Data Assimilation: Initial Conditions and beyond

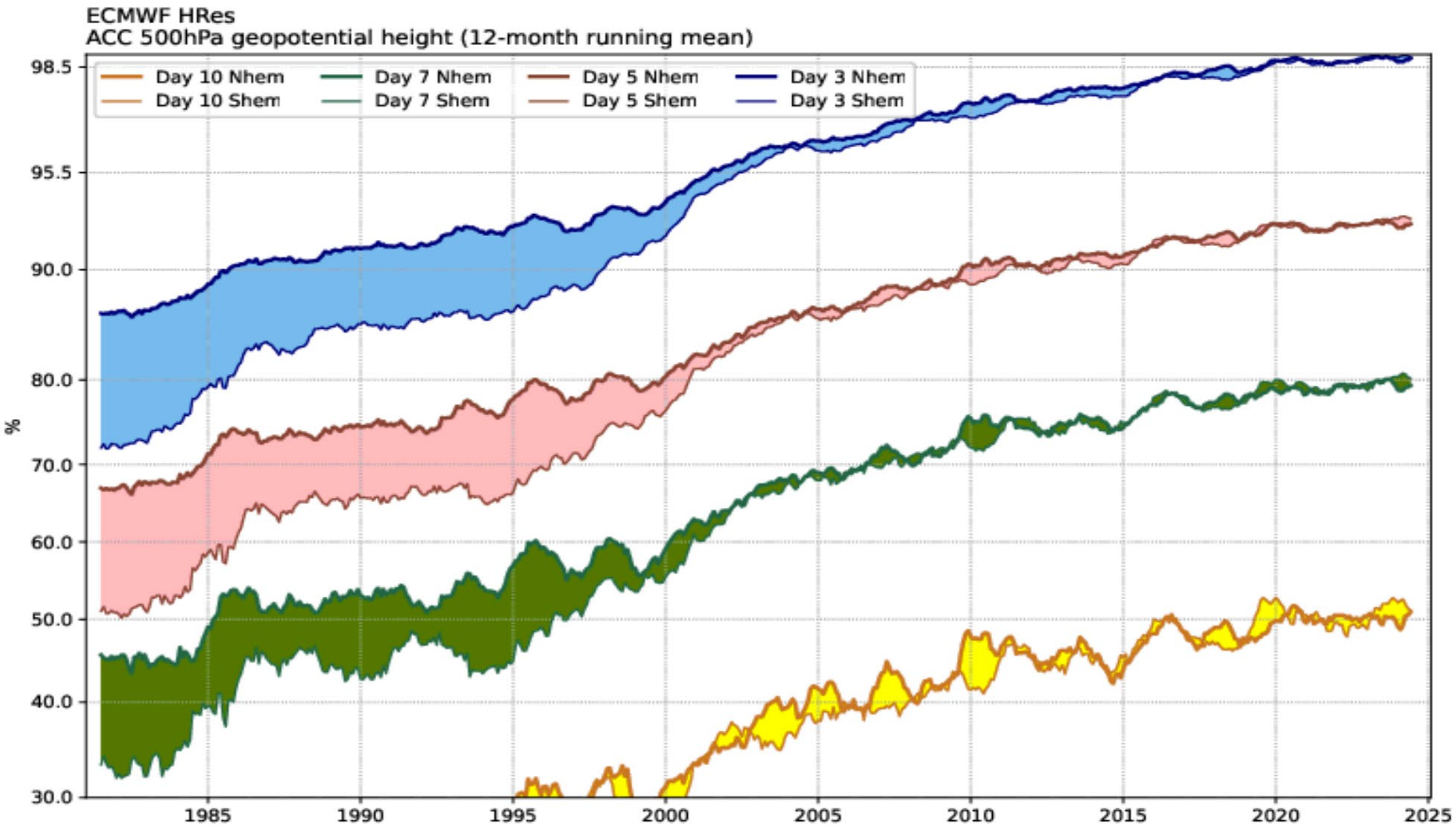
What is the role of DA in the ML Age?

Massimo Bonavita

Principal Scientist
Data Assimilation Method. Team Leader
ECMWF

massimo.bonavita@ecmwf.int

Lead time of anomaly correlation coefficient (ACC) reaching multiple thresholds (High resolution (HRES) 500 hPa height forecasts)



ECMWF

Predictability of the weather and its limits

- **Lorenz, 1963**, qualitatively revealed the essence of a finite predictability within a chaotic system such as the atmosphere.
- However, he did not determine a precise limit for the predictability of the atmosphere.

- The concept of a two-week predictability limit based on error doubling time of five days was proposed in the 60s (**Charney et al., 1966**)
- The two-week predictability limit has since become the consensus doxa in meteorology (Shen et al., 2023)

Predictability of the weather and its limits

- **Zhang et al., 2019**, provide an illustration of current thinking based on evolution of forecast spread from ECMWF Ensemble DA.
- What would be the predictive skill of ECMWF forecasts if we managed to reduce IC errors by a factor of 10?

What Is the Predictability Limit of Midlatitude Weather?

FUQING ZHANG AND Y. QIANG SUN

Department of Meteorology and Atmospheric Science, and Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, Pennsylvania

LINUS MAGNUSSON AND ROBERTO BUIZZA

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

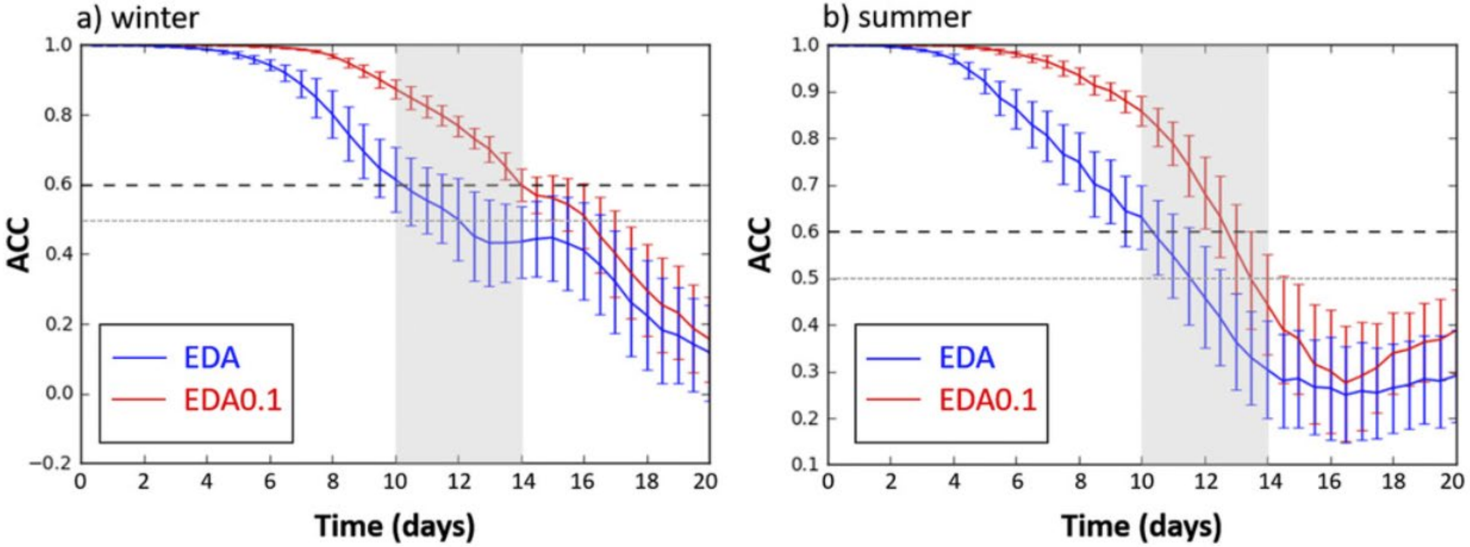
SHIAN-JIANN LIN AND JAN-HUEY CHEN

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

KERRY EMANUEL

Lorenz Center, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 7 September 2018, in final form 9 November 2018)

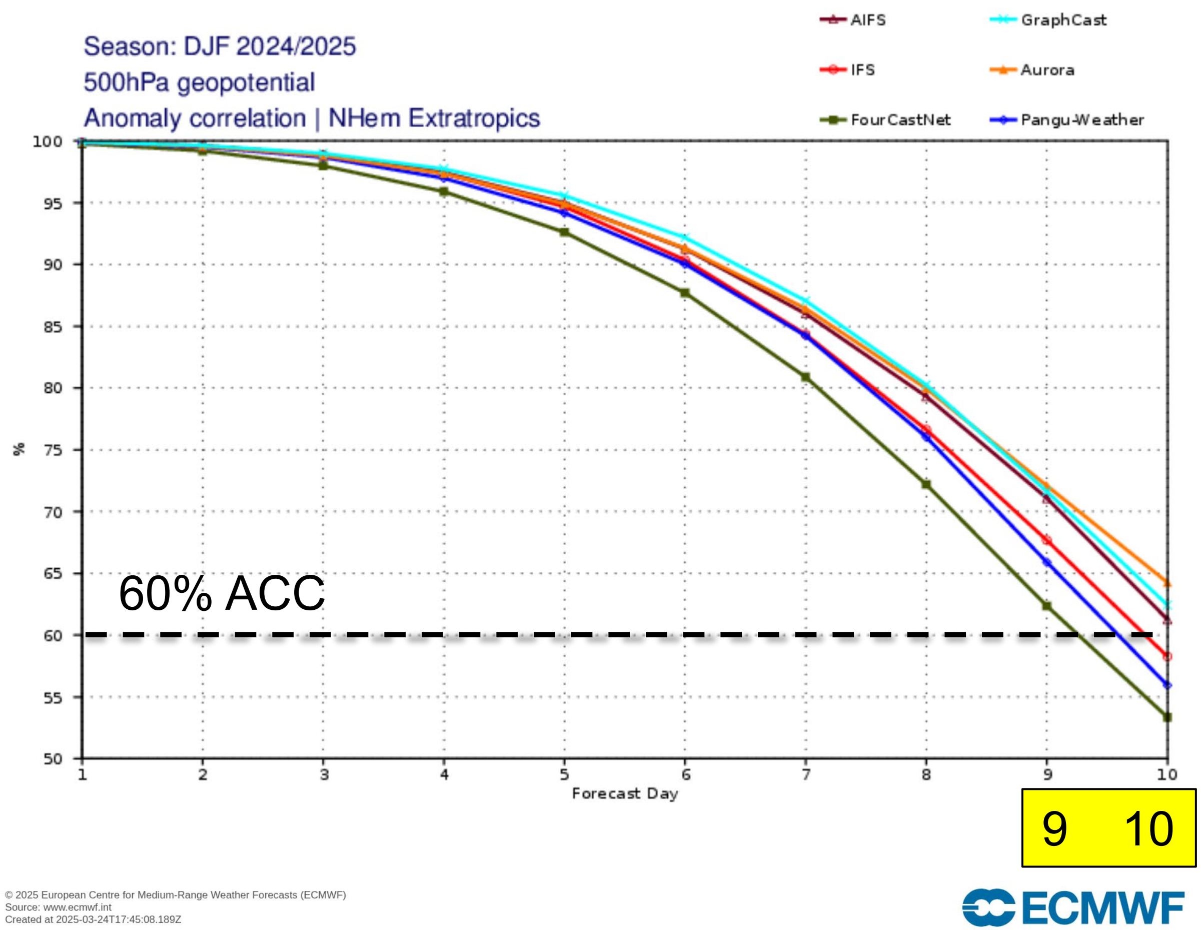


Predictability of the weather and its limits

- We are currently reaching the limit of useful synoptic scale forecasts between fcst day 9 and 10

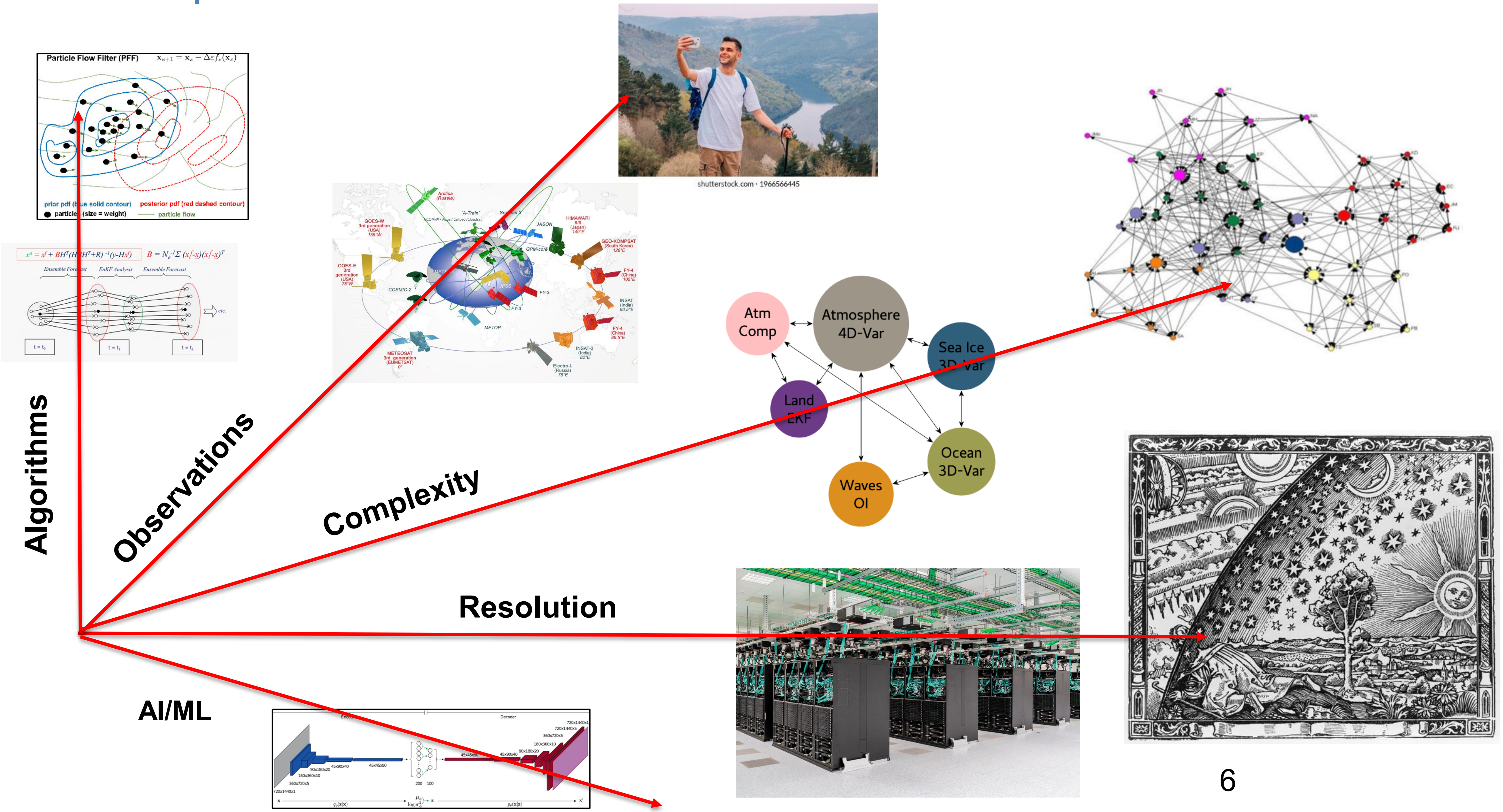
1. What are the missing ingredients to reach the “intrinsic” predictability limit?
2. What is the role of DA in this enterprise?
3. Is the 2-weeks predictability barrier real or is it a feature of our forecast systems?
4. Do ML models have similar/same predictability barrier?

Scores of forecasts of upper-air parameters by experimental machine learning models



