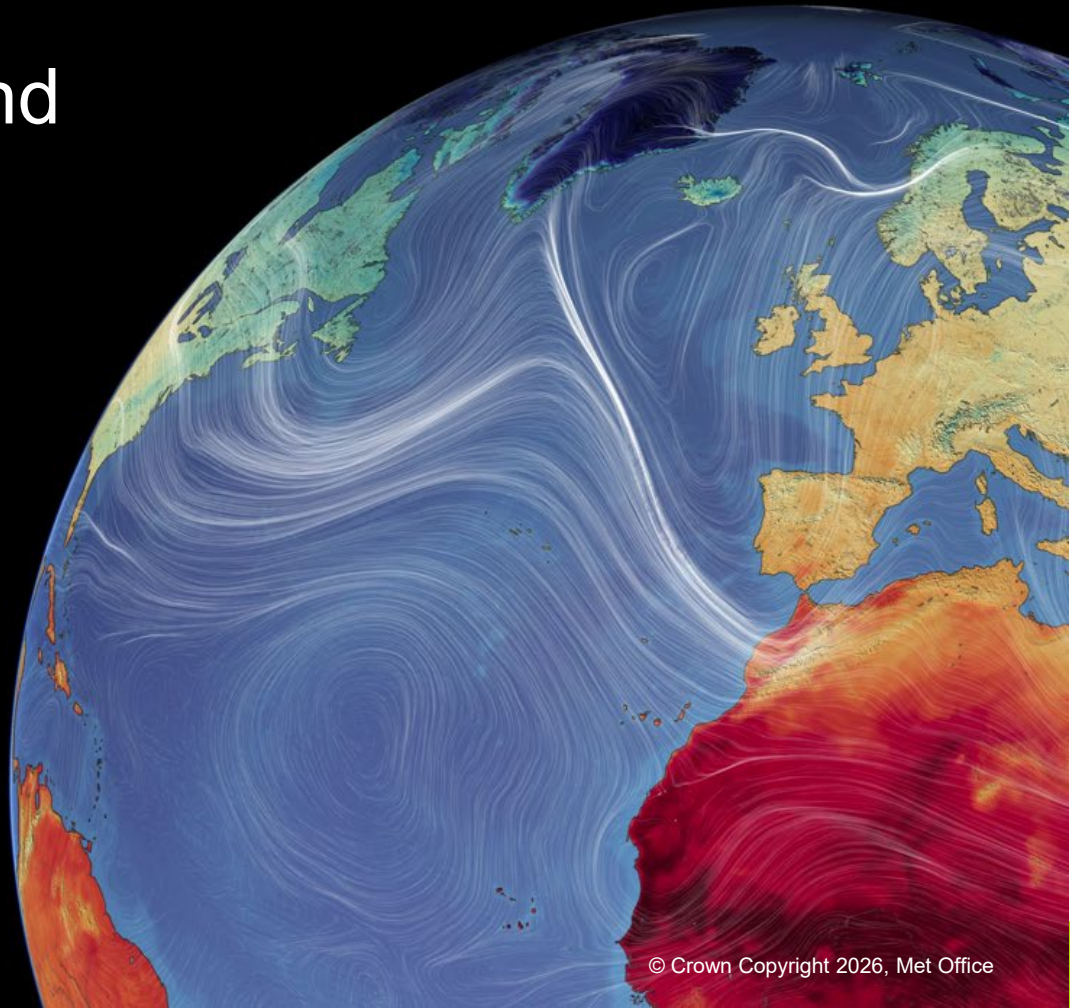
Assessing the value and impact of the current observing system

Fabien Carminati

Met Office

ECMWF Annual Seminar 2026



© Crown Copyright 2026, Met Office

Acknowledgements

These slides were prepared with contributions from:

- Niels Bormann, ECMWF, Reading
- Peter Lean, ECMWF, Reading
- Katrin Lonitz, ECMWF, Reading
- Nahidul Samrat, Bureau of Meteorology, Melbourne
- Brett Candy, Met Office, Exeter
- Chiara Piccolo, Met Office, Exeter
- Philippe Chambon, Météo-France, Toulouse
- Katie Lean, ECMWF, Reading
- Theo Tobin, Met Office, Exeter

Met Office

Part 1

Why do we measure observation impact?

Why we measure observation impact?

Monitor
Is the observing system healthy?
Are all components contributing as expected?

Plan
Where are the gaps?
What should be deployed next?

Justify
Programme decisions cost billions.
Evidence is required.



The societal benefit of weather prediction to EU-27 is estimated at EUR 61.5B/year [1].

Credit: Image generated with Microsoft Copilot

The WMO Impact Workshop series

Expanding Workshop Scope

Global NWP impact (OSE era)

↓

Regional / Limited Area NWP

↓

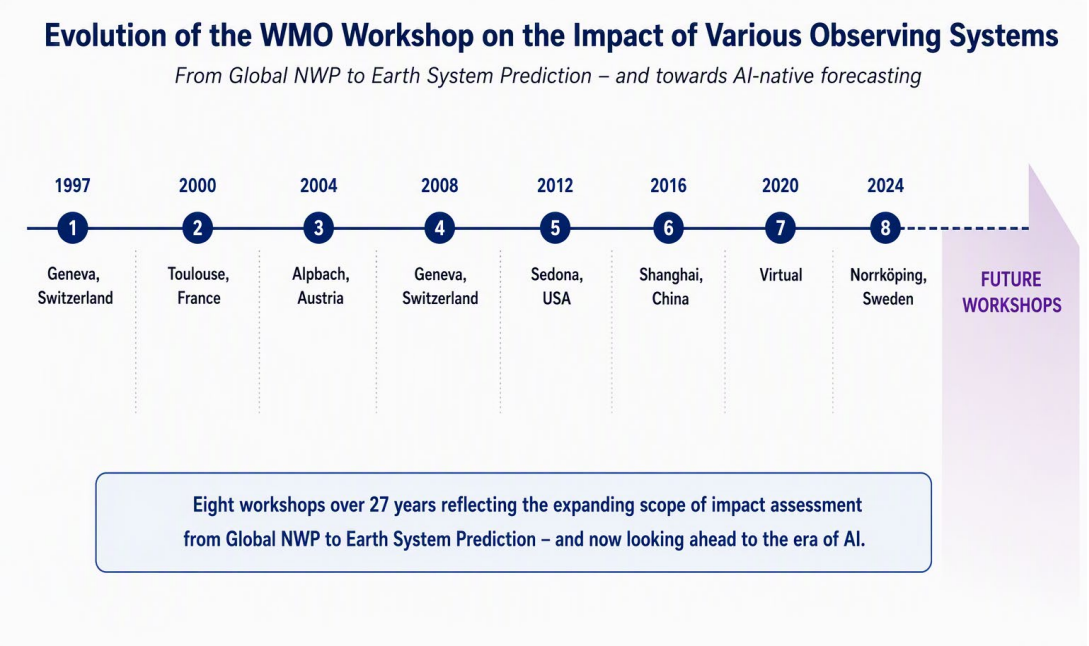
FSOI and advanced impact diagnostics

↓

Earth System Prediction

↓

AI-native forecasting?



Credit: Image generated with Microsoft Copilot

Part 2

Why do we measure observation impact?

→ Because the observing system costs billions and policy decisions need evidence.

How do we measure it?

