Correcting model error with an online Artificial Neural Network

Marcin Chrust, Massimo Bonavita and Patrick Laloyaux

ECMWF

Marcin.chrust@ecmwf.int

ECMWF Earth System Model - IFS

Systematic forecast errors are a limiting factor for predictive skill.

46r1 Improvements in the convection and radiation schemes

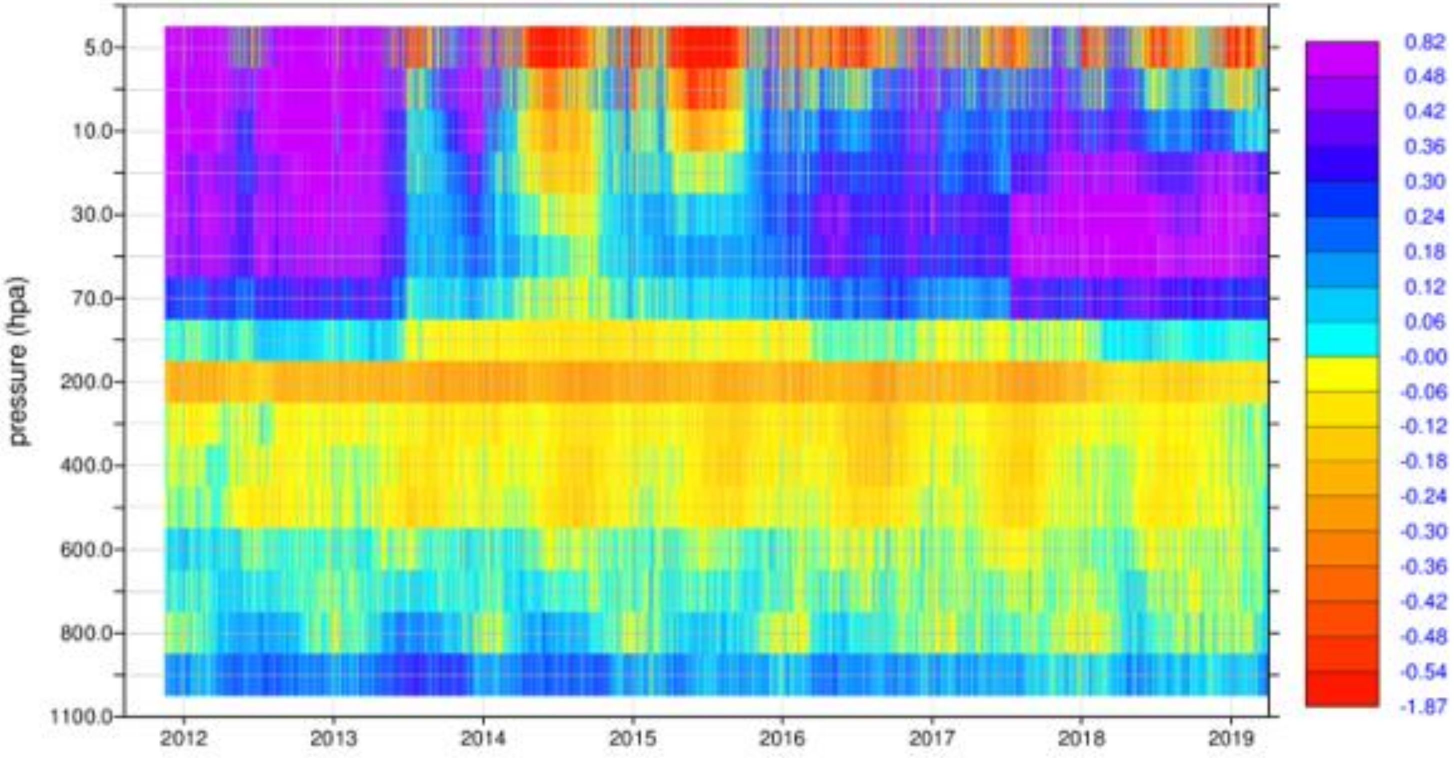
47r1 Quintic vertical interpolation in semi-Lagrangian advection

47r3 Moist physics upgrade

48r1 Hybrid linear ozone, semi-Lagrangian vertical filter and new solar spectrum?

Model bias diagnostics

Difference between the 12-hour model trajectory with reference observations (radiosondes)

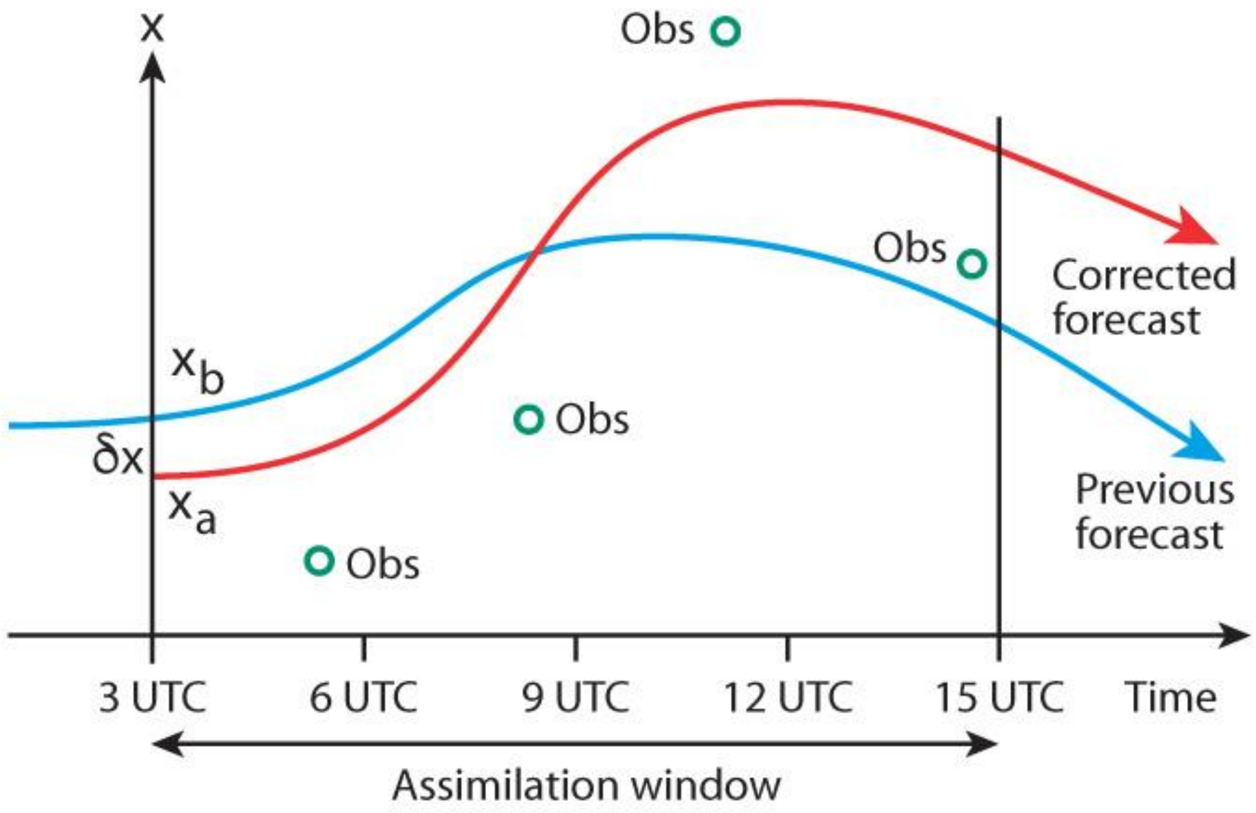


Systematic errors when the atmospheric model is integrated over 12 hours

➔ Cold bias in the mid/lower stratosphere (>0.5C)

➔ Warm bias in the upper stratosphere (>0.5C)