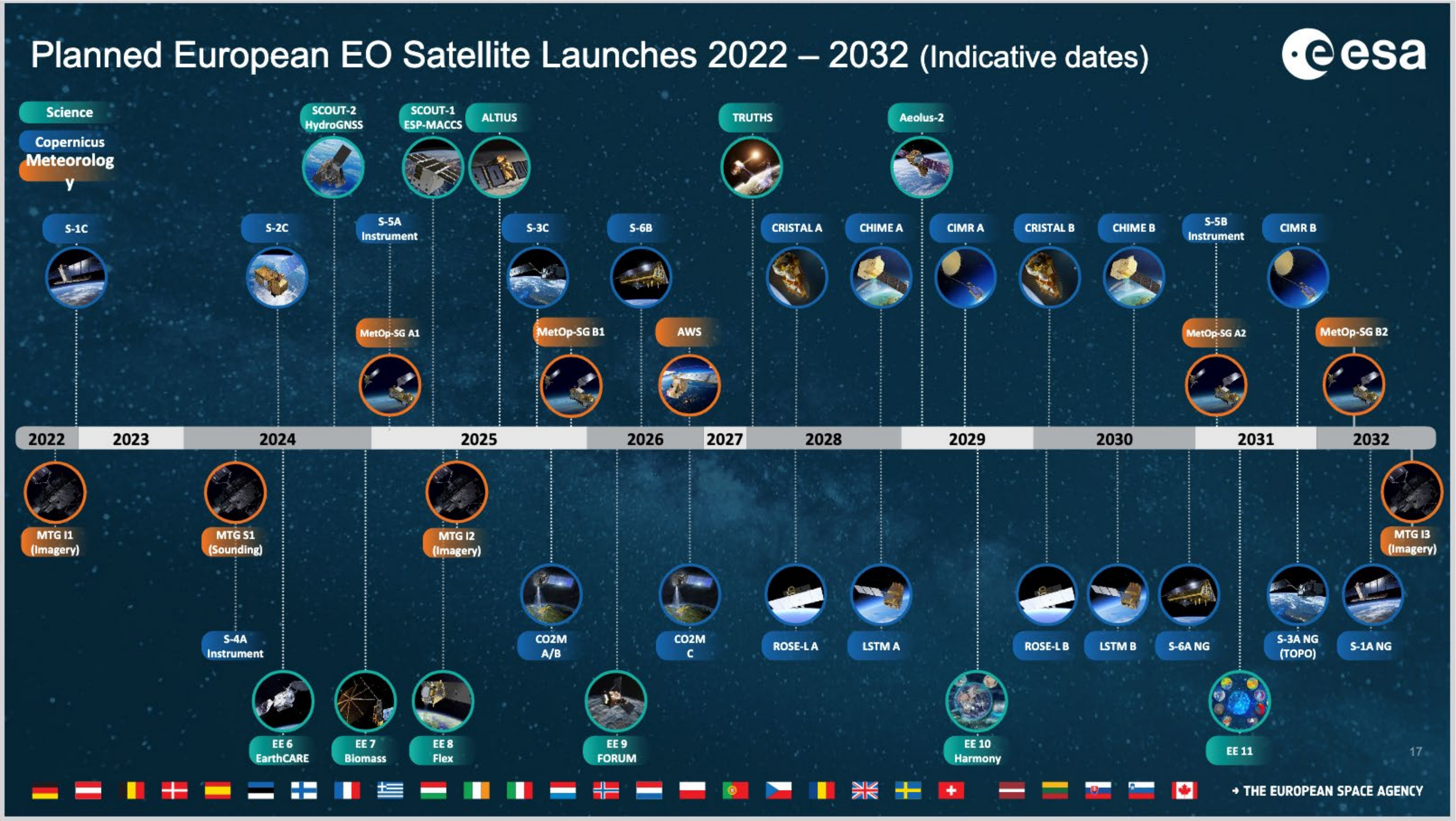
ECMWF

Development axes



Observations

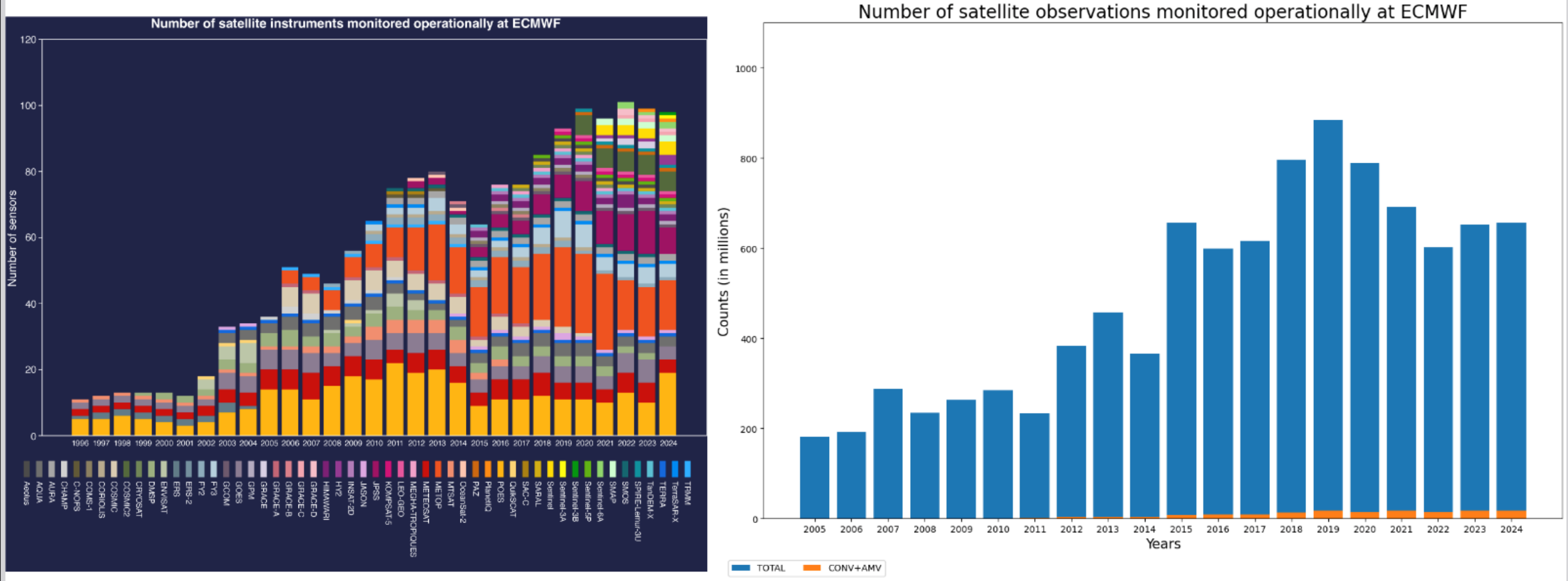
- Forthcoming and future satellite missions (EPS-SG, MTG, Sentinels,…) will add capabilities and improve coverage
- Will any of the future missions be transformational of current observing capabilities for NWP?
- (Angela Benedetti to provide the answers later today…)



Algorithms

- We only use ~5% of current available observations…
- Is the bigger problem how to make use of all this untapped wealth of information?
- (Prof. Sarah Dance has some exciting ideas on how to solve this problem…)

Evolution of data counts and diversity



- ~ 700 million pieces of observations received daily and ~40 millions active. A lot of potential

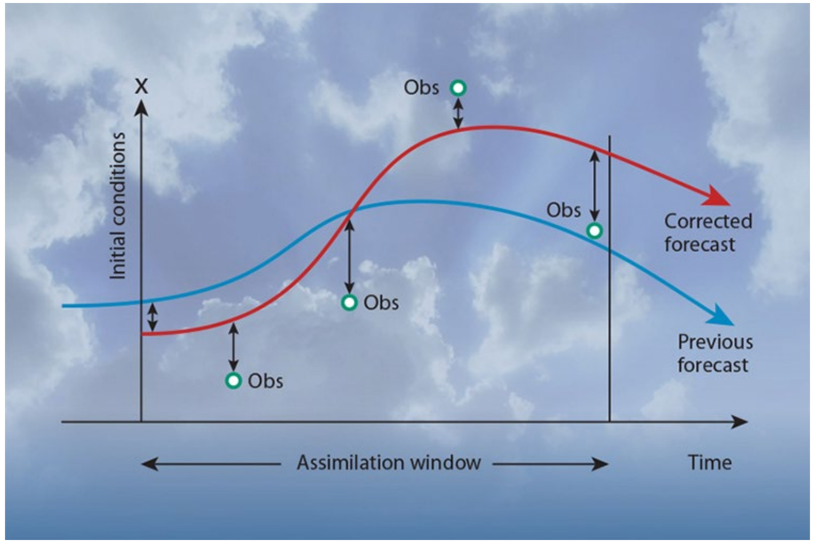
Courtesy of Mohamed Dahoui, ECMWF

Resolution

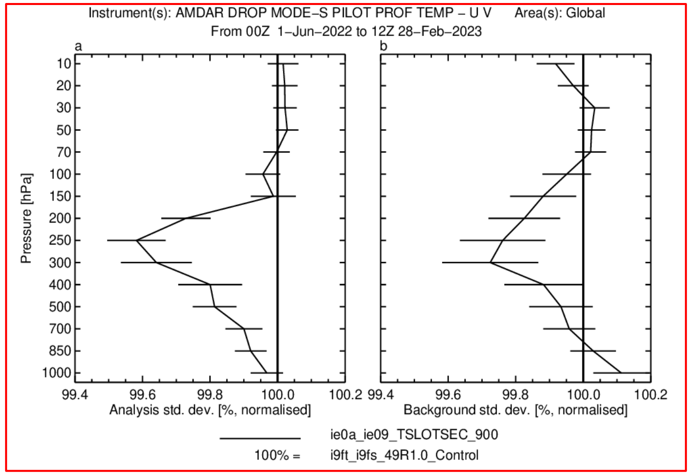
Self-evident Data Assimilation postulates (Ziga and Emiliano):

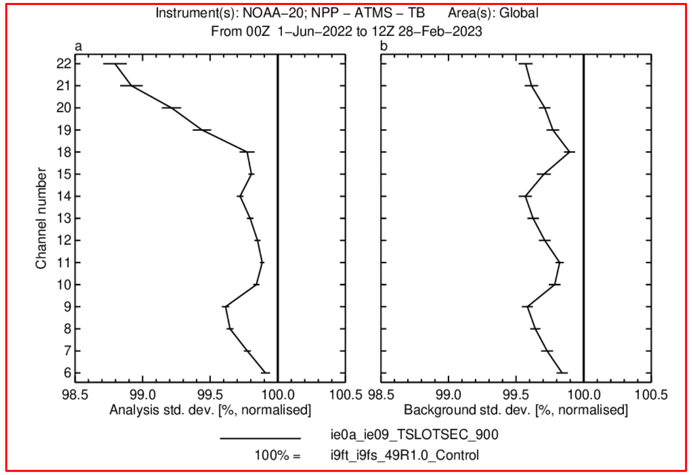
1. *“Effective use of dense observations in space and time requires increased spatial and temporal resolution in the analysis updates”*
2. *“Increased temporal frequency of analysis updates is crucial to control nonlinearity in the hoigh resolution DA”*

Resolution (1)



- $y - H\left(M\left(x_{fg}\right)\right)$ are currently computed aggregating observations in 30’ timeslots
- Soon, observations will be compared to model equivalents every 15’
- 4D-Var can extract more information from frequent observations if the model $\mathrm{H(M())}$ provides an accurate and realistic depiction of the atmosphere

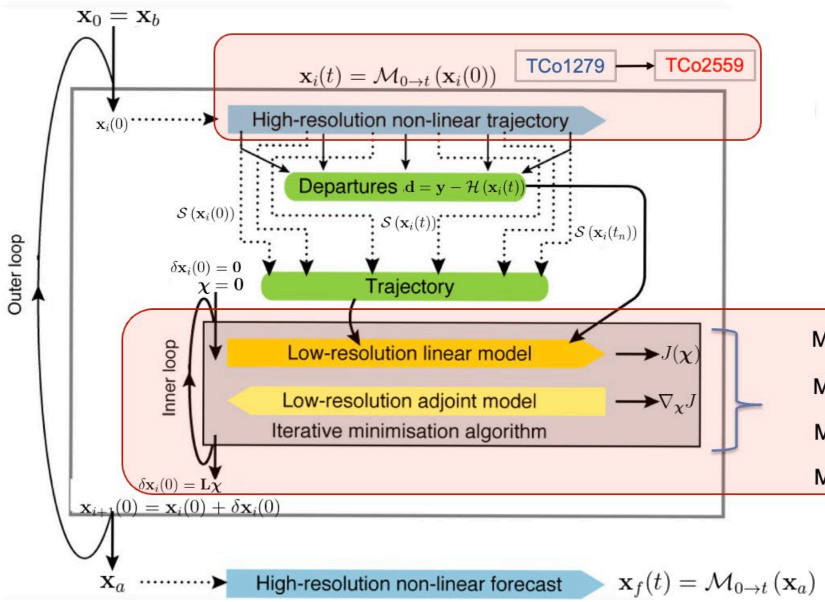




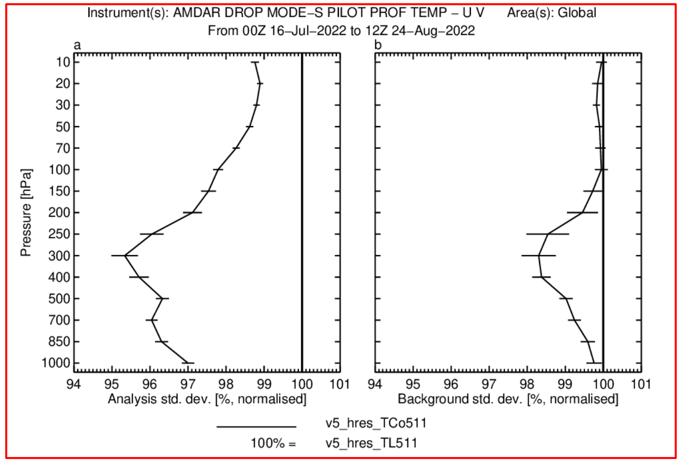
Courtesy of Jorge Bandeiras, ECMWF

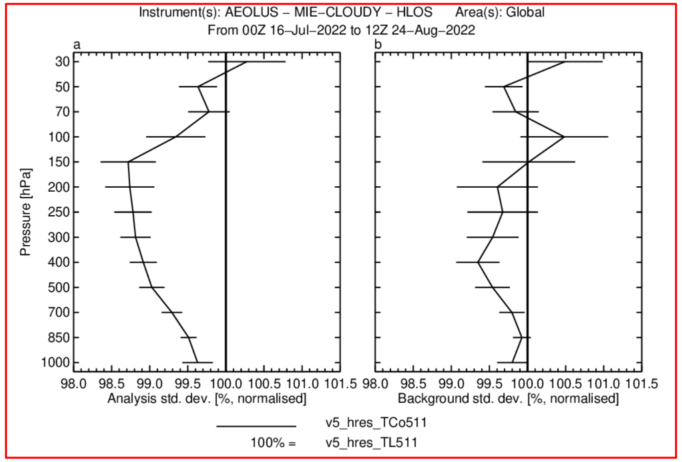
Resolution (2)

4D-Var data assimilation in IFS



- Increasing the 4D-Var inner loop resolution from 40 km to 20 km => doubling the effective resolution of the analysis increments
- 4D-Var can extract more information from observations if the model H(M()) provides an accurate and realistic depiction of the atmosphere



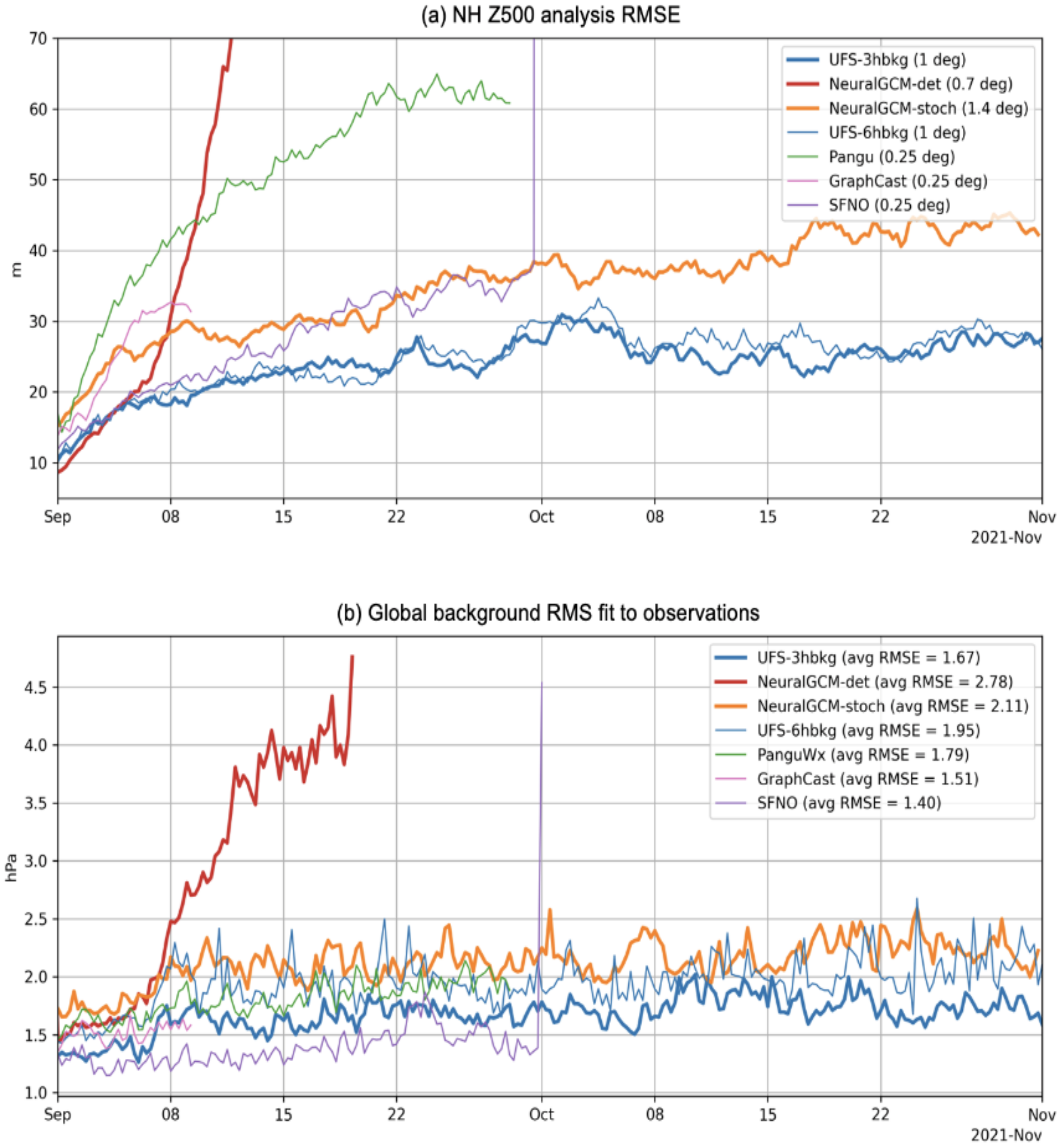


Courtesy of Ziga Zaplotnik, ECMWF

On the relevance of physically consistent models

- What happens if a model is not physically consistent (e.g., Bonavita, 2024) and one tries to use it in Data Assimilation?
- Slivinski et al., 2025, applied different ML models in an EnKF DA cycle assimilating only Surface Pressure observations
- Results: DA cycles blew up (numerically!) after 2-4 weeks of cycling, or went very wrong.