Methodological approaches

Features/methods	VR	DFS	PAI	OSE	OSSE	SOSE	EFSOI	ESA	EDA
Applicable for ensemble forecasting systems	✓	✓	✓	✓	✓	○	✓	✓	✓
Targetable via user-specified verification metric	✓	–	–	✓	✓	✓	✓	✓	✓
Applicable for future observations	○	○	○	✗	✓	✓	○	✓	✓
Computationally inexpensive	✓	✓	✓	✗	✗	✗	✓	✓	✗
Able to consider the nonlinearity of observation operators	○	○	○	✓	✓	✓	○	○	✓
Without the assumption of linear error growth	✗	–	–	✓	✓	✗	○	○	✓
Able to be separated in observation space	✓	✓	✓	✗	✗	✗	✓	✗	✗
Able to be separated in model space	✓	✗	✓	✓	✓	✓	✓	✓	✓
Able to assess the accumulated value of observations	✗	✗	✗	✓	✓	○	✗	✗	✓
Able to assess the value of anchoring	✗	✗	✓	✓	✓	✓	✓	✗	✗

Variance Reduction (**VR**)

Degrees of Freedom for Signal (**DFS**)

Partial Analysis Increment (**PAI**)

Observing System Experiment (**OSE**)

Observing System Simulation Experiment (**OSSE**)

Sensitivity Observation System Experiment (**SOSE**)

Forecast Sensitivity to Observation Impact (**FSOI**) and Ensemble FSOI (**EFSOI**)

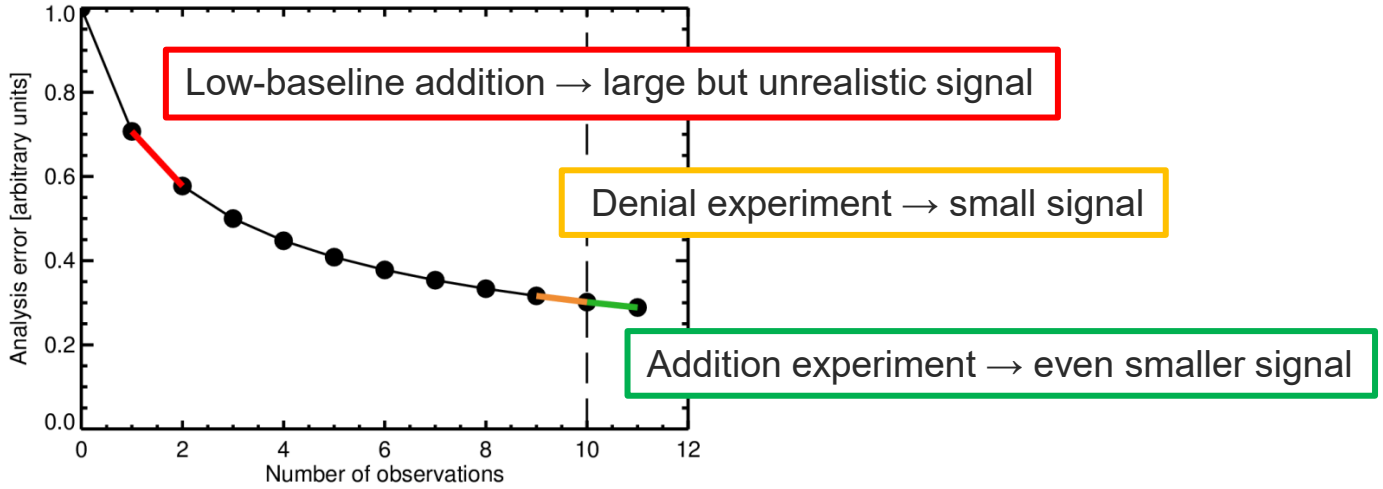
Ensemble Sensitivity Analysis (**ESA**)

Ensemble of Data Assimilations (**EDA**)

See Hu et al. (2025) [10] and Healy et al. (2024) [11] for comprehensive reviews.

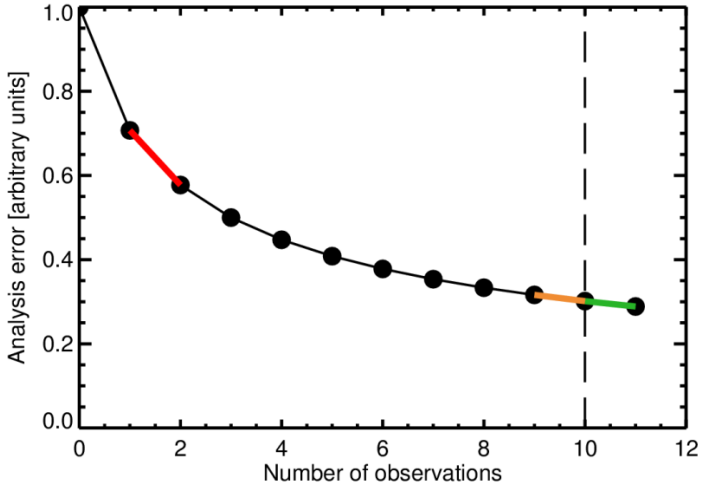
Table 1 from Hu et al. (2025) [10]

OSE: a gold standard … with caveats

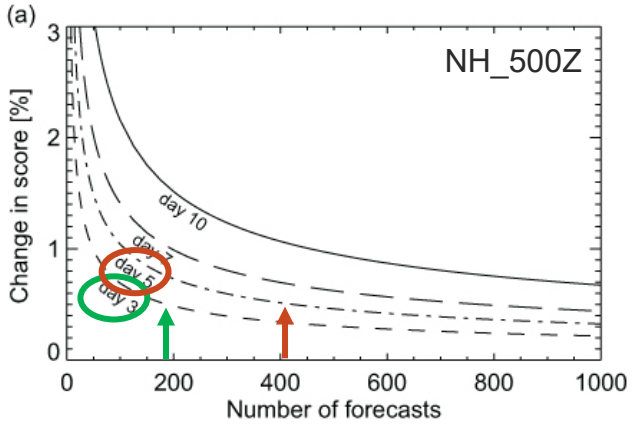


Credit: Healy et al. (2024), figure 1, [11]

OSE: a gold standard … with caveats

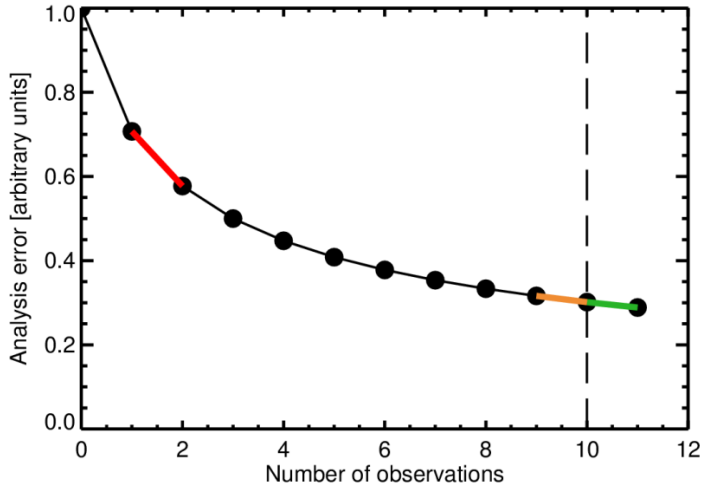


Credit: Healy et al. (2024), figure 1, [11]



