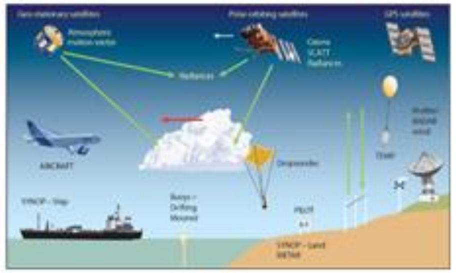
Machine Learning for Earth System Assimilation and Prediction (actually, this talk is about **model error**!)

Massimo Bonavita

Research Department, ECMWF
massimo.bonavita@ecmwf.int

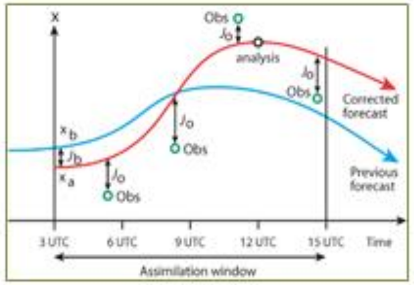
Ack.: Marcin Chrust[1], Patrick Laloyaux[1], Sebastien Massart[1], Alban Farchi[2], Marc Bocquet[2]

1: ECMWF
2: École des ponts ParisTech



Observations

- quality control
- adaptive thinning
- bias correction
- Fast observation models

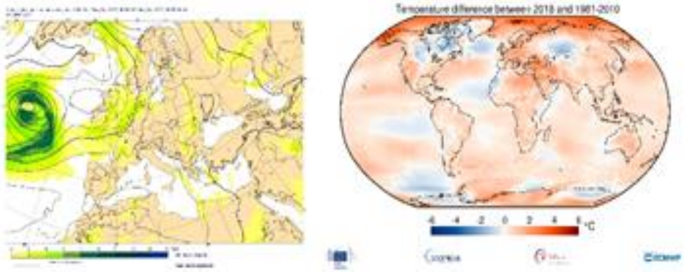


Data assimilation

- Model error estimation
- Model parameter estimation
- Error covariance statistics

Forecast Models

- Surrogate models
- Fast emulator models
- Linearised models for Var



Product generation

- information extraction (resolution, ensembles, weather features)
- data compression

From: Bonavita et al., 2021

Data Assimilation and Machine Learning

- Machine Learning / Deep Learning can be seen as a particularisation of standard Data Assimilation methodology, what is the current excitement about?

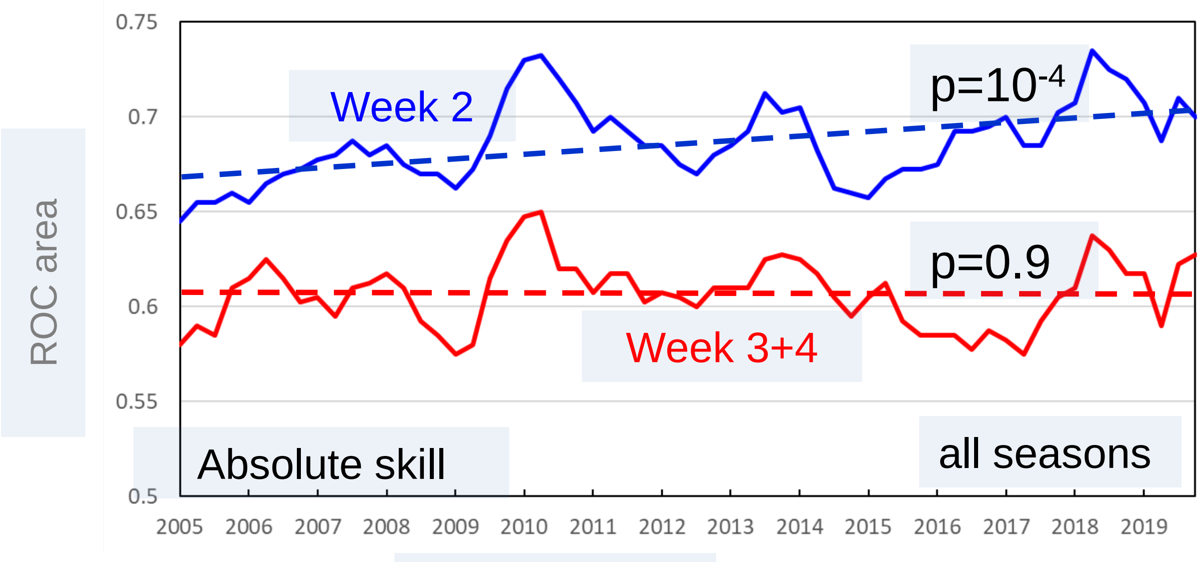
- At least two reasons:

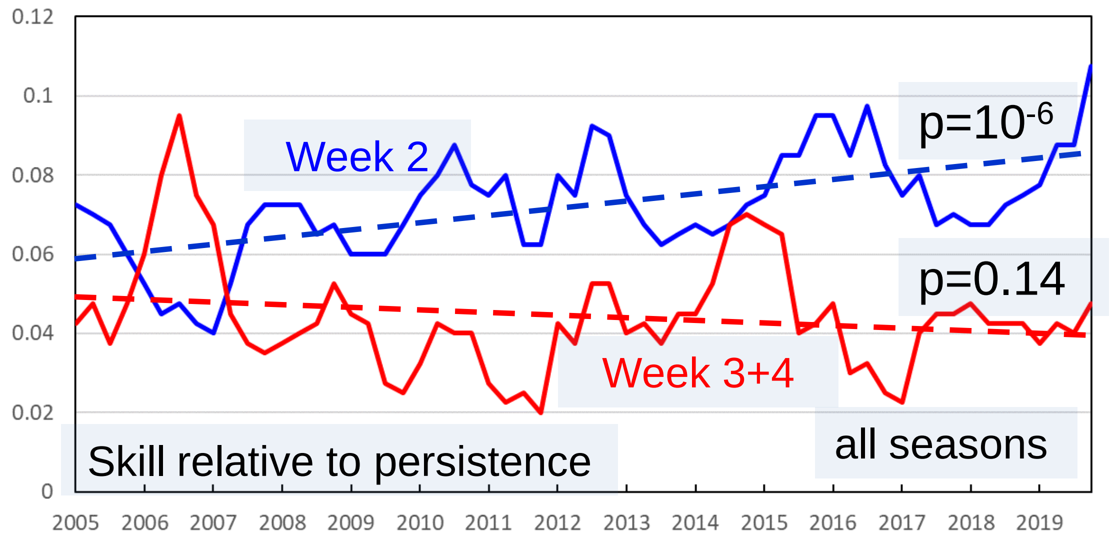
  1. Technically the explosion of ML/DL applications has produced an ecosystem of efficient, easy to use open source software that can be easily repurposed for specific applications: "the democratisation of Machine Learning"

  2. In the Earth Sciences, it has brought into prominence the idea that the current and ever-increasing wealth of **geophysical observations can be directly used to improve/extend/correct/replace current forecast models**

Model Error: Atmosphere

- **Improving forecast models** can be considered one of the most urgent requirements to advance predictive skill, on all time scales
- 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 help break the 2-week Lorenz predictability barrier?