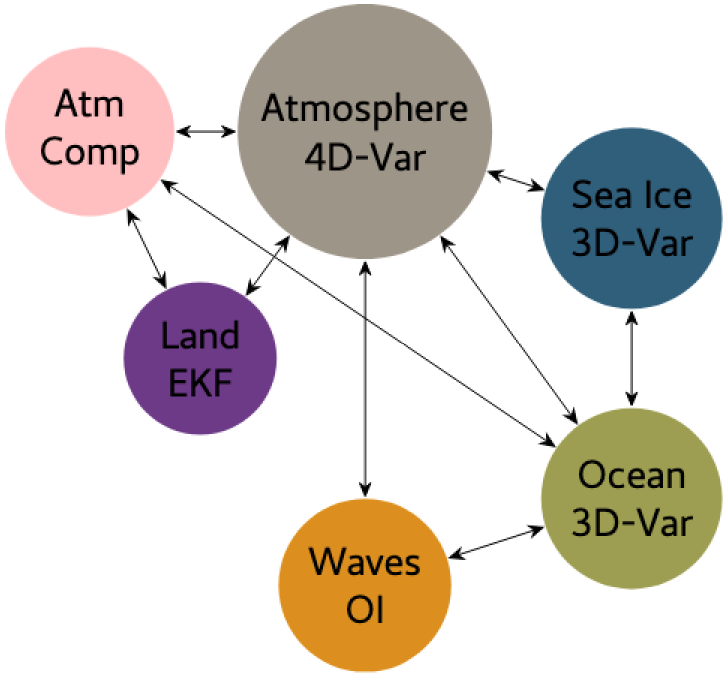
- Success of DA is predicated on the ability of the model to extrapolate observational info to unobserved variables, times and locations in a physically consistent manner



Bonavita, M. (2024). On some limitations of current machine learning weather prediction models. *Geophysical Research Letters*, 51, e2023GL107377. https://doi.org/10.1029/2023GL107377

Slivinski, L.C. J. S. Whitaker, S. Frolov, T. A. Smith, N. Agarwal (2025): Assimilating Observed Surface Pressure into ML Weather Prediction Models. arXiv:2412.18016v1, *GRL* accepted

Complexity



From NWP to Earth System Prediction

A driving theme in the last three ECMWF Strategy documents

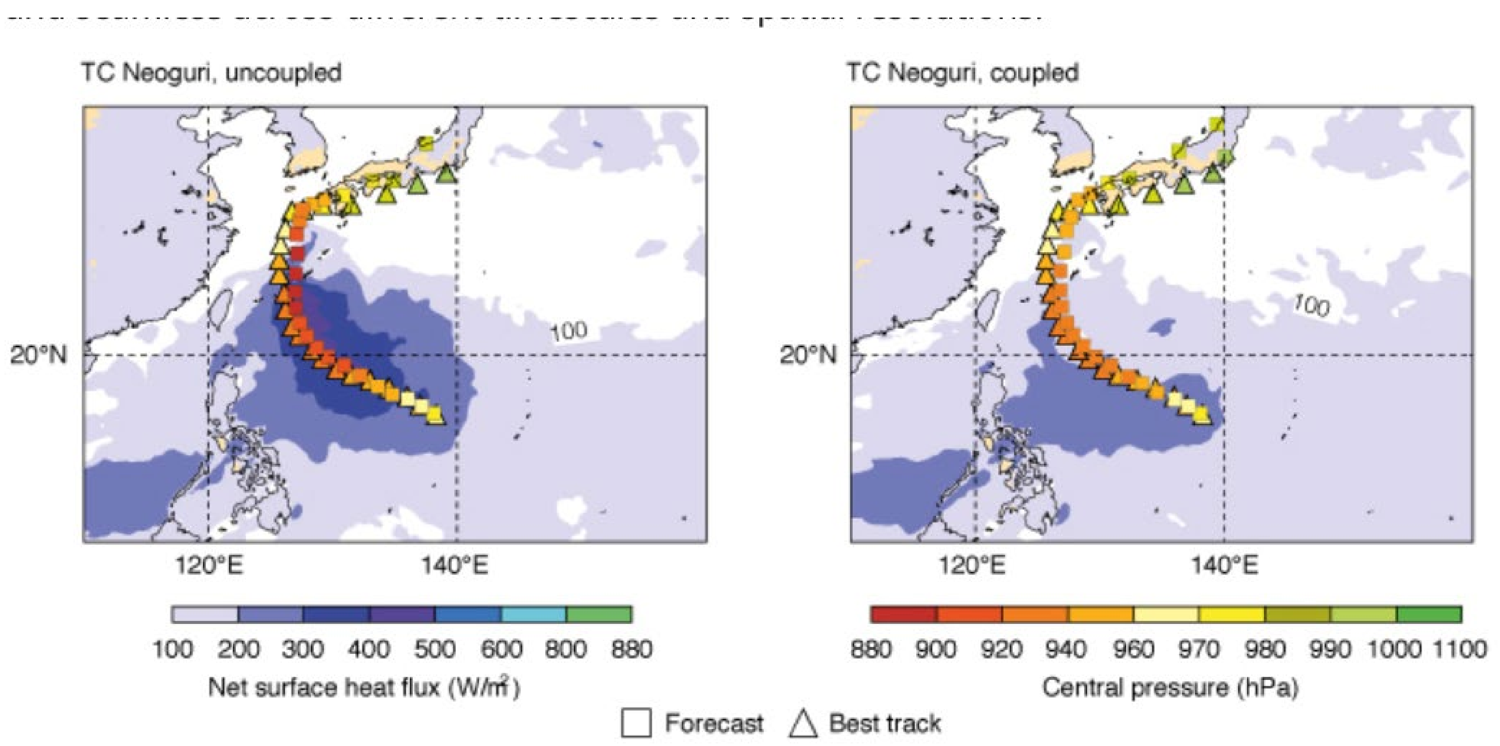
Mission: Deliver global numerical weather predictions focusing on the medium range and monitoring of the Earth system to and with our Member States.

Vision: World-leading monitoring and predictions of the Earth system enabled by cutting-edge physical, computational and data science, resulting from a close collaboration between ECMWF and the members of the European Meteorological Infrastructure, will contribute to a safe and thriving society.

ECMWF Strategy 2025–2034
https://www.ecmwf.int/en/about/what-we-do/strategy

Complexity: Atmosphere-Ocean

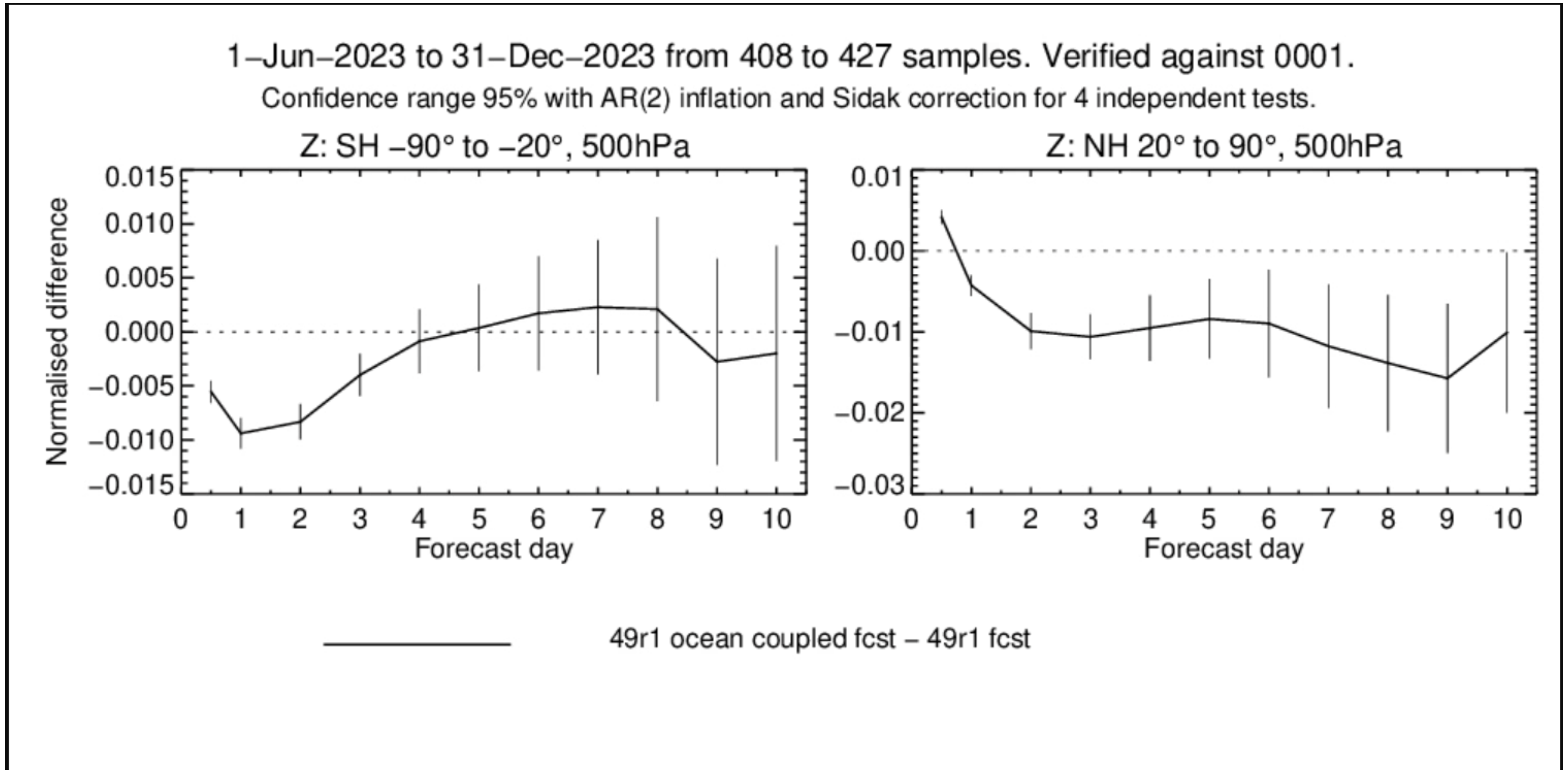
June 2018, IFS Cy45r1: H-RES IFS model coupled to Ocean NEMO model in forecast mode



ECMWF Newsletter No. 156 - Summer 2018

Complexity: Atmosphere-Ocean

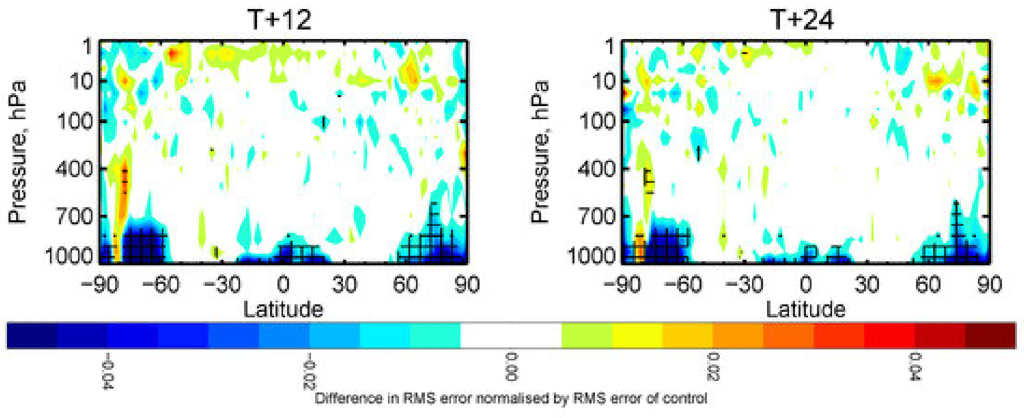
What is the impact of running Atmos-Ocean coupled forecast models on weather?



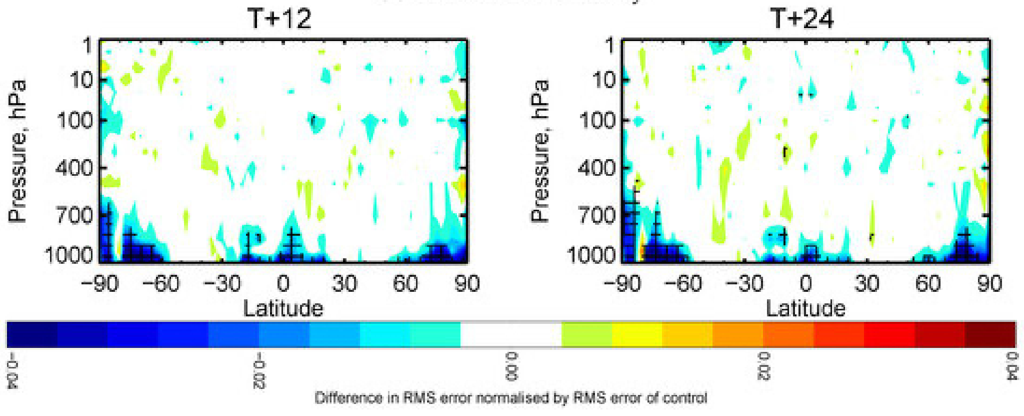
Z500 hPa RMSE reduction from Atmos-Ocean coupled forecast vs Atmos-only forecast. (current operational IFS and NEMO models, June-Dec 2023)

Complexity: Atmosphere-Ocean

Weakly Coupled DA: Atmosphere and Ocean analysis updates are separate; the two systems exchange info through coupled model integration to cycle the analyses



(a) dRMSE of humidity



(b) dRMSE of temperature

"It is clear that the impact of WCDA does not have long-range impacts on the upper troposphere or on the spatial regions where WCDA is not active."

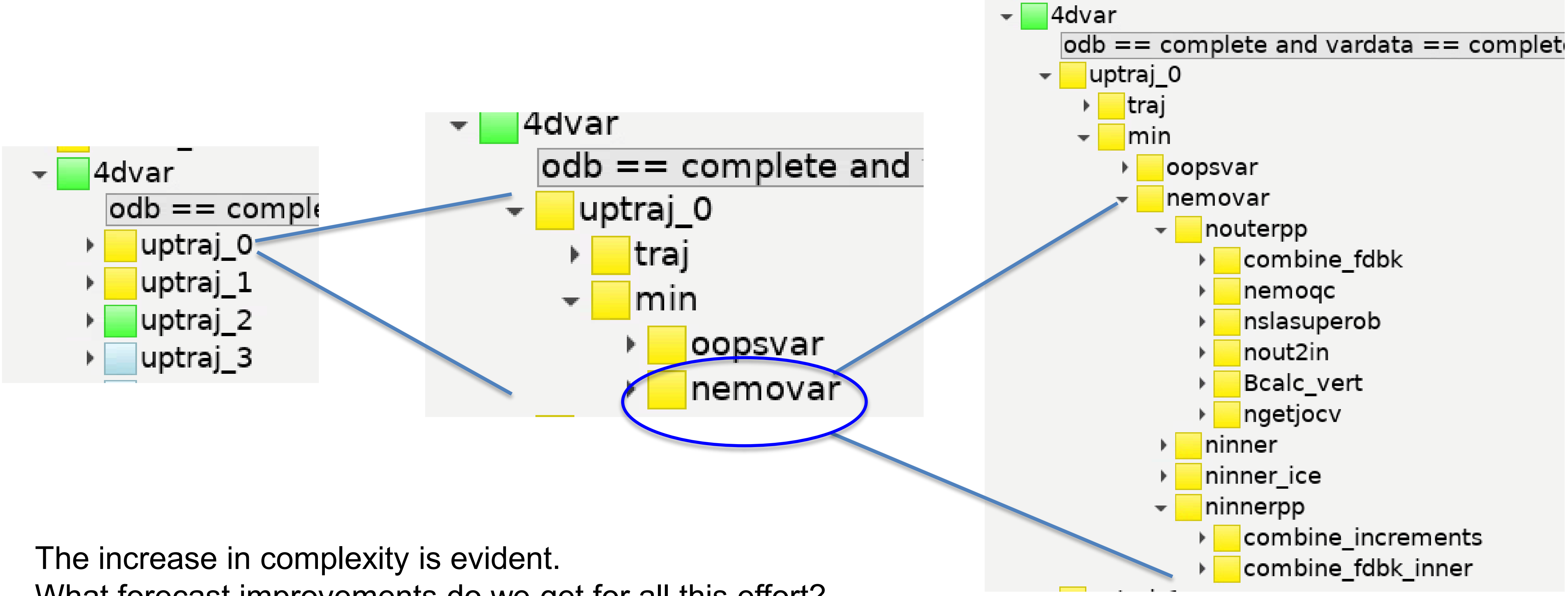
Browne, P. A., de Rosnay, P., Zuo, H., Bennett, A., & Dawson, A. (2019). Weakly Coupled Ocean–Atmosphere Data Assimilation in the ECMWF NWP System. *Remote Sensing*, *11*(3), 234. https://doi.org/10.3390/rs11030234

"A statistically significant difference with confidence above 90 % is only observed for the near-surface air temperature at around 1000 hPa for the 12 and 36 h forecasts (the difference is statistically insignificant elsewhere)."

Skachko, S., Buehner, M., Laroche, S., Lapalme, E., Smith, G., Roy, F., Surcel-Colan, D., Bélanger, J.-M., and Garand, L.: Weakly coupled atmosphere–ocean data assimilation in the Canadian global prediction system (v1), Geosci. Model Dev., 12, 5097–5112,

Complexity: Atmosphere-Ocean

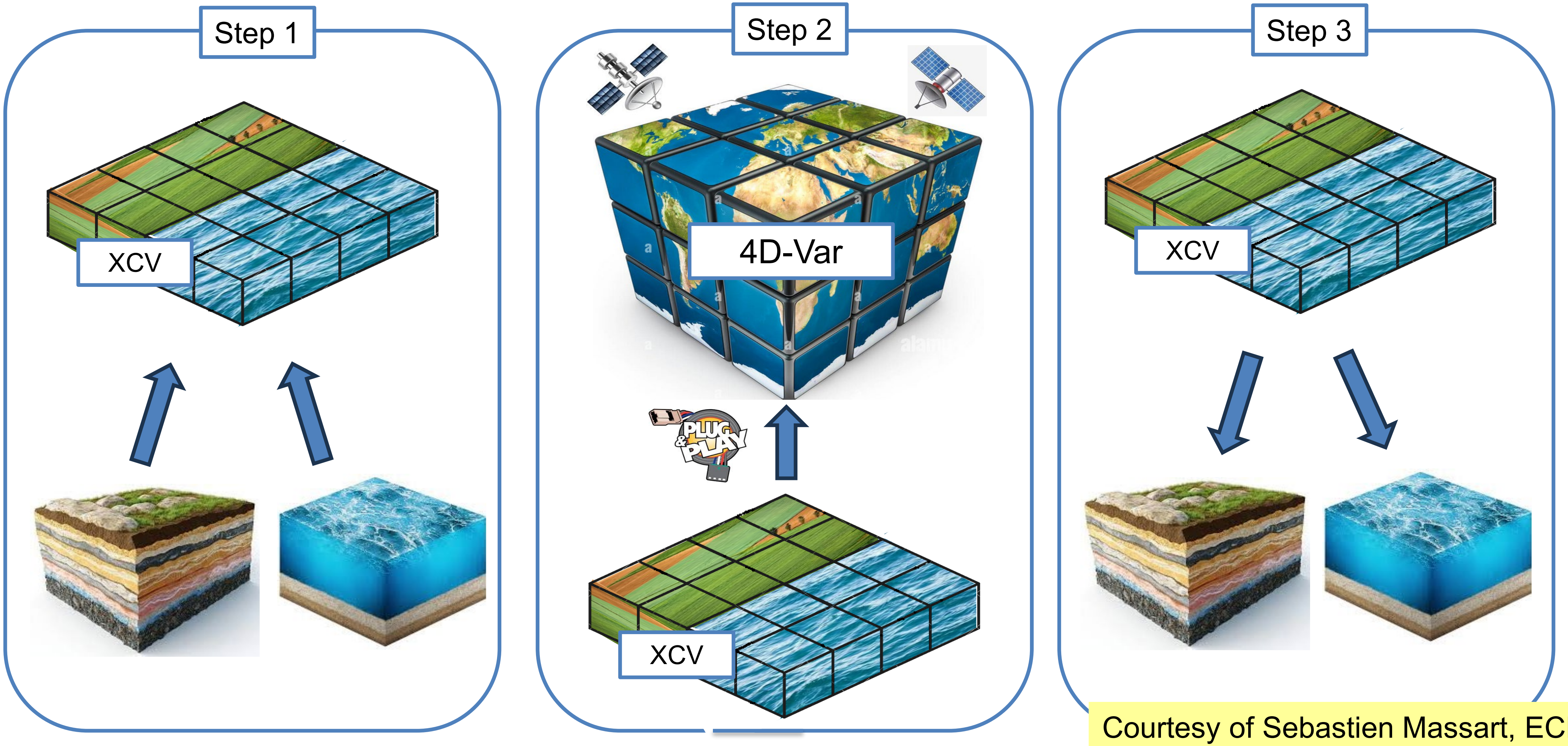
"Quasi Strong" Coupled DA: Outer Loop Coupling, ie 4DVar and Ocean DA do separate minimisations and exchange information between one minimisation and the next



The increase in complexity is evident.
What forecast improvements do we get for all this effort?