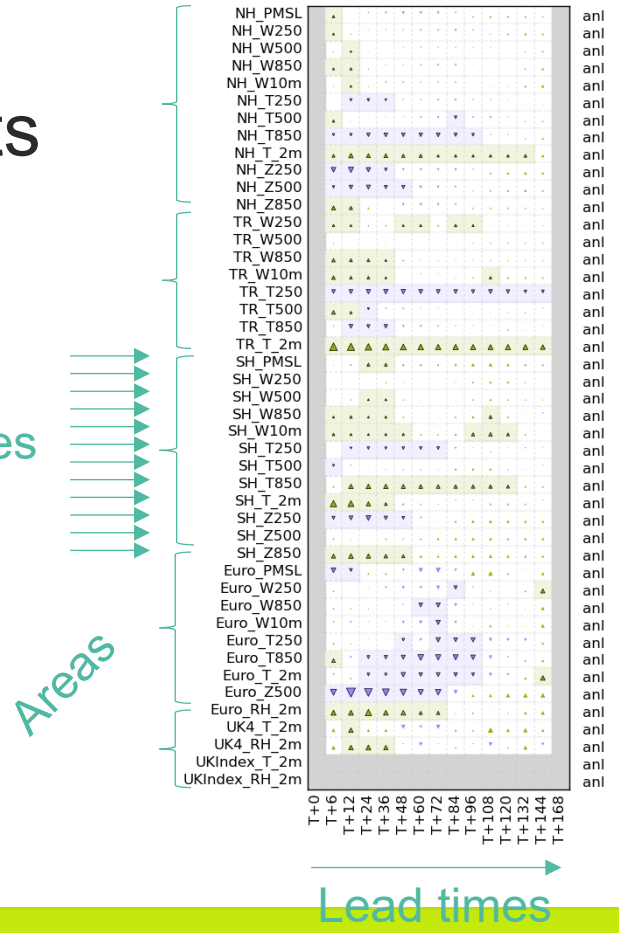
Credit: Geer, 2016, figure 15 [13]. See also Weston, 2014, [12].

OSE: a gold standard, with caveats



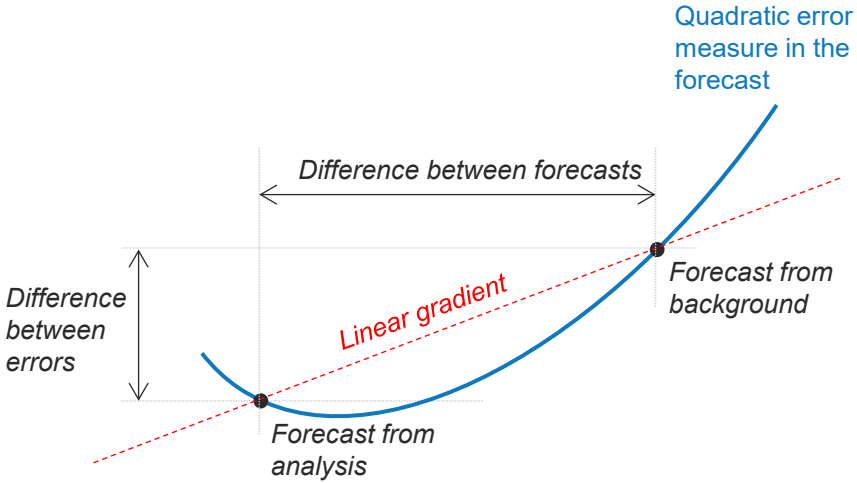
Credit: Healy et al. (2024), figure 1, [11]

Standard Met Office score card



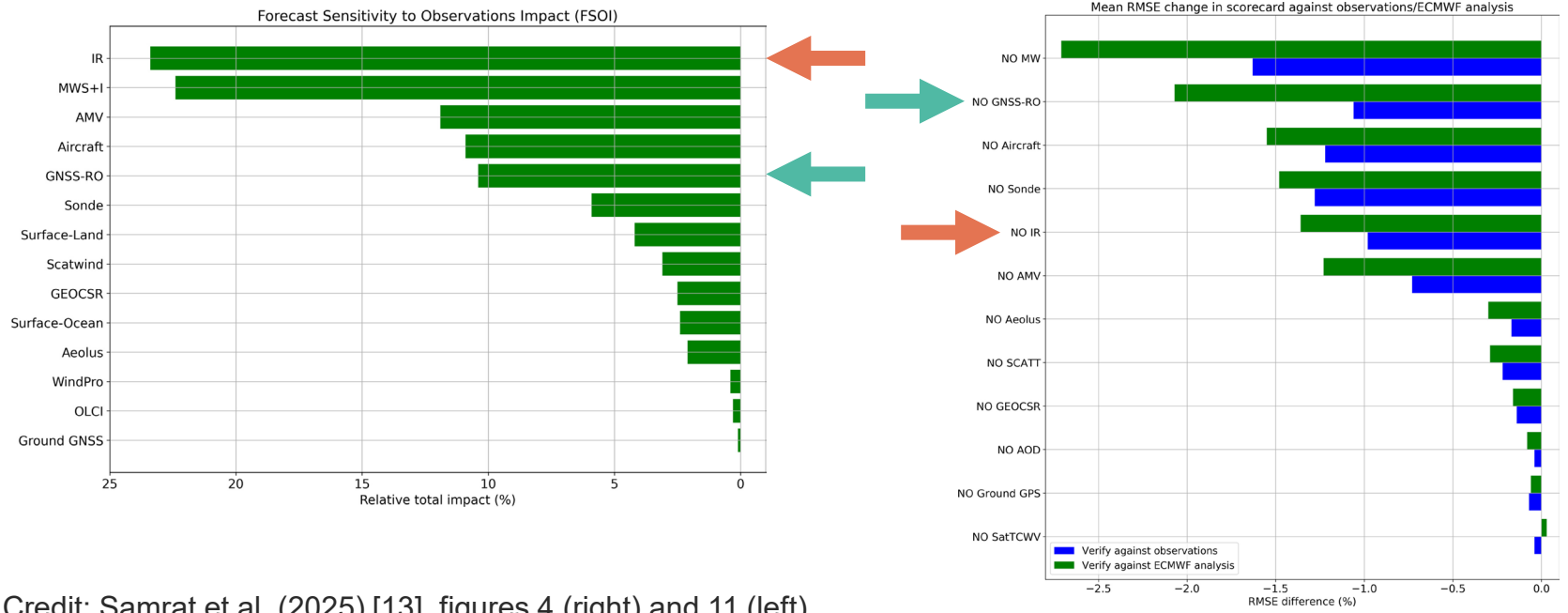
FSOI: cheap but interpret with care

- Use a total energy norm combining errors from wind, temperature, pressure, humidity.
- Efficient, runs routinely, can monitor impact over time.
- Gives relative ranking of all obs types simultaneously.
- It can produce impact by instrument, by channel, by satellite, by station, by region, by observation type.
- Linear assumption, limited to 24-48h forecast range.
- Result depends on choice of energy norm, pressure range, verification reference, and the correct system tuning.



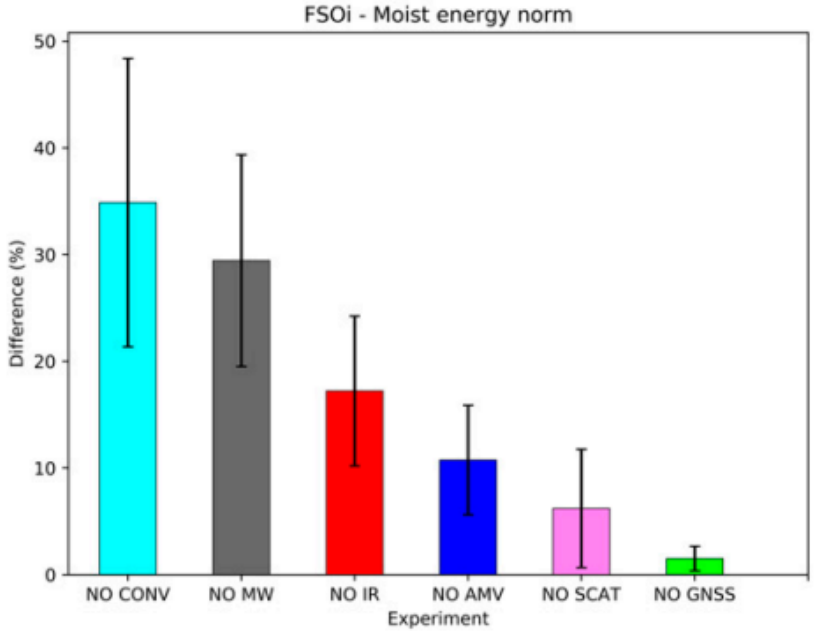
Credit: Illustration based on Tremolet (2007) [12] and Lorenc and Marriott (2014) [3].

