Data Assimilation and Machine Learning Science at ECMWF

Massimo Bonavita

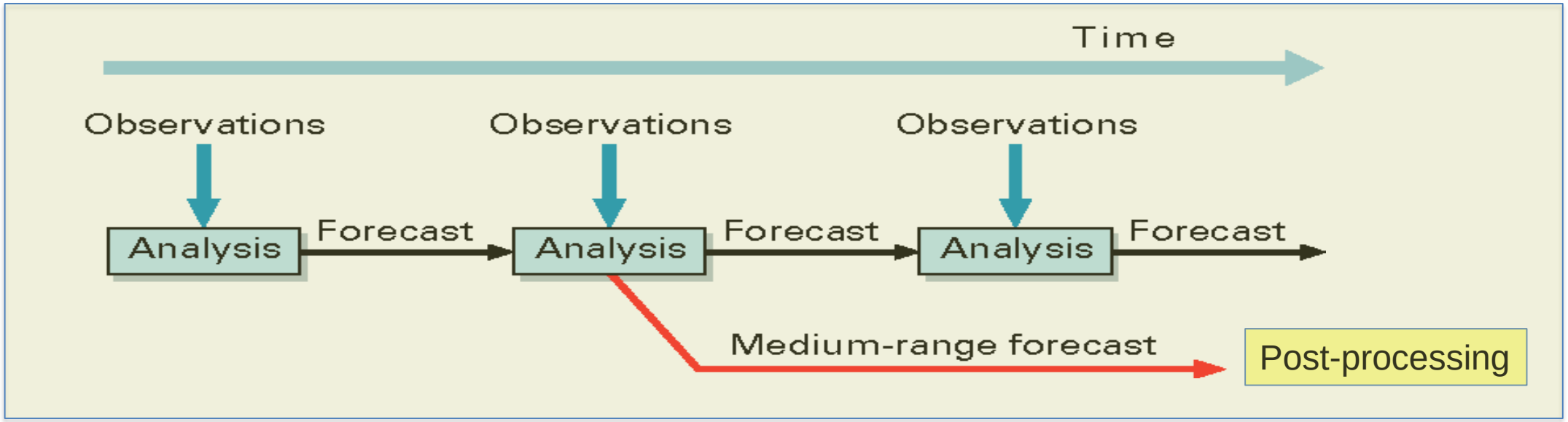
Research Department, ECMWF

massimo.bonavita@ecmwf.int

Contributors: Patrick Laloyaux[1], Sebastien Massart[1], Alban Farchi[2], Marc Bocquet[2]

1: ECMWF

2: École des ponts ParisTech



All components of the NWP workflow can potentially be improved by ML technologies:

a) **Observations**: Quality Control decisions, Bias correction, Observation operator;

b) **Forecast**: Model error correction, Model identification, model development

c) **Analysis**: Model error estimation and correction, Model parameter estimation, development of linearised/adjoint models

d) **Post-processing**: Ensemble/Determ. post-processing, Sig. Weather ident.,

Data Assimilation and Machine Learning

- We all agree ML can be useful in NWP/Climate Prediction, this is uncontroversial (hopefully!)
- Maybe something that is not so uncontentious is the following statement:

*"From a methodological point of view, ML **is not a conceptual jump** from the current Data Assimilation theoretical framework, it is a particularisation of DA methodology"*

- We already do ML in operational Data Assimilation: It is called weak-constraint 4D-Var!

Data Assimilation and Machine Learning

Weak Constraint 4D-Var Loss Function

$$J_{WC4DVar}(\boldsymbol{x}_0, \boldsymbol{\eta}) = J_B + J_O + J_Q =$$

$$\frac{1}{2}\left(x_0 - x_0^b\right)^T B^{-1}\left(x_0 - x_0^b\right)$$

$$+ \frac{1}{2}\sum_{i=0}^{N}\left(H_i(x_i) - y_i\right)^T R_i^{-1}\left(H_i(x_i) - y_i\right)$$

$$+ \frac{1}{2}\sum_{i=1}^{N}\left(x_i - M_i(x_{i-1}, \boldsymbol{\eta})\right)^T Q_i^{-1}\left(x_i - M_i(x_{i-1}, \boldsymbol{\eta})\right)$$

Generic ML/DL Loss Function

$$J_{ML}(\cancel{x_0}, \boldsymbol{\eta}) =$$

$$\cancel{\frac{1}{2}\left(x_0 - x_0^b\right)^T B^{-1}\left(x_0 - x_0^b\right)}$$

$$\cancel{+ \frac{1}{2}\sum_{i=0}^{N}\left(H_i(x_i) - y_i\right)^T R_i^{-1}\left(H_i(x_i) - y_i\right)}$$

$$+ \frac{1}{2}\sum_{i=1}^{N}\left(x_i - M_i(x_{i-1}, \boldsymbol{\eta})\right)^T Q_i^{-1}\left(x_i - M_i(x_{i-1}, \boldsymbol{\eta})\right)$$

Notes: (Brajard et al., 2020; Bocquet et al., 2020; Farchi et al., 2020)

1. ML/DL models are not used for state estimation → $J_B = 0$
2. ML/DL loss functions typically assume full, noiseless observations $(H_i = I, R_i \to 0) \to J_O = 0$
3. ML/DL models can optionally have a regularization term function of the ANN model parameters $\boldsymbol{\mathcal{L}}(\boldsymbol{\eta}) = \boldsymbol{\mathcal{L}}(\boldsymbol{W}, \boldsymbol{b})$, e.g. Tikhonov regularisation, drop-out, etc.

Data Assimilation and Machine Learning

- Given that ML/DL can be seen as a particularisation of standard Data Assimilation methodologies, is the current excitement about it justified?

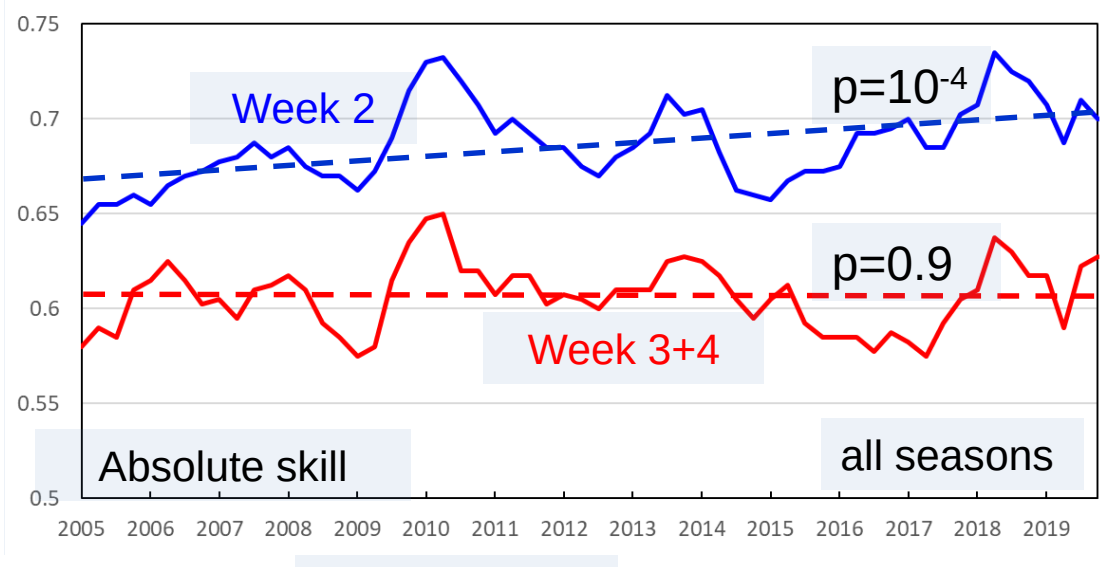
- We believe the answer is yes, for at least two reasons:

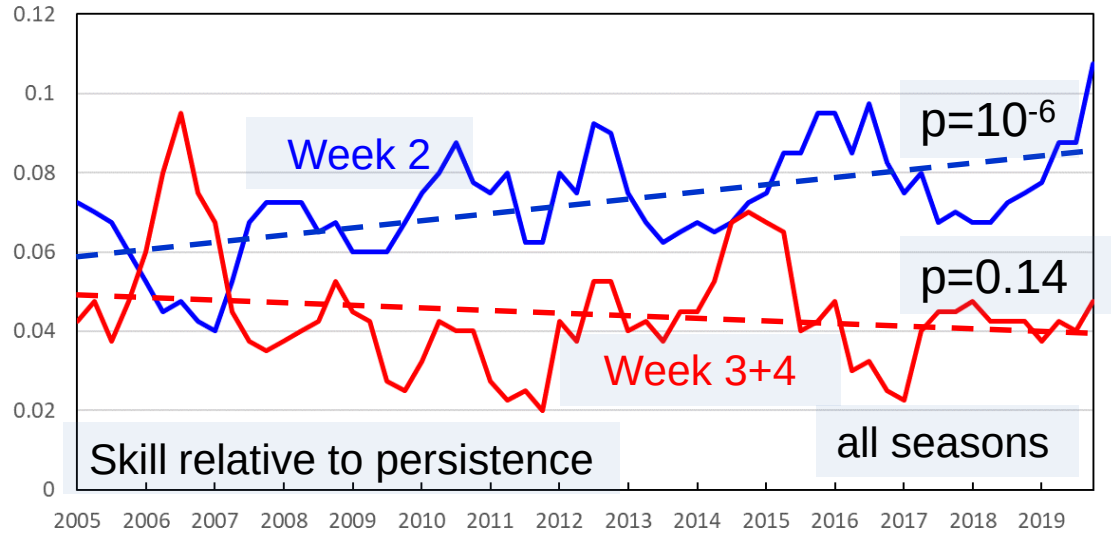
  1. Technically the explosion of ML/DL has provided a wealth of efficient, easy to use open source software that can be easily repurposed for one's applications;

  2. Conceptually, it has brought into prominence the idea that the current and ever-increasing wealth of geophysical **observations can be directly used to improve the forecast models** (and the forward models too, see A. Geer talk in this series)

Data Assimilation and Machine Learning

- Improving forecast models can be considered the most urgent requirement to make progress on extended, sub-seasonal and seasonal predictive limits
- Traditionally, DA algorithms have used a perfect model assumption and aimed at reducing random initial conditions errors: this brings steady progress up to ~2 weeks, but not much is visible beyond two weeks
- Can we deploy ML to break the 2-week Lorenz predictability barrier?





From: Frederic Vitart and Thomas Haiden

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:

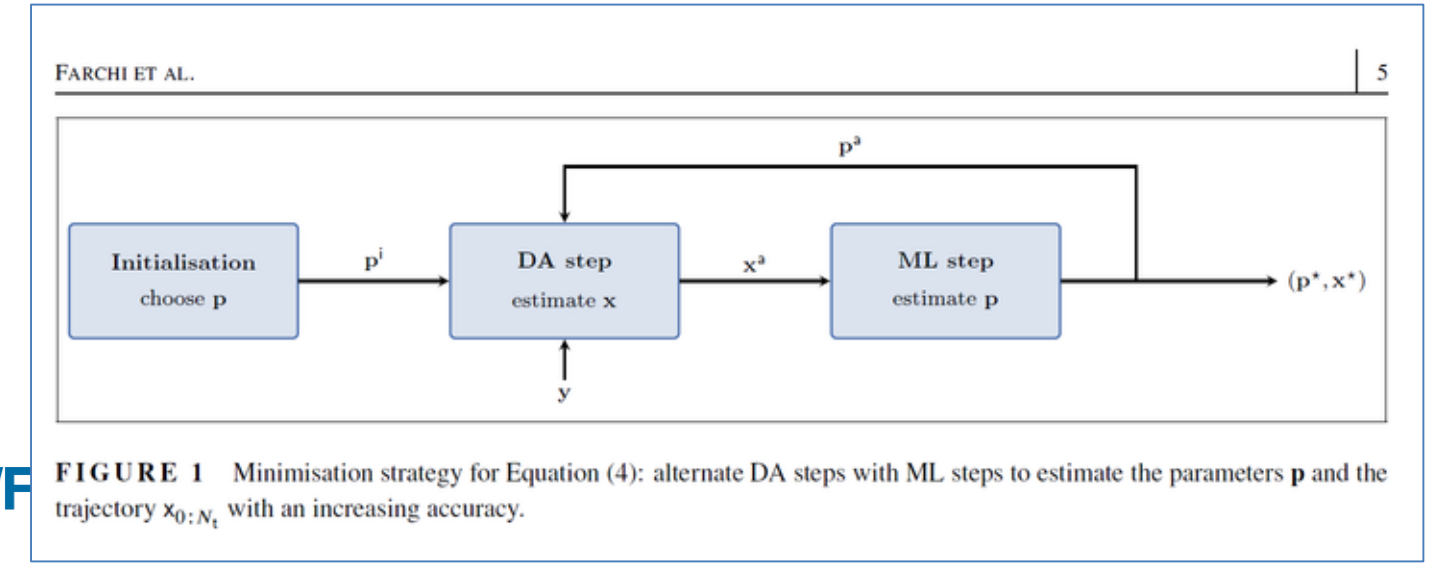


FIGURE 1 Minimisation strategy for Equation (4): alternate DA steps with ML steps to estimate the parameters **p** and the trajectory $\mathbf{x}_{0:N_t}$ with an increasing accuracy.

From: Farchi et al., 2020

Data Assimilation and Machine Learning

- How can we adapt this general recipe to operational NWP and Climate prediction? There are two issues:

  1. We do not have enough time to iterate the DA and ML steps in real time NWP;

  2. We have a much more complex model to deal with ☹, but we do not start from zero, but from a very good model ☺!

- These two features of the NWP/Climate prediction endeavour point to two possible ways of tackling the combined state-model estimation problem:

  1. Model Parameter estimation

  2. Model Error estimation and correction

Combining DA and ML: Model Parameter Estimation

- In Model Parameter estimation we start from a fundamental assumption:

"*the model is structurally sound, model errors (mainly) arise from incorrect settings of its parameters (or parameters of their pdfs, in a stochastic model framework)*".

- This assumption has the big advantage that it drastically reduces the hypothesis space of our ML model to that of the model parameters(~$O(10^2)$)

- Traditionally, these parameters are tuned by hand by comparing short and long forecasts (and ensemble of forecasts) to observations/analyses. This is computationally expensive and labour intensive.

- Assuming the model parameters have a strong correlation with the observed variables (Identifiability condition: Navon, 1997), we can add them to the control vector of the analysis (Augmented Control Vector, see Ruiz et al., 2013, for a review), both in Ensemble and Variational DA

Combining DA and ML: Model Parameter Estimation

- Augmented Control Vector in 4DVar:

$$J_{4DVar}(\boldsymbol{x}_0, \boldsymbol{p}) = J_B + J_P + J_O = \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^T \mathrm{B}^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) + \frac{1}{2}\left(\boldsymbol{p} - \boldsymbol{p}^b\right)^T \mathrm{B}_\mathrm{p}^{-1}\left(\boldsymbol{p} - \boldsymbol{p}^b\right) + \frac{1}{2}\sum_{i=0}^{N}\left(H_i \circ M_i(\boldsymbol{x}_0; \boldsymbol{p}) - \boldsymbol{y}_i\right)^T \mathrm{R}_\mathrm{i}^{-1}\left(H_i \circ M_i(\boldsymbol{x}_0; \boldsymbol{p}) - \boldsymbol{y}_i\right)$$

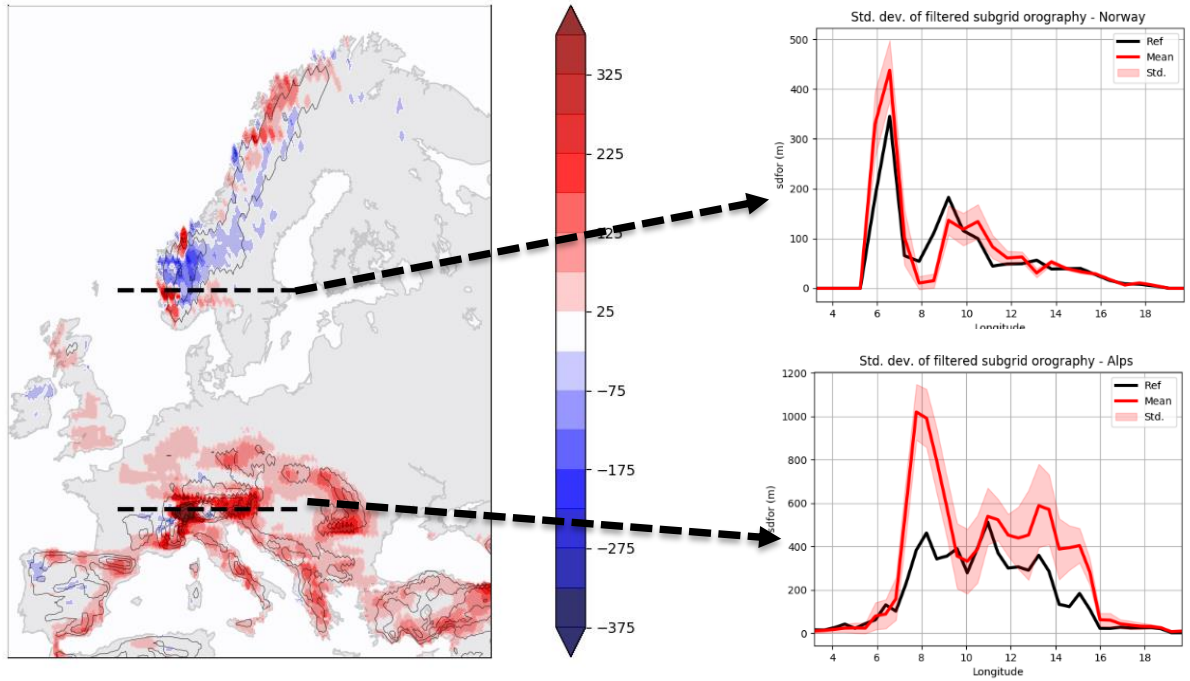
- ($\boldsymbol{p}$) is the set of model parameters (or fields of those) that are estimated together with the state. (This could also be applied to uncertain parameters of the forward model H)