(Less) Complexity: Atmosphere-Land-Ocean-Sea Ice…

The **eXtended Control Variable (XCV™)** approach to coupled DA



Courtesy of Sebastien Massart, ECMWF

Complexity: NWP – Atmospheric Composition Coupling

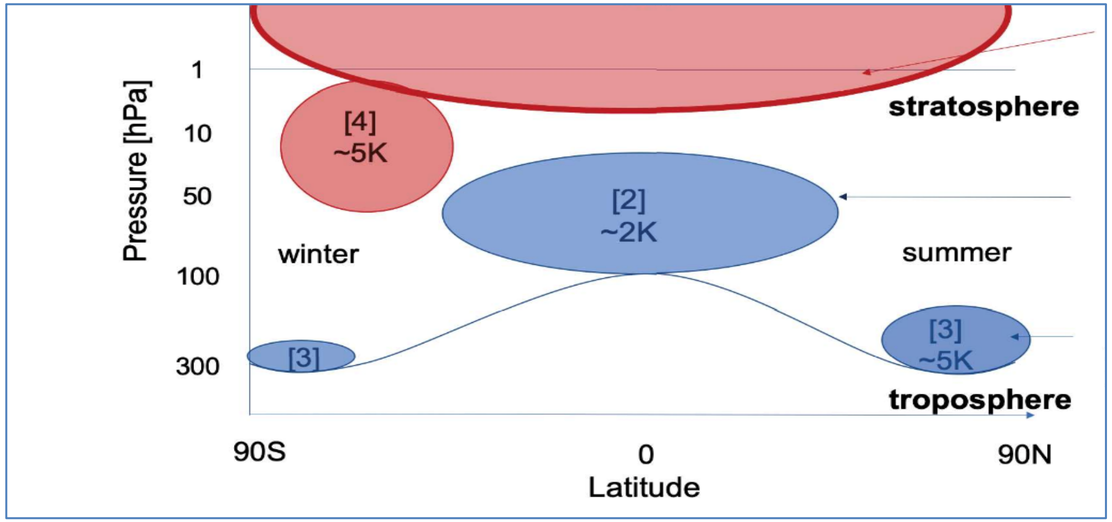
- The synergies and benefits of coupling DA for NWP with Atmospheric Composition DA have not received the attention they deserve
- Antje Inness will set the record straight today!
- A couple of examples to whet your appetite…

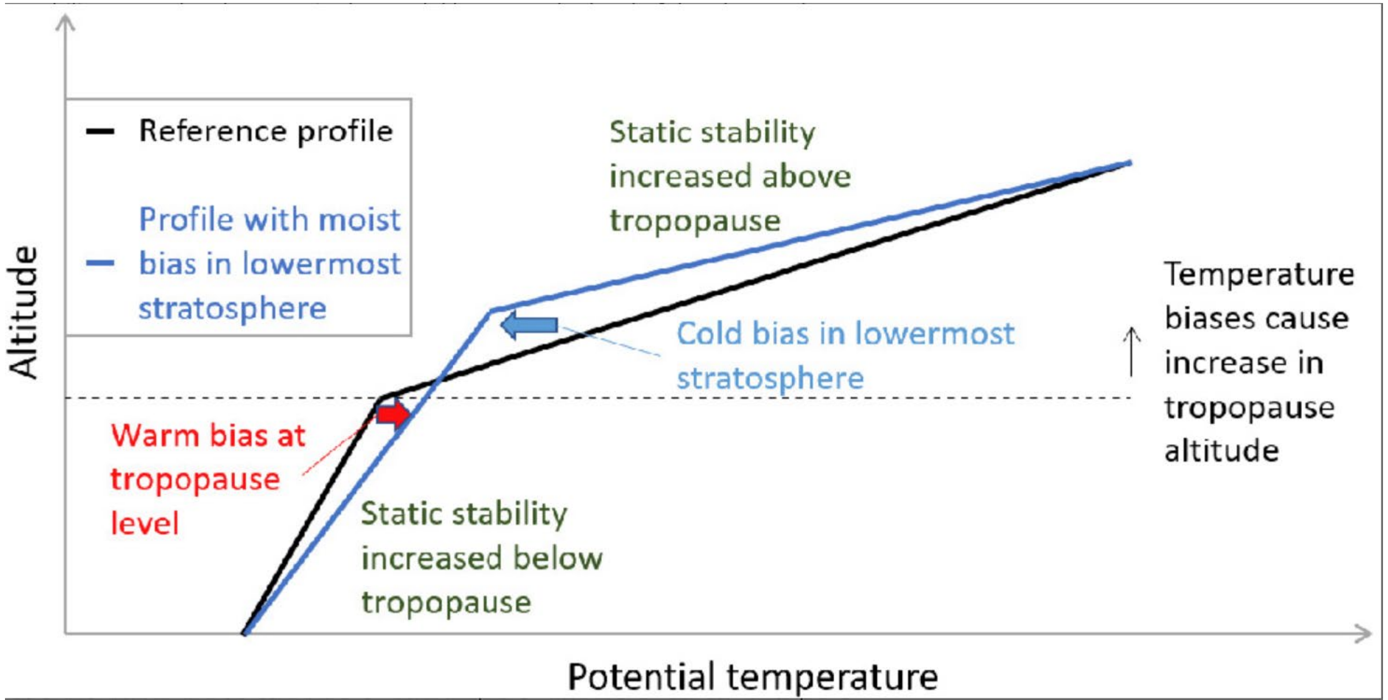
Complexity: NWP – Atmospheric Composition

- Trace gases like H2O, O3, CO, etc, have long lifetimes in the stratosphere with respect to the length of the typical NWP assimilation window (~hours)
- Improved characterisation of their I.C. helps NWP through radiative and dynamic effects and the other way round
- These improvements are long-lived and visible on NWP timescales (2 weeks) and beyond!

Complexity: NWP – Atmospheric Composition (1)

- Cold temperature bias and forecast drift in the lower stratosphere has long been an issue in the ECMWF IFS forecasts (and elsewhere)
- The cold temperature bias originates from long-wave radiative cooling driven by moist stratospheric bias in the initial conditions
- The combination of temperature and humidity biases causes changes in the tropopause and atmospheric static stability



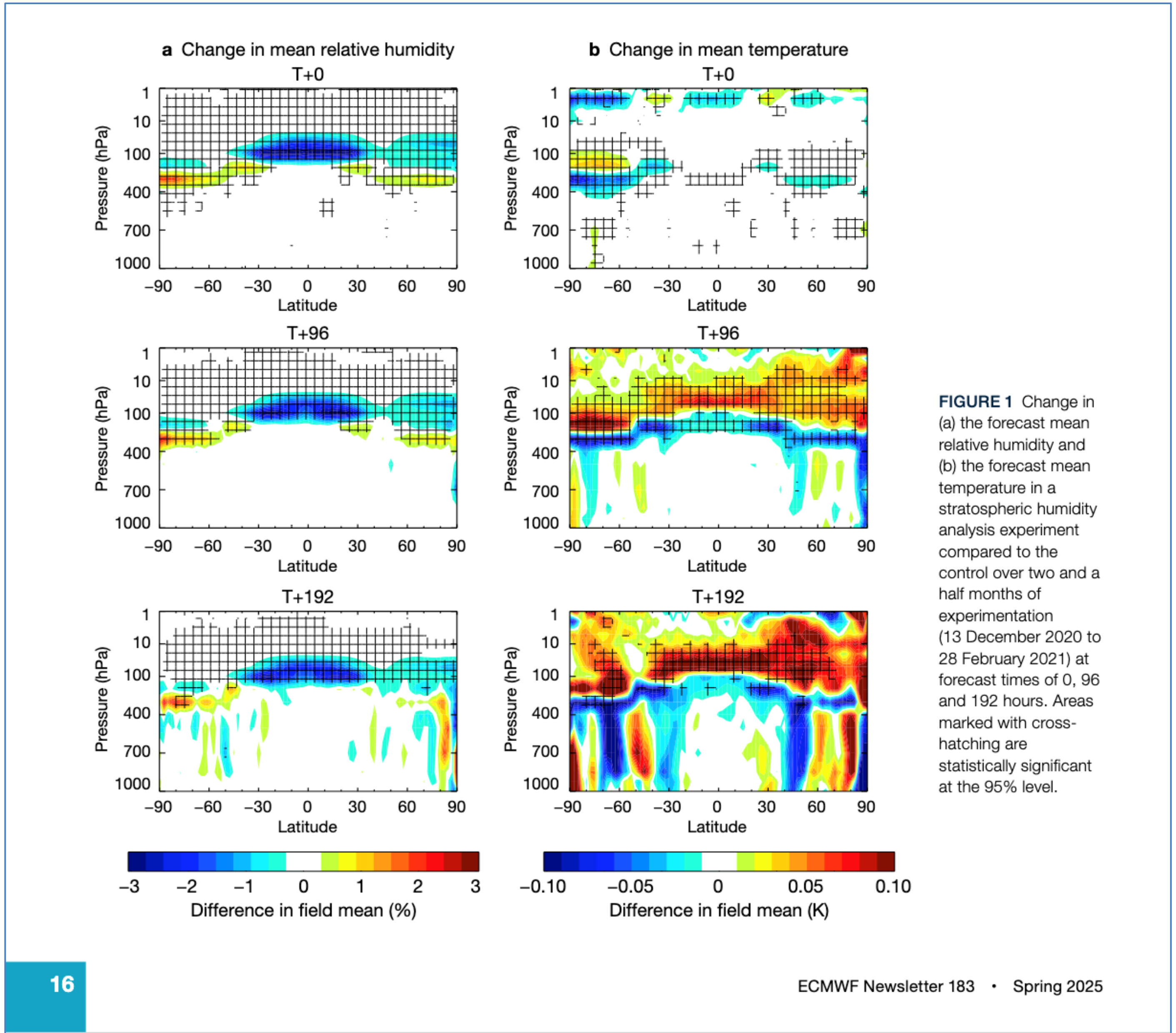


Polichtchouk, I. et al. (2021). Stratospheric Modelling and Assimilation. *ECMWF Tech. Memo.* 877, https://doi.org/10.21957/25hegfoq

Bland, J., et al. (2021). Characterizing extratropical near-tropopause analysis humidity biases and 615 their radiative effects on temperature forecasts, *Q. J. R. Meteorol. Soc.*, 140, 3878-3898

Complexity: NWP – Atmospheric Composition (1)

- Assimilation of humidity in the stratosphere will be re-activated later this year (Cy50r1) after 25 years!
- Large improvement of both humidity and forecast temperature biases
- Made possible by EDA covariances, increased model vertical resolution and improved observational constraint
- Use of current (MLS on Aura) and planned (CAIRT mission) limb sounders gives additional significant improvements



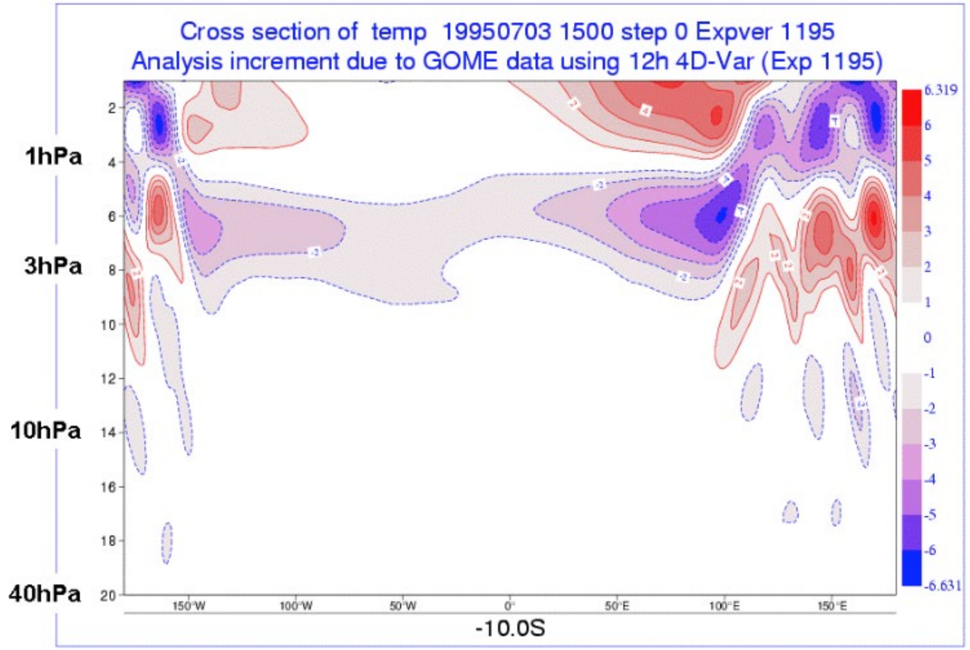
Semane, N., and M. Bonavita (2025). Reintroducing the analysis of humidity in the stratosphere. *ECMWF Newsletter n.183, Spring 2025.*

Complexity: NWP – Atmospheric Composition (2)

Year 2007, Slide courtesy of Dick Dee

4D-Var ozone assimilation

The impact of the ozone data on the temperature analysis at 10S



Ozone assimilation

Can 4D-Var infer stratospheric winds from ozone data?

- The answer is: **Not yet.**
- Assimilation of ozone profile data causes large and unrealistic T/U/V increments near the stratopause to accommodate the observed discrepancies between background and data

Year 2024

$$\frac{\partial \delta \mathbf{v}}{\partial t} = -(\delta \mathbf{v} \cdot \nabla)\mathbf{v} - (\mathbf{v} \cdot \nabla)\delta \mathbf{v}$$

$$\frac{\partial \delta c}{\partial t} = \boxed{-\gamma(x, y, t)}\ (\nabla \cdot (\delta \mathbf{v}\, c)) - \nabla \cdot (\mathbf{v}\delta c)$$

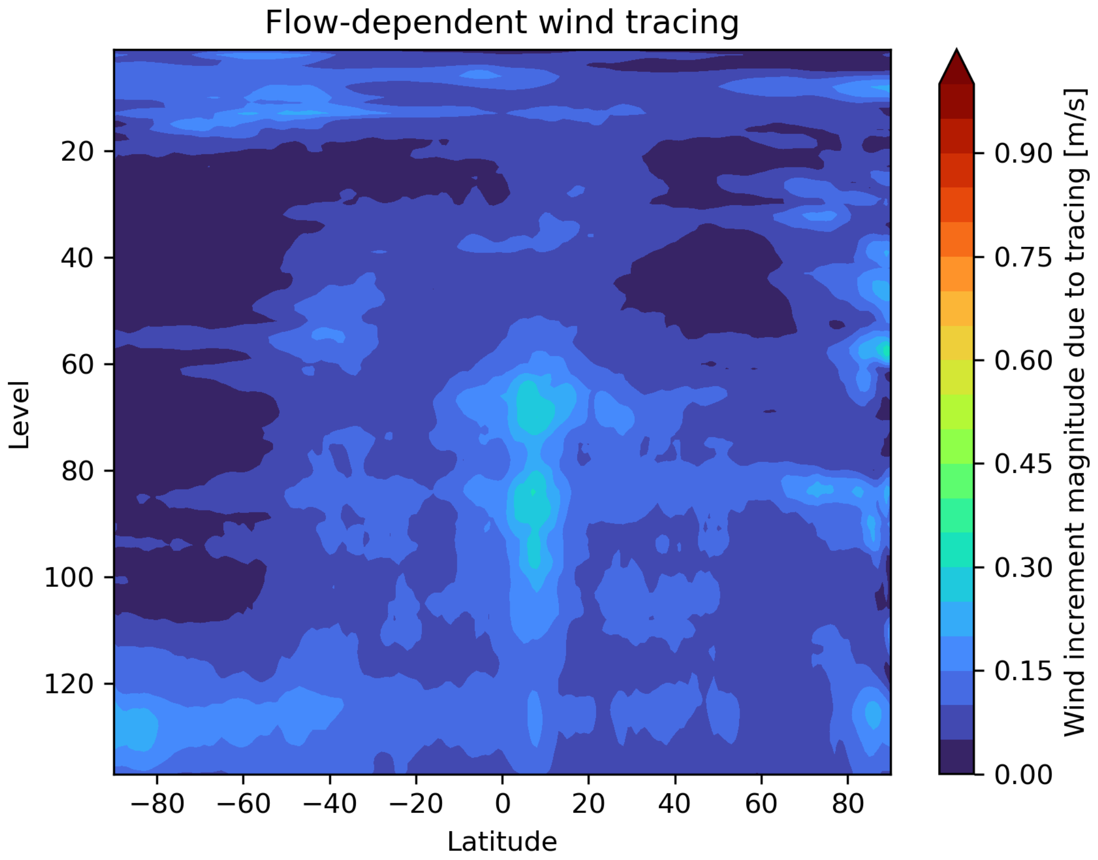
$$-\frac{\partial \delta \mathbf{v}^*}{\partial t} = (\mathbf{v} \cdot \nabla)\delta \mathbf{v}^* - (\nabla \otimes \mathbf{v})\delta \mathbf{v}^* \boxed{+\gamma(x, y, t)}\ \left(-\delta c^* \nabla c + c \nabla \delta c^*\right)$$