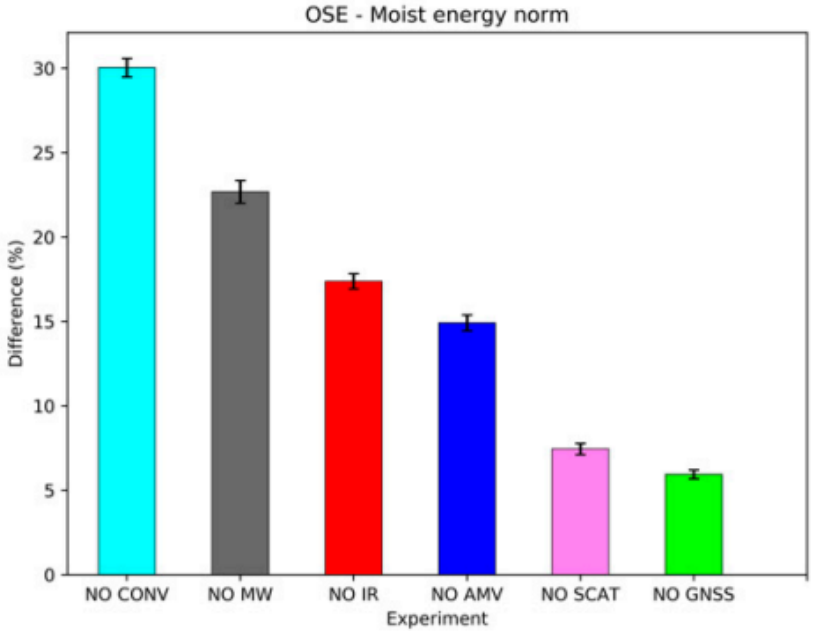
Rankings depend on method and metric



Credit: Samrat et al. (2025) [13], figures 4 (right) and 11 (left).

Rankings depend on method and metric





Credit: Chambon et al. (2023) [14], figure 8.

Part 3

Why do we measure observation impact?

→ Because the observing system costs billions and policy decisions need evidence.

How do we measure it?

→ Multiple tools, each with strengths and blind spots.

What does the current system deliver?

Multi-centre headline: what dominates?

Ranking	Met Office[a]	Météo-France[b]	ECMWF[c]	ECCC[d]	NOAA[e]
1st	$\mathrm{IR}^{\uparrow}_{26\%}$	$\mathrm{MW}^{\uparrow}_{33\%}$	$\mathrm{MW}^{\rightarrow}_{34\%}$	$\mathrm{MW}^{\rightarrow}_{31\%}$	$\mathrm{Conventional}_{34\%}$
2nd	$\mathrm{Conventional}^{\uparrow}_{23\%}$	$\mathrm{IR}^{\uparrow}_{22\%}$	$IR^{\uparrow}_{22\%}$	$\mathrm{Conventional}^{\rightarrow}_{27\%}$	$\mathrm{MW}_{29\%}$
3rd	$\mathrm{MW}^{\downarrow\downarrow}_{22\%}$	$\mathrm{Conventional}^{\downarrow\downarrow}_{20\%}$	$\mathrm{Conventional}^{\downarrow}_{21\%}$	$\mathrm{AMV}^{\uparrow}_{15\%}$	$\mathrm{IR}_{20\%}$
4th	$\mathrm{AMV}^{\rightarrow}_{12\%}$	$\mathrm{AMV}^{\rightarrow}_{11\%}$	$\mathrm{RO}^{\uparrow}_{10\%}$	$\mathrm{IR}^{\downarrow}_{13\%}$	$\mathrm{AMV}_{9\%}$
5th	$\mathrm{RO}^{\uparrow}_{10\%}$	$\mathrm{RO}^{\uparrow}_{10\%}$	$\mathrm{AMV}^{\downarrow}_{6\%}$	$\mathrm{RO}^{\rightarrow}_{10\%}$	$\mathrm{RO}_{8\%}$

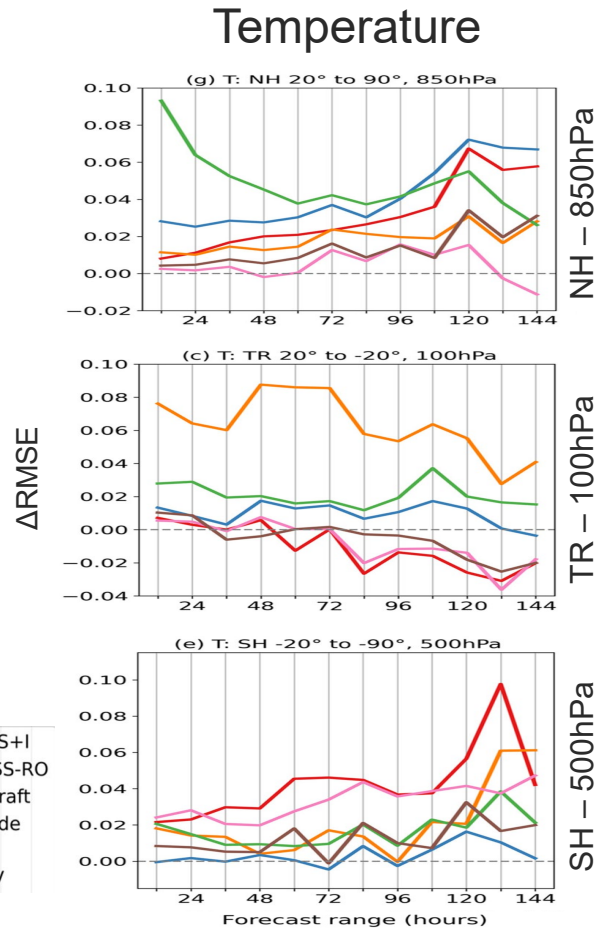
a: Samrat 2023 vs Candy 2019
b: Chambon 2025 vs Chambon 2020
c: Bormann 2018 vs 2021
d: Du 2022 vs 2024
e: Collard 2016 (?); EFSOI