- This approach has been around in atmospheric data assimilation since at least the mid ’90s, but results have been rather inconsistent in realistic applications

- With the current and projected wealth of available observations (ECMWF DA ingests ~40 million obs every 12h!) we believe time is ripe for revisiting this idea

Combining DA and ML: Model Parameter Estimation

Example: Online Estimation of the “standard deviation of subgrid orography” in the IFS Orographic Drag scheme (S. Massart, ECMWF)



- Identifiability: Surface pressure is sensitive to orography, 4DVar can constrain the parameter through surf. press. observations
- Preliminary results show improved surface pressure forecast skill!

Analysis increments of estimated StDev orography (left)
Cross section of default (black) and estimated (red, mean + spread) parameter values (right)

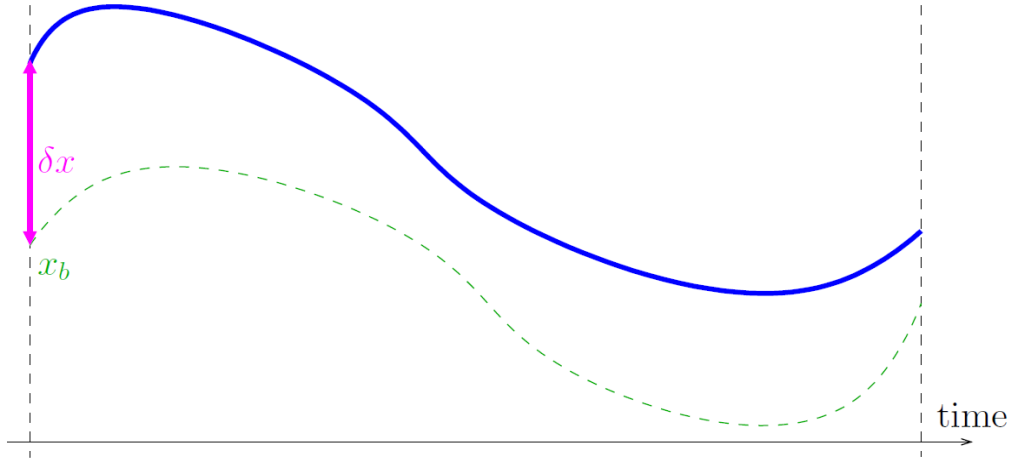
Combining DA and ML: Model Error Estimation and Correction

- In Model Parameter estimation we have made the assumption that a perfect model is a model with optimal parameter values;
- However models (model parameterisations) have also non-negligible structural uncertainties (e.g., BL turbulence, cloud convection), which cannot be resolved by tuning of the closure parameters;
- An alternative approach is to abandon the perfect model assumption and consider the evolution of the state given by the knowledge-based model plus a residual model error component:

$$\boldsymbol{x}_k = M_k(\boldsymbol{x}_{k-1}) + \eta_k(\boldsymbol{x}_{k-1}) = M_k(\boldsymbol{x}_{k-1}) + M_k^{ML}(\boldsymbol{x}_{k-1})$$

- In Data Assimilation this idea is called weak constraint 4DVar (Sasaki, 1970)