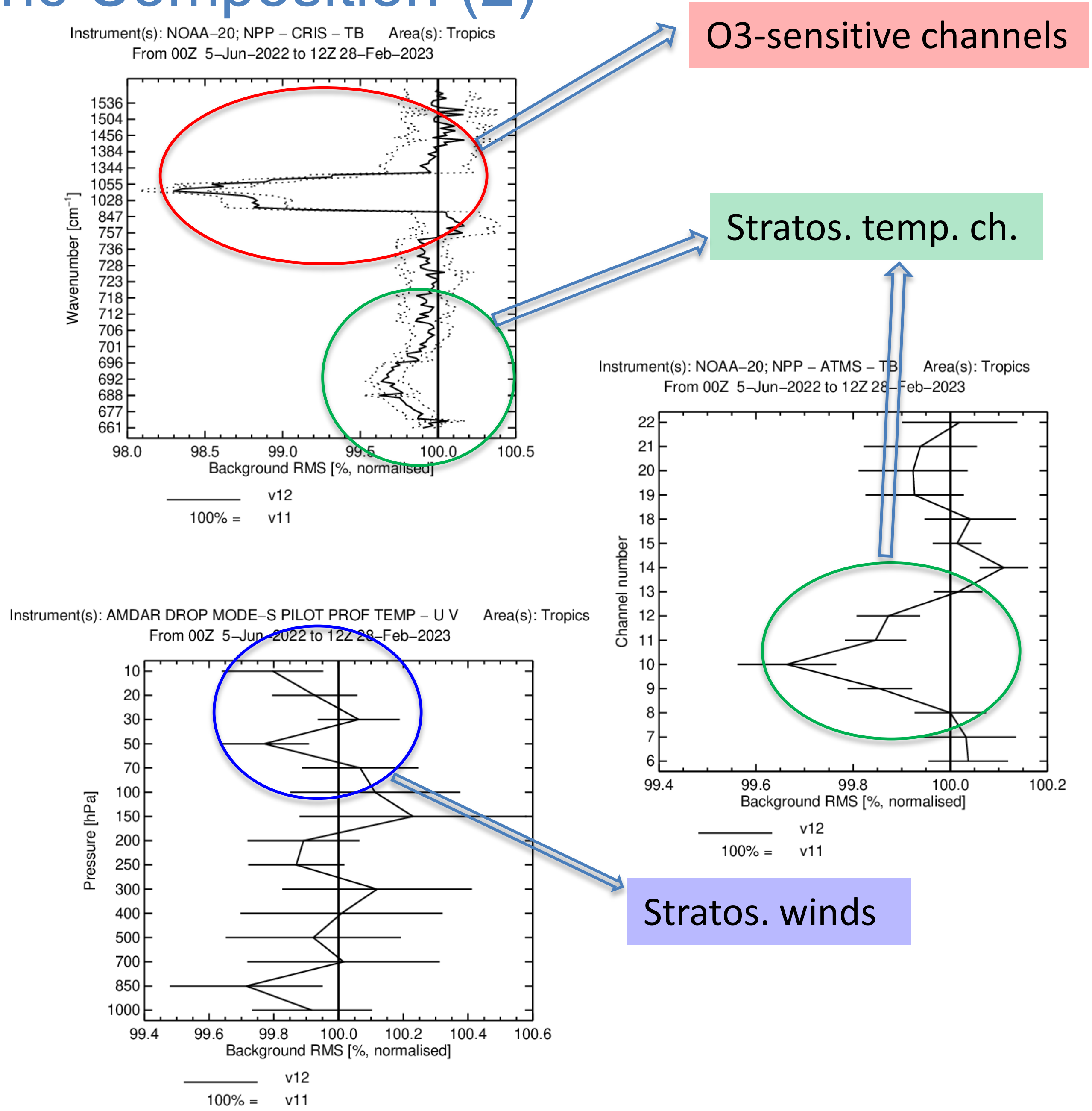
$$-\frac{\partial \delta c^*}{\partial t} = -\nabla \cdot (\delta c^* \mathbf{v}) + \frac{\partial J}{\partial c}.$$

Parameter γ controls the “level” of tracer-wind coupling in the assimilation.

Zaplotnik, Ž., Žagar, N. & Semane, N.(2023) Flow-dependent wind extraction in strong-constraint 4D-Var. *Q. J. R. Meteorol. Soc.*, 149(755, 2107–2124)

Complexity: NWP – Atmospheric Composition (2)

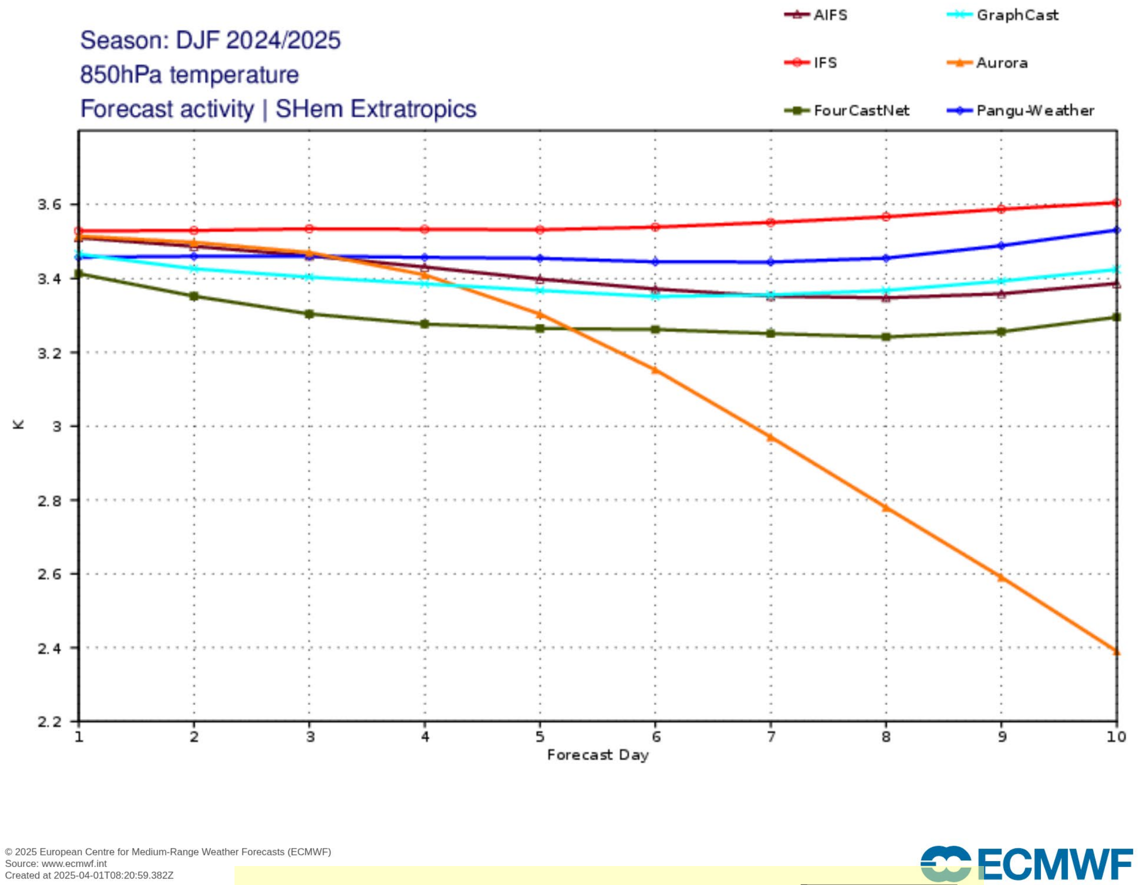




Courtesy of Ziga Zaplotnik, Sebastien Massart & Noureddine Semane, ECMWF

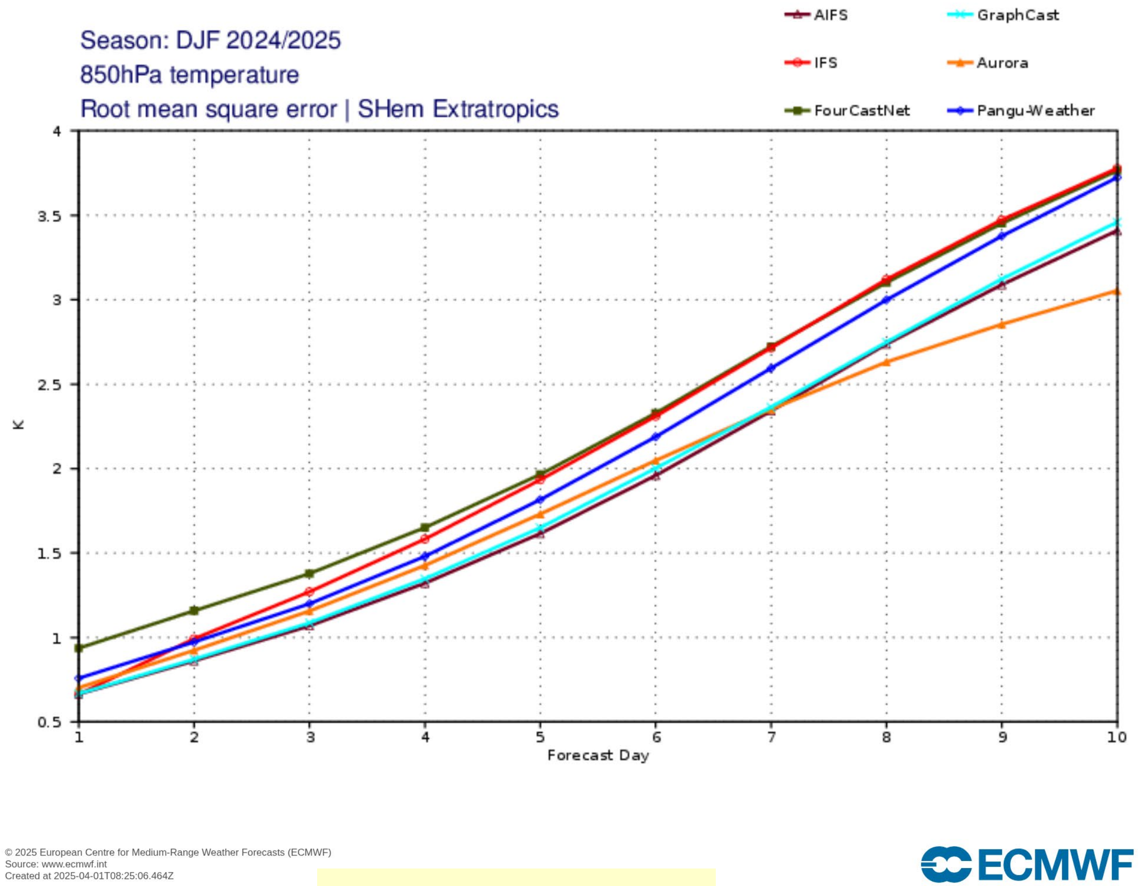
A short digression into Forecast Verification

Scores of forecasts of upper-air parameters by experimental machine learning models



Forecast activity = SDAF = $\sqrt{\overline{(f-c)^2}}$

Scores of forecasts of upper-air parameters by experimental machine learning models



RMSE = $\sqrt{\overline{(f-a)^2}}$

Available at https://charts.ecmwf.int/products/

A short digression into Forecast Verification

A.H. Murphy, 1988*:

$$MSE(f,a) = ME^2 + SDAF^2(1 - 2ACC * SDAV/SDAF) + SDAV^2$$

- MSE can be improved by working on conditional/unconditional model biases (ME^2, calibration/reliability)
- MSE can be improved by working on Forecast Activity (SDAF, smoothing/damping)

MSE = $\overline{(f-a)^2}$

ME = $(\bar{f} - \bar{a})$

Forecast activity = SDAF = $\sqrt{\overline{(f-c)^2}}$

Observed activity = SDAV = $\sqrt{\overline{(a-c)^2}}$

Anom. Corr. Coeff. = ACC = $\frac{\overline{(f-c)(a-c)}}{SDAF*SDAV}$

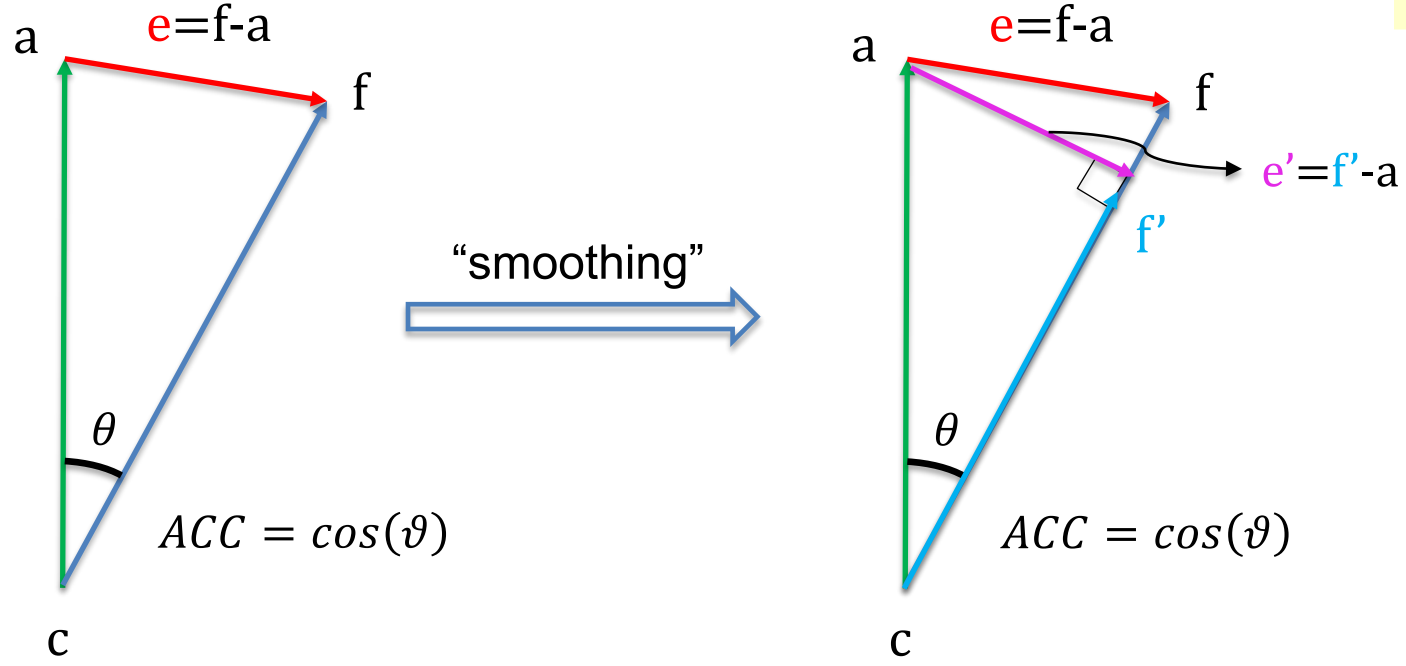
*__Murphy, A. H. (1988__). Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *Mon. Wea. Rev.*, **116**, 2417–2424

A short digression into Forecast Verification (1)

Playing with Forecast Activity (SDAF)…

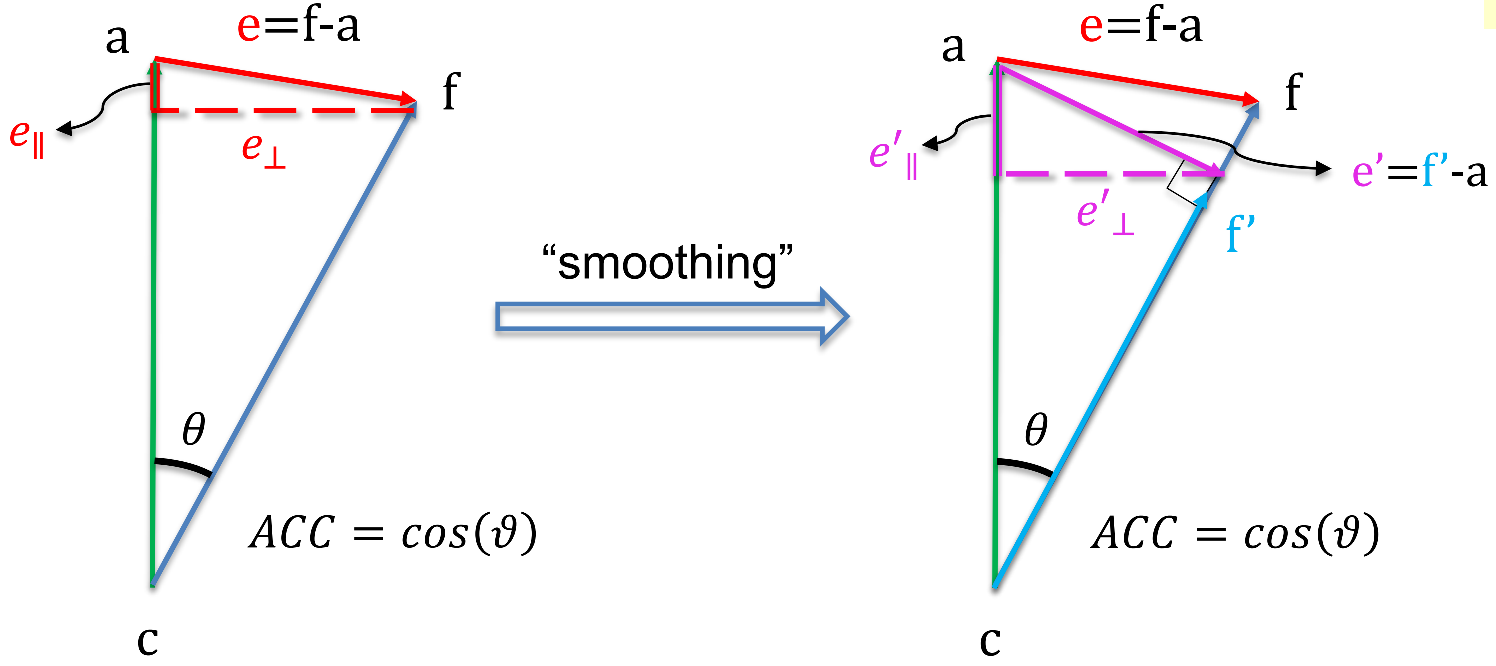
$$|f' - c| = SDAF(f') < SDAF(f) = |f - c|$$

$$|f' - a| = RMS(e') < RMS(e) = |f - a|$$



A short digression into Forecast Verification (1)

Playing with Forecast Activity (SDAF)…



$$|f' - c| = SDAF(f') < SDAF(f) = |f - c|$$

$$|f' - a| = RMS(e') < RMS(e) = |f - a|$$

however:

$$RMS(e'_{\parallel}) > RMS(e_{\parallel}) \quad (1)$$

$$RMS(e'_{\perp}) < RMS(e_{\perp}) \quad (2)$$

The smoothed forecast is less noisy (2) but it is **less informative on the observed anomaly** (1), i.e. it has **lower statistical resolution***

Bonavita and Geer (2025) call:

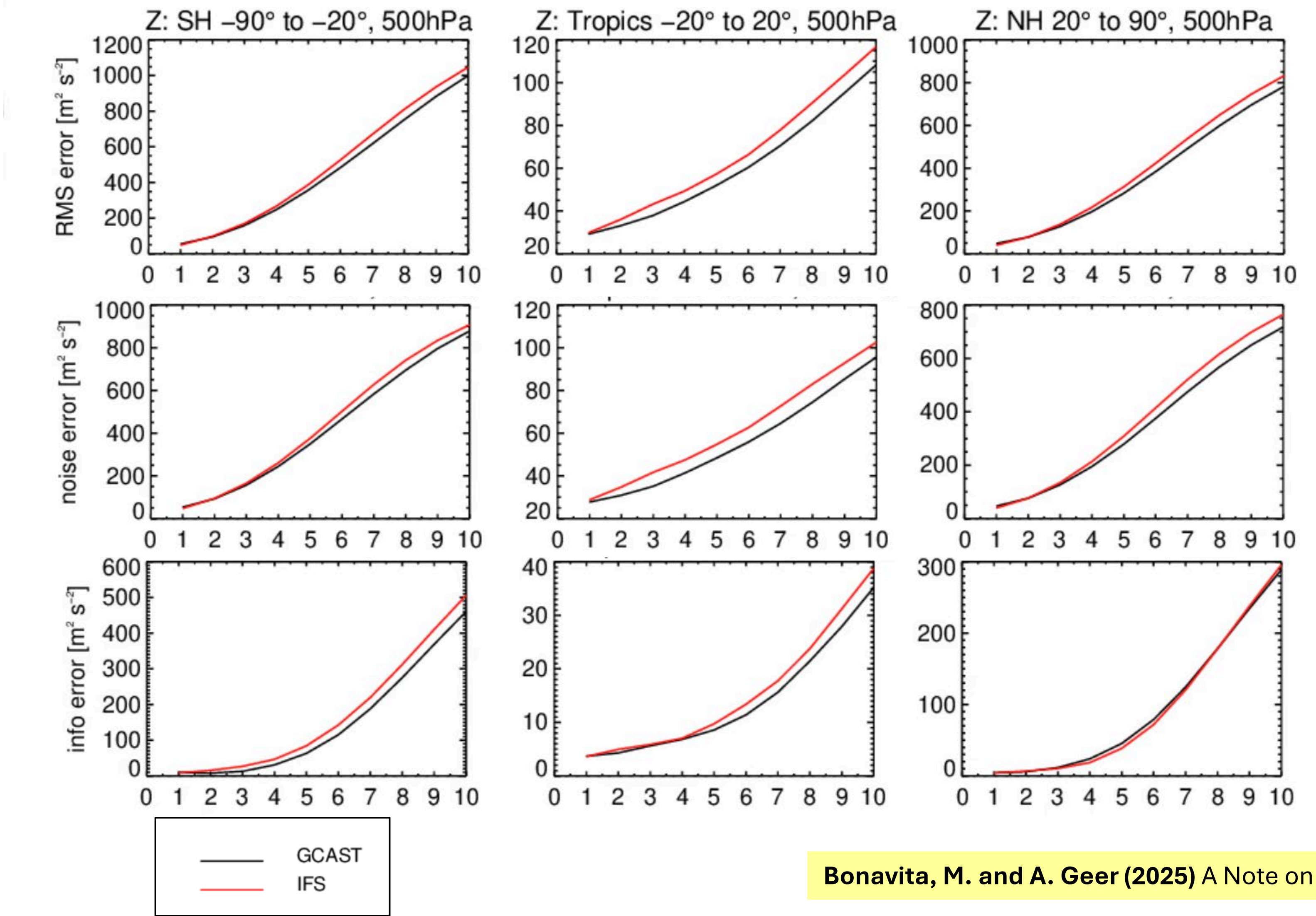
$e_{\parallel}$=Information Error (IE)
$e_{\perp}$=Noise Error (NE)

$$MSE = {e_{\parallel}}^2 + {e_{\perp}}^2 = IE^2 + NE^2$$

***Toth, Z., Talagrand, O. & Zhu, Y. (2005)** The attributes of forecast systems: a framework for the evaluation and calibration of weather forecasts. In: Palmer, T.N. & Hagedorn, R. (Eds.) Predictability of weather and climate, Cambridge University Press, pp. 584–595.

A short digression into Forecast Verification (1)

GraphCast (Lam *et al.*, 2023, *Science* 382,1416-1421)



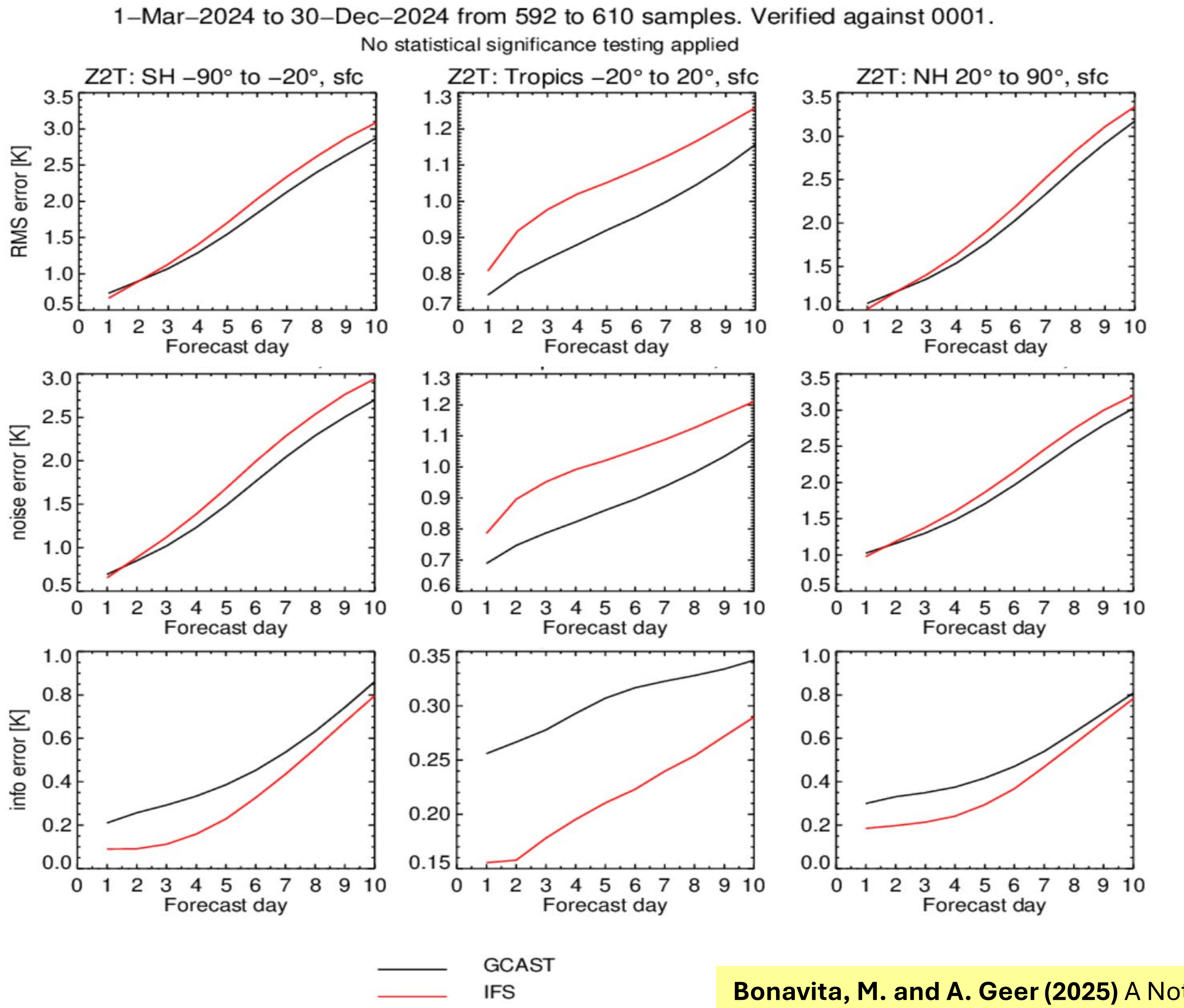
Z500 RMSE

Z500 Noise Error

Z500 Inform. Error

Bonavita, M. and A. Geer (2025) A Note on Forecast Information and Noise, under review.

A short digression into Forecast Verification (1)



T2m RMSE

T2m Noise Error

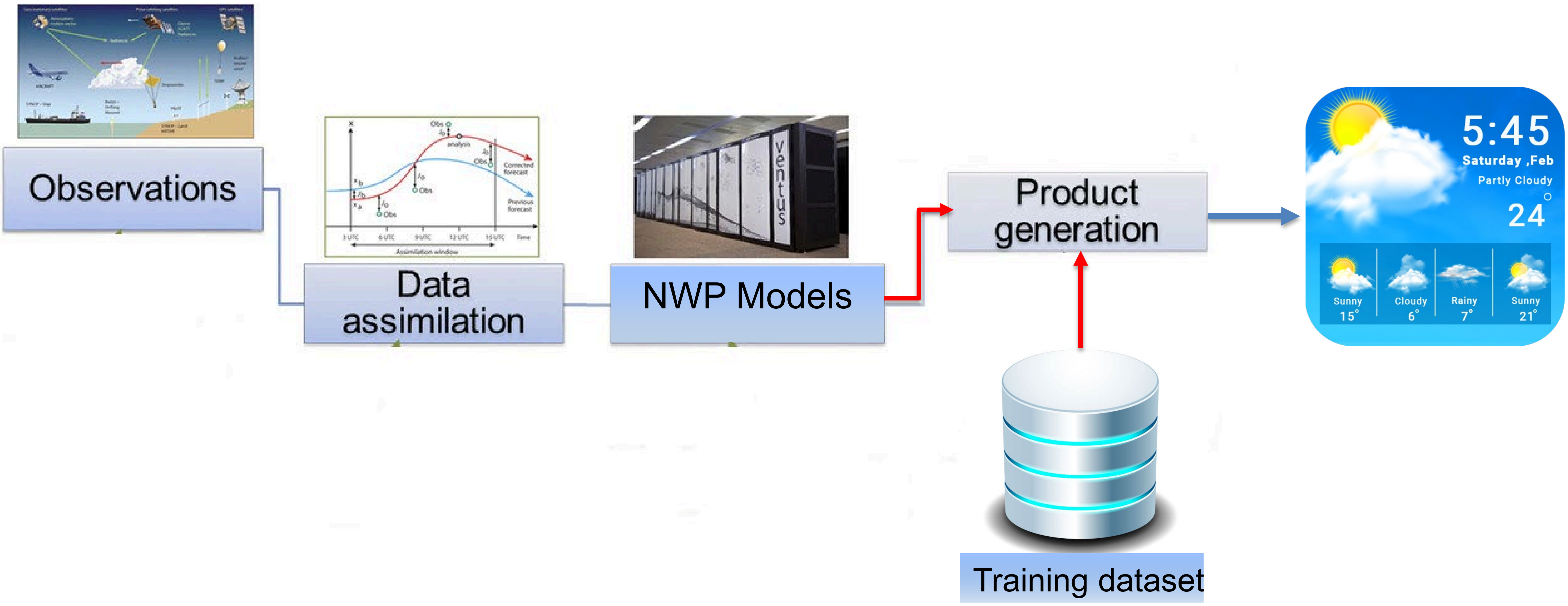
T2m Inform. Error

Bonavita, M. and A. Geer (2025) A Note on Forecast Information and Noise, under review.

A short digression into Forecast Verification (2)

$$MSE(f,a) = ME^2 + SDAV^2 + SDAF^2 - 2ACC * SDAV * SDAF$$

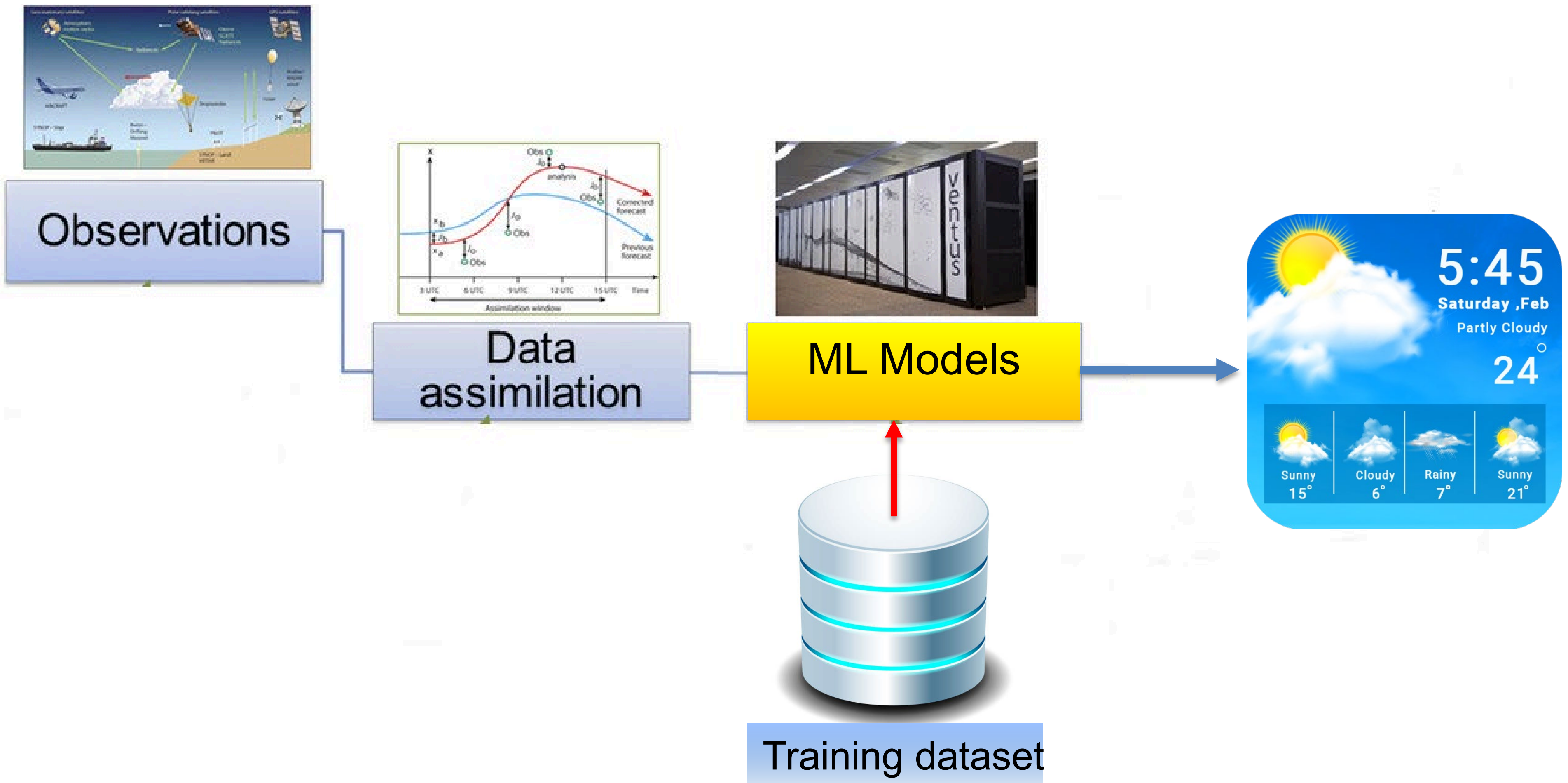
- MSE can be improved by working on conditional/unconditional model biases (ME^2, calibration/reliability)



A short digression into Forecast Verification (2)

$$MSE(f,a) = ME^2 + SDAV^2 + SDAF^2 - 2ACC * SDAV * SDAF$$

- MSE can be improved by working on conditional/unconditional model biases (ME^2, calibration/reliability)



A short digression into Forecast Verification (2)

$$MSE(f,a) = ME^2 + SDAV^2 + SDAF^2 - 2ACC * SDAV * SDAF$$

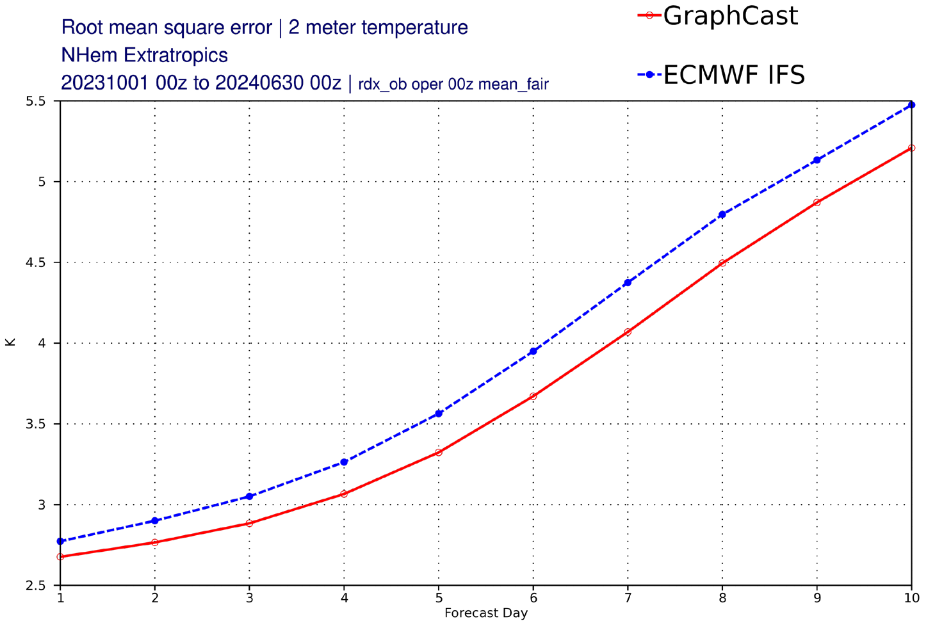
- MSE can be improved by working on conditional/unconditional model biases (ME^2, calibration/reliability)
- NWP model output has not seen Training Dataset(s)
- Comparing NWP model output to ML model output is comparing 

A short digression into Forecast Verification (2)

$$MSE(f,a) = ME^2 + SDAV^2 + SDAF^2 - 2ACC * SDAV * SDAF$$

- MSE can be improved by working on conditional/unconditional model biases (ME^2, calibration/reliability)
- NWP model output has not seen Training Dataset(s): What if it does?