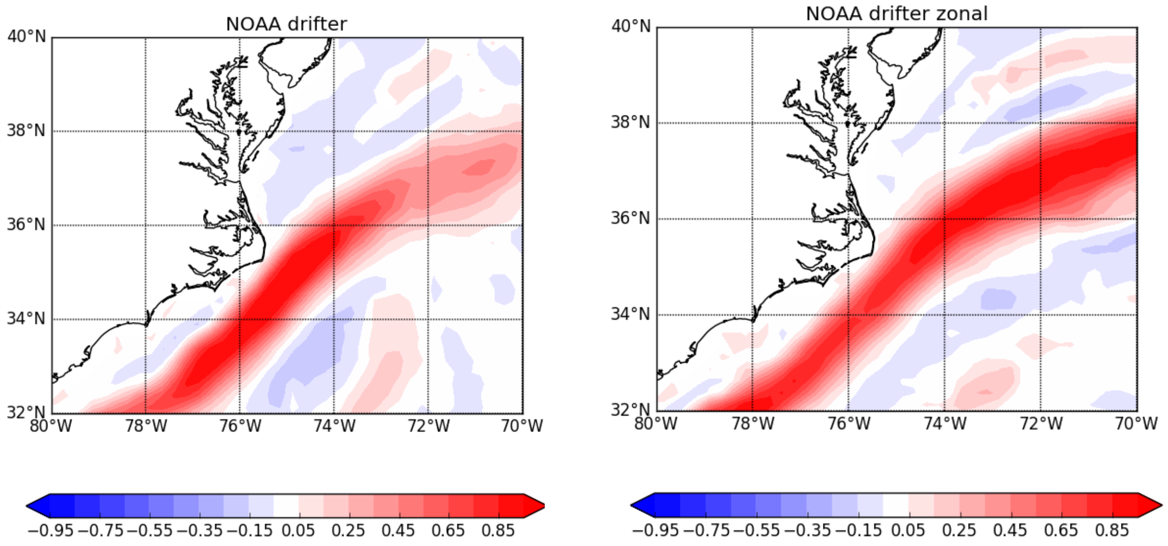


From Frederic Vitart and Thomas Haiden

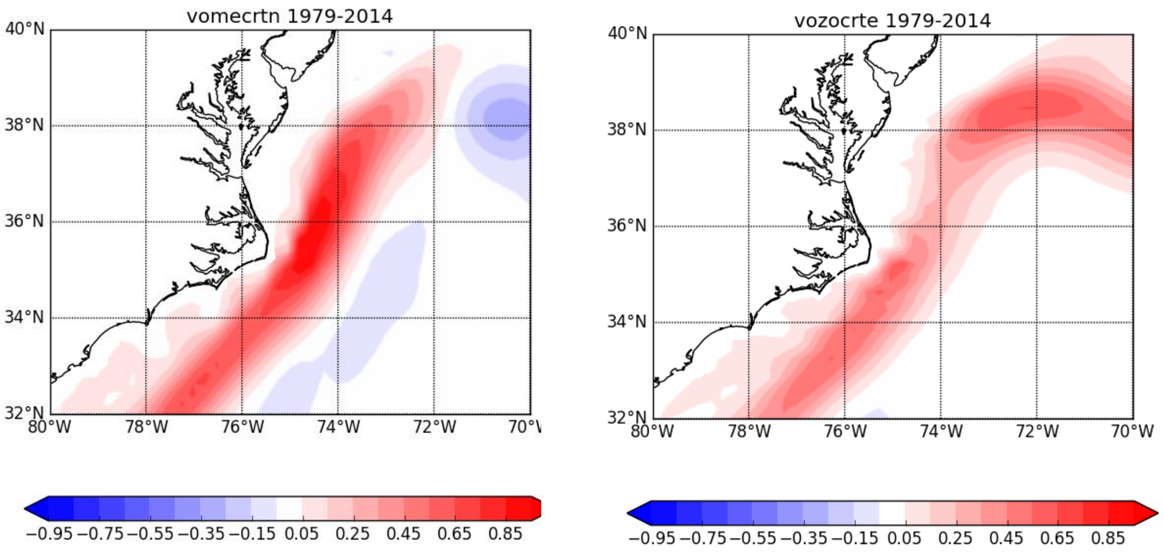
Model Error: Ocean

- **Improving forecast models** can be considered one of the most urgent requirements to advance predictive skill also in the Ocean!

NOAA drifter climatology



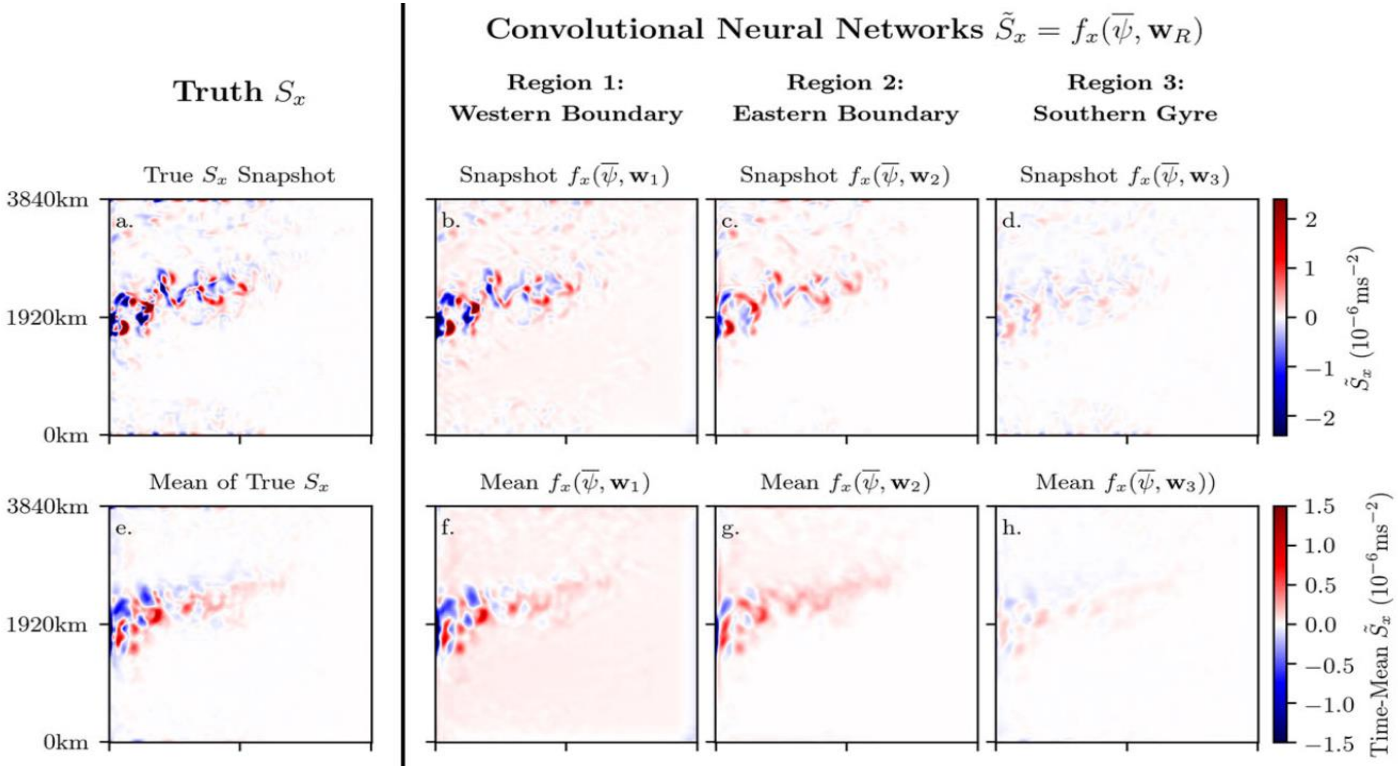
ORAS5 1985-2012.