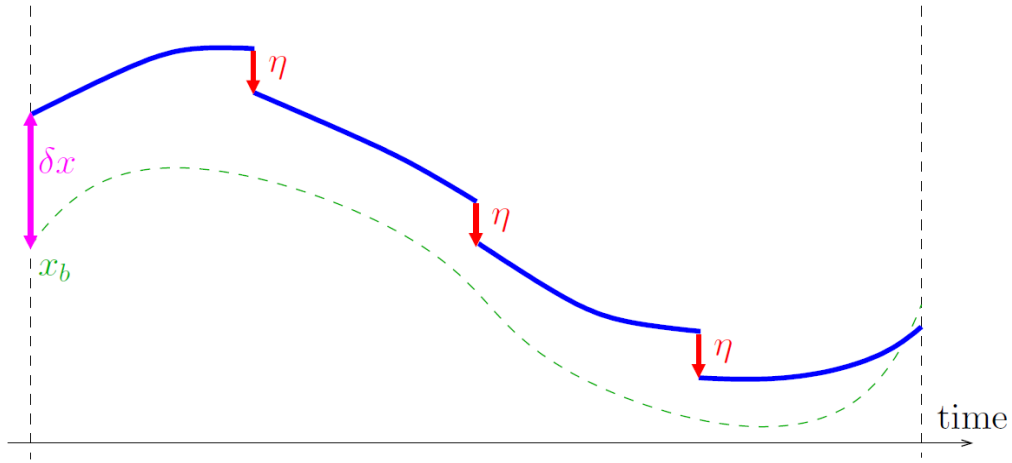
Strong Constraint 4D-Var



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

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

Weak Constraint 4D-Var



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

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

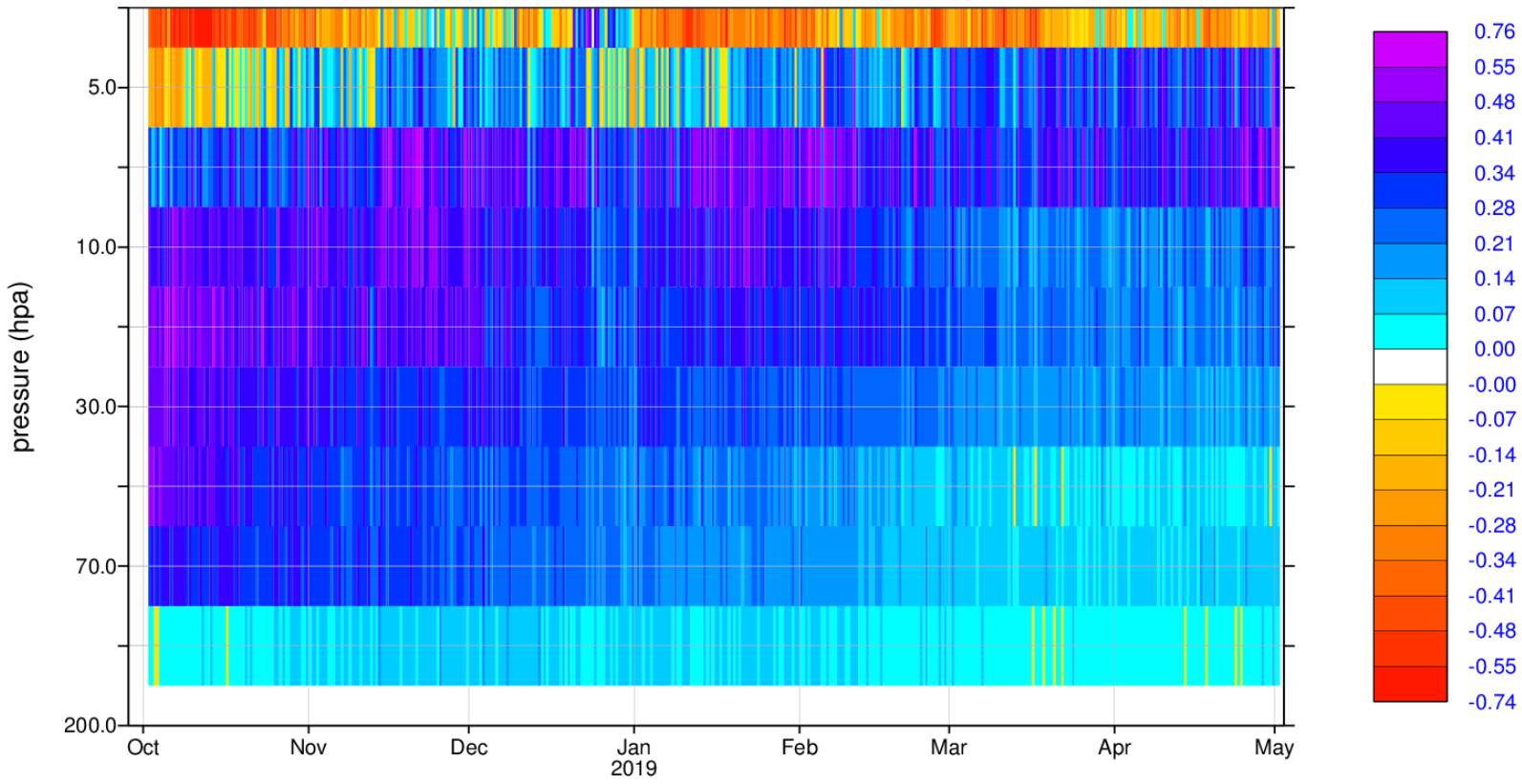
Weak Constraint 4D-Var

- The ECMWF version of WC-4DVar works by estimating (“learning”) a constant model error tendency over an assimilation window (12h)
- The background (and first guess) model error is the previous window’s estimate (we do not have a predictive model of model error!)
- This means that our model error formulation targets errors which are slowly evolving on the timescale of the assimilation window (ie, 12 hours) and large scale with respect to typical background errors
- These insights are the basis of the success of the latest WC-4DVar implementation in dealing with stratospheric model biases (Laloyaux et al., 2020a, b)

Weak Constraint 4D-Var

- 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

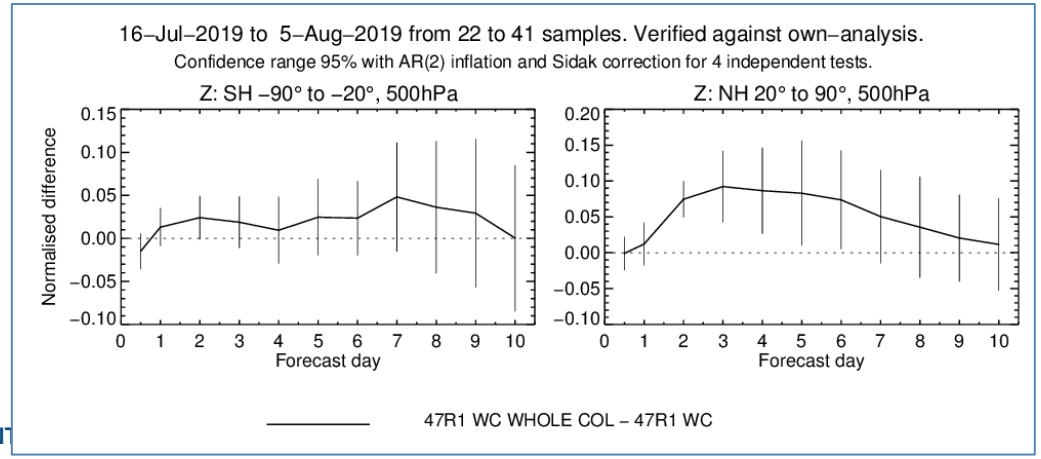
Mean first-guess departure with respect to GPS-RO temperature retrievals



From Laloyaux et al, 2020b

Weak Constraint 4D-Var

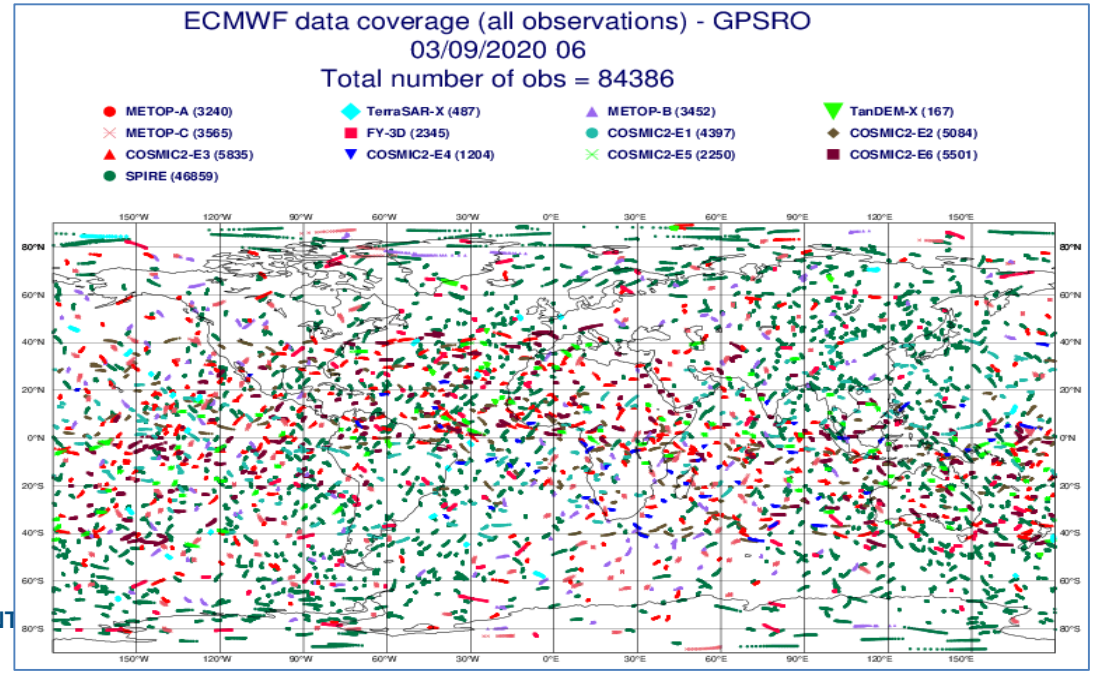
- Current version of WC-4DVar is a step change wrto previous versions, but it is not the end of the story (so far):
  - WC-4DVar reduces stratospheric fcst temperature bias ~ completely against BC corrected obs (radiances), by 30-40% against non-bias corrected obs (Radiosondes, GPS-RO)
  - Current WC-4DVar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



Combining DA and ML: Model Error Estimation and Correction

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

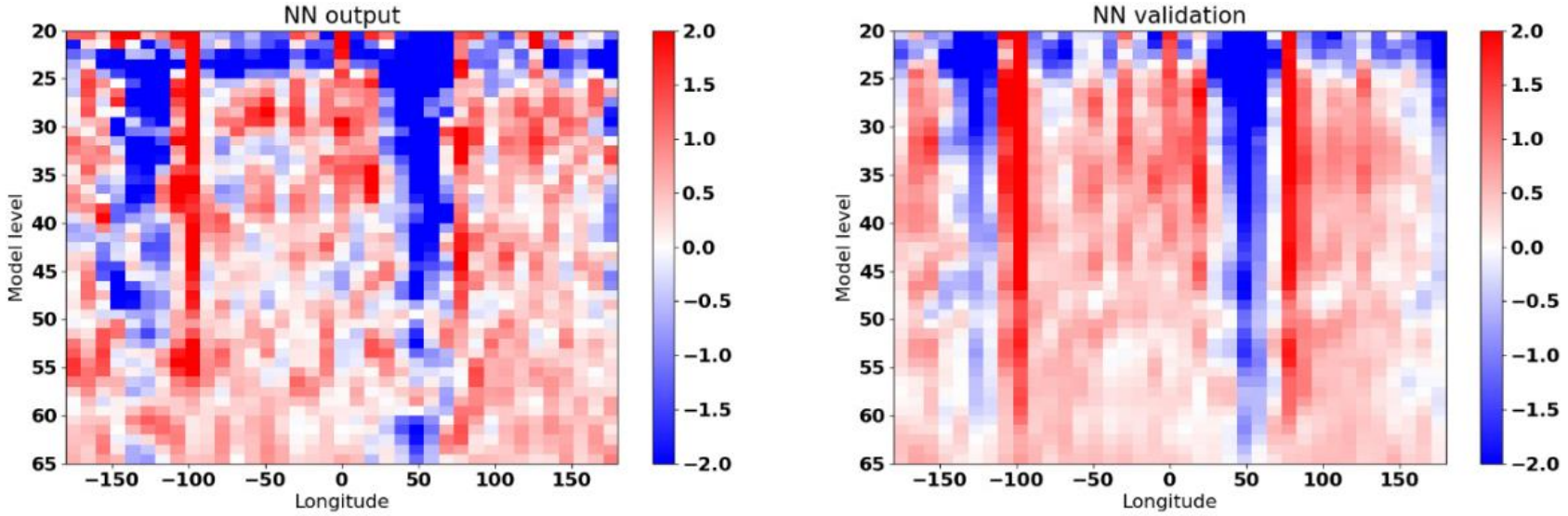
- Basic idea: Train an Artificial Neural Network to “learn” the statistically predictable part of model error
- This idea can be realised in at least two ways
- The most direct way is to train the ANN on a database of observation-background (O-B) departures, choosing an approx. homogeneous, dense and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



Combining DA and ML: Model Error Estimation and Correction

- The idea of training the ML model from observation departures has the advantage of by-passing the assimilation cycle, which avoids the systematic errors present in the analyses
- On the other hand it introduces other complications, e.g. lack of homogeneous coverage, need of long timeseries, how to extrapolate the corrections to unobserved locations and unobserved state variables?