Legend: $ObsType^{\frac{Ranking\ change}{FSOI\ \%\ impact}}$

Met Office

Beyond the big picture: Regional breakdowns matter

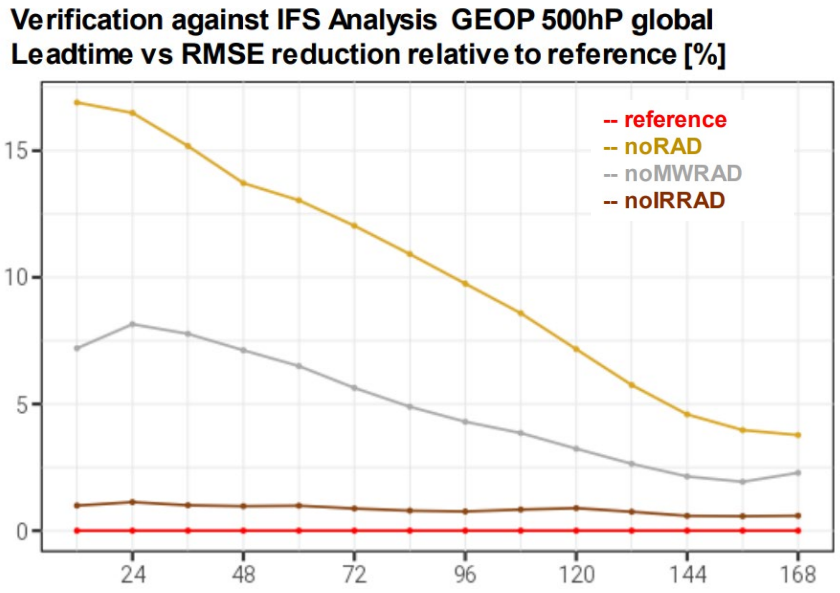
- Northern hemisphere: impact dominated by conventional data, lower to mid-troposphere, for most variables
- Tropics: impact dominated by RO in upper troposphere / lower stratosphere, mostly for temperature and wind
- Southern hemisphere: impact dominated by MW, throughout the troposphere, for most variables

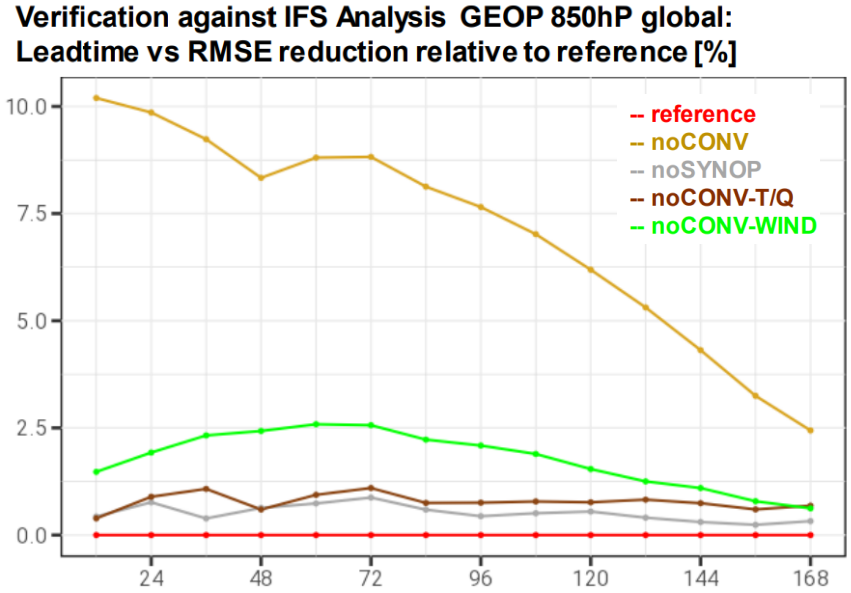


Credit: Samrat et al. (2025) [13], figures 5-8.

Met Office

Stronger together





Credit: Cress et al., 2024, 8th_WMO_Impact_Workshop-Poster-P2.23-Alexander-Cress

Regional NWP: what dominates at convective scale

	Met Office UKV (1.5 km, hourly 4D Var)	Météo-France AROME (1.3 km, hourly 3DEnVar)[a]	Bureau of Meteorology (1.5 km, hourly 4D Var)[b]
What dominates?	Lateral boundary conditions from global model	Mode-S aircraft, radar (reflectivity + Doppler), conventional	Satellite data (radiances, wind, g-b GNSS)
Satellite impact	Radiance denial → (nearly) neutral scorecards	LEO/GEO sounders secondary to in-situ	Large with radiances driving the impact up to T+24. Strong additional impact from newly added high-frequency GEO IR
Key finding	Sounders not adding substantial benefit in hourly-cycling limited area model	Regional model driven by different obs than global	(Conventional) data-sparse regions are more satellite-dependent → satellite impact is amplified

[a] Chambon 2025 EMETSAT conference
[b] Smith 2024 WMO workshop

Part 4

Why do we measure observation impact?

→ Because the observing system costs billions and policy decisions need evidence.

How do we measure it?

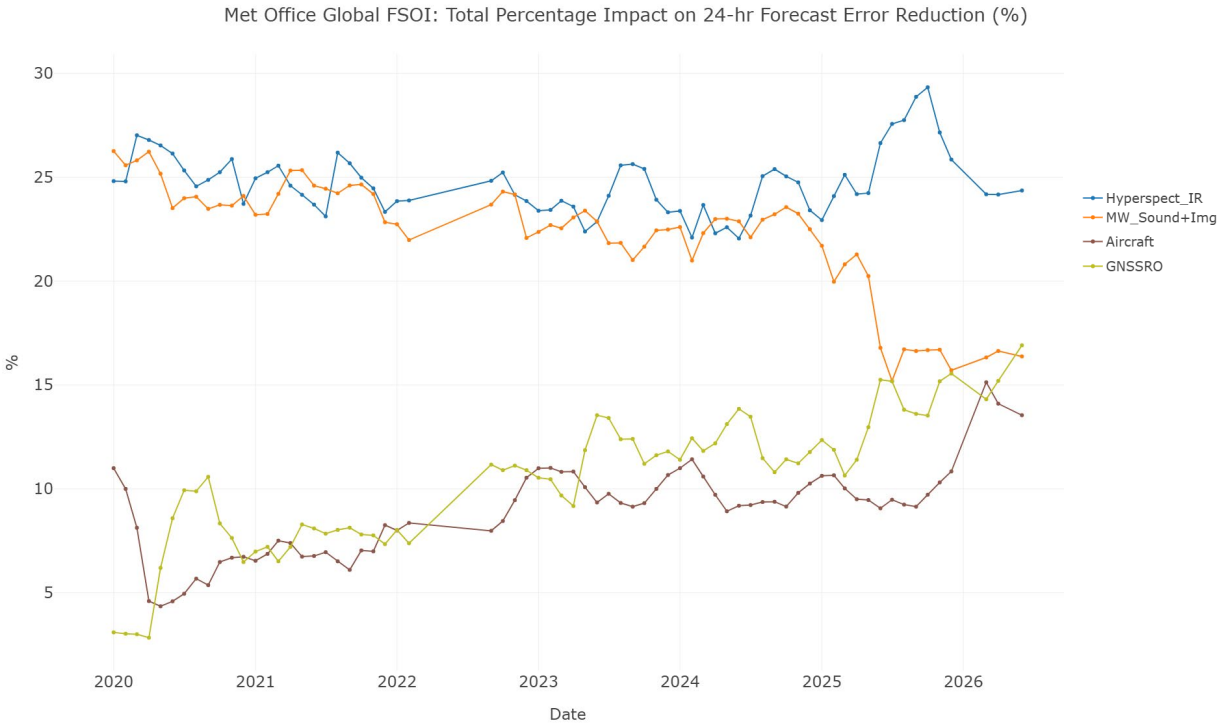
→ Multiple tools, each with strengths and blind spots.

What does the current system deliver?

→ A rather consistent multi-centre picture, with some differences.

What's shifting?

Emerging signals: the observing system is not static



Met Office

MW sounders: the backbone

- The impact comes from sounding channels [13].
- MW radiances have more pronounced impact on long range and humidity than IR, despite being less numerous (10% vs 80% of satellite data) [14].
- No sign of saturation [15].

Policy: MW sounding capability must be sustained
→ this is WMO recommendation 3.2.1 [16].