Model bias can be estimated and accounted for by 4D-Var



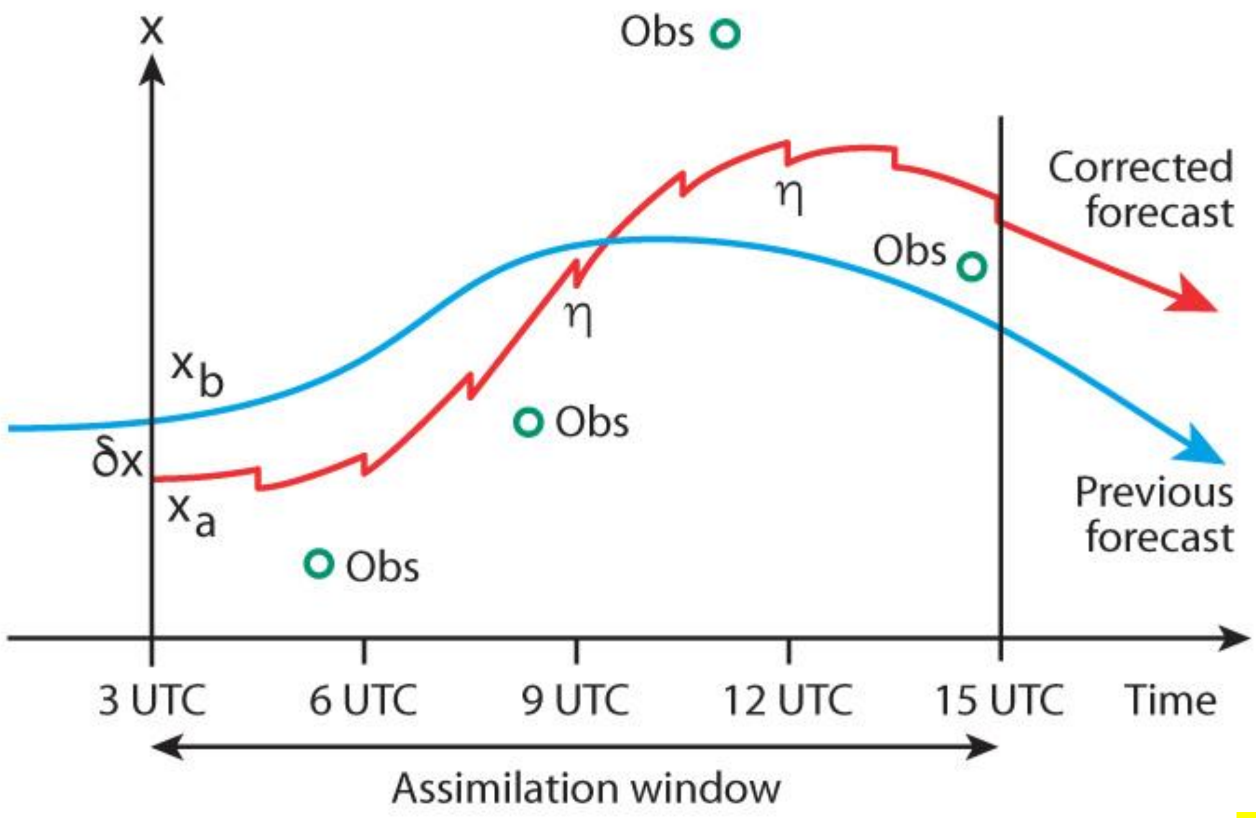
Model equation:

$$\mathbf{x}_k = \mathcal{M}_{k:0}(\mathbf{x}_0)$$

Strong constraint 4D-Var cost function:

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

Model bias can be estimated and accounted for by 4D-Var



Model equation:

$$\mathbf{x}_k = \mathcal{M}_{k:0}(\mathbf{x}_0, \eta)$$

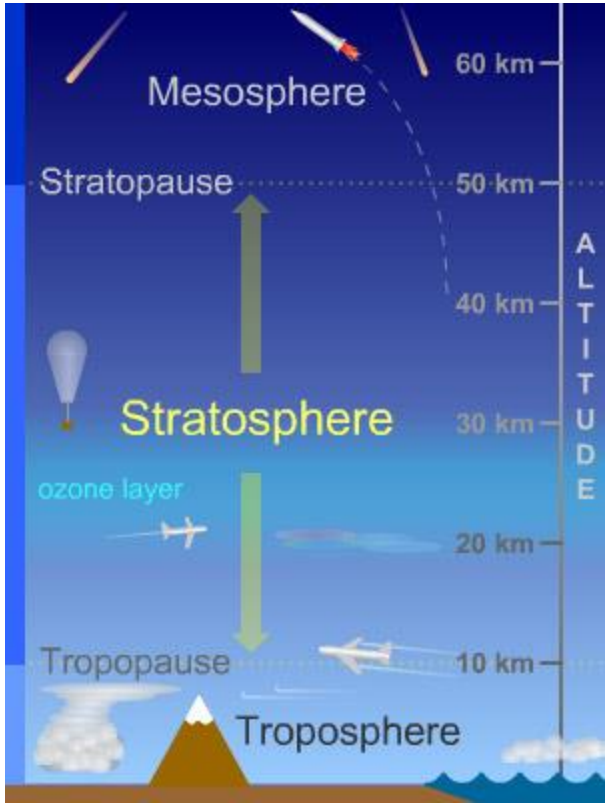
Weak constraint 4D-Var cost function:

$$\begin{aligned}\mathcal{J}(\mathbf{x}_0, \eta) \triangleq\ & \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}^{-1}} + \\ & \frac{1}{2}\|\eta - \eta^b\|^2_{\mathbf{Q}^{-1}} + \\ & \frac{1}{2}\sum_{k=0}^{L}\|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0, \eta)\|^2_{\mathbf{R}_k^{-1}},\end{aligned}$$

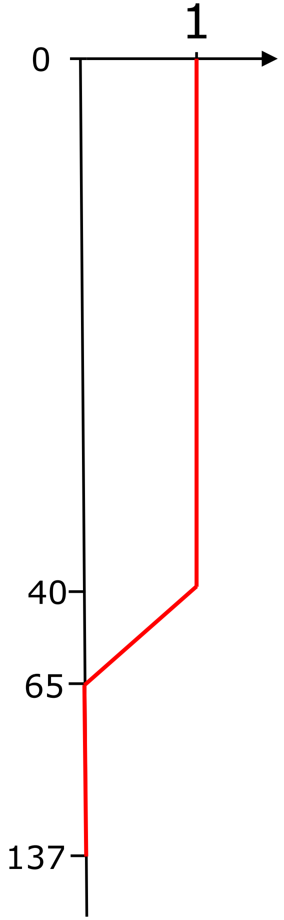
Trémolet, Model-error estimation in 4DVar, 2007

Weak constraint 4D-Var implementation in the IFS

- ➔ The model error forcing estimate η contains 4 fields (temperature, surface pressure, vorticity and divergence);
- ➔ This introduces additional degrees of freedom for 4D-Var to fit background and observations;
- ➔ A WC-4DVar analysis trajectory is entirely determined by its initial condition and the <u>constant</u> model error forcing;
- ➔ Scale separation between background and model errors is a crucial assumption;
- ➔ The model error forcing fields are only estimated in the Stratosphere. Applied to remove large-scale temperature stratospheric biases.



Credit: Randy Russell, UCAR

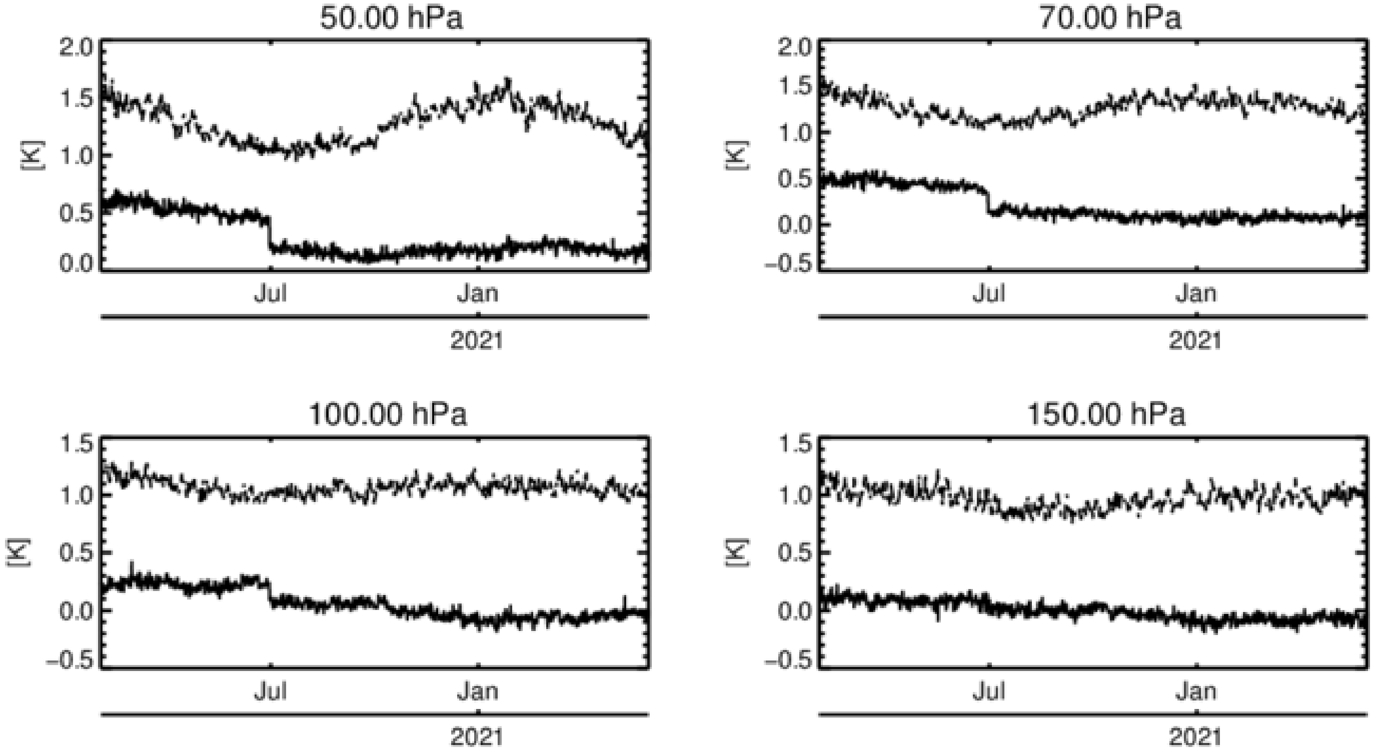


Model Levels

Laloyaux et al., Exploring the potential and limitations of weak-constraint 4D-Var, 2020