O-B 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

Combining DA and ML: Model Error Estimation and Correction

- Another possibility is to train the ML model on a database 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*

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

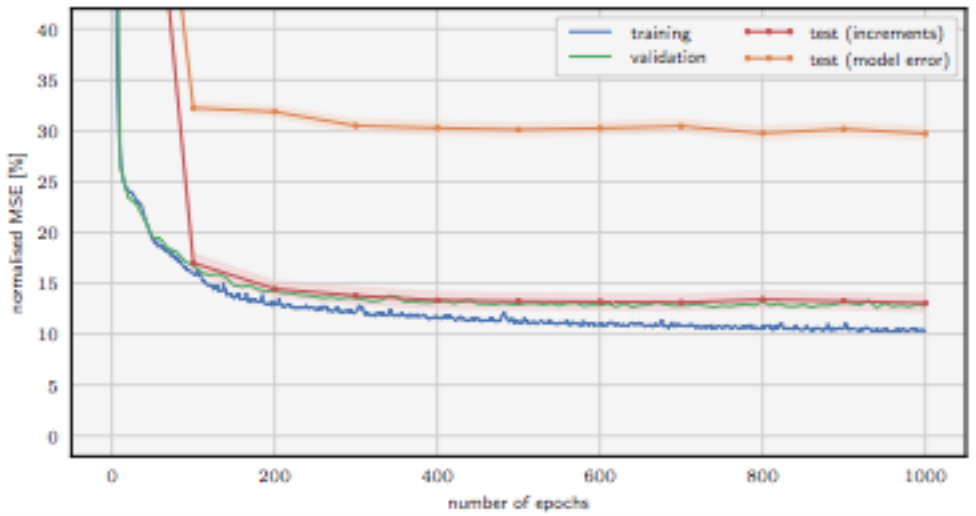
$$\mathbf{dx}_k = \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{h}\left(\mathbf{x}_k^b - \mathbf{dx}_k^p\right)\right) \qquad (44)$$

will correct the model background and produce unbiased analyses."

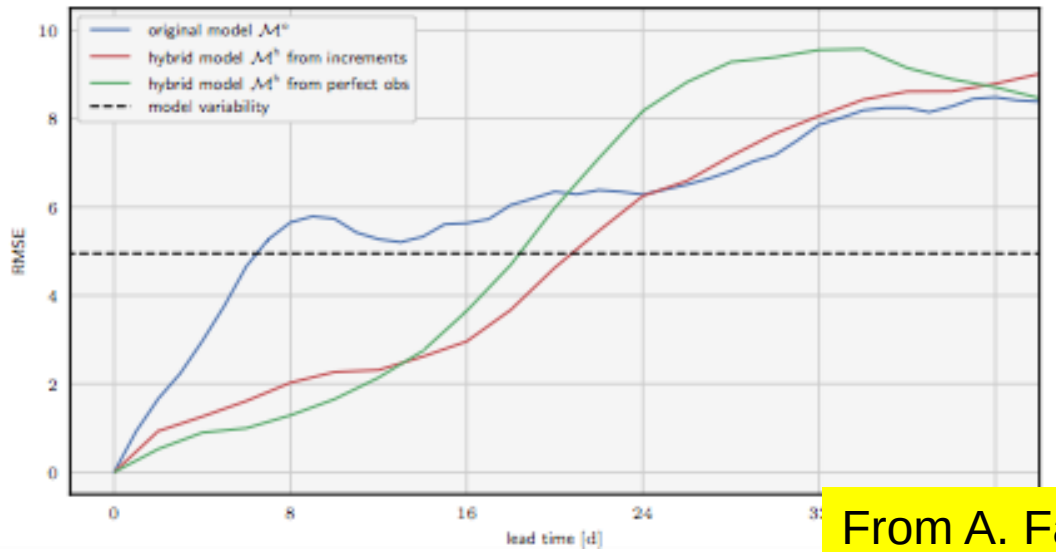
Combining DA and ML: Model Error Estimation and Correction

- We have tested this approach in a low order 2-level, 1600 variables QG model (Farchi et al., 2020)
- In this setup, the trained ANN can explain ~70% of the model error variance and when used in forecast mode to correct model errors can ~double the predictability horizon of the forecasts!
- When the trained ANN is used in the DA cycle only, it reduces RMS analysis errors by ~25%
- How do these surprising results translate into operational-grade assimilation and modelling systems?

Training and test curves for the ANN



Evolution of the fcst RMSE



From A. Farchi

Combining DA and ML: Model Error Estimation and Correction

- We are testing this approach in the operational version of the ECMWF IFS, initial results are available in Bonavita and Laloyaux, 2020

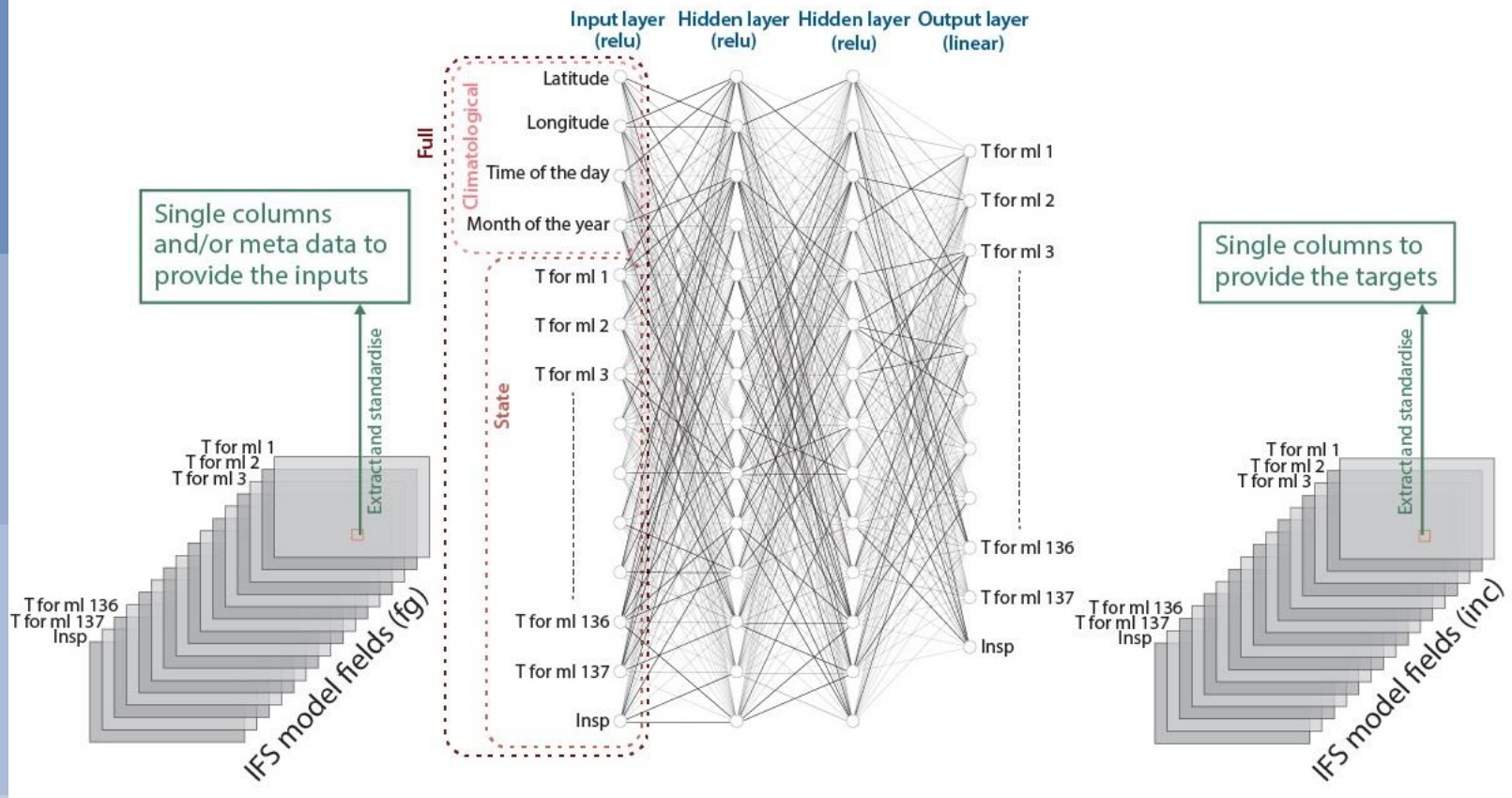
- The size of the state vector is $\mathcal{O}(10^{10})$. Even assuming a much lower effective dimension for the model error vector, a primary design consideration for the ANN has been to reduce its size to manageable dimensions

- 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 main 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 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*.

