

A Consortium for COnvection-scale modelling
Research and Development

SUMMARY REPORT

EWGLAM DATA ASSIMILATION 2025

Magnus Lindskog

General DA session

Session on Data Assimilation Tuesday afternoon (Chair: Reima Eresmaa; co-chair: Loïk Berre)

14:00-14:20	Christoph Schraff: “Update on data assimilation in COSMO” (online)
14:20-14:40	Marco Milan: “LAM-DA impact during the forecast time”
14:40-15:00	Magnus Lindskog: “Algorithmic developments in ACCORD”
15:00-15:20	Antonín Bučánek: “Advances in the use of observations in ACCORD”
15:20-15:40	Valerie Vogt: “The new AROME-France E-suite: data assimilation aspects and general performances”

Parallel DA session

Parallel session on Data Assimilation Wednesday (Chair: Magnus Lindskog; co-chair: Benedict Strajnar)

10:40-11:00 – Benjamin Menetrier: “A modern implementation of the spectral background error covariance matrix”

11:00-11:20 – Swapan Mallick: "Importance of remote sensing observation in reanalysis data"

11:20-11:40 – Xu Xu: “High-resolution data assimilation in global and regional ICON for the GLORI Digital Twin”

11:40-12:00 – Lee Hawkness-Smith: “Improvements to direct assimilation of radar reflectivity in Met Office UKV, including use of French and German radars”

12:00 – Thomas Gastaldo, Poster intro: “High-Resolution data assimilation with ICON and KENDA at Agenzia ItaliaMeteo and Arpae “

12:00-12:10 – Discussion

Update on Data Assimilation in COSMO

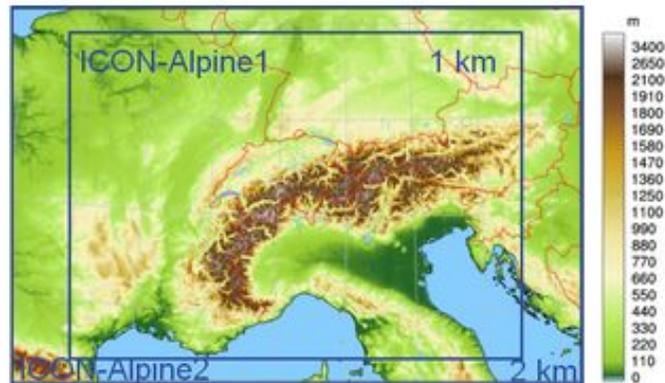
Christoph Schraff

- Conventional DA: KENDA, based on LETKF / DACE code:
 - “DA resources” diverted to AI (DA + model emulators)
→ very little algorithmic developments (LETKF, EnVar; 4D-EnVar (global), PF)
 - observations (crowd-sourced surface, tower, satellite, DIAL) (DWD, ARPAE, CNMCA, MeteoSwiss)
- AI → AI-Var (Keller and Potthast, 2024 (arXiv:2406.00390v1))
(& reanalyses)



Regional High Resolution Assimilation (ICON-LAM)

GIORI-Alps
Basic Setup



High resolution assimilation (1 km)



- Deterministic Run: 2 km
- Ensembles: 2 km
- Assimilation: LETKF
- Ensemble members: 40

- Nest (two-way-nested)
- Deterministic Run: 1 km
- Ensembles: 1 km

- Deterministic Run: 1 km
- Ensembles: 1 km
- Assimilation: LETKF
- Ensemble members: 40
- Nest (two-way-nested)
- Deterministic Run: 500m
- Ensembles: 500m