Operational weak constraint 4D-Var performance and limitations

Mean first-guess departure with respect to temperature measurements from radiosondes

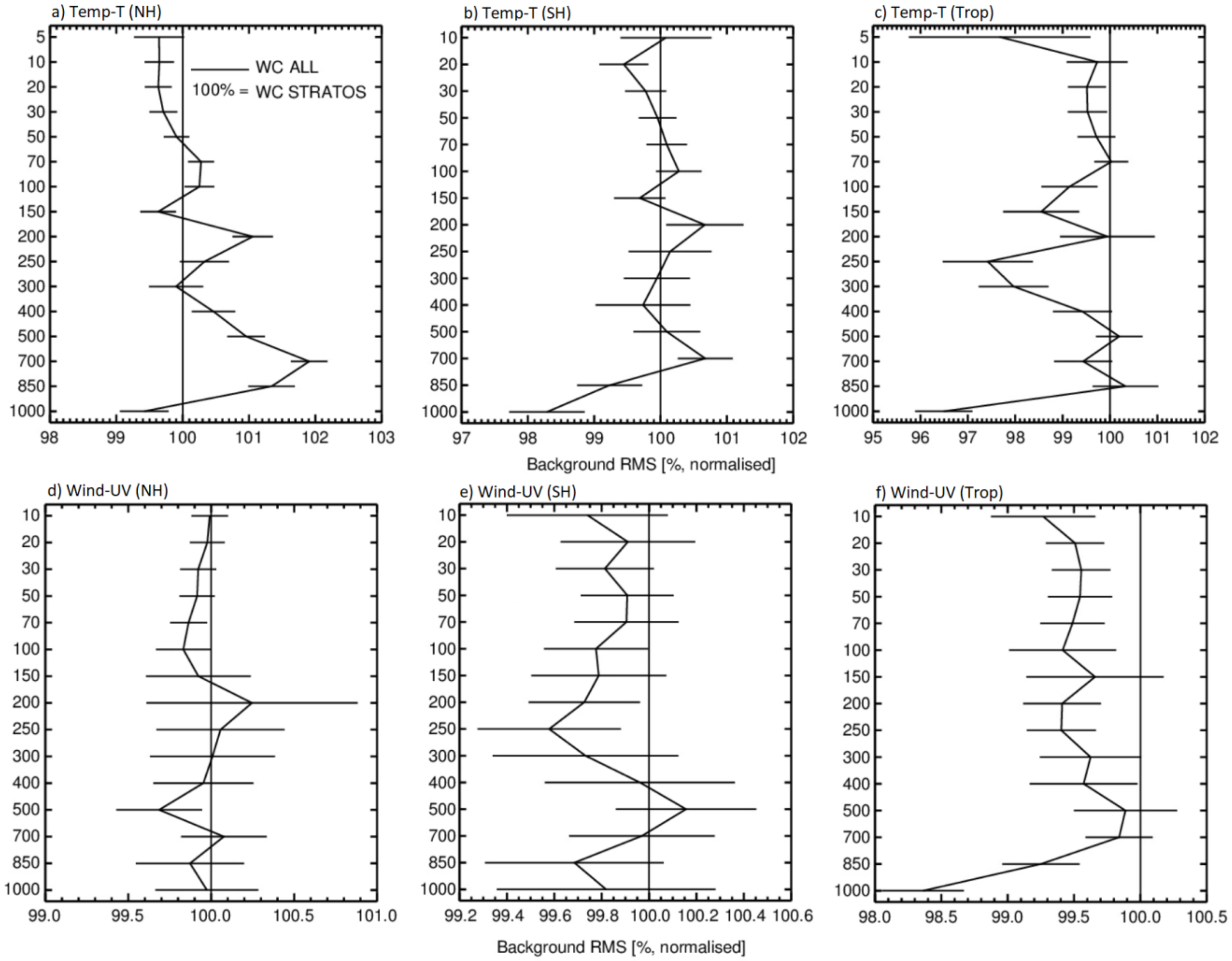


→Bias reduced up to 50%;

→Weak-constraint 4D-Var in EDA has been implemented in 47r3 (similar reduction in biases).

This technique is used operationally since 30 June 2020 to correct the stratospheric biases in the HRES system.

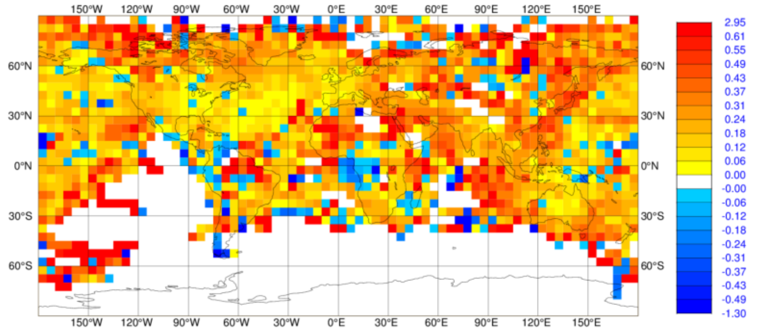
Ongoing work to extend the model error to troposphere



Significant wind improvements in Tropics

Significant temperature improvements in Tropics and SH

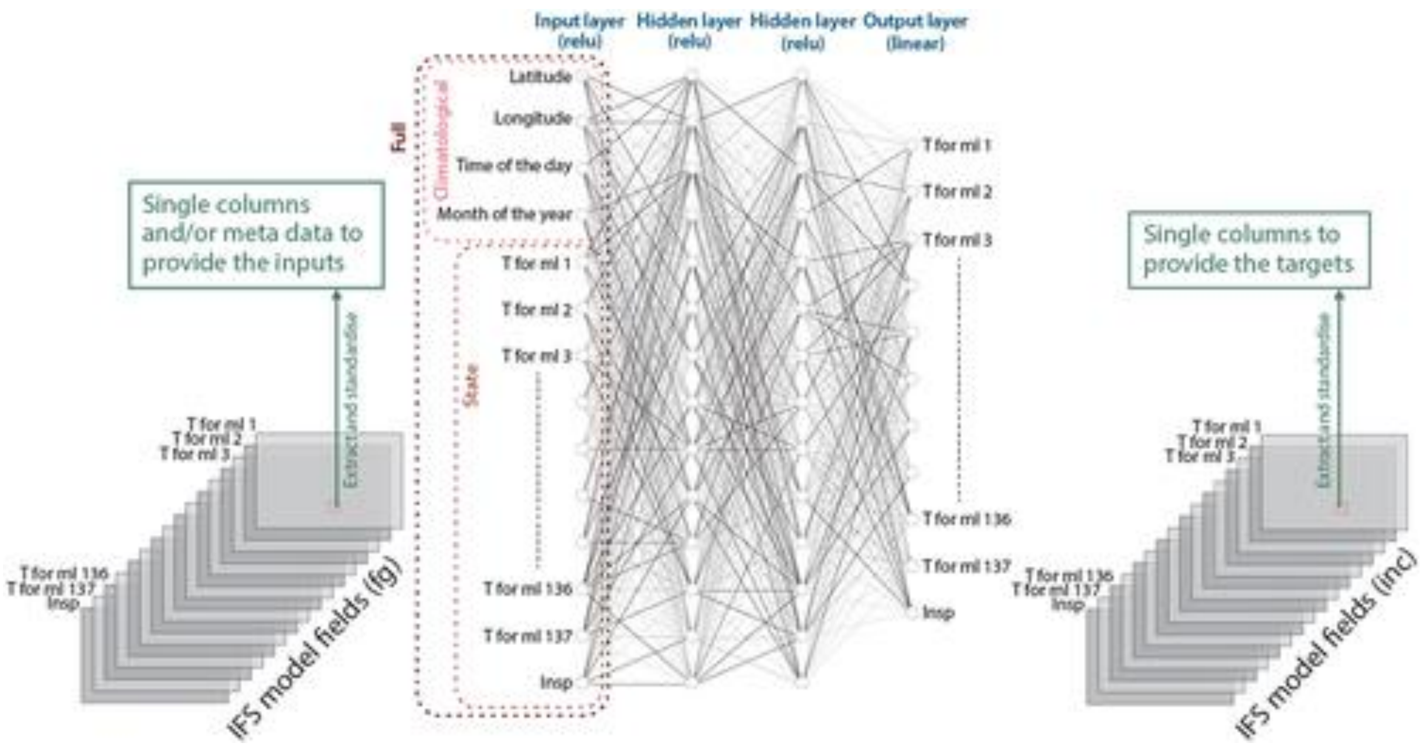
Issue in the NH at 200hPa and 700hPa



Aircraft background departures at 200hPa show a residual bias that is absorbed by WC-4DVar

Neural network based model of the model error

- Bonavita & Laloyaux, 2020: Train a NN model on a database of analysis increments under the assumption that the learnable part of the increments is due to systematic model errors;
- Dimension reduction: input to a Neural Network consists of set 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)
- Training: 1 year (2018) of analysis/bkg fields valid every 12h, valid at nominal analysis times (00, 12 UTC);



Dense Neural Network with Relu activation, 3 layers, numbers of trainable parameters ~6*10^4, size of training dataset ~10^6.