From Hao Zuo

Model error in Ocean DA and prediction

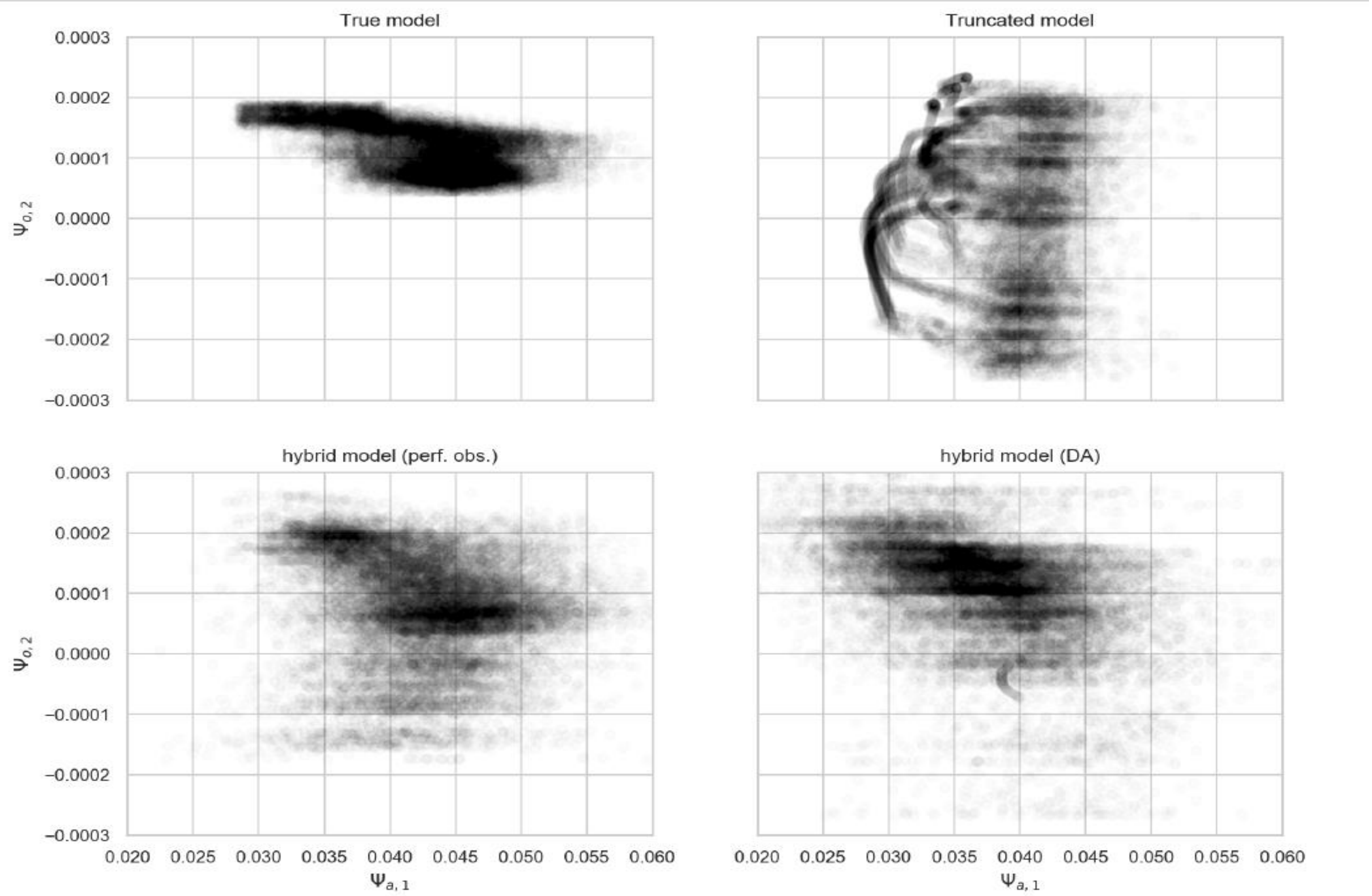
- In Ocean Modelling one of the main sources of errors come from insufficient resolution -> need to account for unresolved physical processes
- This problem has been tackled by using ML tools (CNN, DNN, RVM) to build offline data-driven parameterizations of the unresolved dynamics for a fully observed system (Bolton and Zanna, 2019, Zanna and Bolton, 2020, Kutz, 2017, Ling et al., 2016)



From: Bolton and Zanna, 2019

Model error in Ocean DA and prediction

- A proof of concept with a simplified coupled Ocean-Atmosphere model has been given in Brajard et al., 2020, to obtain an online data-driven parameterizations of the unresolved dynamics using incomplete and noisy observations (i.e., using data assimilation for state estimation)



From: Brajard et al., 2020

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 (e.g. Brajard et al., 2020, Bocquet et al., 2020, Bocquet et al., 2019) using a coordinate descent approach:

FARCHI ET AL. 5

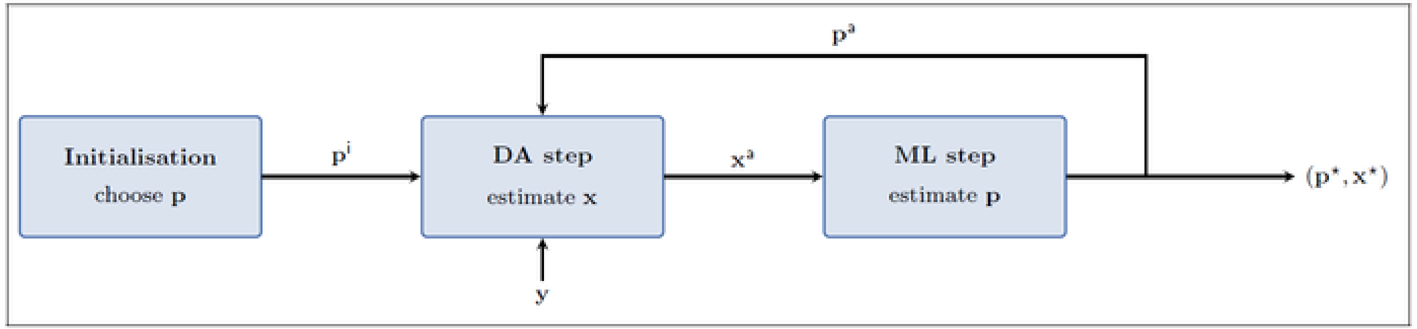


FIGURE 1 Minimisation strategy for Equation (4): alternate DA steps with ML steps to estimate the parameters $\mathbf{p}$ and the trajectory $\mathbf{x}_{0:N_t}$ with an increasing accuracy.

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. Typically we do not have enough time to iterate to convergence the DA and ML steps in operational NWP;
  2. We have a much more complex model, but typically a very good model!
- These two features of the NWP/Climate prediction system point to two ways of attacking the combined state-model estimation problem in a data assimilation framework:
  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 pdfs of these parameters, 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 hand-tuned 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:

$$J_{4DVar}(\boldsymbol{x}_0, \boldsymbol{p}) = J_B + J_P + J_O = \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^T \mathrm{B}^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) + \frac{1}{2}\left(\boldsymbol{p} - \boldsymbol{p}^b\right)^T \mathrm{B}_\mathrm{p}^{-1}\left(\boldsymbol{p} - \boldsymbol{p}^b\right) +$$

$$\frac{1}{2}\sum_{i=0}^{N}(H_i \circ M_i(\boldsymbol{x}_0; \boldsymbol{p}) - \boldsymbol{y}_i)^T \mathrm{R}_\mathrm{i}^{-1}(H_i \circ M_i(\boldsymbol{x}_0; \boldsymbol{p}) - \boldsymbol{y}_i)$$

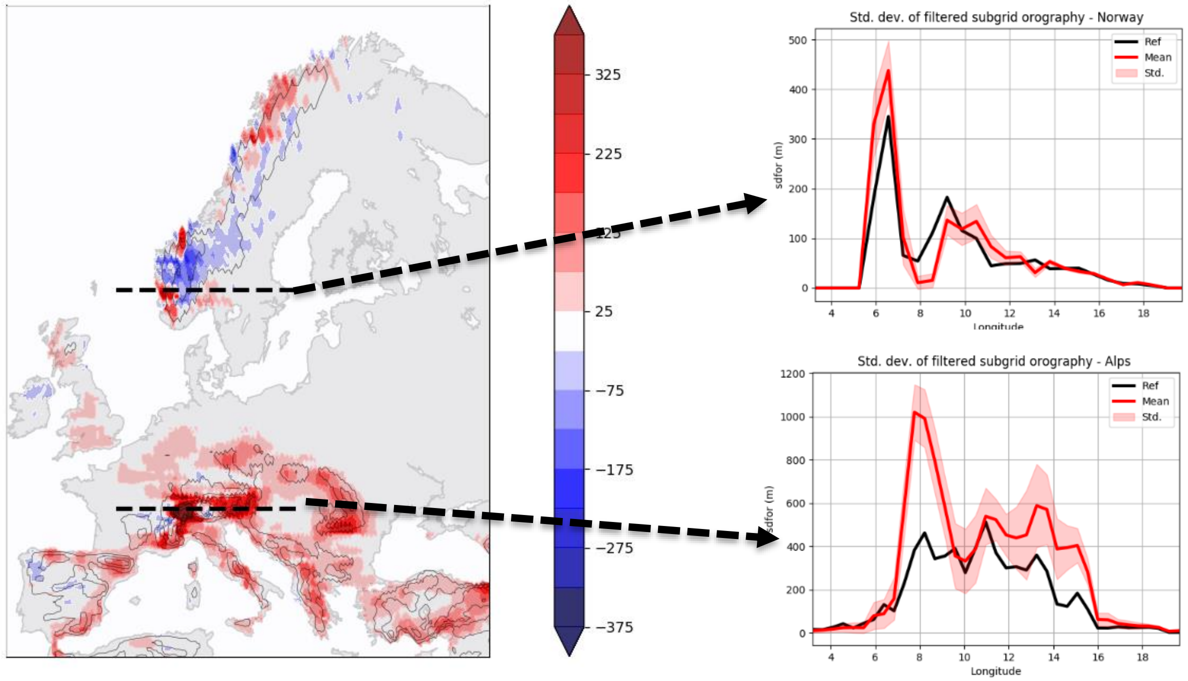
- ($\boldsymbol{p}$) is the set of model parameters (or fields of those) that are estimated together with the state. (This could also be applied to uncertain parameters of the forward model H)