In use at the Met Office

Metop-B, -C **AMSU-A** & **MHS**
SNPP, N20, N21 **ATMS**
FY-3D **MWHS-2** & **MWRI**
DMSP-17 **SSMIS**
~~GCOM-W1 **AMSR-2**~~
GMP **GMI**

Under investigation

AWS **MWR**
Metop-SGA **MWS**
FY-3E **MWHS-2** & **MWTS-3**
Spire **HyMS**

Future candidates

Metop-SGB **MWI** & **ICI**
GOSAT-GW **AMSR-3**
ESP-Sterna **MWR**
NOAA22, QuickSounder **ATMS**

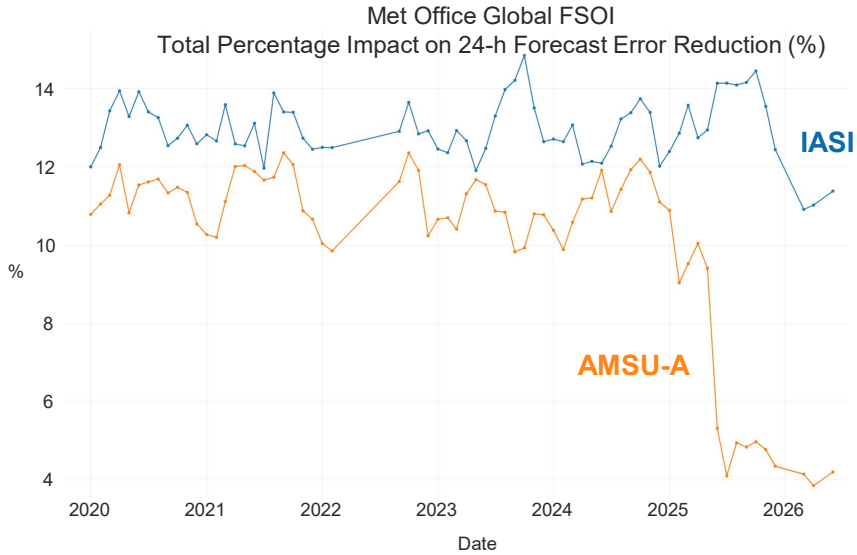
MW sounders: what the system is losing

Decommissioning:

2021 – Metop-A AMSU-A, MHS
2024 – FY-3C MWHS-2, MWTS-2, MWRI
2025 – NOAA-15/18/19 (POES) AMSU-A, -B/MHS
2026.09 – DMSP F17 SSMIS, GCOM W1 AMSR-2
2026.11 – SNPP ATMS
2027.06 – Aqua AMSU-A (not in use)

Degradations:

AMSU-A Metop-B channel 15; Metop-B 6; Metop-C 3.
MHS Metop-C channels 3, 4
SSMIS F17 channels 2, 3, 4, 5

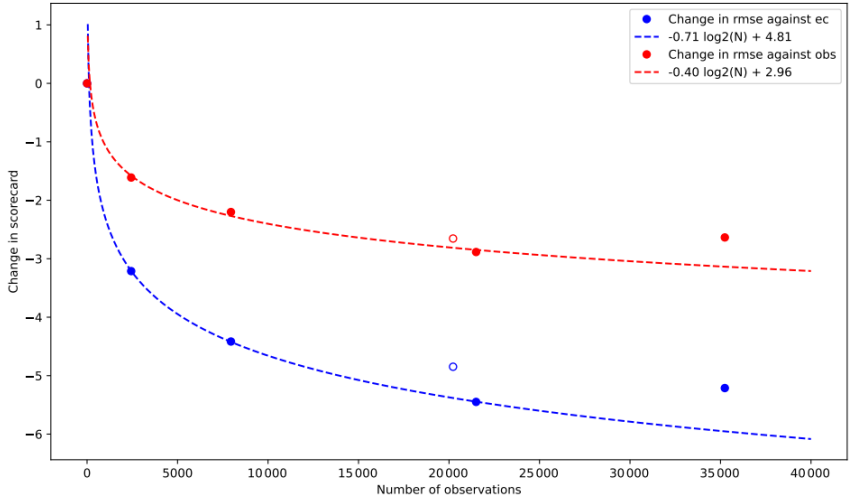


GNSS-RO: consistent rise across centres

- Rise driven by data volume: COSMIC-2, Spire, PlanetIQ, Sentinel-6.
- The Met Office ROMEX study found that going from operational to 20,000 profiles/day gives substantial benefit [17].
- At ECMWF, K. Lonitz showed EDA spread effectively reduced with > 30,000/day.
- GNSS-RO is also an important anchor for VarBC.

Policy: RO sounding capability must be sustained & continue coordinated activities assessment like ROMEX→ this is also WMO recommendation 3.2.1 & 3.3.1 [16].

ROMEX data impact at the Met Office

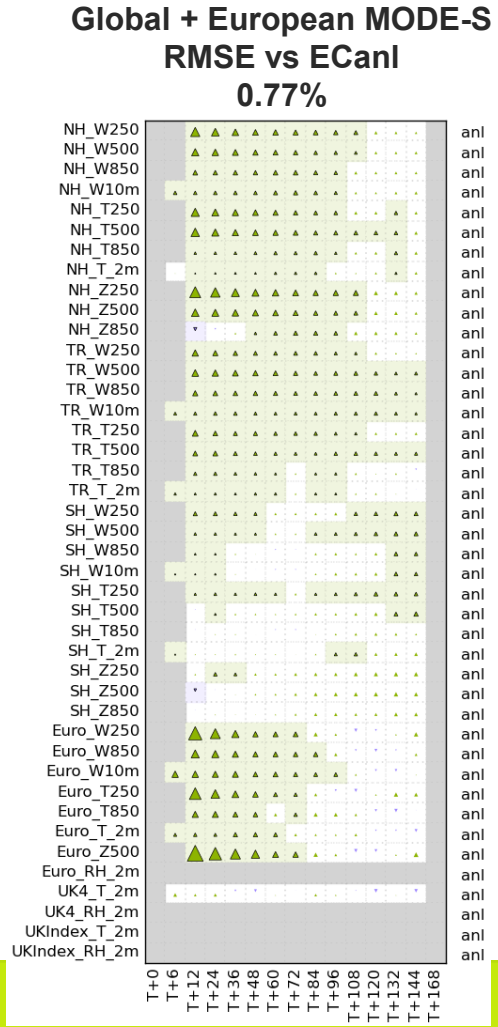


Credit: Bowler and Lewis, 2026 [17], figure 18.

Conventional: two pillars

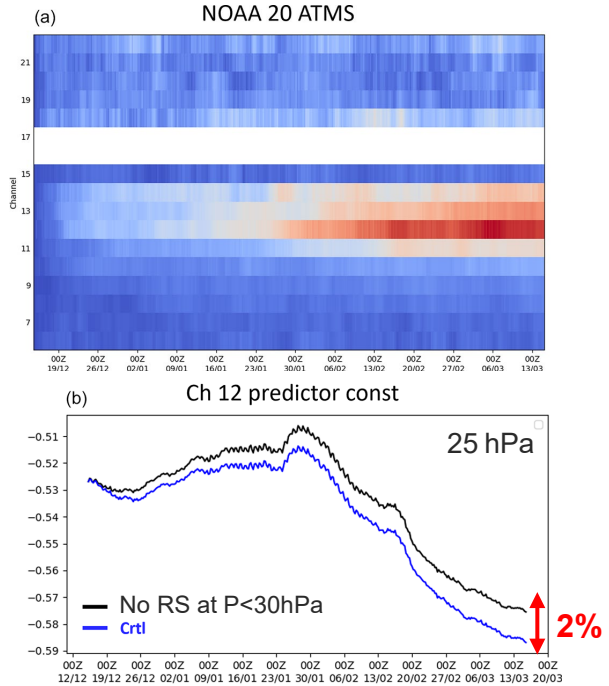
- Radiosondes: consistently high ranking; 1.66% overall degradation when removed; most impact in the NH troposphere but also very important to constrain stratospheric wind; important anchor for VarBC [18].
- Aircraft: similarly important, with dominant signal in NH; rise of Mode-S; AMDAR at risk [19].