From offline to online model error estimation with a Neural Network

1. Estimate offline model error with a NN (REF=0001):

$$\eta^{b,REF} = \mathcal{G}(\mathbf{p}, \mathbf{x}_0^{REF})$$

2. Allow weak constrain to adjust model error

$$\mathcal{J}(\mathbf{x}_0, \eta) \triangleq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\eta - \eta^{b,REF}\|^2_{\mathbf{Q}^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0, \eta)\|^2_{\mathbf{R}_k^{-1}}$$

1. Estimate online model error with a NN:

$$\eta^{b} = \mathcal{G}(\mathbf{p}, \mathbf{x}_0^{b})$$

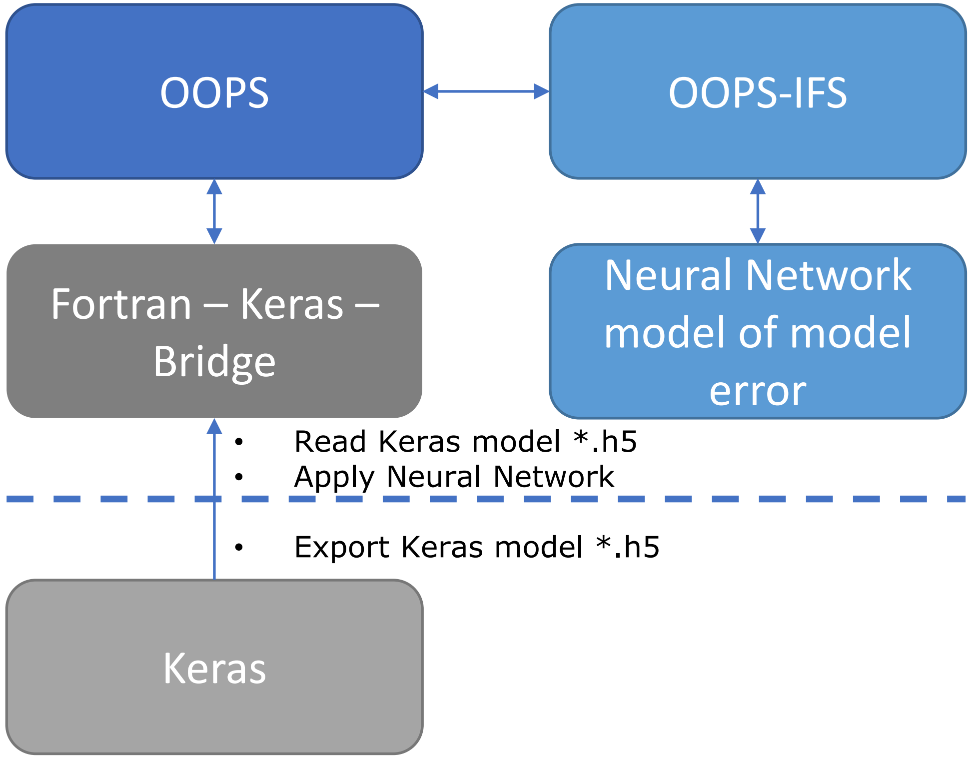
2. Allow weak constrain to adjust model error

$$\mathcal{J}(\mathbf{x}_0, \eta) \triangleq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\eta - \eta^{b}\|^2_{\mathbf{Q}^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0, \eta)\|^2_{\mathbf{R}_k^{-1}}$$

Online model error estimation with a Neural Network

➔ Online model error estimation implemented via interfacing Fortran-Keras-Bridge library to OOPS

$$\mathcal{G}(p, x_0) = \mathcal{I}_{T21:TINNER}(\mathcal{T}_{gp->sp}(\mathcal{D}(\tilde{\mathcal{G}}(p, \mathcal{N}(\mathcal{T}_{sp->gp}(\mathcal{I}_{TINNER:T21}(x_0)))))))$$



- FKB Layers:
  - Dense
  - Dropout
  - Batchnorm
- FKB Activation functions:
  - Relu
  - Leaky relu
  - Sigmoid
  - Gaussian
  - Linear
  - Tanh
  - Step

- $\mathcal{I}_{TINNER:T21}$

  is an interpolation in spectral space (truncation) from the inner loop resolution to T21
- $\mathcal{T}_{sp\to gp}$

  is a spectral transform from spectral space to grid-point space (inverse spectral transfrom);
- $\mathcal{N}$

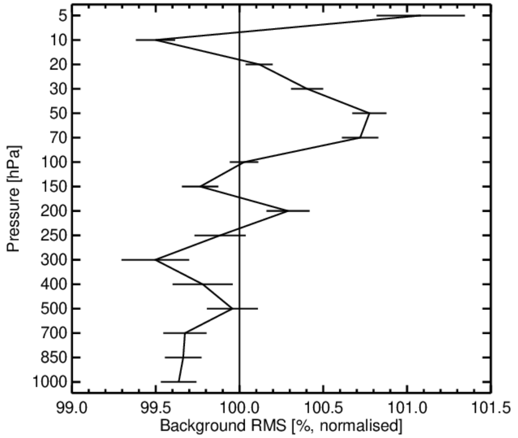
  is a normalization operator (remove the mean and divide by standard deviation)
- $\tilde{\mathcal{G}}$

  is the neural network engine
- $\mathcal{D}$

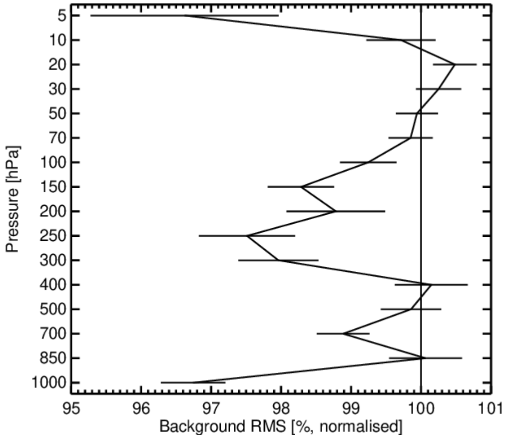
  is the de-normalization operator (multiply by standard deviation and add the mean)