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

- With the current and projected wealth of available observations (ECMWF DA ingests ~40 million obs every 12h) 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;

- A different 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 general idea is called weak constraint 4DVar (Sasaki, 1970)

Weak Constraint 4D-Var

- The ECMWF version of WC-4DVar works by estimating ("learning") a constant model error tendency over the assimilation window (12 hours)

- The background (and first guess) model error is the previous window's estimate (we do not have a predictive model of model error!)

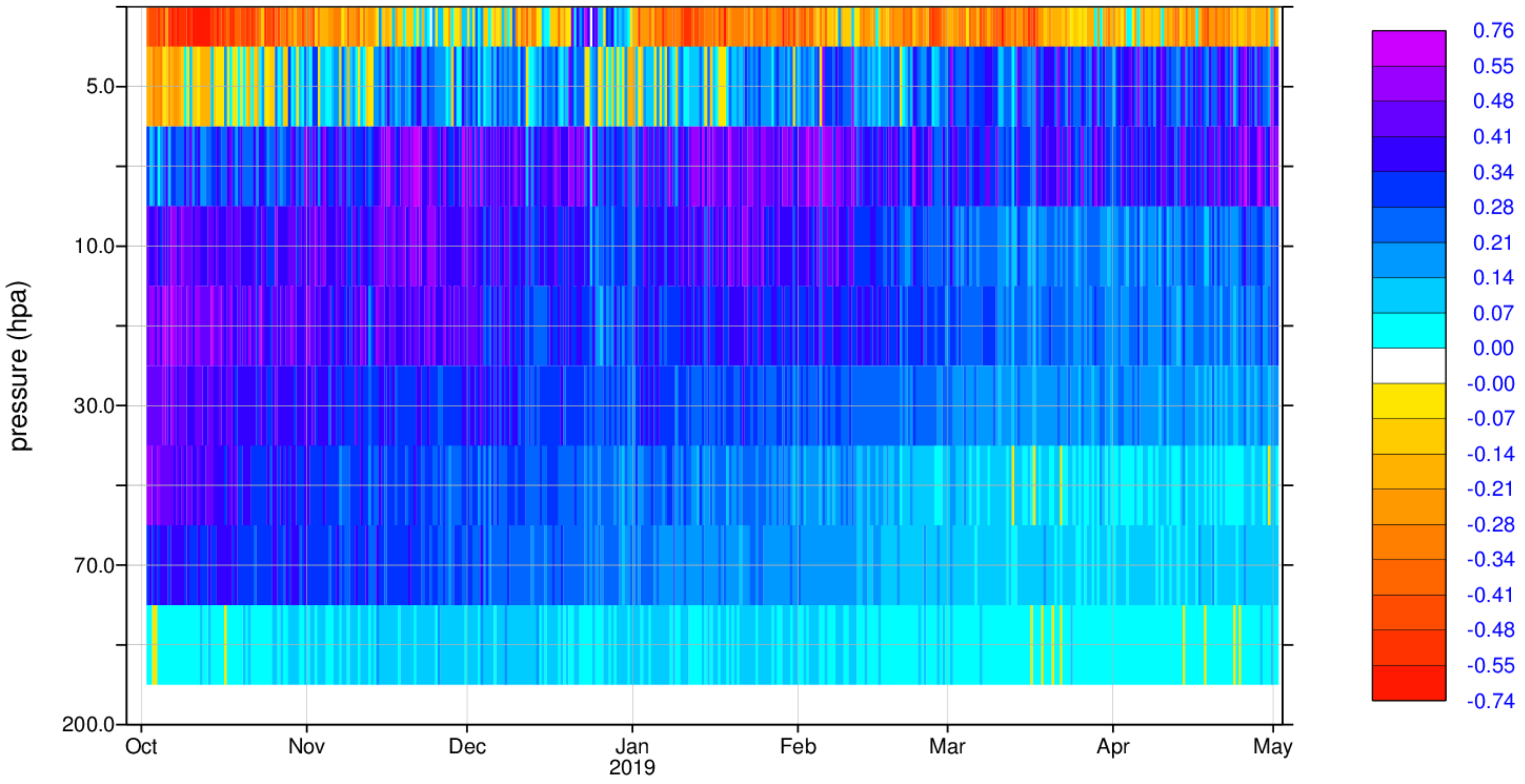
- This means that our model error formulation targets errors which are slowly evolving on the timescale of the assimilation window (ie, 12 hours) and large scale with respect to typical background errors ("errors of the day")

- 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

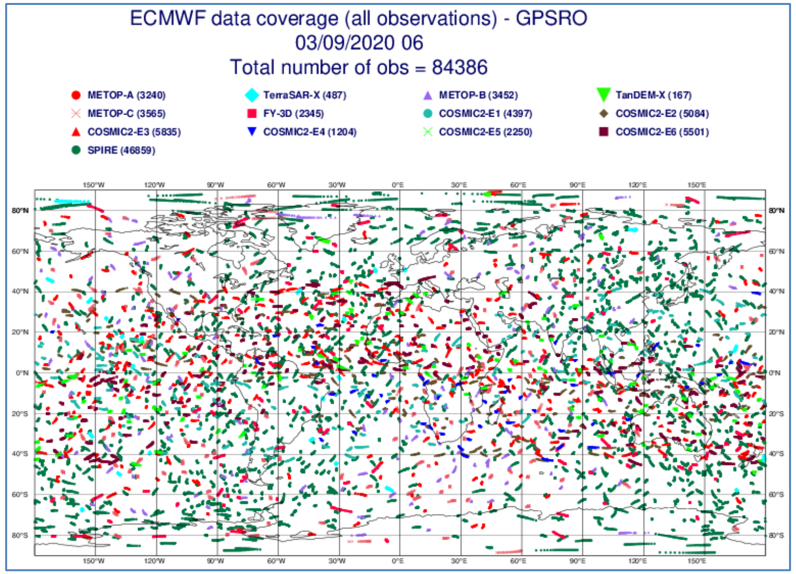
Weak Constraint 4D-Var

- Current version of WC-4DVar is a step change wrto previous versions, but it is not the end of the story:
    - 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 in the stratosphere (above ~100hPa). Letting it loose on full atmospheric column leads to some forecast skill degradation in troposphere