Boundary condition is interpolated from 6.5km EU nest



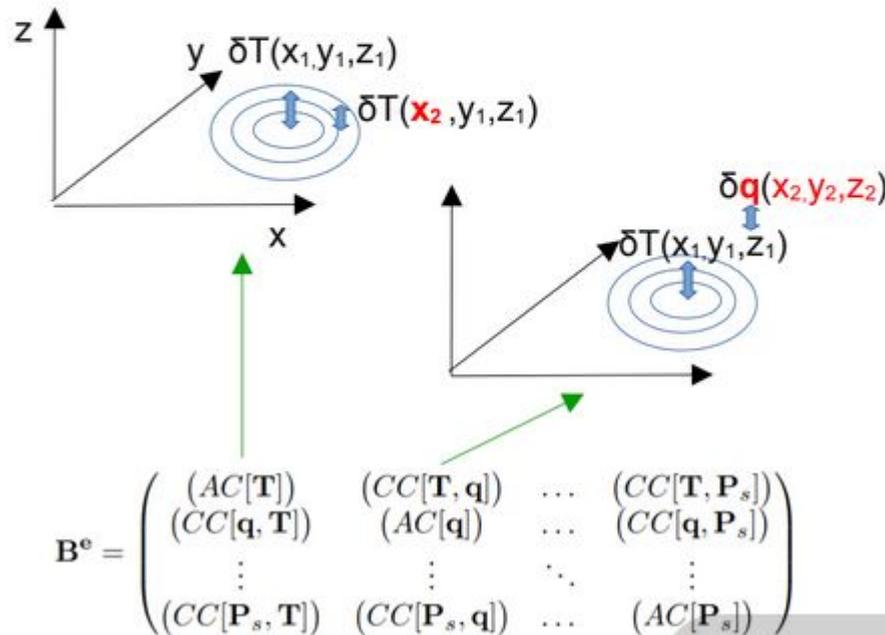
Algorithms



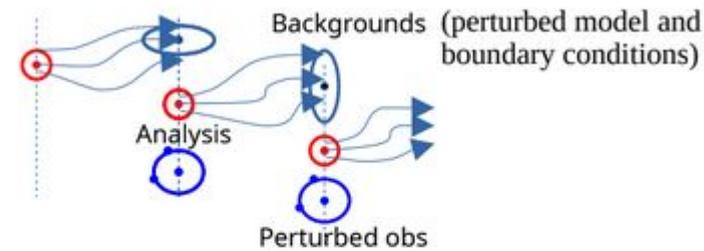
Introduction : 3DENVar operational at Météo-France since October 2024

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x})^T \mathbf{B}^e \delta\mathbf{x} + \frac{1}{2}(\mathbf{d} - \mathbf{H}\delta\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\delta\mathbf{x})$$

Control Variable : $\mathbf{Cv}=(\mathbf{T},\mathbf{q},\mathbf{U},\mathbf{V},\mathbf{P}_s)$



- **Estimation of \mathbf{B}^e in EnVar** : at Météo-France, \mathbf{B}^e is estimated using forecasts from a 50-member EDA to sample uncertainties of the assimilation cycle



- **Flow dependent perturbations** are deduced from this ensemble :

$$\epsilon_l^b = \frac{1}{\sqrt{N_e - 1}} (\tilde{\mathbf{x}}_l^b - \langle \tilde{\mathbf{x}}^b \rangle)$$

N_e nb of members
 $\tilde{\mathbf{x}}_l^b$ Forecast of member l
 $\langle \tilde{\mathbf{x}}^b \rangle$ Mean of the forecasts

- Background error covariances are directly sampled from forecast perturbations, with a **localisation step**