Performance in 4D-Var

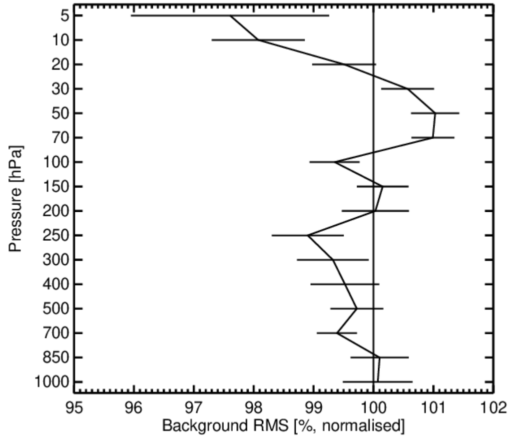
TEMP-T N.Hemis



TEMP-T Tropics



TEMP-T S.Hemis



From 00Z 1–Dec–2019 to 00Z 31–Mar–2020

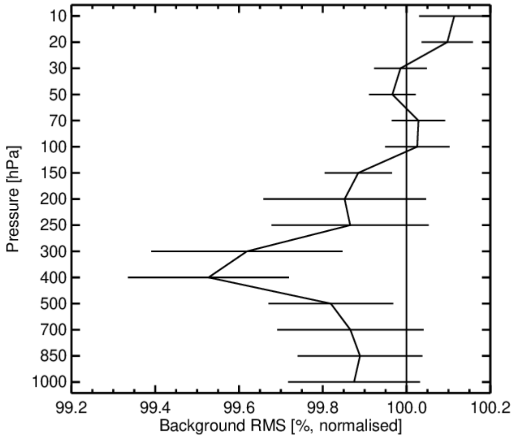
Neural–Network–Model–Bias–FG

100% = REF

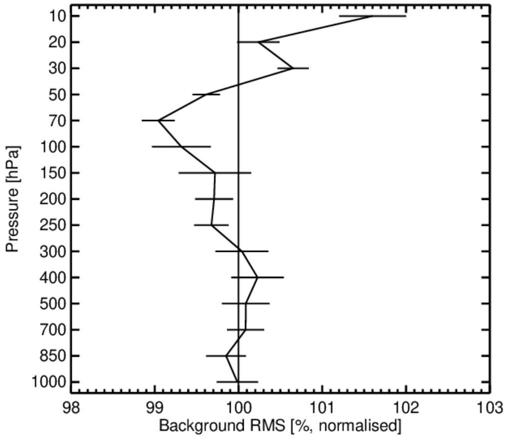
Moderate wind improvements

Significant temperature improvements except NH and SH stratosphere

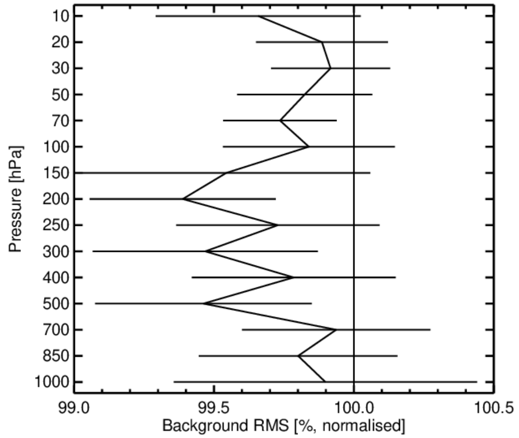
WIND-UV N.Hemis



WIND-UV Tropics

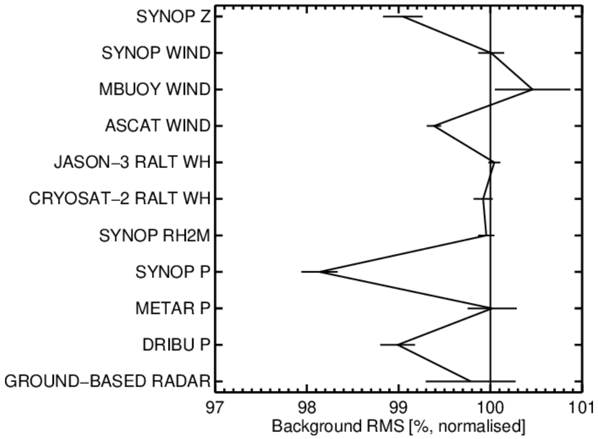


WIND-UV S.Hemis

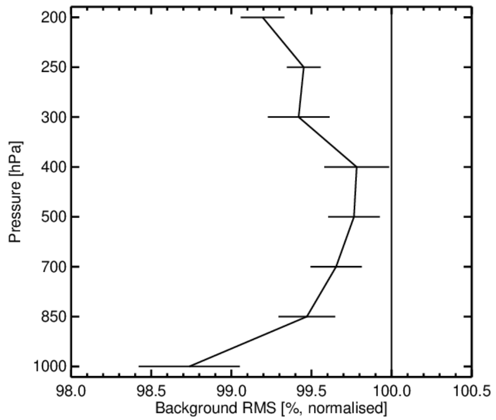


Performance in 4D-Var

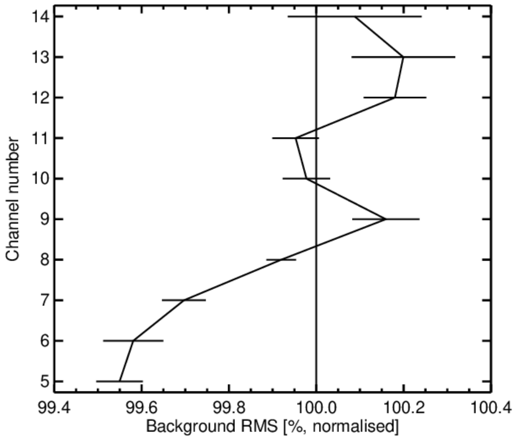
SFC Global



Aircraft-T Global



AMSU-A Global



From 00Z 1–Dec–2019 to 00Z 31–Mar–2020

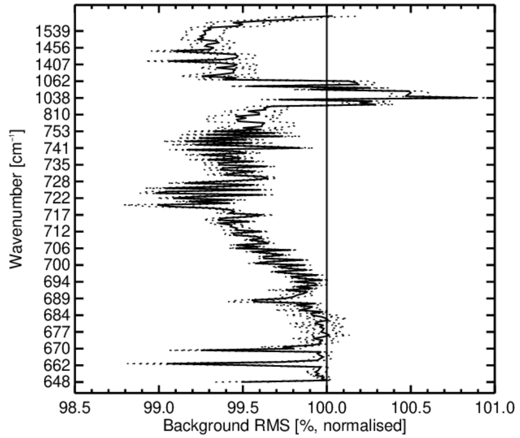
Neural–Network–Model–Bias–FG

100% = REF

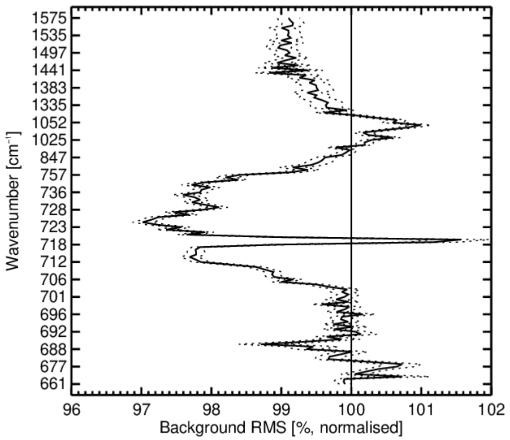
Significant surface pressure improvements