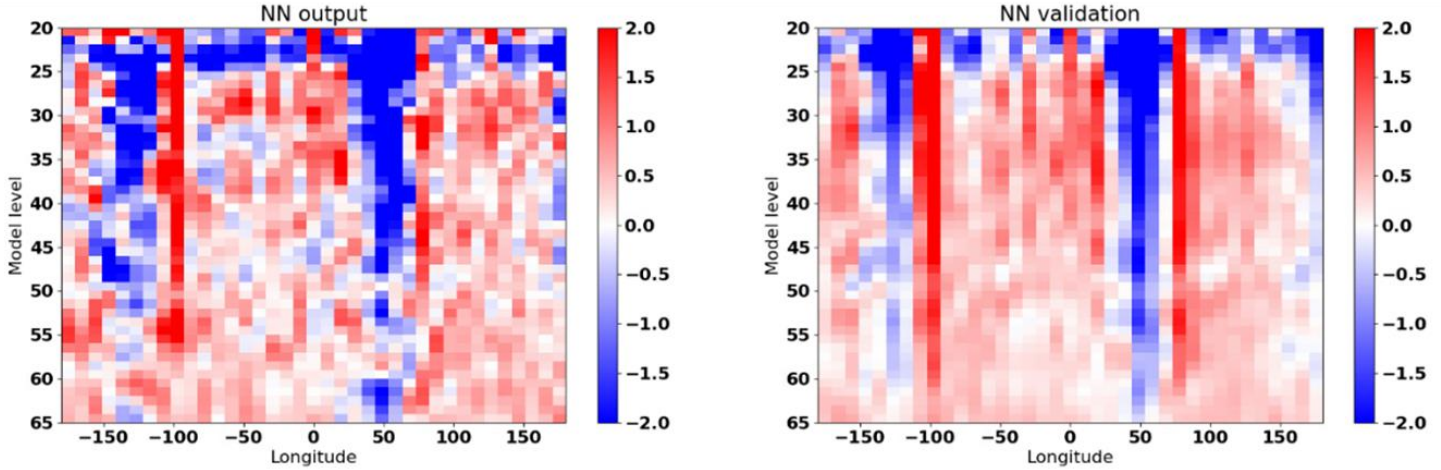
Combining DA and ML: Model Error Estimation and Correction

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

- Basic idea: Train an Artificial Neural Network to “learn” the deterministic. predictable part of model error
- The most direct way is to train the ANN on a database of observation-background (O-B) departures, choosing an ~ homogeneous, dense and unbiased observing system GPS-RO



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

- 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*

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

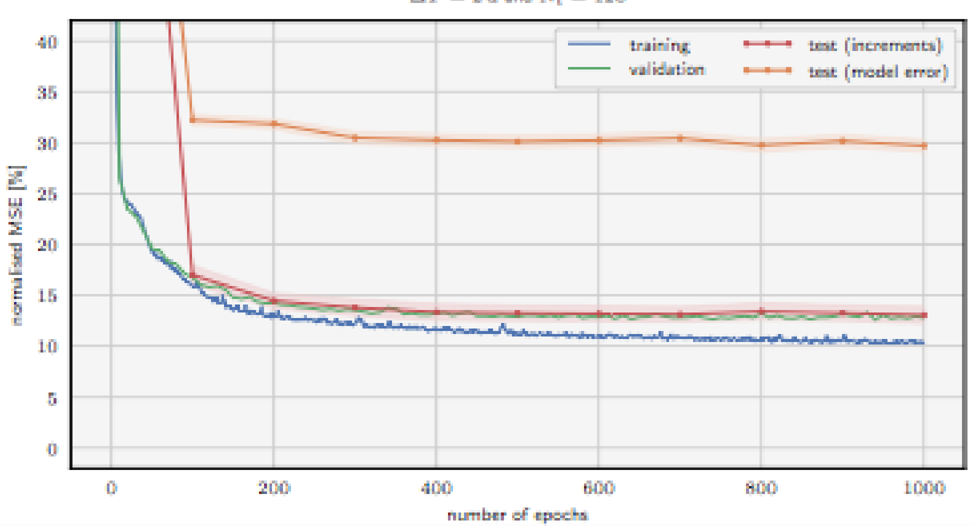
$$\mathbf{dx}_k = \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{h}\left(\mathbf{x}_k^b - \mathbf{dx}_k^p\right)\right) \qquad (44)$$

will correct the model background and produce unbiased analyses."

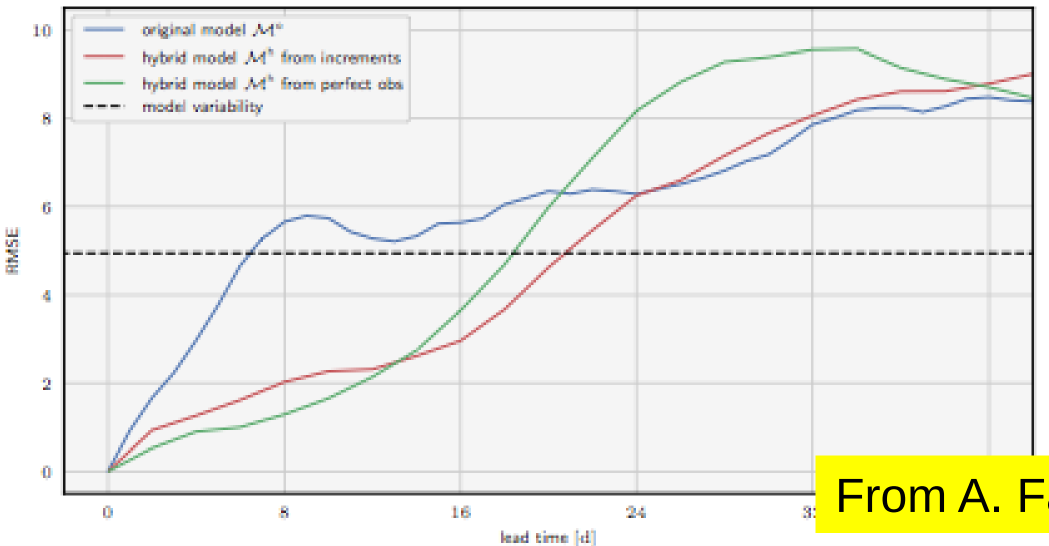
Combining DA and ML: Model Error Estimation and Correction

- This approach ahs been tested in a low order 2-level, 1600 variables QG model (Farchi et al., 2021)
- 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 relevant are these amazingly good results for operational-scale assimilation and modelling systems?

Training and test curves for the ANN



Evolution of the fcst RMSE



ORECAST

From A. Farchi

Combining DA and ML: Model Error Estimation and Correction

- This approach has also been demonstrated in the operational version of the ECMWF IFS, initial results are available in Bonavita and Laloyaux, 2020

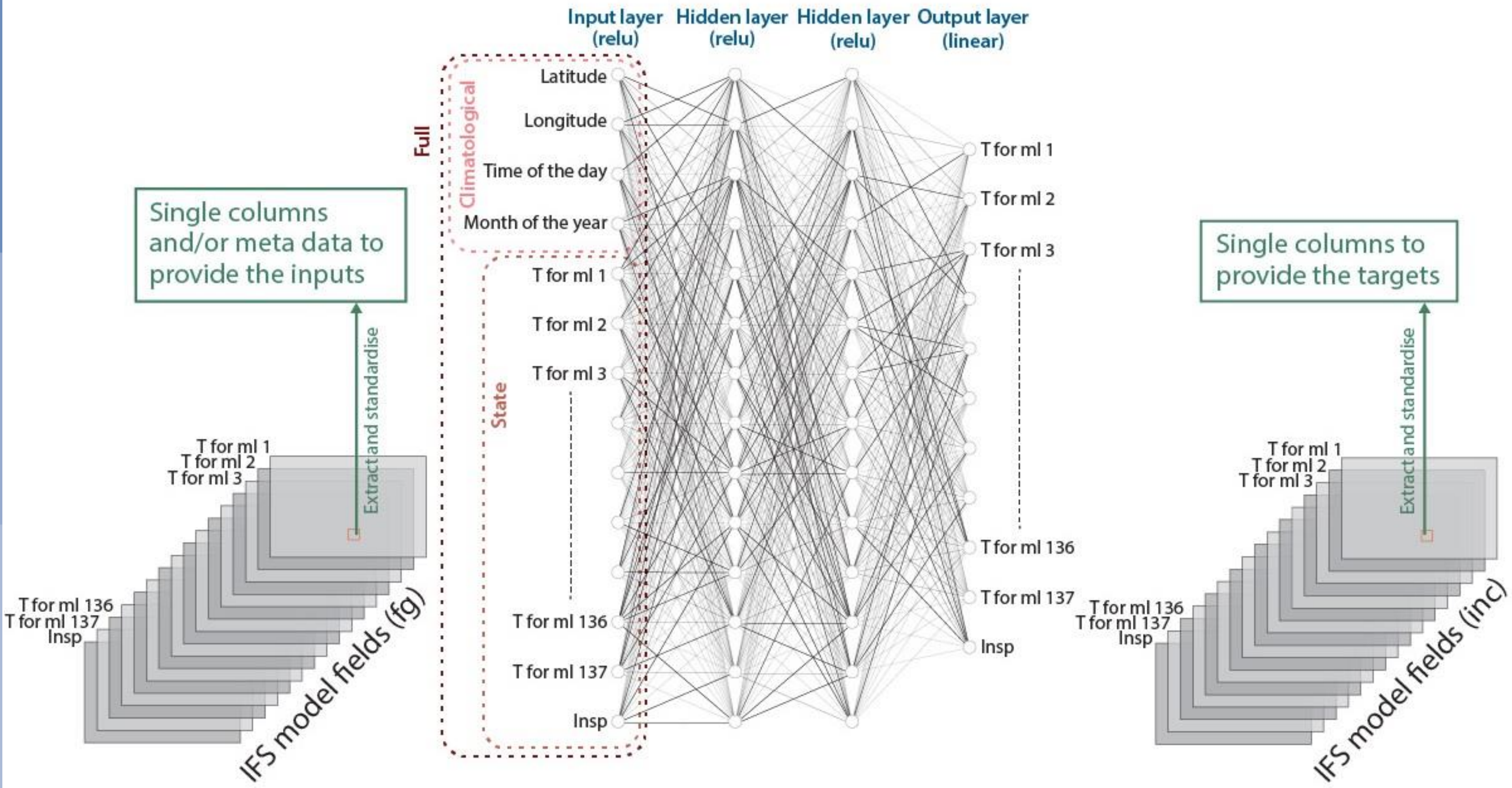
- The size of the state vector is $O(10^{10})$. Even though model error has much lower effective dimension, a primary design consideration for the ANN has been to reduce its size