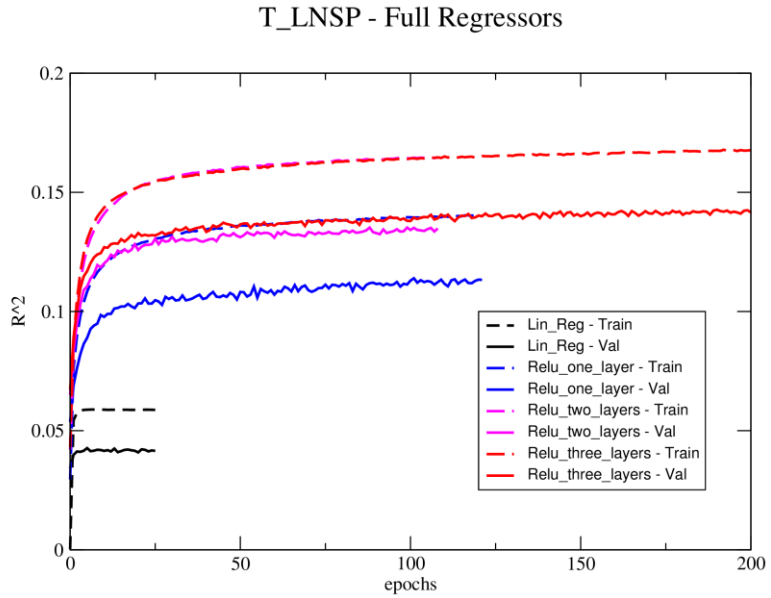
Combining DA and ML: Model Error Estimation and Correction

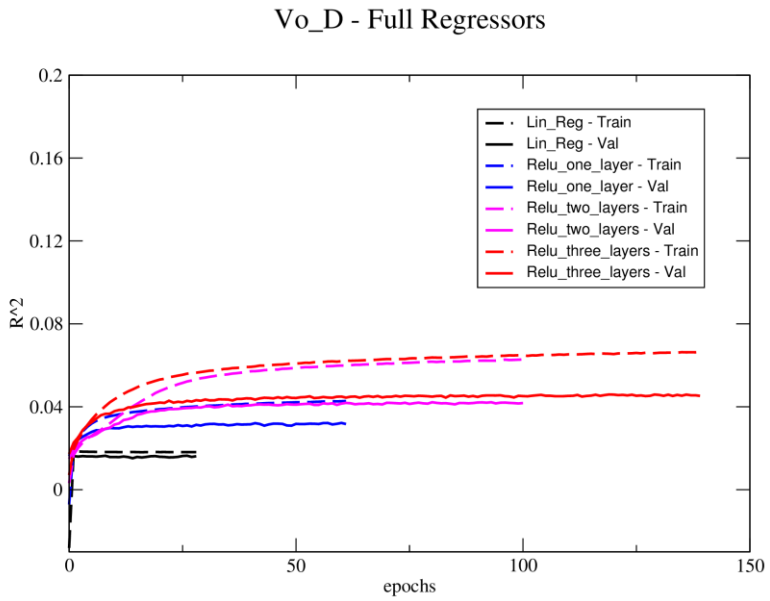


- Dense Neural Network with Relu activation
- Three layers with nonlinear activations give best results: problem with only moderate nonlinearities
- Dropout layers used to control overfitting, input/outputs pre-normalised for training, Adam minimiser
- Number of trainable parameters ~$6*10^4$, size of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

From Bonavita & Laloyaux, 2020

Combining DA and ML: Model Error Estimation and Correction





- 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

Combining DA and ML: Model Error Estimation and Correction

- 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:

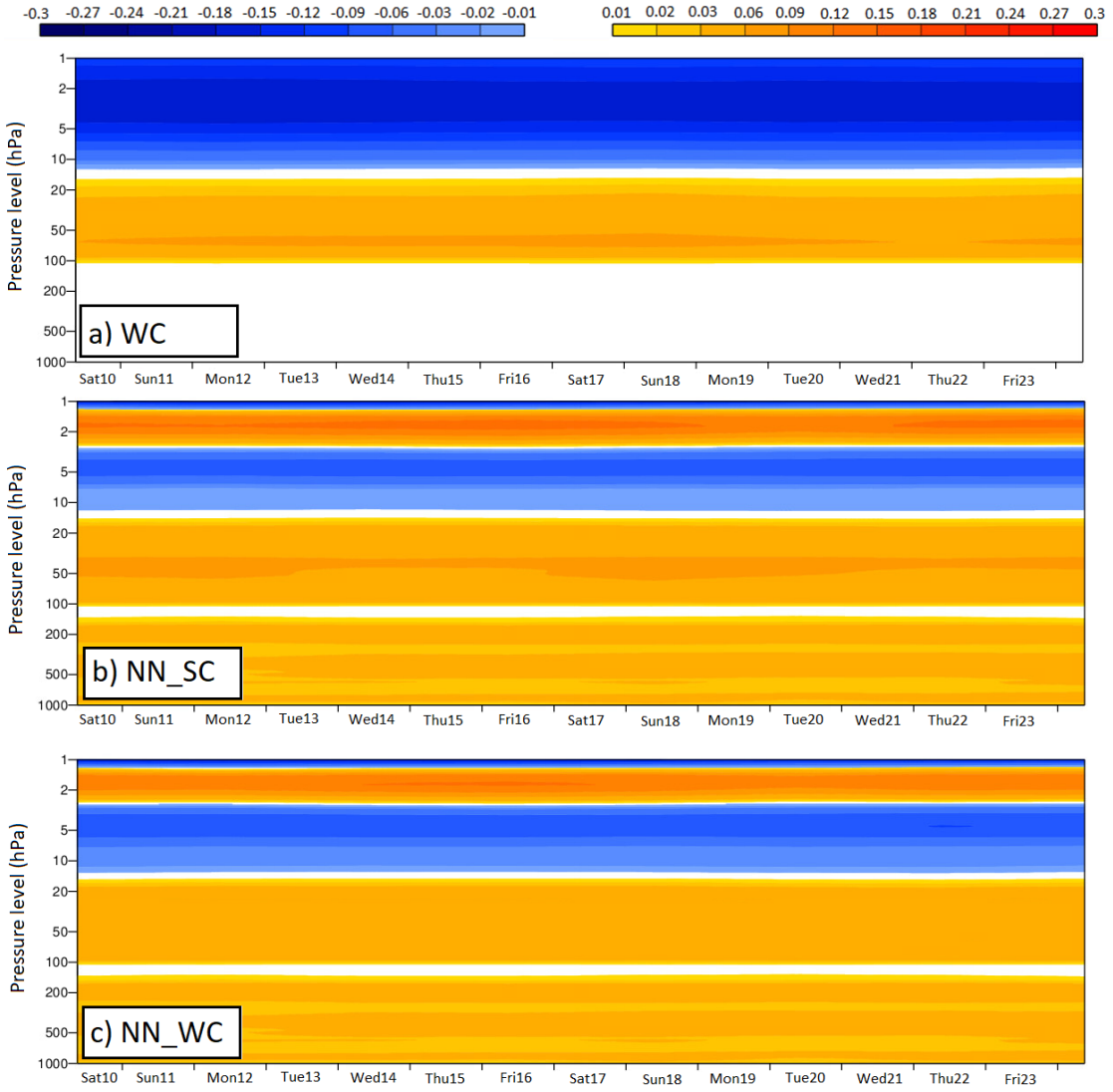
  a) Use the NN model errors instead of the Weak Constraint 4D-Var model errors (NN_SC in the following);