Significant improvements seen by MW and IR instruments

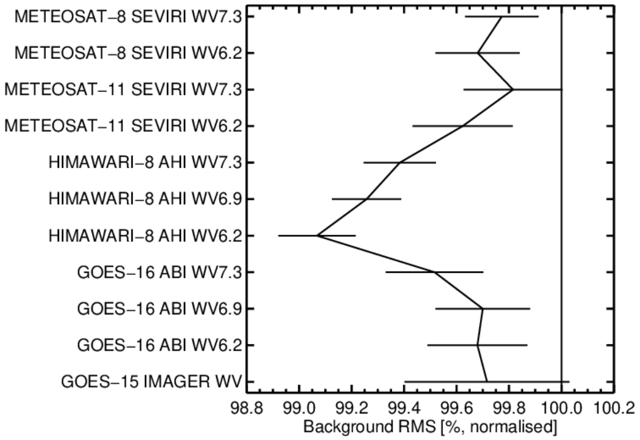
IASI Global



CRiS Global

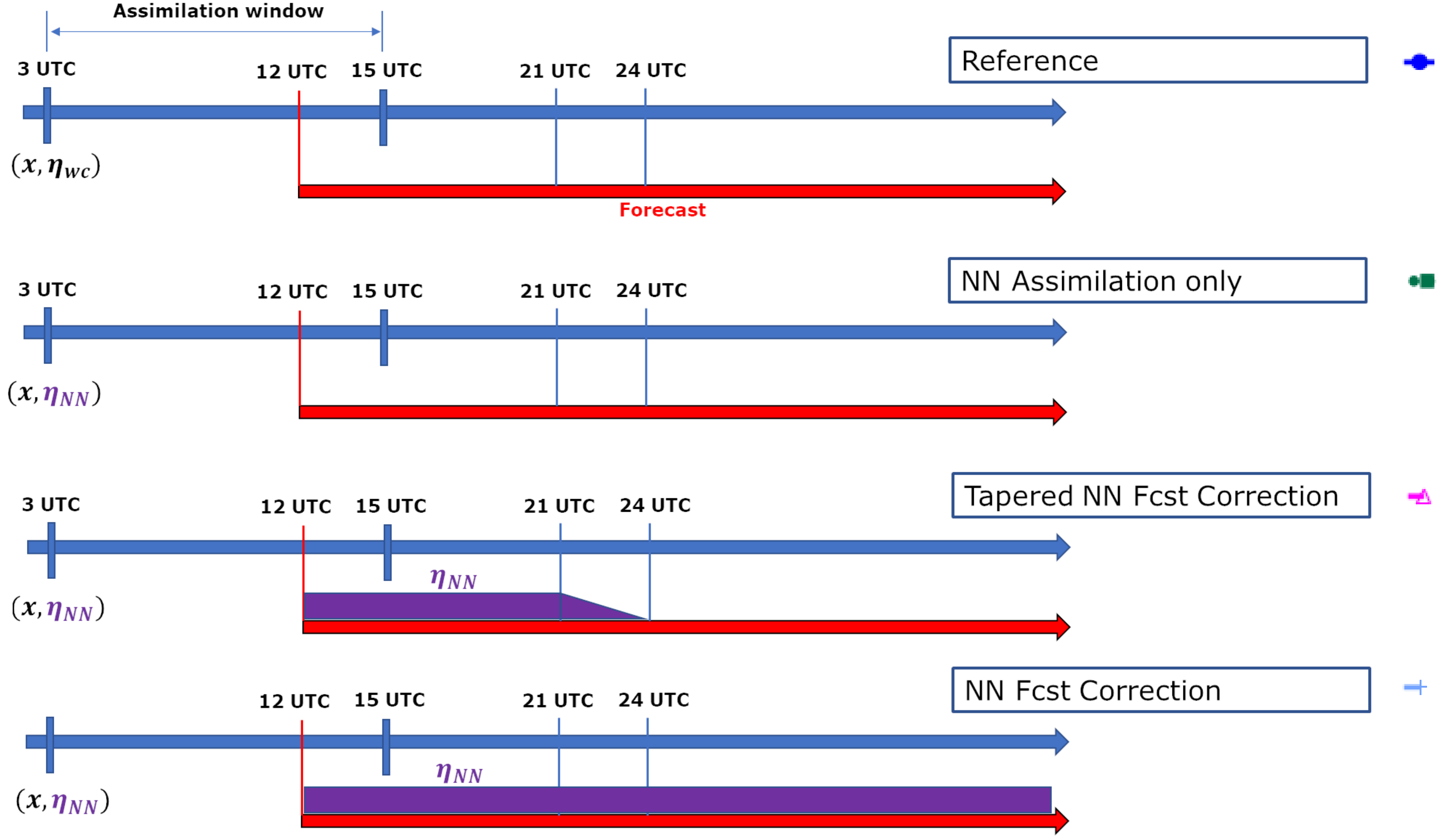


GEOS Global

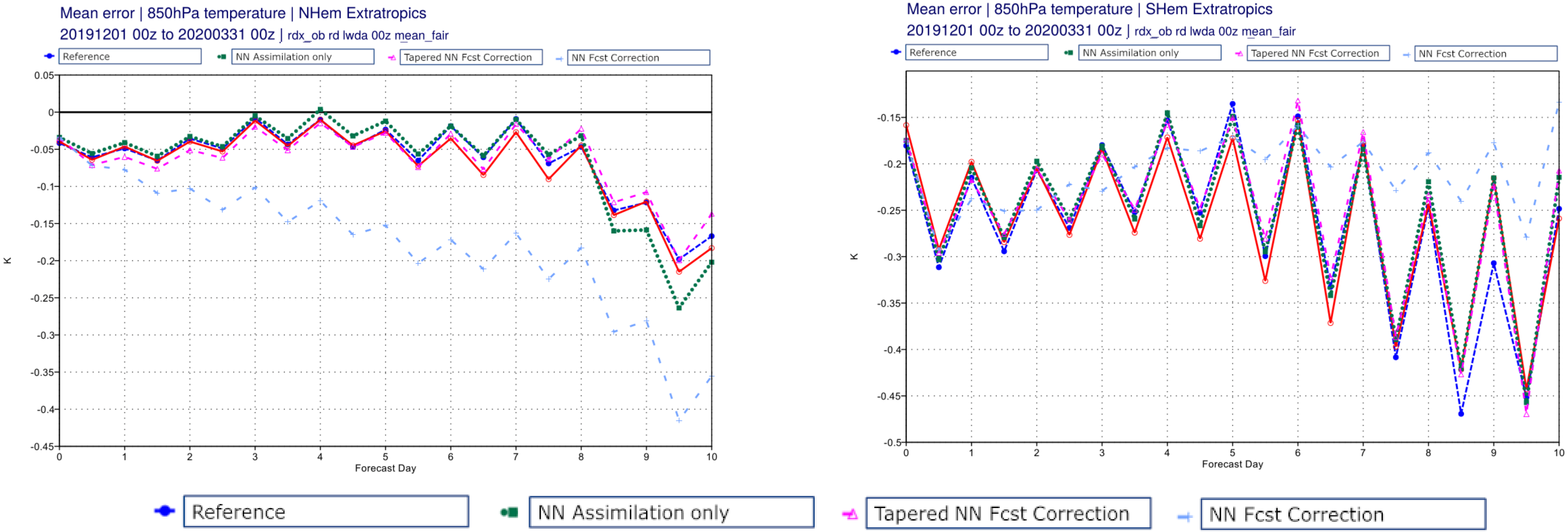


Performance in Forecast in various configurations

From 00Z 1–Dec–2019 to 00Z 31–Mar–2020

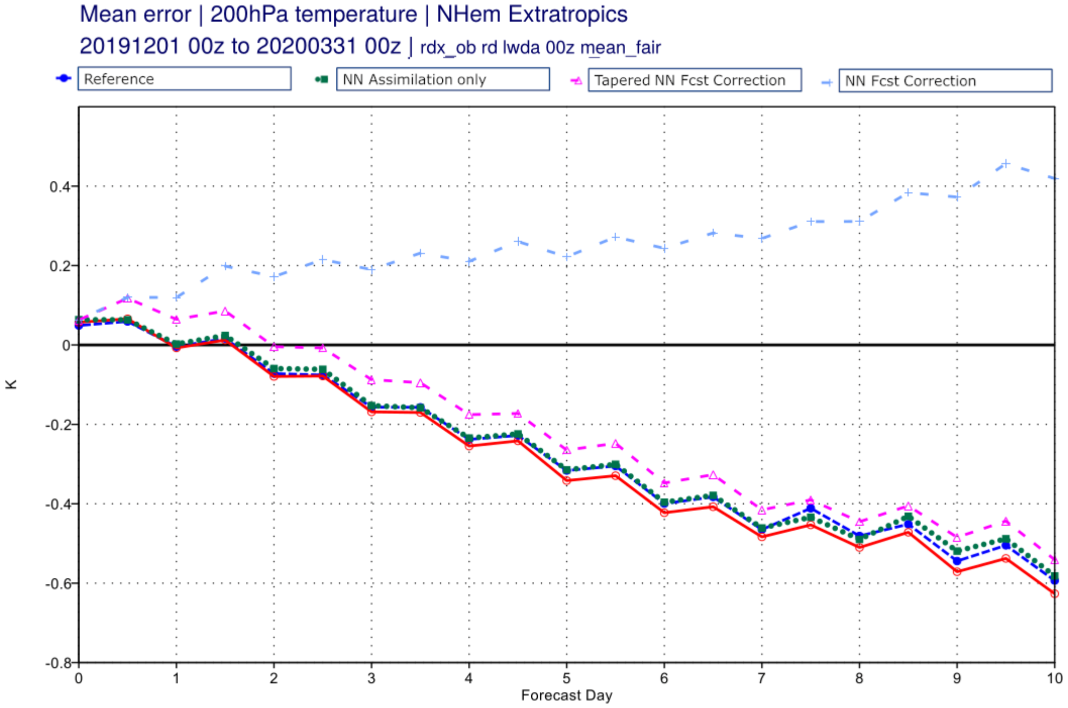


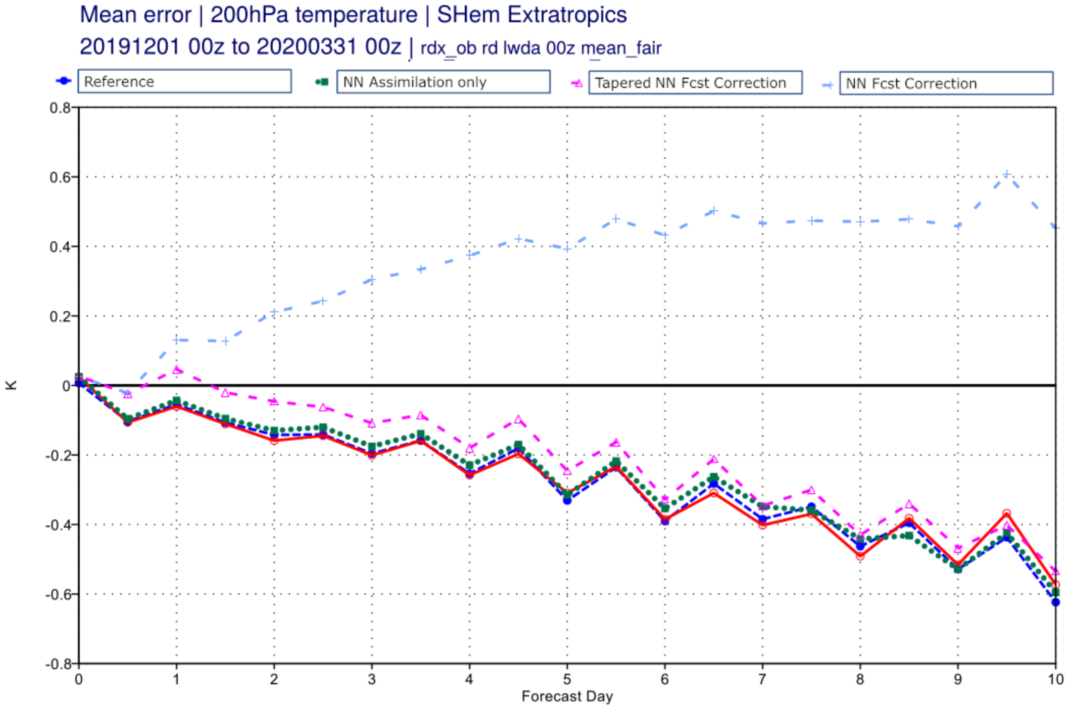
Performance in Forecast: NH, SH, 850hPa (lw00)