- 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*.

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

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

- This work focussed on two different configurations:

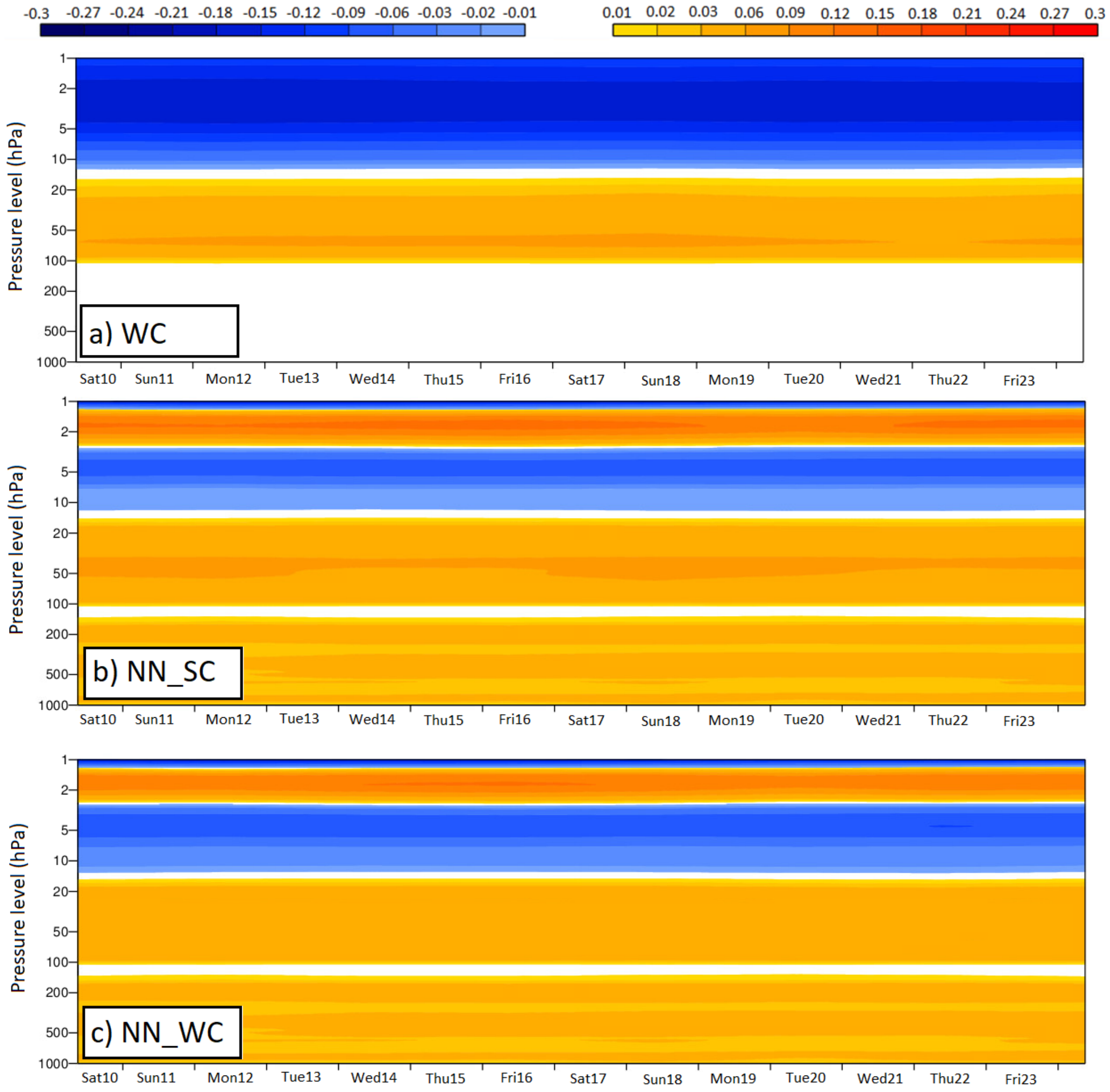
  a) Use the NN model errors instead of the Weak Constraint 4D-Var model errors (NN_SC in the following);

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

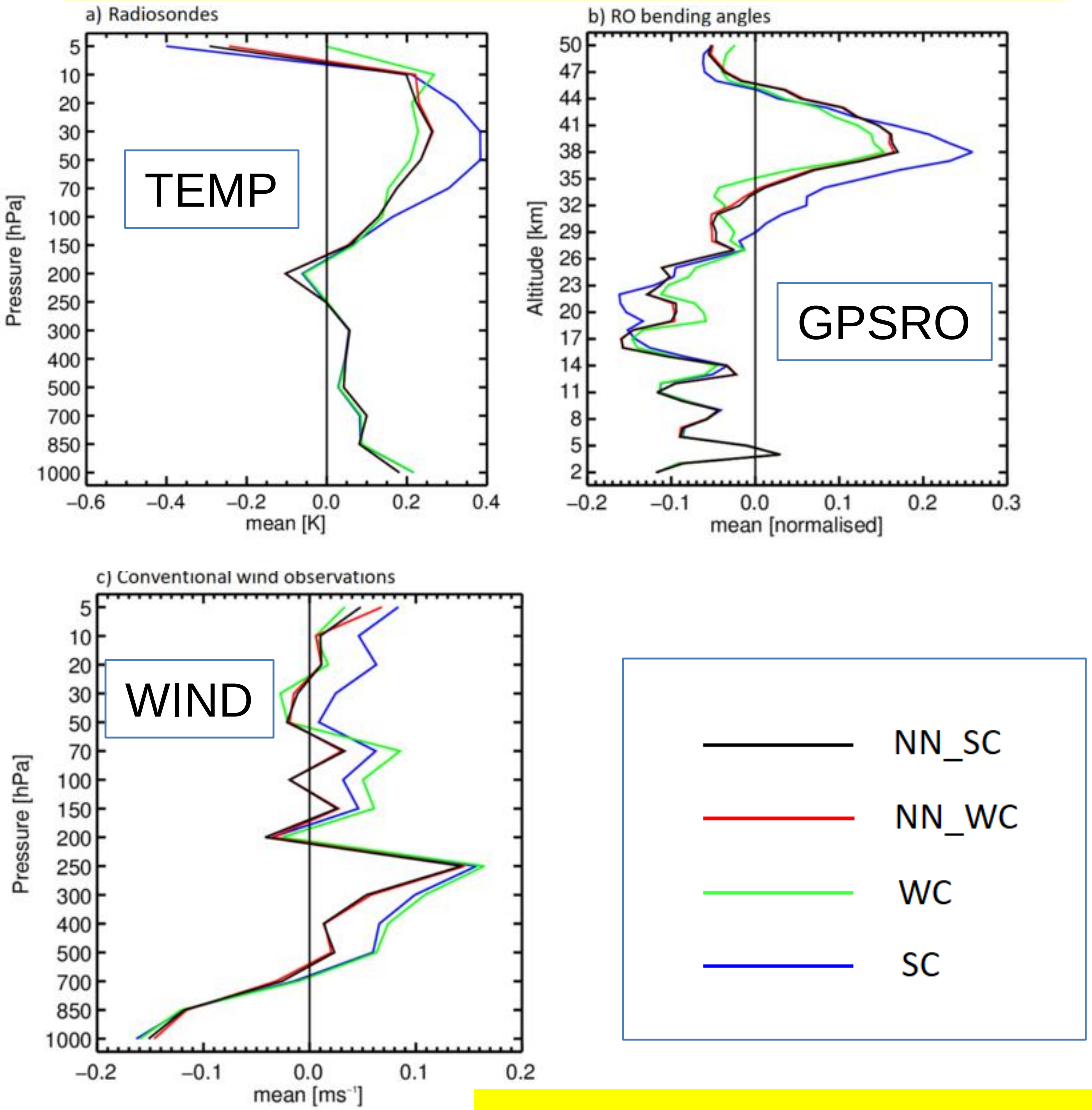
- Baseline configuration was the currently operational version of Weak Constraint 4D-Var at full resolution (T1279, ~9km grid spacing) (WC) and the previously operational Strong Constraint 4D-Var (SC)