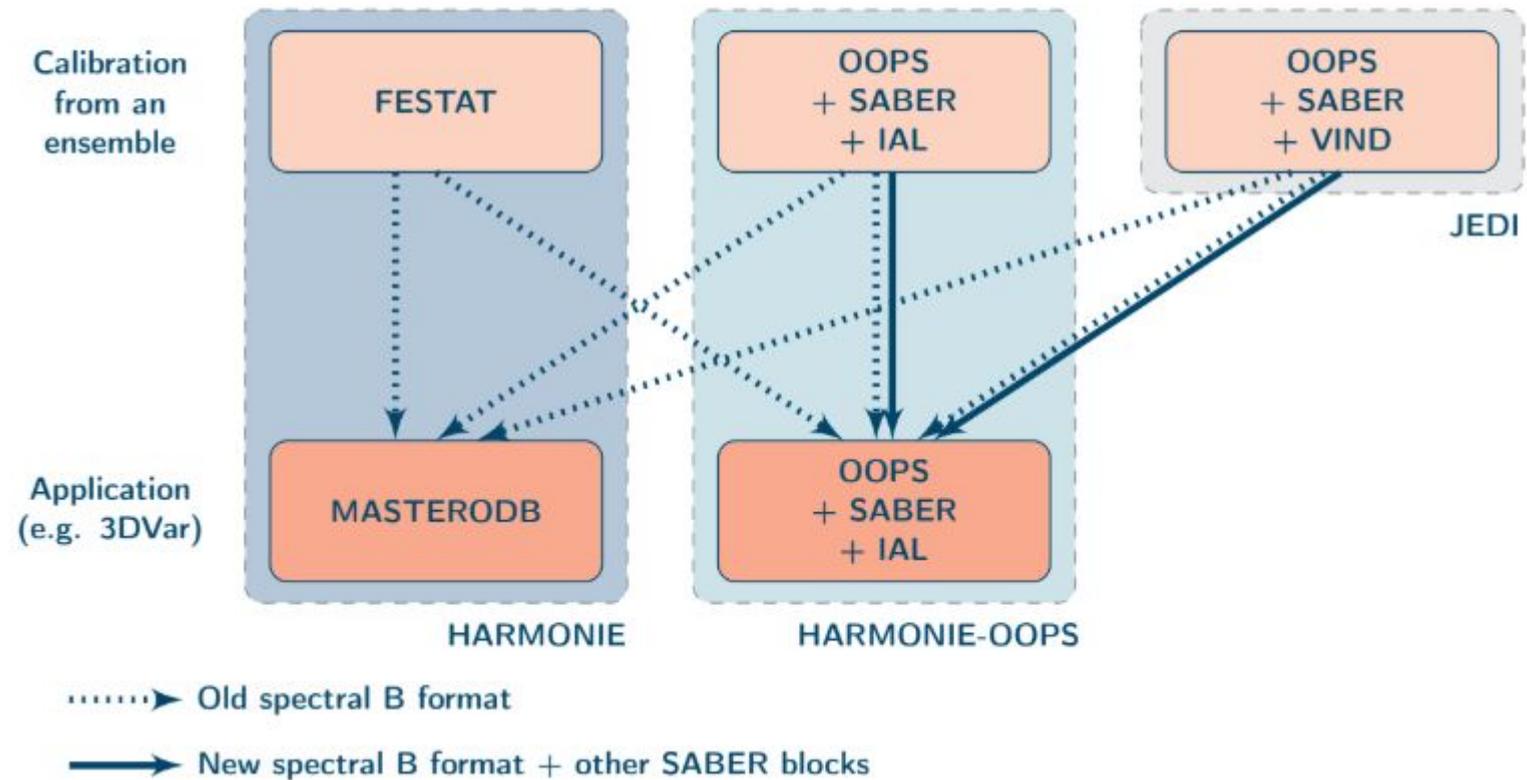
$$\mathbf{X}^b = [\epsilon_1^b, \dots, \epsilon_{N_e}^b] \quad \mathbf{B}^e = \mathbf{X}^b \mathbf{X}^{bT} \circ \mathbf{C}$$

Algorithms

Spectral B ooo	Old implementation oo	SABER oooo	New implementation ooo	Interoperability oo●	Advantage oooo	Conclusions o
-------------------	--------------------------	---------------	---------------------------	-------------------------	-------------------	------------------

Combining calibration and application steps

Interoperability between environments:



Bias-correction of microwave radiance observations using ML

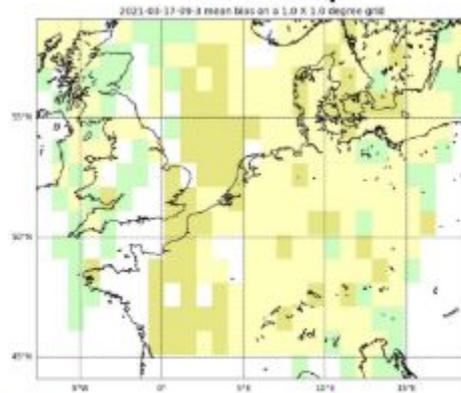
- VarBC updates the bias coefficients β 's at every cycle:
- Update β 's with a NN instead of VarBC
 - Training: DINI domain
 - Testing: Dutch domain
- Use the NN predictions of bias coefficients in a Harmonie-Arome run, and compare:
 - Traditional H-A
 - ML bias correction H-A
 - Denial H-A

$$b(x, \beta) = \beta_0 + \beta_1 p_1(x) + \beta_2 p_2(x) + \dots + \beta_n p_n(x)$$