- Model does not have clear forecast bias trend, but daily cycle is visible;
- Applying to the forecast NN correction estimated during one 12h cycle is not optimal, evolving ME correction needed;

Performance in Forecast: NH, SH, 200hPa (lw00)

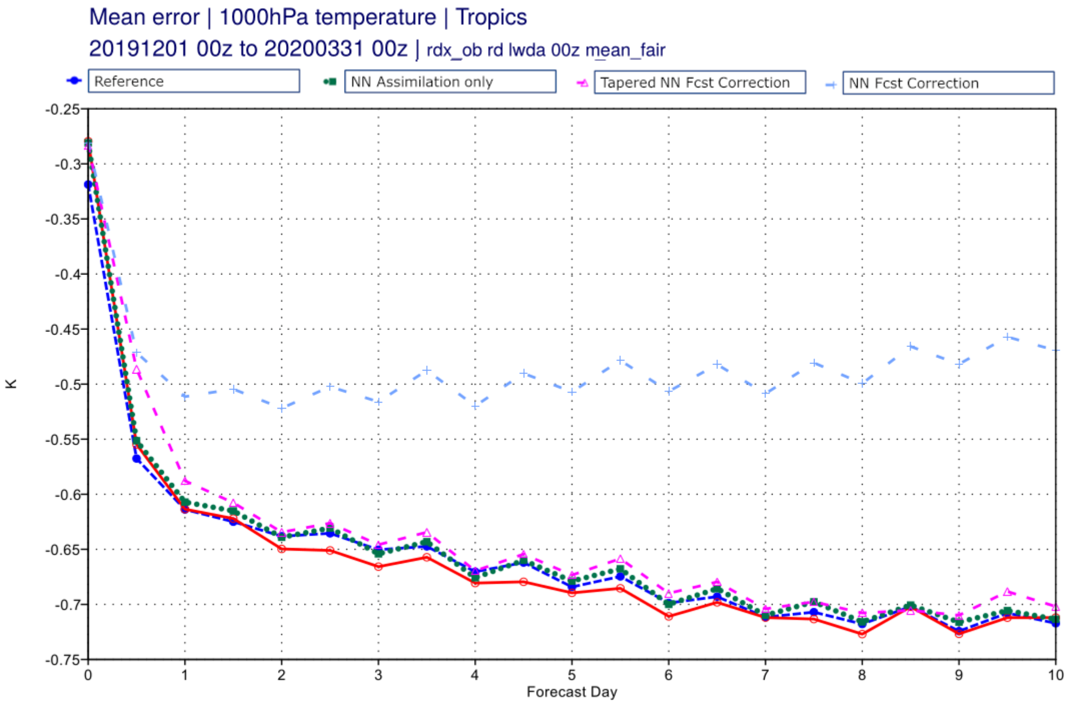


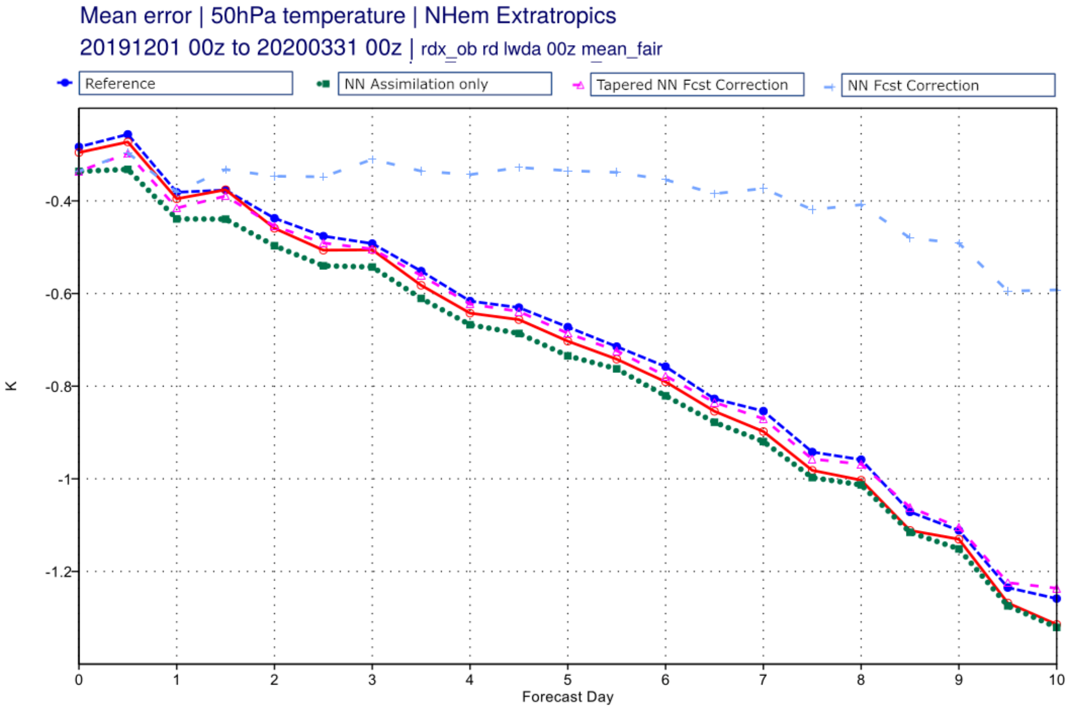


Reference | NN Assimilation only | Tapered NN Fcst Correction | NN Fcst Correction

- Model has clear forecast bias trend, daily cycle is superimposed;
- Applying to the forecast NN correction estimated during one 12h cycle reduces mean error, evolving ME correction would be useful.

Performance in Forecast: Tropics, 1000hPa (lw00) and NH, 50hPa (lw00)

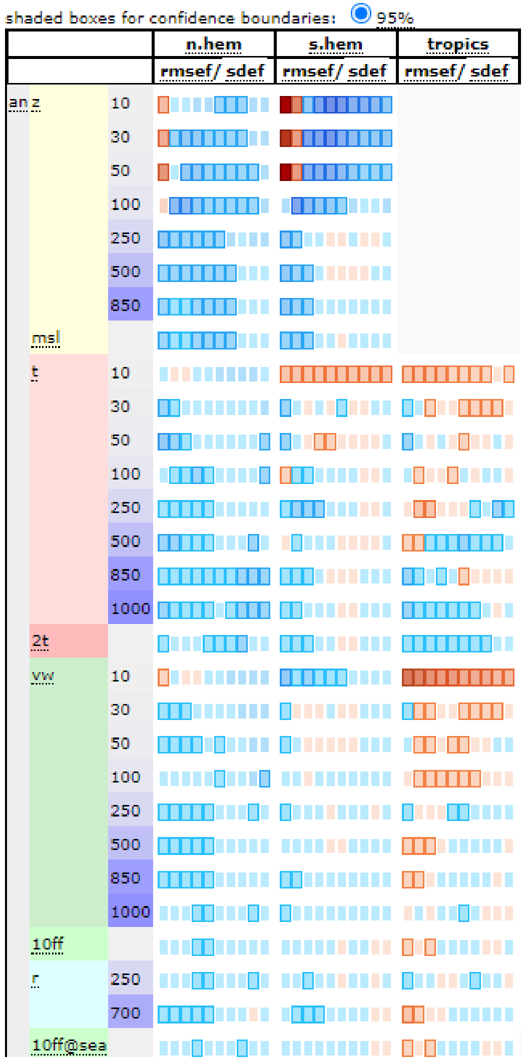




Reference | NN Assimilation only | Tapered NN Fcst Correction | NN Fcst Correction