Model Error Estimation and Correction in the IFS: Mean errors

Globally-averaged Mean Temperature Correction



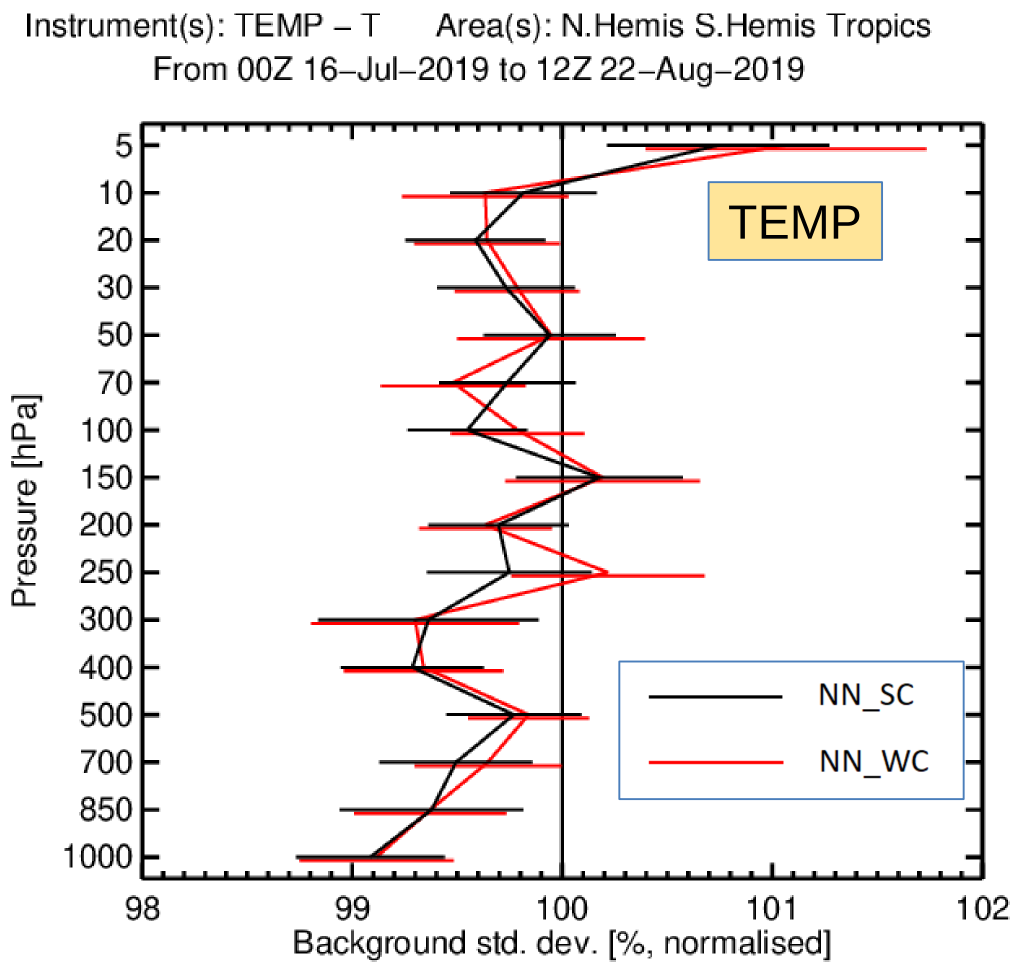
Globally-averaged Obs-Fg Mean Difference