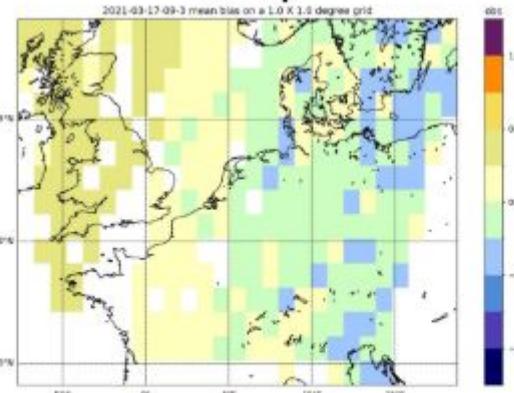


Mean bias from AMSU-A 09:00 UTC

Default exp



ML exp





CONCLUSIONS

- We analysed a winter (December 2021) and a summer (July 2022) month using different sensitivity tests.
- The DA impact => using dump retrievals from a fully cycling DA suite.
- Cycle-by-cycle experiments affects equally each forecast generated. Differences can be small and more difficult to interpret as the dump file contains information of the increments added at T-1 in the fully cycling DA suite. The Control retrieves and uses the dump file and all increments.
- Two main sensitivity experiments were made: ALL_OFF and ALL_OFF+LSB. The impact of having LSB on is non-negligible and can affect the decisions regarding lead time impact of the DA forcing.
- From the reproducibility suite analysis, our findings for the lead time impact of UK-DA are:
 - Can vary strongly between seasons and between fields (summer more sensitive and temperature more sensitive).
 - Does not seem to vary strongly with regimes or cases.
 - Small differences were shown to persist during the forecast.
 - On average, based on the Hinton plots, ALL_OFF vs CONTROL max critical lead time: summer = T+10-T+15 ; winter = T+18 – T+24.
 - On average, based on the Hinton plots, ALL_OFF+ LSB vs CONTROL max critical lead time: summer = T+3-T+4; winter =T+5-T+8.
- From a full cycling DA suite analysis, the differences associated with ALL_OFF can persist well into the end of the run. Further analysis shows that the strongest impact is up to T+12. Beyond this, this is more case and regime dependent where the impact can last up to T+30.
- The longevity of DA impact is likely to be related to the nature and magnitude of the model error.

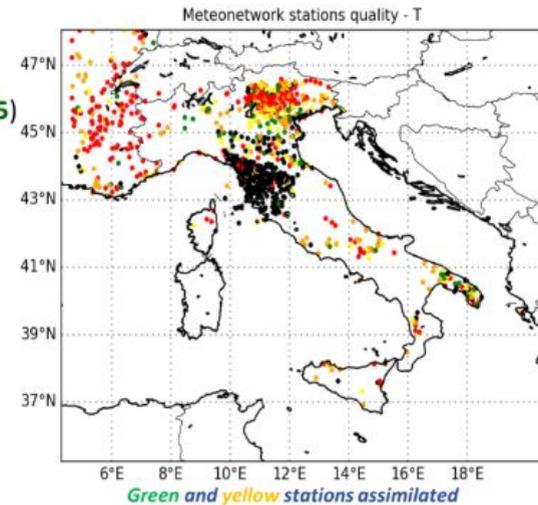
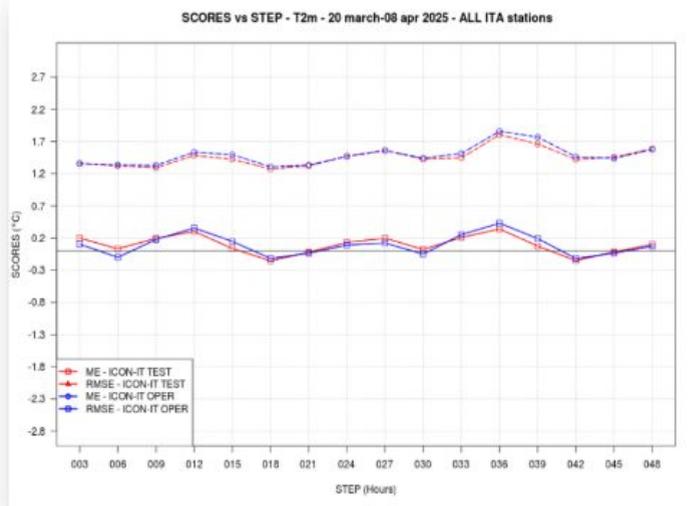
Confirming benefit of data assimilation and also more details.