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

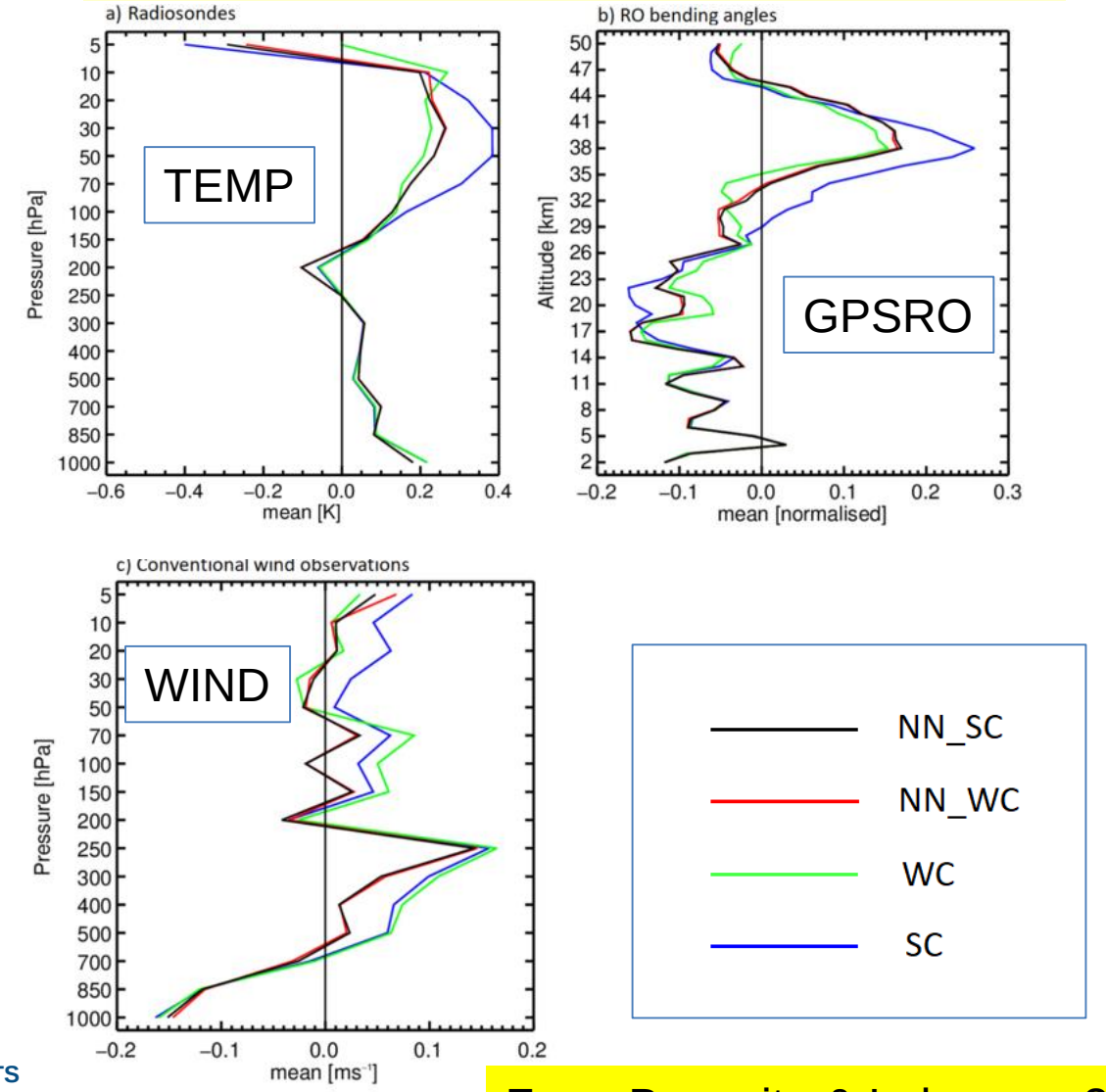
- Baseline configuration was the currently operational version of Weak Constraint 4D-Var at full resolution (T1279, ~9km grid spacing) (WC) and the previously operational Strong Constraint 4D-Var (SC)

Model Error Estimation and Correction in the IFS: Mean errors

Globally-averaged Mean Temperrature Correction



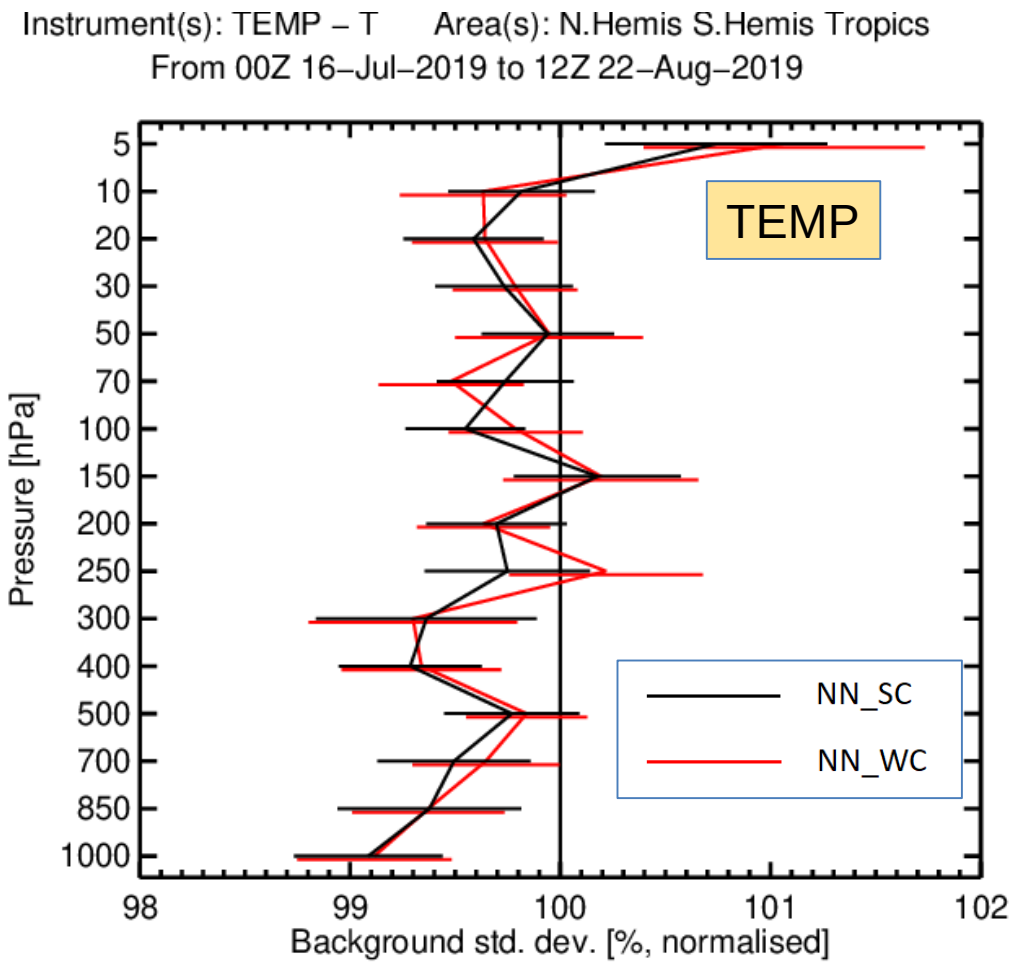
Globally-averaged Obs-Fg Mean Difference