Policy: Sustain and improve the geographic coverage radiosonde observations, especially in the remote areas and isolated island stations, and aircraft-based observations → this is also WMO recommendation 3.1.1 [16].



Bias correction stability: the anchor problem

- Left unchecked, variational bias correction gradually corrects observations towards model biases [20].
- After 3 months without radiosondes above 30hPa level, ~2% VarBC coeff drift for ATMS ch12; for GNSS-RO denial, even larger (4.6% for adjacent channels) [18].
- No visible detrimental effects in this 3-month OSE, but long-term concern.
- Sustaining anchor observations is essential for long-term system health, not just for direct forecast benefit.

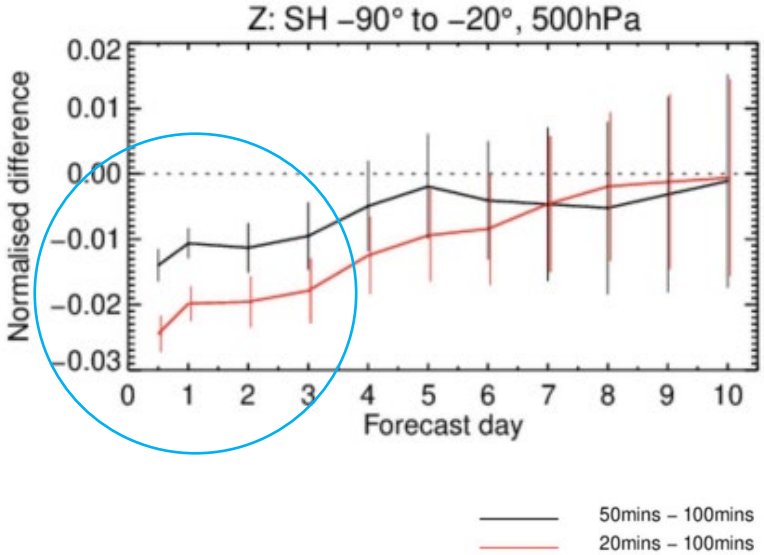


Credit: Fujiwara et al., 2025 [18], figure 14.

Data timeliness: eroding impact silently

- ~25% of MW sounder impact at day 3 lost if timeliness is 100 min.
- Going from 100→20 min recovers ~2% Z500 RMSE at day 3 in SH.
- Important cost driver: faster delivery means more impact, but more expensive ground segment.

Policy: To build in the possibility to reduce data delivery times where possible in the development phase of new instruments and systems → WMO recommendation 3.2.4 [16].



Credit: Lean et al., 2024, 8th WMO workshop, O5.11.

Part 5

Why do we measure observation impact?

→ Because the observing system costs billions and policy decisions need evidence.

How do we measure it?

→ Multiple tools, each with strengths and blind spots.

What does the current system deliver?

→ A rather consistent multi-centre picture, with some differences.

What's shifting?

→ Rankings change, capabilities are lost, and factors we previously ignored now matter.

Where does all this lead?

WMO 8th Impact Workshop: key recommendations

For observing network operators:

- Sustain radiosonde, aircraft, near-surface and GBON-mandated observations, especially in remote areas (3.1.1)

For space agencies:

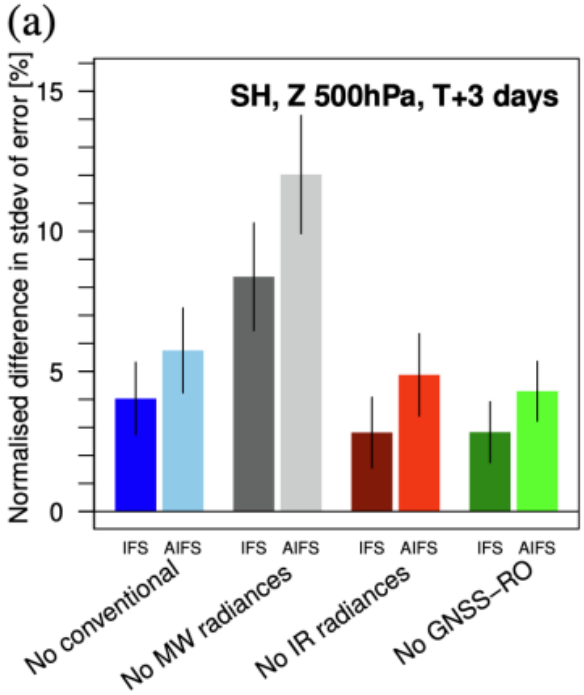
- Sustain MW, IR and GNSS-RO soundings (3.2.1)
- Improve data timeliness at design stage of new instruments (3.2.4)
- Consider high-frequency MW/IR soundings, and more GNSS-RO (3.2.3)

For the scientific community:

- Continue coordinated on observing system impact assessment activities like ROMEX (3.3.1)
- Assess impact by multiple methods (FSOI, OSE/DDE, EDA, ESA, PAI) (3.3.4)
- Embrace ML as a potential tool to improve ESP, including NWP (3.3.8)

Observation denial in physics-based and AI forecasts

- Both systems degrade when observations denied, but …
- **AIFS degradation is larger**, especially in the Tropics
- AIFS remains more skilful than IFS even with denied data
- AIFS advantage is ~½ day NH, ~1 day SH, ~1½ days Tropics



Credit: Bormann et al., 2027 [22]

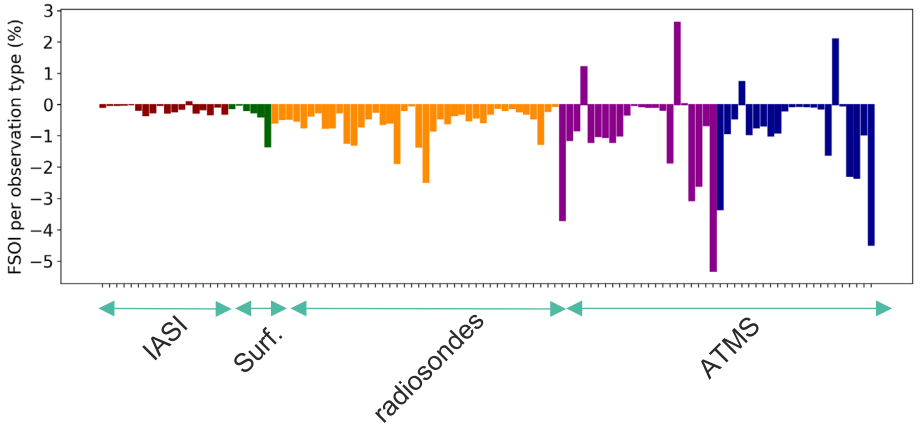
FSOI for AI models: new tools for a new paradigm

GraphDOP FSOI:

- FSOI computed via automatic differentiation.
- MW channels show particularly large FSOI values.
- Some channels detrimental (mixed surface/atmosphere signals).
- Sensitivity maps show physically meaningful patterns (storm tracking, advection).

Institutional framing:

- ESA ITT ARCH-AI: defining EO architecture requirements for AI-native forecasting.
- Key question: What should future observing systems look like if AI is the forecasting engine?



"*FSOI could help build high-quality training datasets, avoiding the introduction of confusing information*"
Credit: Laloyaux et al., 2025, [21]

Summary and take-home messages

Met Office

We have mature tools, rich evidence, and a clear picture of what the current observing system delivers.