Observation usage

Assimilation of Meteonetwork T2m and wind speed obs Valerio Cardinali

- ☐ Meteonetwork: crowd sourced synoptic stations (approx. 1030)
- ☐ Quality check based on bias and stdv of the obs increments
- 176 stations used for T, 257 for wind speed (operational in May 2025)

surface temperature verification vs synop

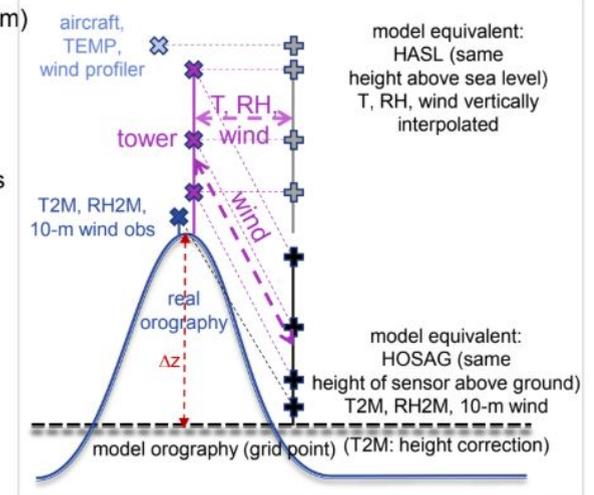


2026

- ☐ DA of Meteonetwork stations (pressure and rh2m obs)
- ☐ DA of surface obs from:
 - MeteoMont (Army Alpine stations): 46 stations
 - Autostrade: 250 stations
 - DPC (Civil Protection): approx. 5500 stations

PBL profiler obs: ICOS towers

- 2 mountain station ($\Delta z \approx 100m$), 2 hill stations ($\Delta z \approx 30m$)
- tower T, RH: upper-air "HASL" operator always better
- tower wind: 2 towers, surface "HOSAG" operator is better for lower level (60m, 93m), "HASL" operator better for all other levels
- need to look at each individual tower / level
- ICOS towers: wind speed is underestimated at many towers / levels (→ wind blacklisted)
- assimilated data:
 - T: 34 obs levels from 9 towers
 - RH: 28 obs levels from 9 towers
 - wind: 8 obs levels from 4 + 2 towers



christoph.schraff_at_dwd.de

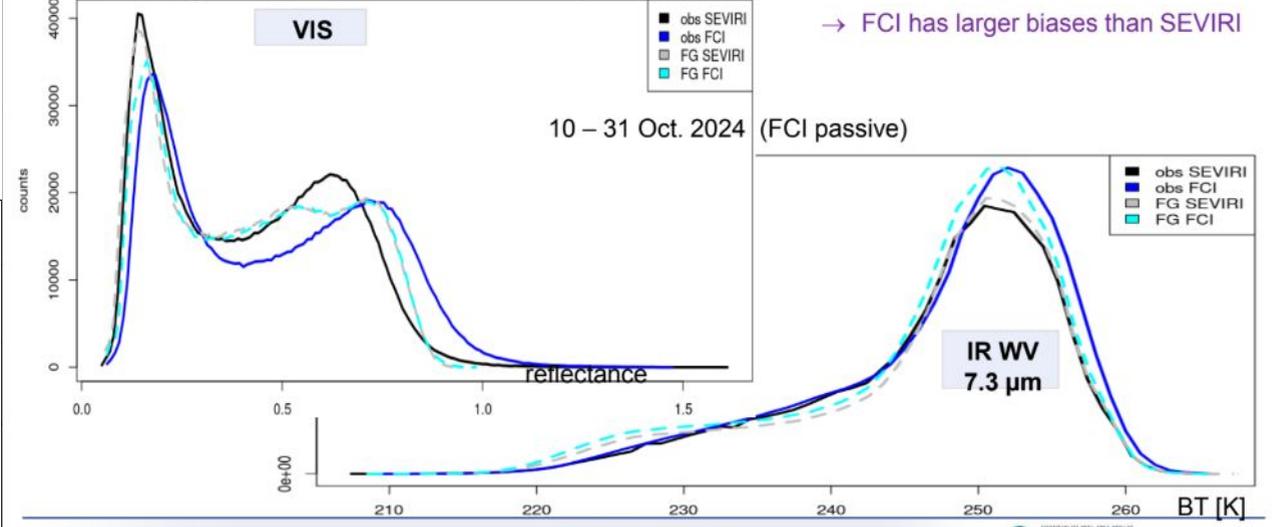
Update on DA in COSMO
EWGLAM / SRNWP Meeting, Norrköping, 22 – 25 Sept. 2025