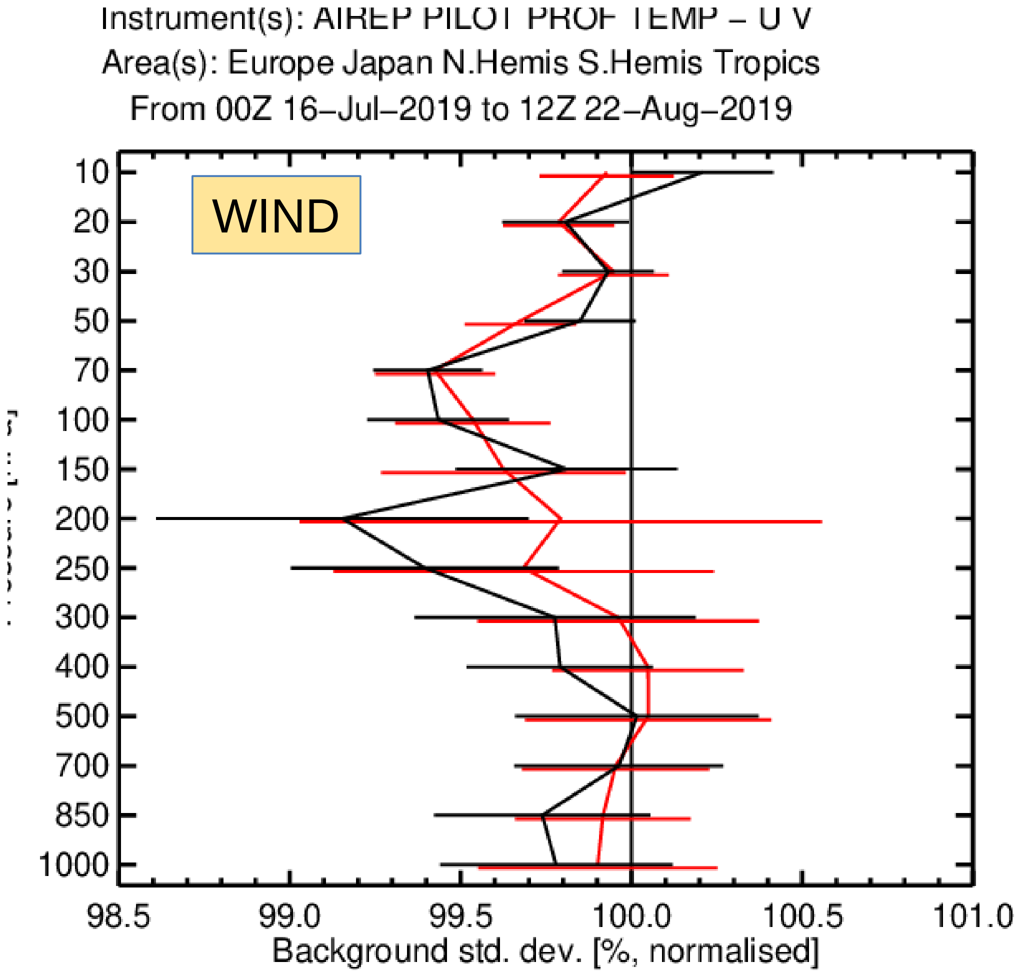


From Bonavita & Laloyaux, 2020

Model Error Estimation and Correction in the IFS: Random errors

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



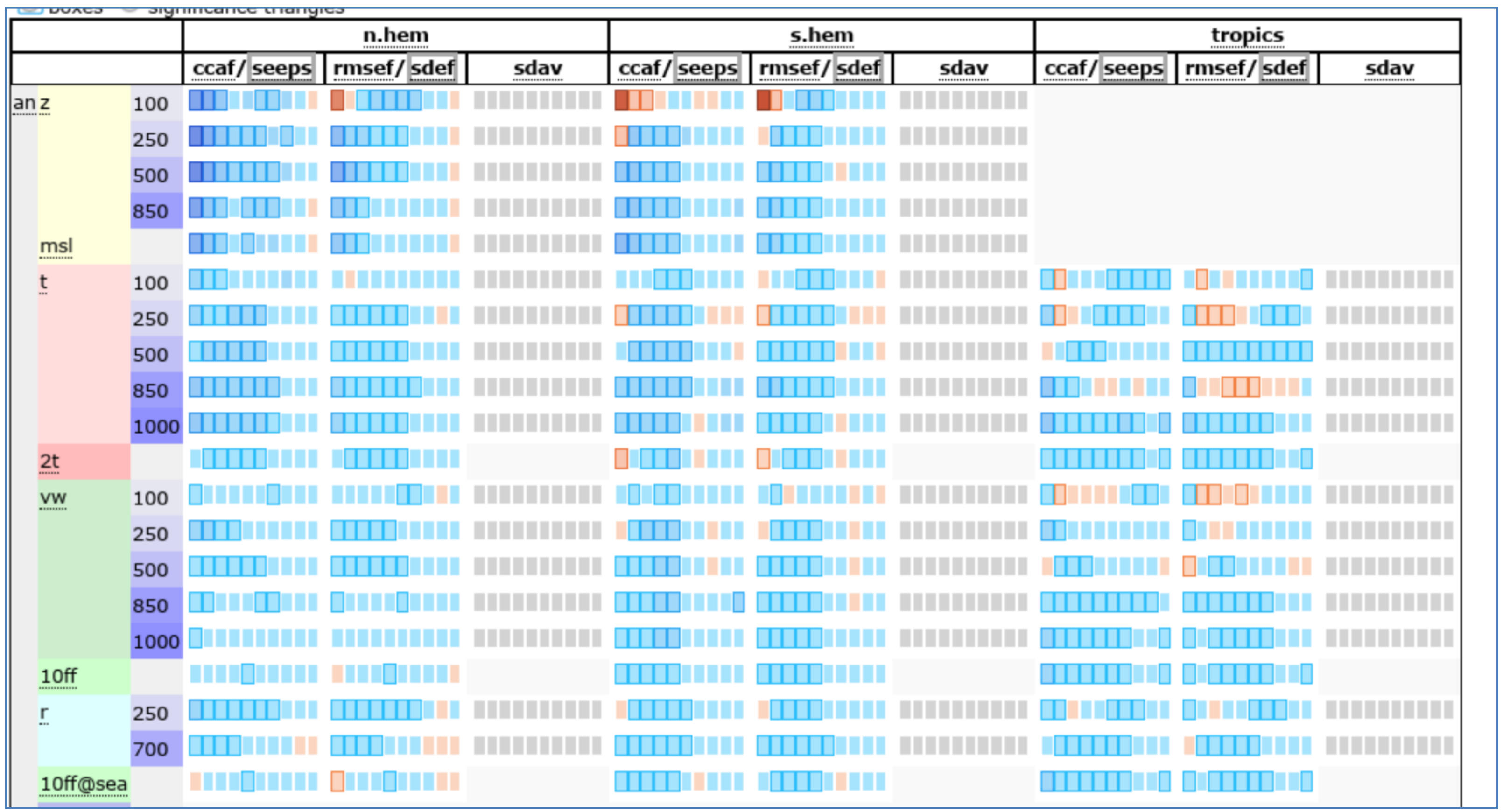


From Bonavita & Laloyaux, 2020

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

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?



Improvement

Degradation

From Chrust, Bonavita et al., in preparation

Model Error Estimation and Correction for Ocean DA

- The IFS examples of dealing with model error in DA are based on:
    1. Conceptually separating predictability errors, which we model through ensemble DA, and systematic model errors which have longer (in time and space) decorrelation lengths and are dealt with by a model error term in the cost function
    2. Estimating and correcting systematic model errors with WC-4DVar or hybrids of WC-4DVar and Machine Learning
- Wrto error scale separation, similar ideas have been proposed in the Ocean DA literature, e.g. dual (or multiple) B matrix in 3DVar (Mirouze et al., 2016, Lin et al., 2015)
- Mirouze et al., 2016 find that “The main impact of the dual length scale formulation can be seen on the bias of the innovations”, which is significantly reduced. This results is interpreted as coming from the “the usefulness of fatter tails in the background error correlation functions”
- Another interpretation is that these longer-range errors come from model/boundary conditions forcing and thus could be cured in terms of corrections to the model/BC error tendencies instead of corrections to the initial state

Model Error Estimation and Correction for Ocean DA

- Interestingly, in NEMOVAR implementations the analysis increments are applied as an additional forcing term to the prognostic equations ((IAU; Bloom et al.,1996)
- This is similar to how model error estimates from WC-4DVar are used in the DA cycle
- Additionally, there is an explicit treatment of slow-evolving model biases in NEMOVAR based on Balmaseda et al., 2007, Mogensen et al., 2012 where the model bias is parameterised as:

$$\mathbf{b}_c = \bar{\mathbf{b}} + \mathbf{b}'_c \qquad (1)$$

$$\mathbf{b}'_c = \alpha\mathbf{b}'_{c-1} - \mathbf{A}\delta\mathbf{x}^a_{c-1} \qquad (2)$$

- Model bias computed as sum of an a-priori (offline) bias term ($\bar{\mathbf{b}}$), which contains seasonal dependencies, plus an online term ($\mathbf{b}'_c$) which is updated based on the previous analysis increments
- The bias correction is then used to modify the tendencies of the nonlinear model in the DA cycle, as in WC-4DVar!

Take-home Messages

- We are in the Big Data Era! There is an unprecedented and rapidly growing amount of geophysical data ready to be used (at ECMWF we use ~80 million obs a day in 4DVar: that is less than 5% of the total amount of obs that reach the building every day!)
- This suggests that we have enough data not only to improve the initial state, but also the models (both forecast model and forward models)
- In atmospheric DA we are already on this path, using WC-4DVar to correct model systematic errors in the stratosphere.
- WC-4DVar to tackle systematic errors in the troposphere: prospects look good, both with pure WC-4DVar and with a hybrid Machine Learning-4DVar approaches
- Other applications of ML are also gaining traction in DA-NWP: model parameter optimisation, correcting model errors in extended/seasonal/climate prediction, observation operators etc
- All these ideas are very relevant to Ocean DA because problems are similar in nature and so similar methodological solutions can be successfully applied

Thank you for your attention!

ECMWF-ESA Workshop on ML for ESOP, 5-8 October 2020
(https://events.ecmwf.int/event/172/; Bonavita et al., 2021, BAMS)

ESA-ECMWF Workshop on ML for ESOP, 15-18 November 2021

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

Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. Journal of Advances in Modeling Earth Systems, 11, 376– 399. https://doi.org/10.1029/2018MS001472

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

Bonavita, M., Arcucci, R., Carrassi, A., Dueben, P., Geer, A. J., Le Saux, B., Longépé, N., Mathieu, P., & Raynaud, L. (2021). Machine Learning for Earth System Observation and Prediction, Bulletin of the American Meteorological Society, 102(4), E710-E716

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.

Continued…

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

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, Q.J.R. Meteorol. Soc., under review.

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

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.