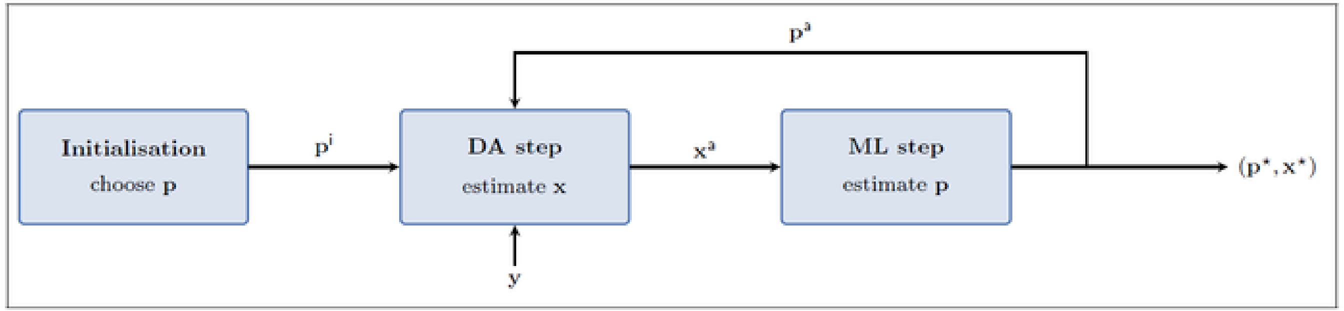


FIGURE 1 Minimisation strategy for Equation (4): alternate DA steps with ML steps to estimate the parameters **p** and the trajectory $\mathbf{x}_{0:N_t}$ with an increasing accuracy.

From: Farchi et al., 2020