Annika Schomburg a.o.

- monitoring: comparison to SEVIRI: 2D images, time series of RMSE + bias, histograms

→ FCI has larger biases than SEVIRI



christoph.schraff_at_dwd.de

Update on DA in COSMO
EWGLAM / SRNWP Meeting, Norrköping, 22 – 25 Sept. 2025



Outline

- Radar reflectivity & wind dealiasing
- GNSS-derived data
- Satellite datasets in clear-sky mode:
 - Polar satellites
 - Lambertian vs Specular reflection assumptions (microwave)
 - Assimilation of AWS
 - CrIS and IASI assimilation and inter-channel correlations
 - Geostationary satellites
 - SEVIRI horizontal error correlations
 - SEVIRI use in polar regions
- All-sky assimilation
- MTG Lightning Imager
- Conclusions

"How AI Sees a Weather Nerd's Journey: From the Earth to the Sky and Back"

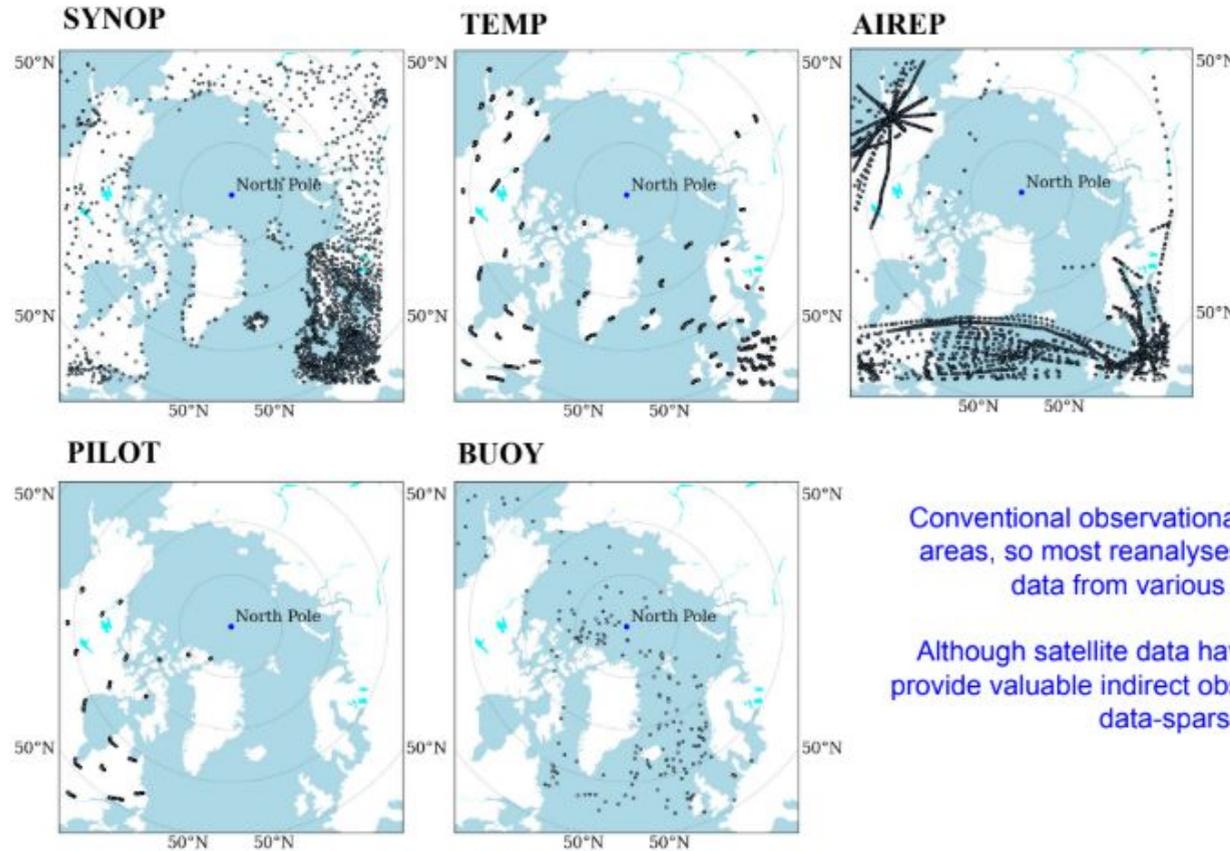
- **We start on the ground** – where the wind lies and weather stations tattletale (AWS and wind dealiasing).
- **Up we go** – GNSS signals bending like they're dodging taxes.
- **Still clear?** – Polar satellites checking in with their best guesses (and some questionable assumptions about how shiny the Earth is).
- **Microwaves get philosophical** – Lambertian vs Specular: flat Earth vs disco ball.
- **Satellites, assemble!** – CrIS, IASI, SEVIRI... and their complicated inter-channel drama.
- **Geostationary spies** – always watching, even where they're not supposed to (looking at you, SEVIRI in the polar regions).
- **Clouds crash the party** – All-sky assimilation: because real weather isn't always clear.
- **When lightning strikes** – MTG Lightning Imager joins the plot with some flash.
- **And back to Earth** – conclusions drawn, data assimilated, sanity... questionable.