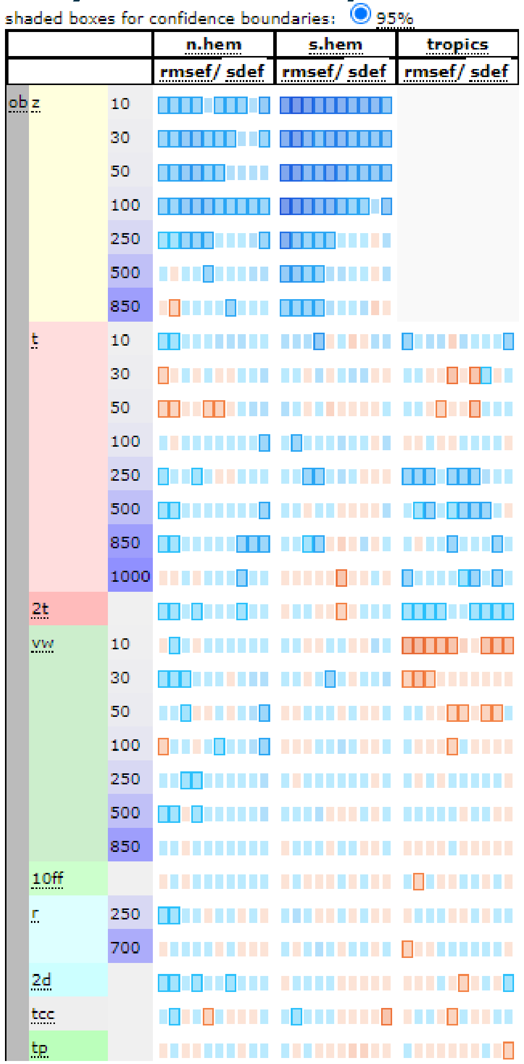
- Model error forecast bias trend dominates the mean forecast error;
- Applying NN correction gives visible forecast improvements.

Score cards with 12h verification (0001 and obs) starting from T+00h/12h



Tapered NN Fcst Correction

Verified against 0001



Verified against obs

Summary and Outlook

- Neural network model of the model bias trained on a sufficiently long record of analysis increments is capable of correcting systematic errors in 4D-Var;
- Applying the NN in the forecast:
  - is beneficial in presence of a monotonic model bias (e.g. extra-tropical Strato., Tropical troposphere) ;
  - Is challenging when the model bias is small and non-monotonic (e.g. NH, 850hPa and Tropics 200hPa);
  - Requires evolving ME correction when the model bias has significant diurnal variability (e.g. NH and SH 850hPa and SH 200hPa);
- The IFS model is being continuously developed – retraining the Neural Network is considered cumbersome -> need to develop online NN training methodologies inside the 4DVar analysis update (see A. Farchi talk in this WS)
- We are not applying an evolving, flow dependent model bias correction in the forecast - the NN correction does not have a chance to adjust to flow conditions;

Summary and Outlook

- These results also confirm conclusions in Bonavita, 2021 regarding the need to take diurnal component of model error into account;
- This can be done in **WC-4DVa**r by simply changing the cycling model error strategy or, more generally, relaxing the hypothesis of constant model error in the assimilation window
- In terms of the **NN** approach to model error estimation and correction, the results indicate the need for:
    1. Implement an evolving, flow-dependent NN engine so that the correction adapts during the forecast (to daily cycle and slowly evolving conditions)
    2. Increase the time frequency of the NN training dataset (from 12h to 6/3h) to get a better handle on evolution of daily error cycle
- Caution: Systematic increments can be a consequence of biased observations rather than biased model which may make both WC-4DVar and the NN learn the wrong corrections!

ECMWF Strategy – future directions

‘Science and Technology’ strategic actions

The ‘Science and Technology’ strategic actions are linked to enhancements in the exploitation of observations, data assimilation, modelling and exploitation of new technologies, computational science and operational processes.

Strengthen leadership in Earth system data assimilation

ECMWF will strengthen its leadership position in data assimilation by progressing in coupled assimilation, algorithmic development and integration of approaches. 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.

From online model error estimation with a Neural Network to estimation of the parameters of Neural Network model of model error in 4D-Var*

1. Estimate online model error with via a Neural Network model:

$$\eta^b = \mathcal{G}(\mathbf{p}, \mathbf{x}_0^b)$$

2. Allow weak constrain to adjust model error

$$\mathcal{J}(\mathbf{x}_0, \eta) \triangleq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\eta - \eta^b\|^2_{\mathbf{Q}^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0, \eta)\|^2_{\mathbf{R}_k^{-1}}$$

1. Estimate parameters of a Neural Network based model of the model error:

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

***With Alban Farchi and Marc Bocquet**

Implementation in incremental 4D-Var*

- Requires the tangent linear of a neural network w.r.t. its parameters G^p and its input G^x as well as their adjoints;
- A new Fortran library has been developed to provide the required operators and interfaced to OOPS in place of FKB for dense neural networks;
- Extension of the existing weak constraint framework to accommodate joint estimation of the neural network parameters is straight-forward and requires only few additional steps (in red); the work to extend the OOPS control layer is under-way;

Input: $\delta \mathrm{p}, \delta \mathrm{x}_0$

$\Delta_0 \leftarrow \mathrm{R}_0^{-1}[\mathrm{d}_0 + \mathrm{H}_0 \delta \mathrm{x}_0];$

$\delta \eta \leftarrow \mathrm{G}^{\mathrm{p}} \delta \mathrm{p} + \mathrm{G}^{\mathrm{x}} \delta \mathrm{x}_0;$

for $k = 1$ **to** L **do**

  $\delta \mathrm{x}_k \leftarrow \mathrm{M}_{k:k-1} \delta \mathrm{x}_{k-1} + \delta \eta;$

  $\Delta_k \leftarrow \mathrm{R}_k^{-1}[\mathrm{d}_k + \mathrm{H}_k \delta \mathrm{x}_k];$

end

$\delta \tilde{\mathrm{x}}_L \leftarrow 0;$

$\delta \tilde{\eta}_L \leftarrow 0;$

for $k = L$ **to** 1 **do**

  $\delta \tilde{\mathrm{x}}_k \leftarrow \mathrm{H}_k^\top \Delta_k + \delta \tilde{\mathrm{x}}_k;$

  $\delta \tilde{\eta}_{k-1} \leftarrow \delta \tilde{\mathrm{x}}_k + \delta \tilde{\eta}_k;$

  $\delta \tilde{\mathrm{x}}_{k-1} \leftarrow \mathrm{M}_{k:k-1}^\top \delta \tilde{\mathrm{x}}_k;$

end

$\delta \tilde{\mathrm{p}} \leftarrow [\mathrm{G}^{\mathrm{p}}]^\top \delta \tilde{\eta}_0;$

$\delta \tilde{\mathrm{x}}_0 \leftarrow [\mathrm{G}^{\mathrm{x}}]^\top \delta \tilde{\eta}_0 + \delta \tilde{\mathrm{x}}_0;$

$\delta \tilde{\mathrm{x}}_0 \leftarrow \mathrm{H}_0^\top \Delta_0 + \delta \tilde{\mathrm{x}}_0;$

$\nabla_{\delta \mathrm{x}} \mathcal{J}^{\mathrm{i}} \leftarrow \delta \tilde{\mathrm{x}}_0;$

$\nabla_{\delta \mathrm{p}} \mathcal{J}^{\mathrm{i}} \leftarrow \delta \tilde{\mathrm{p}};$

Output: $\nabla_{\delta \mathrm{x}} \mathcal{J}^{\mathrm{i}}, \nabla_{\delta \mathrm{p}} \mathcal{J}^{\mathrm{i}}$

***With Alban Farchi and Marc Bocquet**

Towards correcting observation bias with a Neural Network in 4D-Var

Variational Bias Correction: Estimate parameters of a multi-linear regression bias model:

$$\mathcal{J}(\mathbf{x}_0, \beta) \triangleq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\beta - \beta^b\|^2_{\mathbf{B}_\beta^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\|\mathbf{y}_k - \tilde{\mathcal{H}}_k(\mathbf{x}_k, \beta)\|^2_{\mathbf{R}_k^{-1}}$$

$$\triangleq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\beta - \beta^b\|^2_{\mathbf{B}_\beta^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k) - \beta^T\mathbf{P}_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}}$$

Estimate parameters of a Neural Network model of observation bias:

$$\mathcal{J}(\mathbf{x}_0, \gamma) \triangleq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\gamma - \gamma^b\|^2_{\mathbf{B}_\gamma^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\|\mathbf{y}_k - \tilde{\mathcal{H}}_k(\mathbf{x}_k, \gamma)\|^2_{\mathbf{R}_k^{-1}}$$

$$\triangleq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\gamma - \gamma^b\|^2_{\mathbf{B}_\gamma^{-1}} + \frac{1}{2}\sum_{k=0}^{L}\|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k) - \mathbf{L}(\gamma, \mathbf{P}_k(\mathbf{x}_k))\|^2_{\mathbf{R}_k^{-1}}$$

Vector of bias parameters

Vector of bias predictors

Parameters of a Neural Network model of observation bias

Thank you