But the picture is not static - rankings shift, capabilities are lost, hidden factors (timeliness, anchoring) matter more than realised.

And, we have forward-looking challenges. The tools and evidence we've built now face a paradigm shift (AI forecasting) and we need new tools and answers.

References

[1] EUMETSAT, 2014, https://www.eumetsat.int/media/16881

[2] Bouttier and Kelly, 2001, https://doi.org/10.1002/qj.49712757419

[3] Lorenc and Marriott, 2014, https://doi.org/10.1002/qj.2122

[4] Baker and Daley, 2001, https://doi.org/10.1002/qj.49712656511

[5] Tan et al., 2007, https://doi.org/10.1002/qj.43

[6] Masutani et al., 2010, https://doi.org/10.1007/978-3-540-74703-1_24

[7] Diefenbach et al., 2023, https://doi.org/10.1002/qj.4419

[8] Torn and Hakim, 2008, https://doi.org/10.1175/2007MWR2132.1

[9] Kalnay et al., 2012, https://doi.org/10.3402/tellusa .v64i0.18462

[10] Hu et al., 2025, https://doi.org/10.1002/qj.4933

[11] Healy et al., 2024, https://www.ecmwf.int/en/elibrary/81556-methods-assessing-impact-current-and-future-components-global-observing-system

[12] Tremolet, 2007, https://doi.org/10.1127/0941-2948/2007/0258

[13] Samrat et al., 2025, https://doi.org/10.1002/qj.5002

[14] Chambon et al., 2023, https://doi.org/10.1175/MWR-D-22-0087.1

[15] Duncan et al., 2021, https://doi.org/10.1002/qj.4149

[16] WMO 8th Impact Workshop, 2024, https://wmoomm.sharepoint.com/:b:/s/wmocpdb/EQKjuBmDcO5GhC0nobUuc8YB6MU7YsrHdtB5JNS0tzMPvg?e=47Xtet

[17] Bowler and Lewis, 2026, https://doi.org/10.5194/amt-19-2479-2026

[18] Fujiwara et al., 2025, https://doi.org/10.5194/amt-18-2919-2025

[19] Warren et al., 2026, https://doi.org/10.1002/qj.70099

[20] Francis et al., 2023, https://doi.org/10.1002/qj.4482

[21] Laloyaux et al., 2025, https://arxiv.org/html/2510.27388v1

[22] Bormann et al., 2026, https://www.ecmwf.int/en/elibrary/81744-comparison-observing-system-experiments-physics-based-ifs-and-data-driven-aifs

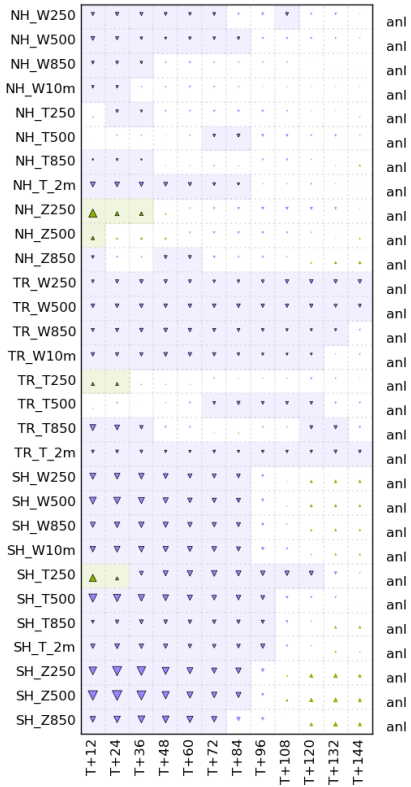
[23] Hu et al., 2025, https://doi.org/10.1029/2024MS004809

Thank you! Any questions?

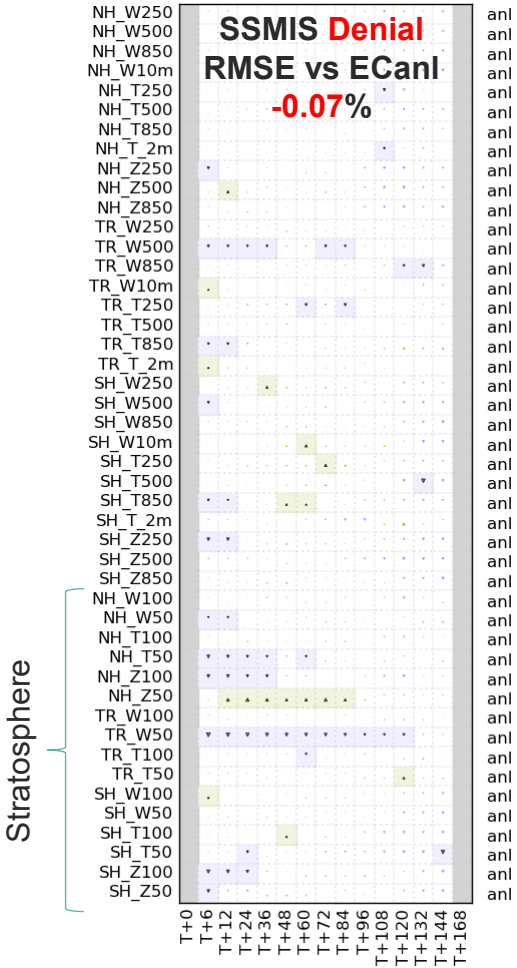
And thanks again to all contributors: Niels Bormann, Peter Lean, Katrin Lonitz, Nahidul Samrat, Brett Candy, Chiara Piccolo, Philippe Chambon, Katie Lean, & Theo Tobin

Additional Material

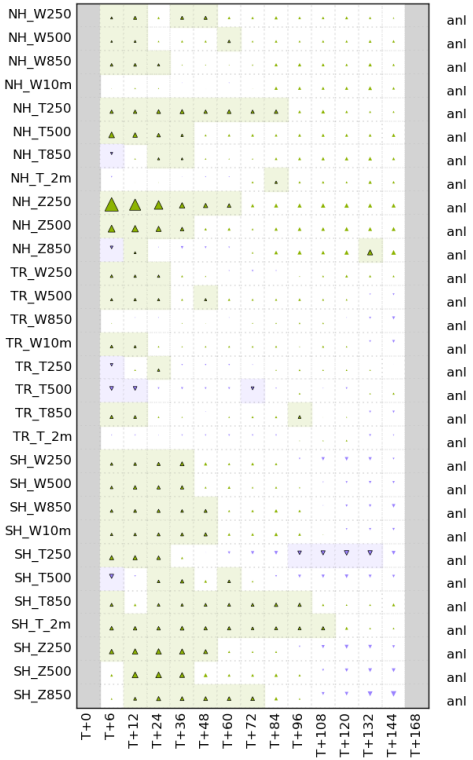
NOAA POES Denial
RMSE vs ECanl
-0.34%



SSMIS Denial
RMSE vs ECanl
-0.07%



AWS Addition
RMSE vs ECanl
+0.21%



MetOp-C life extension: from evidence to decision

The evidence:

- MetOp-C contributes ~10% of total FSOI and ~20% of assimilated observations.
- Consistent benefit over multiple years of FSOI monitoring.
- Metop-SG-A1 MW water vapour band known to be degraded.
- Loss of MetOp-C would reduce MW sounding redundancy at a time when POES fleet is already gone.

The recommendation:

- Recommendation to EUMETSAT to extend MetOp-C operations (Sept Council)
- Cost of extension << cost of replacement
- Operational continuity preserved until Metop-SG constellation fully established