T2m Nhem RMSE, Oct 2023 – Jun 2024

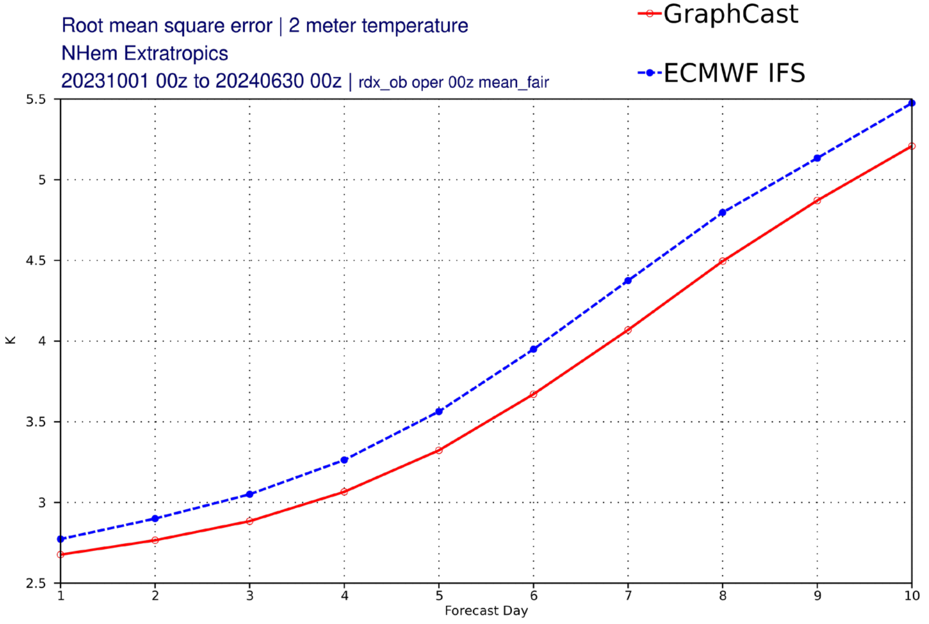


A short digression into Forecast Verification (2)

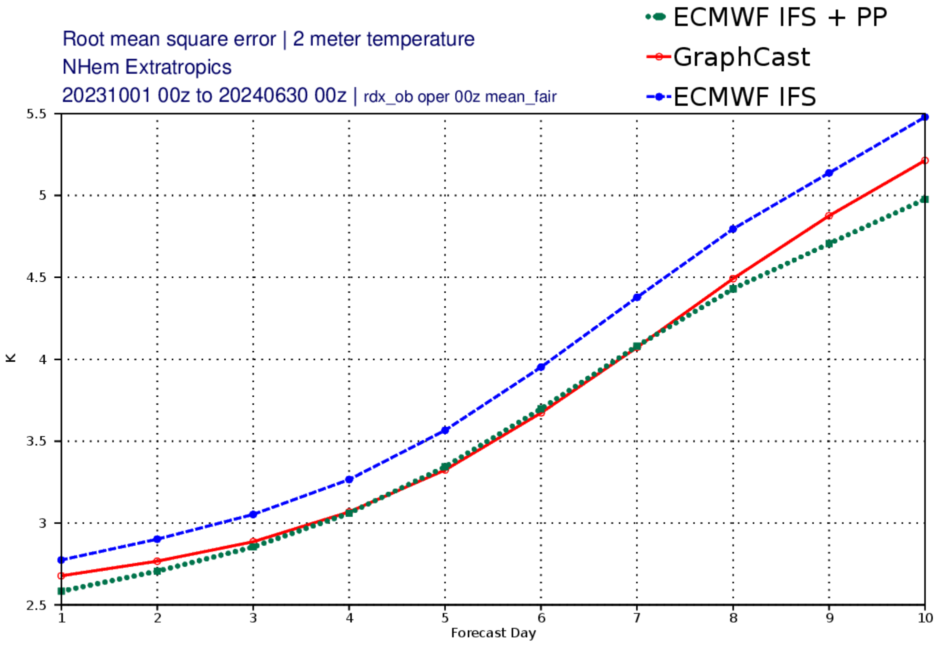
$$MSE(f,a) = ME^2 + SDAV^2 + SDAF^2 - 2ACC * SDAV * SDAF$$

- MSE can be improved by working on conditional/unconditional model biases (ME^2, calibration/reliability)
- NWP model output has not seen Training Dataset(s): What if it does?

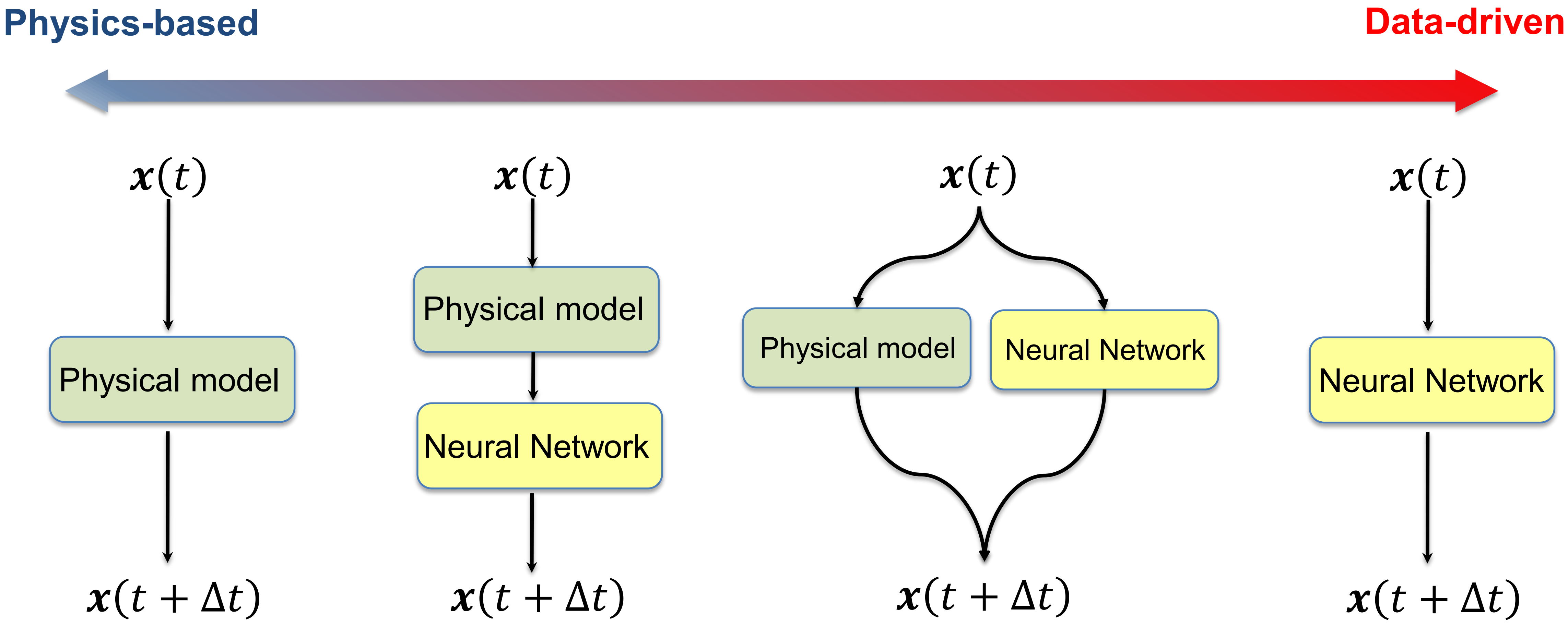
T2m Nhem RMSE, Oct 2023 – Jun 2024



Post-process ECMWF IFS **T2m** forecast output with toy NN (**7.3K trainable parameters**) using **4 years** of IFS T2m operational analyses/forecasts



The way forward



NWP models
(IFS, ICON, ARPEGE, UKMO,…)

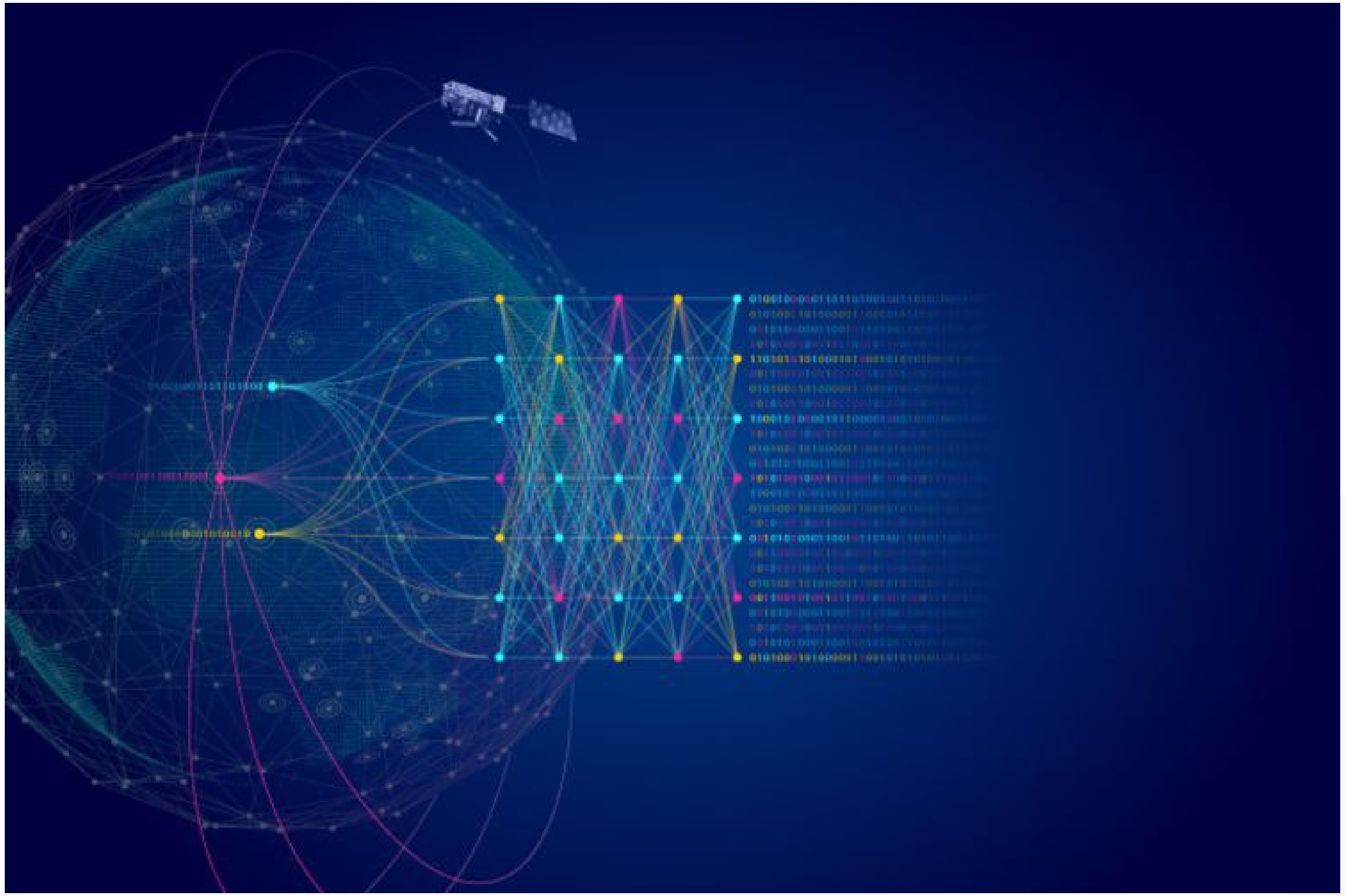
Hybrid models
(e.g. Bonavita and Laloyaux, 2020; Farchi et al., 2023, 2024; …)

Partial model emulation
(e.g. Hatfield at al., 2021, Chantry et al., 2021; Geer, 2024 …)

Full emulators
(e.g. Keisler, 2021, FourCastNet, NVIDIA, 2022; PanguWeather, 2022; GraphCast, 2022, AIFS, 2024,…)

ECMWF

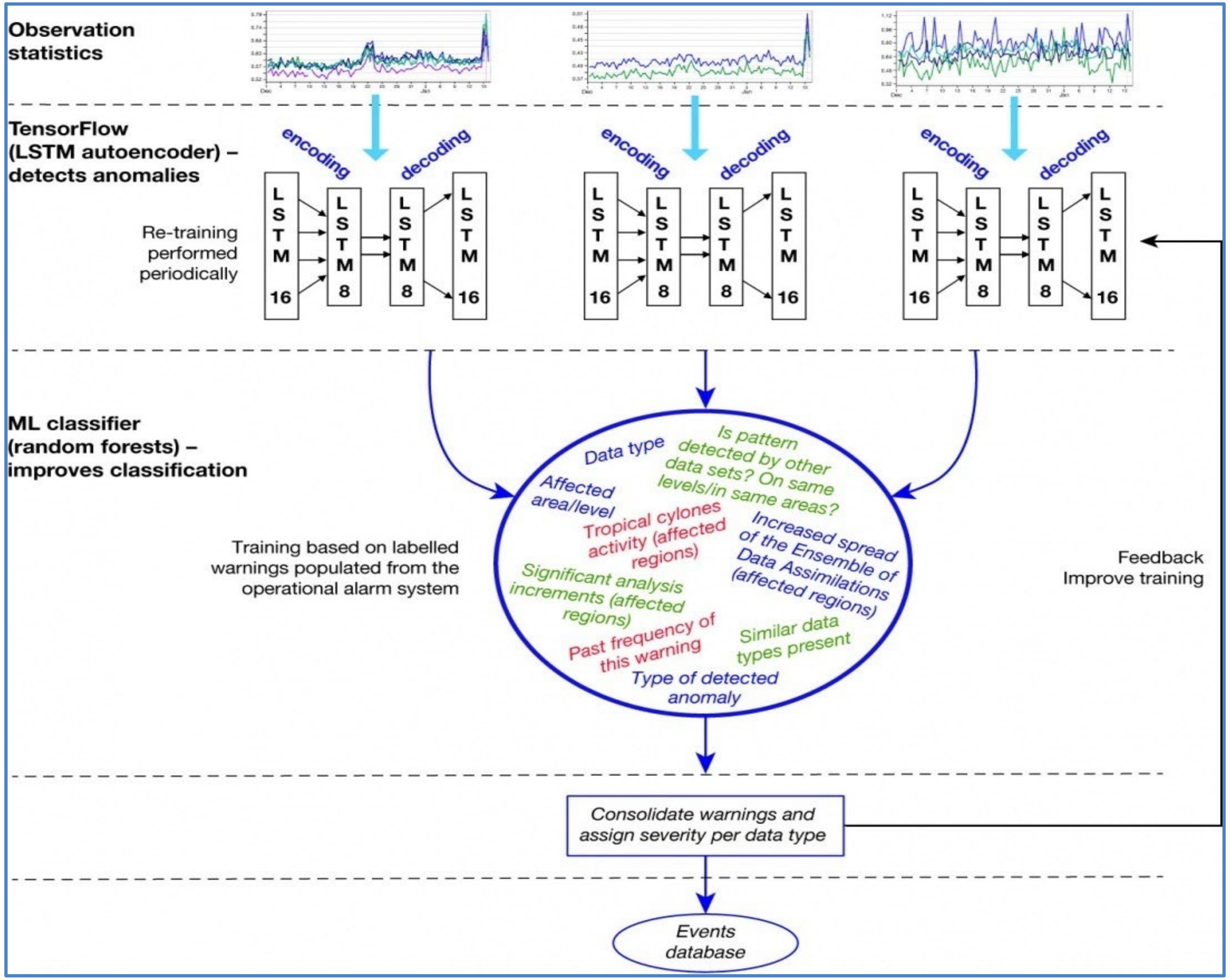
Thanks!



ECMWF

Additional slides

The evolution path: Observation and DA System Monitoring



O-B departures timeseries

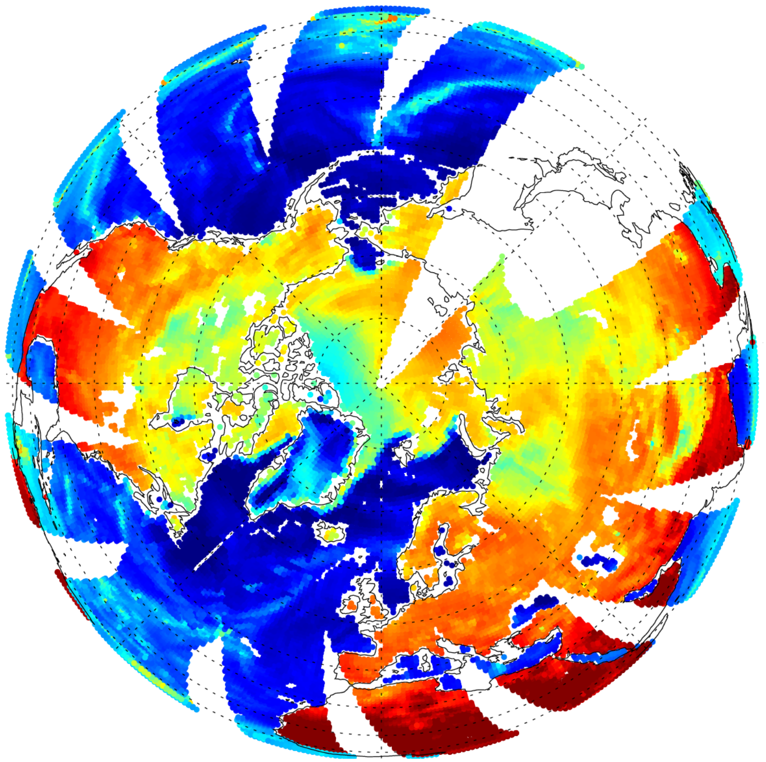
Anomaly detection module
(LSTM autoencoder**)**

Event classifier
(Random Forest**)**

Dahoui, M. (2023). Use of machine learning for the detection and classification of observation anomalies, *ECMWF Newsletter N. 174, Winter 2023*

The evolution path: Hybrid Physical-ML Observation models (H)

Satellite observations



SSMIS F-17 channel 13 (19 GHz, v)
Microwave brightness temperatures
3rd December 2014

Geophysical variables

Equations & parameters – where sometimes knowledge is quite uncertain

$$y = h\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_{\dots} \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_{\dots} \end{bmatrix} \right)$$

Forward function

Gas spectroscopy

Scattering from hydrometeors

Cloud and precipitation micro and macro-structure

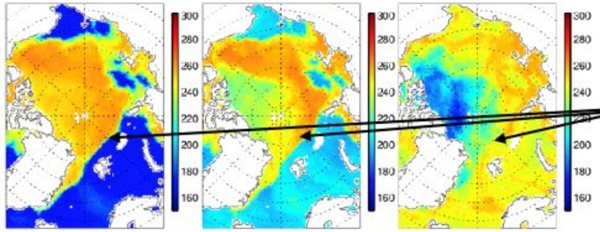
Snow / ice grain size and structure

…..