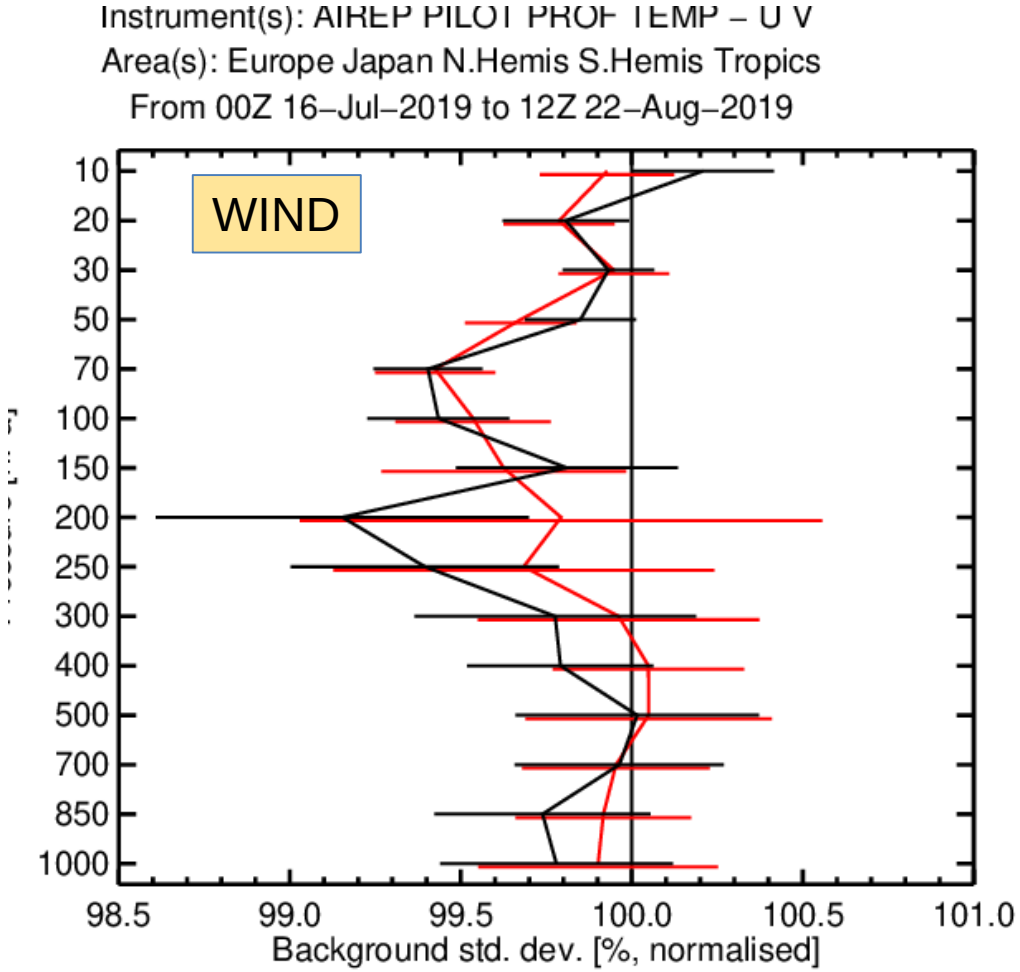


From Bonavita & Laloyaux, 2020

Model Error Estimation and Correction in the IFS: Random errors

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





ECMWF

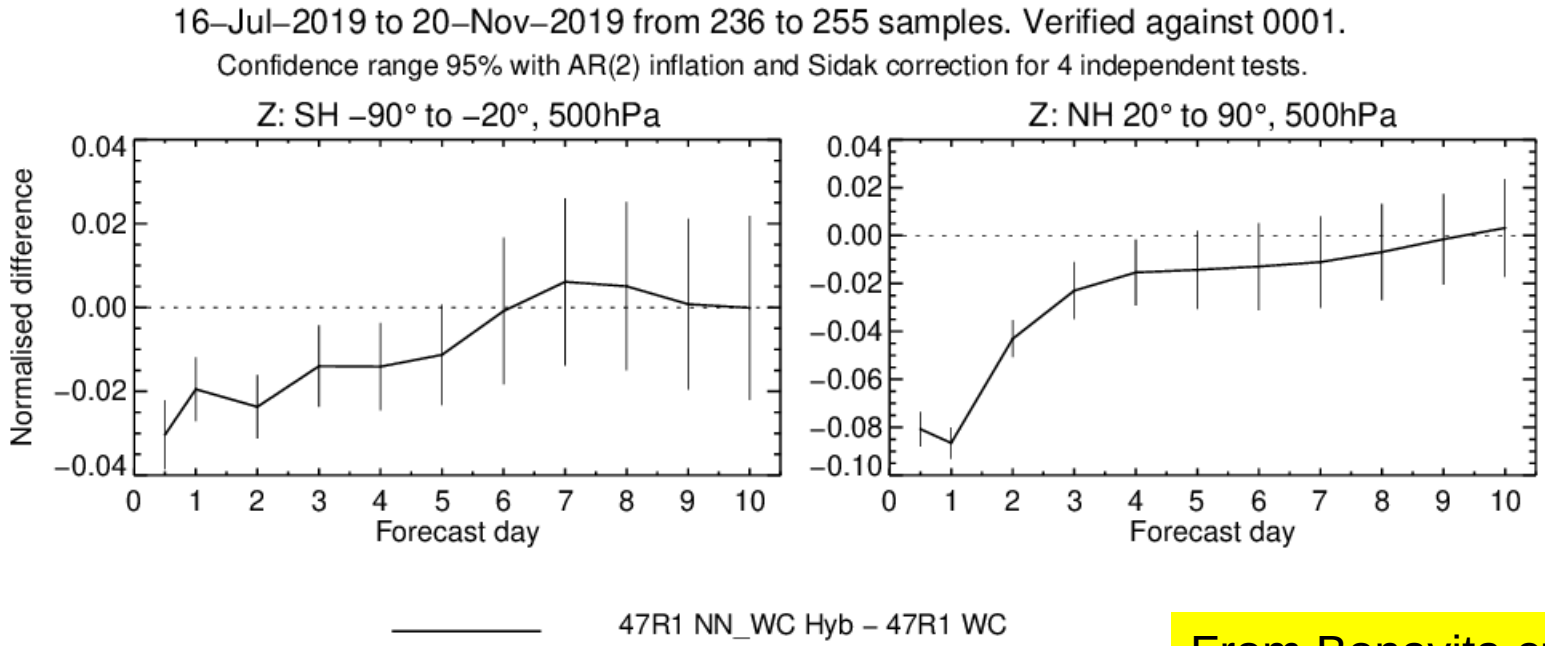
Note: 100% Baseline is current operational Weak Constraint 4D-Var

Model Error Estimation and Correction in the IFS: Forecast Skill

- We have seen that using a NN in combination with Weak Constraint 4DVar improves the fit of observations to the model, both in the mean and in the random part.

- What can the NN bring to tropospheric forecast skill?

Reduction in Z 500 hPa RMS forecast error of NN-WC_4DVar hybrid vs current operational IFS.



From Bonavita et al., in preparation

Take-home Messages

- Modern ML/DL can be interpreted in the more general framework of Data Assimilation Methodology. While ML/DL is not a conceptual innovation, we can usefully integrate some of its ideas and tools into DA and, more generally NWP/Climate prediction:

> "*ECMWF will strengthen its leadership position in data assimilation,…, 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*."
>
> ECMWF Strategy 2021-2030

- ML/DL can have multiple applications in NWP/Climate. Today we have presented ideas and initial results on how to hybridise current DA with ML/DL to either improve or correct the forecast model inside the DA cycle
- We have shown that significant improvements are already achievable today in analysis accuracy and medium-range forecast skill

Take-home Messages

- These are initial results and many research avenues are open: Can we use more sophisticated NN configurations? Can we use more/better predictors? Can we gain predictive power by increasing resolution?, etc.
- The results presented today were obtained using NN model error models only inside the Data Assimilation window. What happens if we apply it during the forecast on extended, sub-seasonal, climatic scales? To be continued…
- Improving the forecast model(s) is the key tool to try to increase predictability from medium-range to climate scales. Many in the NWP and Climate communities pin their hopes on step changes in model resolution, e.g. km scale NWP and climate simulations
- Data Assimilation and ML harness the huge and ever increasing amount of Earth System observations to provide a different, cost effective, evidence-based way of tackling these important societal problems

Thanks for your attention!

Bocquet, M., Brajard, J., Carrassi, A. and Bertino, L. , 2019: Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. Nonlinear Processes in Geophysics, 26 (3), 143–162. doi:10.5194/npg-26-143-2019.

Bocquet, M., Brajard, J., Carrassi, A. and Bertino, L. (2020). Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. Foundations of Data Science, 2 (1), 55–80. doi:10.3934/fods.2020004

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

Brajard, J., Carrassi, A., Bocquet, M. and Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the lorenz 96 model. Journal of Computational Science, 44, 101171. doi:10.1016/j.jocs.2020.101171.

Farchi, A., Laloyaux, P., Bonavita, M. and M. Bocquet, 2020: Using machine learning to correct model error and application to data assimilation with a quasi-geostrophic model, in preparation.

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

Continued…

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

Navon, I.M., 1998: Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography, Dynamics of Atmospheres and Oceans, Volume 27, Issues 1–4, 55-79

Ruiz, J.J., Pulido, M. and T. Miyoshi, 2013: Estimating Model Parameters with Ensemble-Based Data Assimilation: A Review, Journal of the Meteorological Society of Japan. Ser. II, 91, 2, 79-99. https://doi.org/10.2151/jmsj.2013-201

Sasaki, Y., 1970: Some basic formalisms on numerical variational analysis. Mon. Wea. Rev., 98, 875–883.