Observation usage

Conventional Assimilated Observation



Valid on 20230111 at 00 UTC



Conventional observations are sparse over ocean areas, so most reanalyses rely on remote sensing data from various satellite sensors.

Although satellite data have their own biases, they provide valuable indirect observations that cover many data-sparse regions.

6

Satellites

RA2



Satellites	Period	More Information (https://space.oscar.wmo.int)	
Aqua, Metop-A, Metop-C	1999 -	<ul style="list-style-type: none"> Advanced Microwave Sounding Unit - A 15 channels Resolution: 48 km 	
NOAA-21	Sep, 2012 -	<ul style="list-style-type: none"> Advanced Technology Microwave Sounder 22 channels Resolution: 16 km s.s.p. for channels 165-183 GHz, 32 km for channels 50-90 GHz, 75 km for channels 23-32 GHz 	
Metop-B, Metop-C	2007 -	<ul style="list-style-type: none"> Infrared Atmospheric Sounding Interferometer 8461 channels Resolution 4 x 12-km IFOV (sampling distance: 24 km) 	
NOAA-21	Sep, 2012 -	<ul style="list-style-type: none"> Cross-track Infrared Sounder Resolution 1305 Channels 3 x 3 14 km IFOV covering a 48 x 48 km² cell (average sampling distance: 15 km) 	
MSU	NOAA-11, NOAA-12, NOAA-14	1991-2007	Microwave Sounding Unit
MWHS-2	FY-3C, FY-3D, FY-3E	Dec, 2015 -	Microwave Humidity Sounder -2
AMSU-B/ MHS	NOAA 16-19, Metop-A, Metop-B, Metop-C	2001 -	Advanced Microwave Sounding Unit - B

Observation data retrieval process for CARRA collects these 6-hour datasets from the archives and combines observation files for intermediate cycles at 03, 09, 15, and 21 UTC. For instance, for the 09 UTC cycle, observations from 06 and 12 UTC are merged.