From Alan Geer, ECMWF

The evolution path: Hybrid Physical-ML Observation models (H)

AMSR2 observations



$$J_{\text{obs}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(y_{\text{obs},ij} - y_{\text{sim},ij})^2}{r_j^2}$$

Known atmosphere radiative transfer

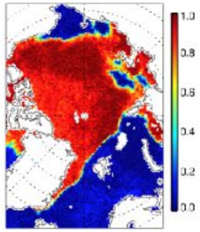
$$\mathbf{y} = f_{\text{atmos}}(\mathbf{e}, T_S, \mathbf{z}_A)$$

Mixed surface emissivity

$$\mathbf{e} = C_{\text{ice}}\mathbf{e}_{\text{ice}} + (1 - C_{\text{ice}})\mathbf{e}_{\text{wat}}$$

Known water surface emissivity

Map of sea ice fraction to be estimated

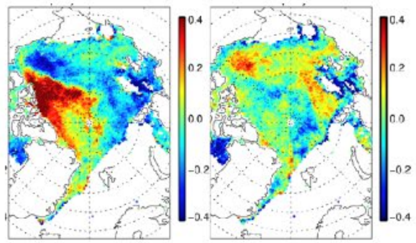


$$C_{\text{ice}} = h(\mathbf{C}_{\text{ice}}, \mathbf{z}_G)$$

$$\mathbf{e}_{\text{ice}} = f_{\text{empirical}}(\mathbf{w}, \mathbf{x}_{\text{ice}}, \mathbf{z}_B)$$

Geolocation

Maps of empirical parameters representing **unknown** sea ice state including microstructure



$$\mathbf{x}_{\text{ice}} = h(\mathbf{X}_{\text{ice}}, \mathbf{z}_G)$$

w – trainable weights of NN model for sea ice

h() Interpolation operator: map to observation location in time and space

Fixed parameter

Dependent parameter

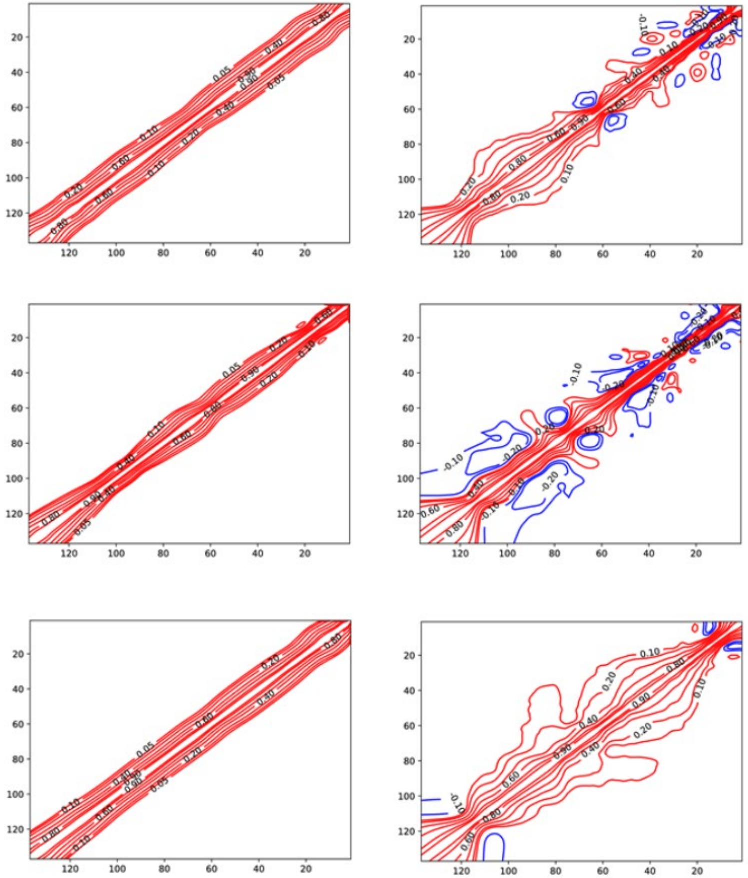
Trainable parameter

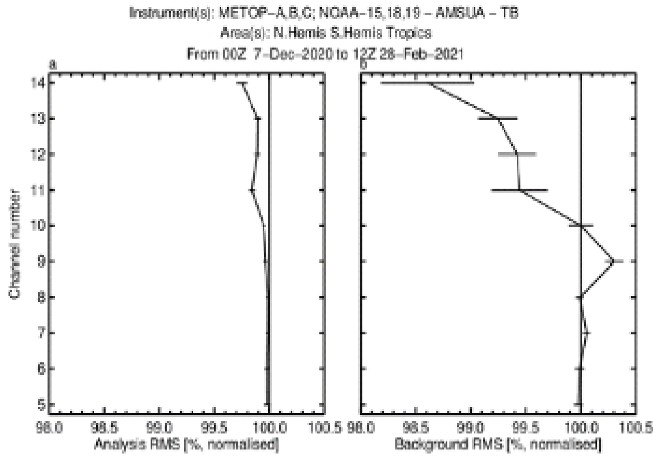
Geer, A. J. (2024). Simultaneous inference of sea ice state and surface emissivity model using machine learning and data assimilation. *JAMES*, https://doi.org/10.1029/2023MS004080

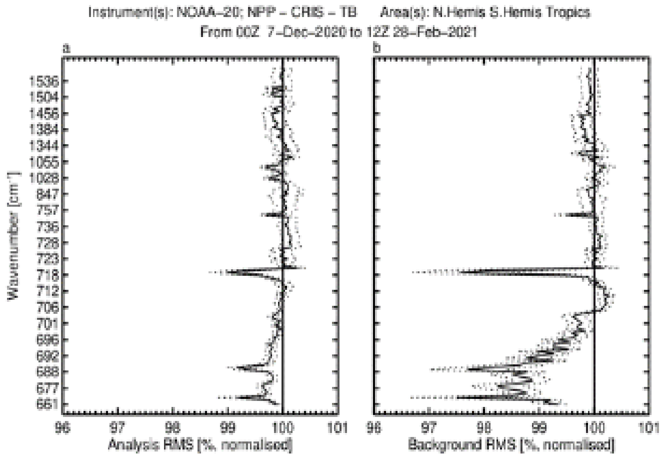
The evolution path: Data Assimilation (1)

Estimation of **model error covariance** $\mathbf{Q}$ for weak constraint 4D-Var using ANN:

$$J_{wc4DVar}(\boldsymbol{x}_0, \boldsymbol{\eta}) = J_B + J_O + J_Q = J_B + J_O + \frac{1}{2}\sum_{i=1}^{N}\Big(x_i - M_i(x_{i-1}, \boldsymbol{\eta})\Big)^T \mathbf{Q}^{-1}\big(x_i - M_i(x_{i-1}, \boldsymbol{\eta})\big)$$







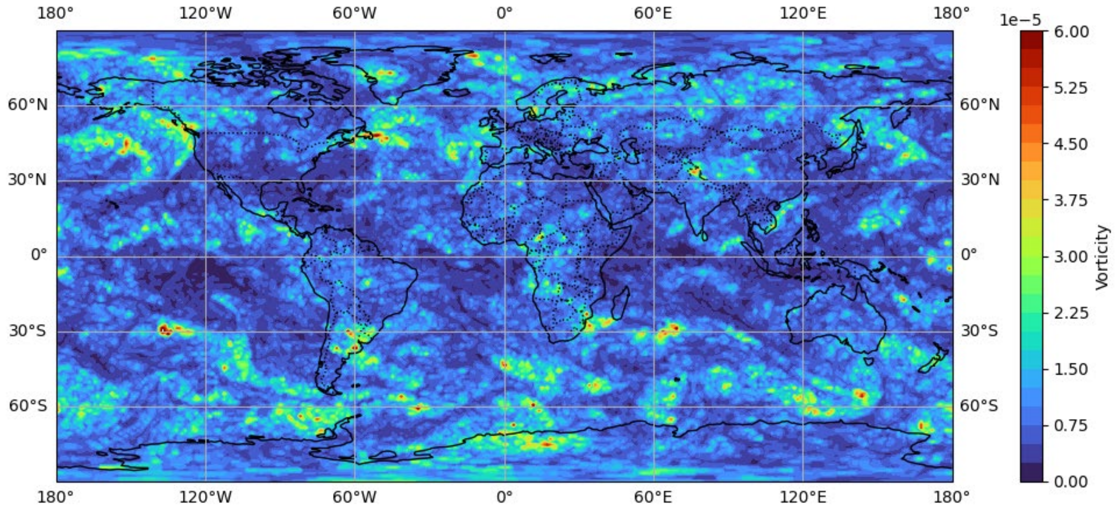
Bonavita, M. and P. Laloyaux (2022). Estimating Model Error Covariances with Artificial Neural Networks https://doi.org/10.48550/arXiv.2209.11510

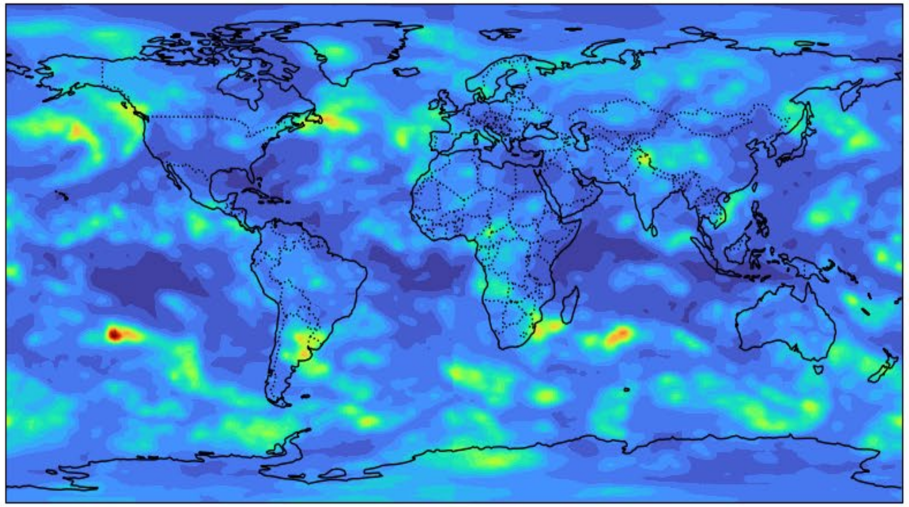
The evolution path: Data Assimilation (2)

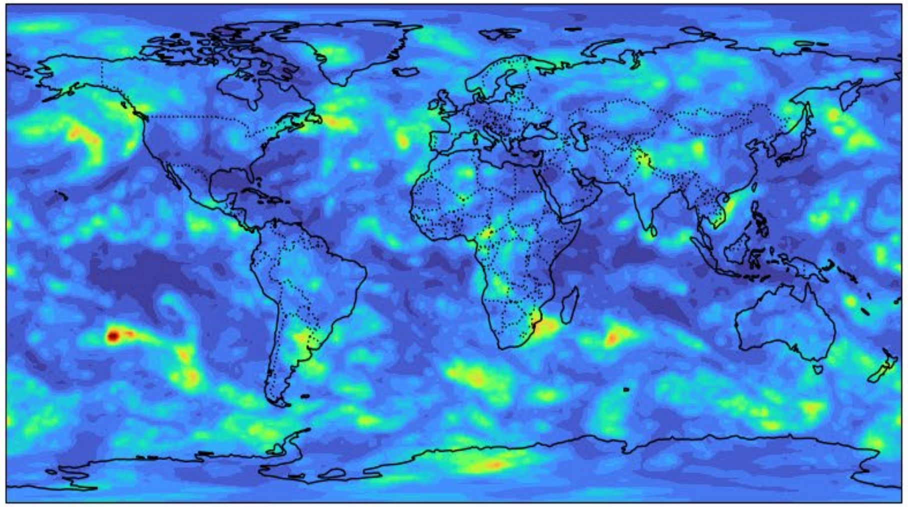
Emulation of Ensemble DA **analysis/background error Cov A/B** using generative AI/ML:

Input: **5-member** EDA sampled BG-error stdev field
Output: Emulation of operational **50-member** EDA sampled BG error stdev

$$\hat{\Sigma}_N \xrightarrow{\text{VED}} \mathbb{P}_\theta(\Sigma \,|\, \hat{\Sigma}_N) \xrightarrow{\text{samples}} \bar{\hat{\Sigma}} \approx \Sigma.$$



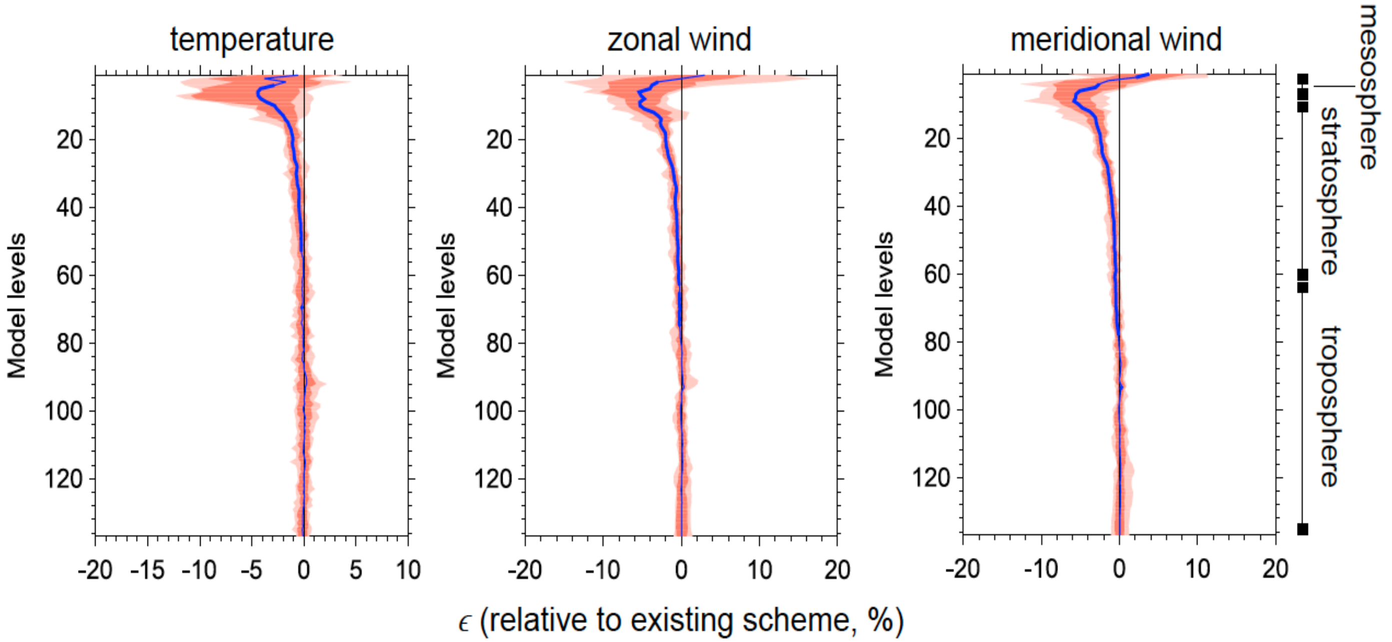




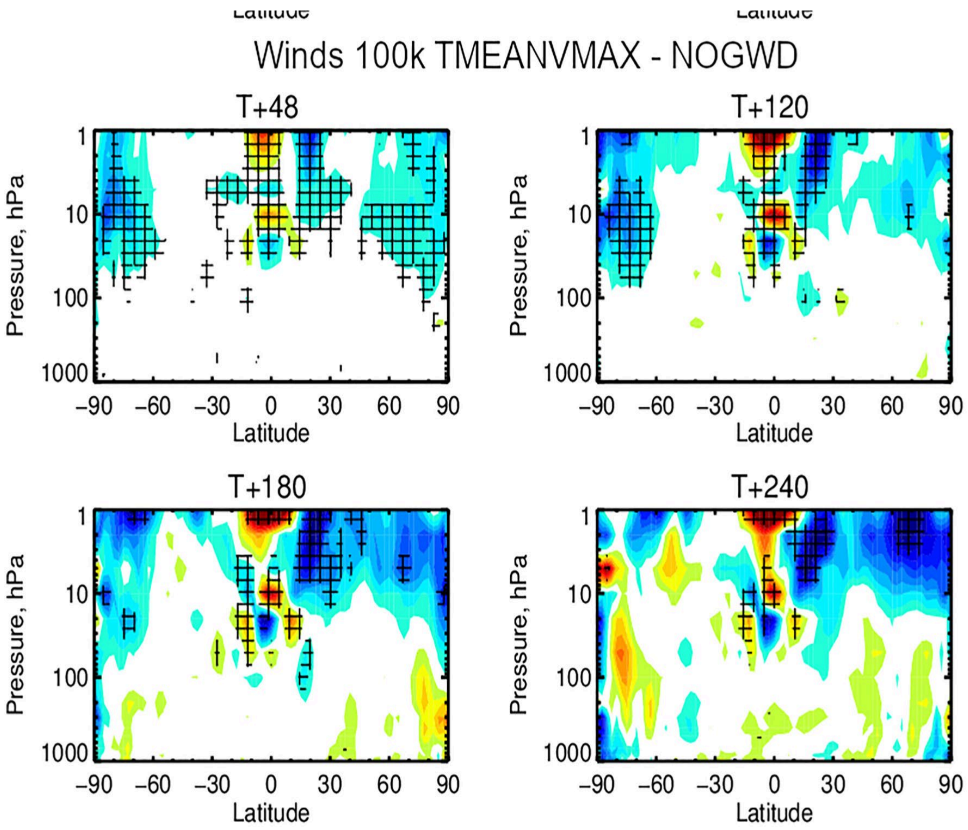
Wei Pan,
ECMWF

The evolution path: Emulation of Model Components

- Emulation of TL/ADJ of ECMWF *nonorographic gravity wave drag scheme* by Hatfield et al., 2021



Accuracy of the Neural Network TL model wrto manually coded TL model



Hatfield, S., et al. (2021). Building tangent-linear and adjoint models for data assimilation with neural networks. *JAMES*, 13, e2021MS002521. https://doi.org/10.1029/2021MS002521

The evolution path: Model Error estimation and correction

- **Prognostic model as hybrid** of physical model ($\mathcal{M}$, IFS) and data-driven state-dependent model error tendency correction ($\mathcal{F}$, ANN):

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

- **Train the model error model directly on observations**: The 4D-Var cost function is adapted to estimate the parameters p (weights and biases) of the ANN $\mathcal{F}$:

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

- The cost function $\mathcal{J}^{nn}$ (p) is minimised using the standard incremental 4D-Var formulation

From Marcin Chrust and Alban Farchi, ECMWF