Observation usage

Met Office

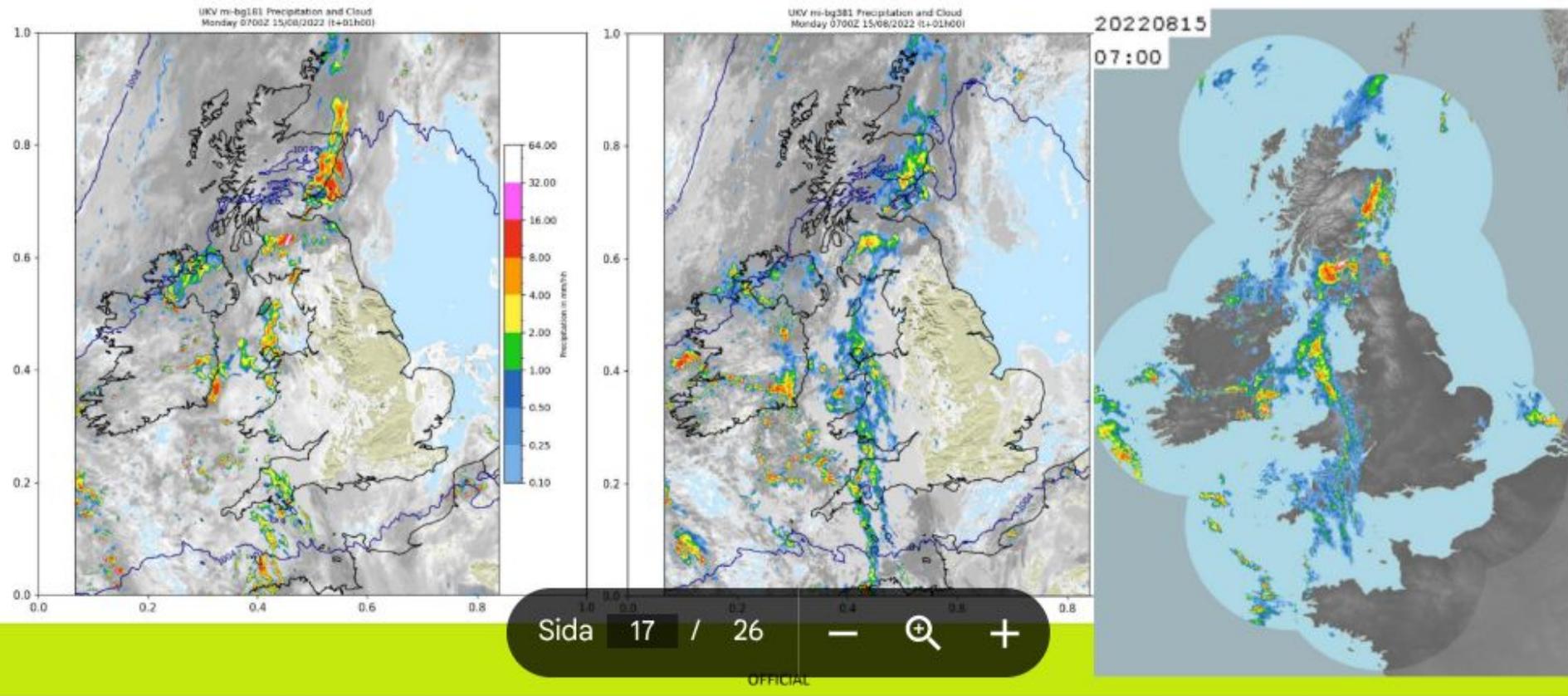
OFFICIAL
15 Aug 2022 07UTC (T+1)

RADAR

Control

Final Conservative GC4

Radar composite



Summary

- Refining flow-dependent algorithms, moving to more portable object oriented systems and introduction of AI/ML.
- Exploiting/Optimising use of Radar data and feeding back to OPERA (sharing of wind-optimised scans, move to direct assimilation of refl.)
- Use of new satellites/instruments and move to all sky.
- GNSS STD from more processing centres
- Need improved/extended exchange of surface observations

I will continue as EWGLAM/SRNWP Data assimilation coordinator until the end of 2025, then someone else will take over.