

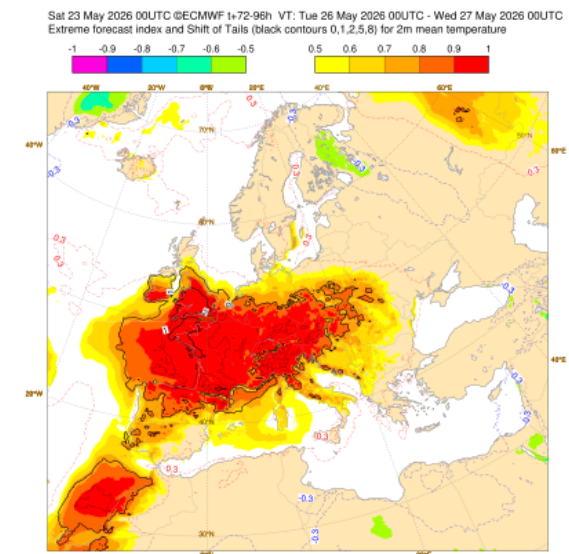
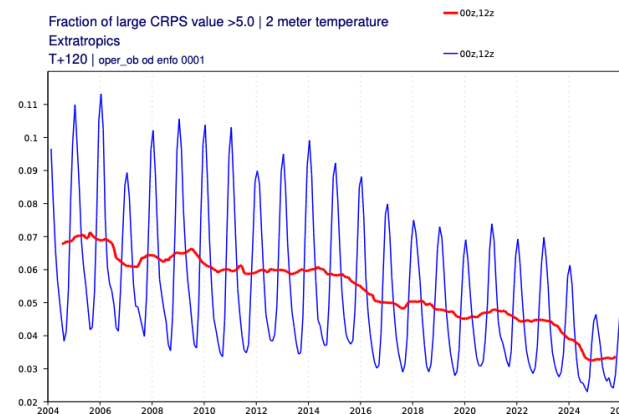
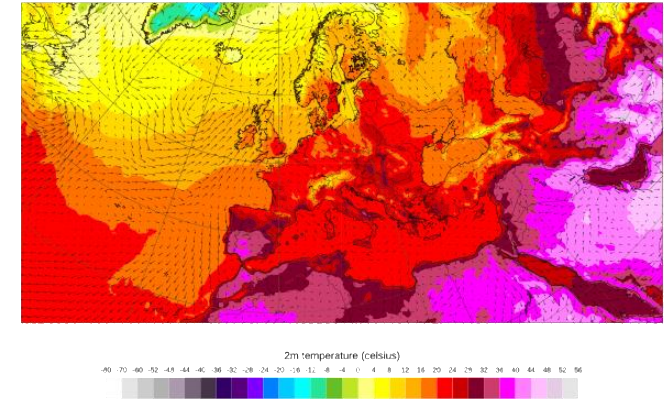
Medium-range forecast skill of 2m temperature in the IFS and AIFS

UEF 2026, Reading

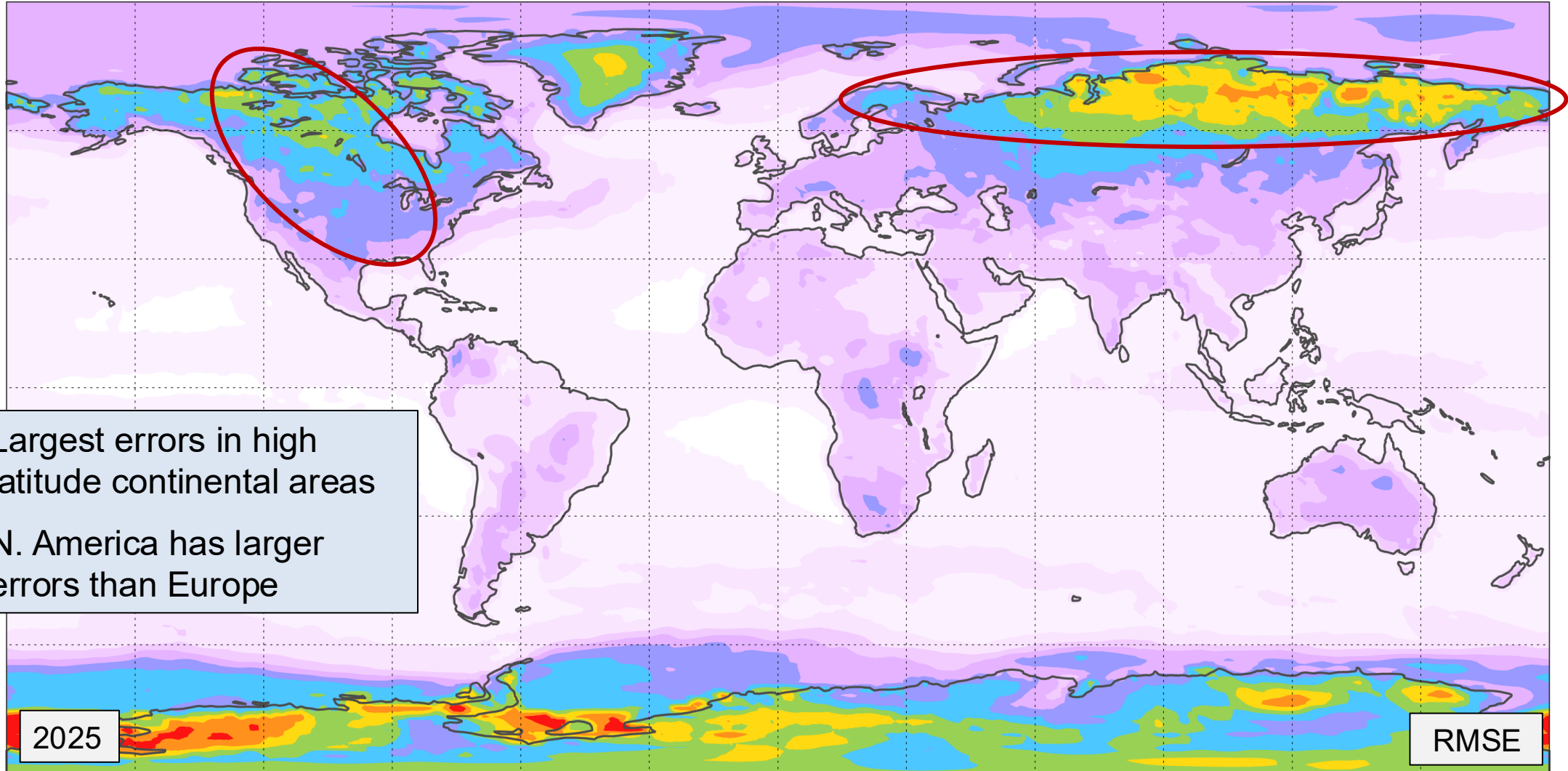
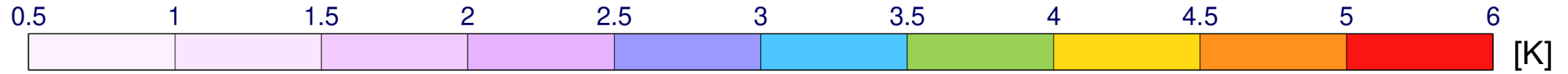
Thomas Haiden

Content

- T2m forecast errors
- The role of T850
- AIFS vs IFS
- Forecast jumpiness
- Spread reliability
- Comparison with other global models

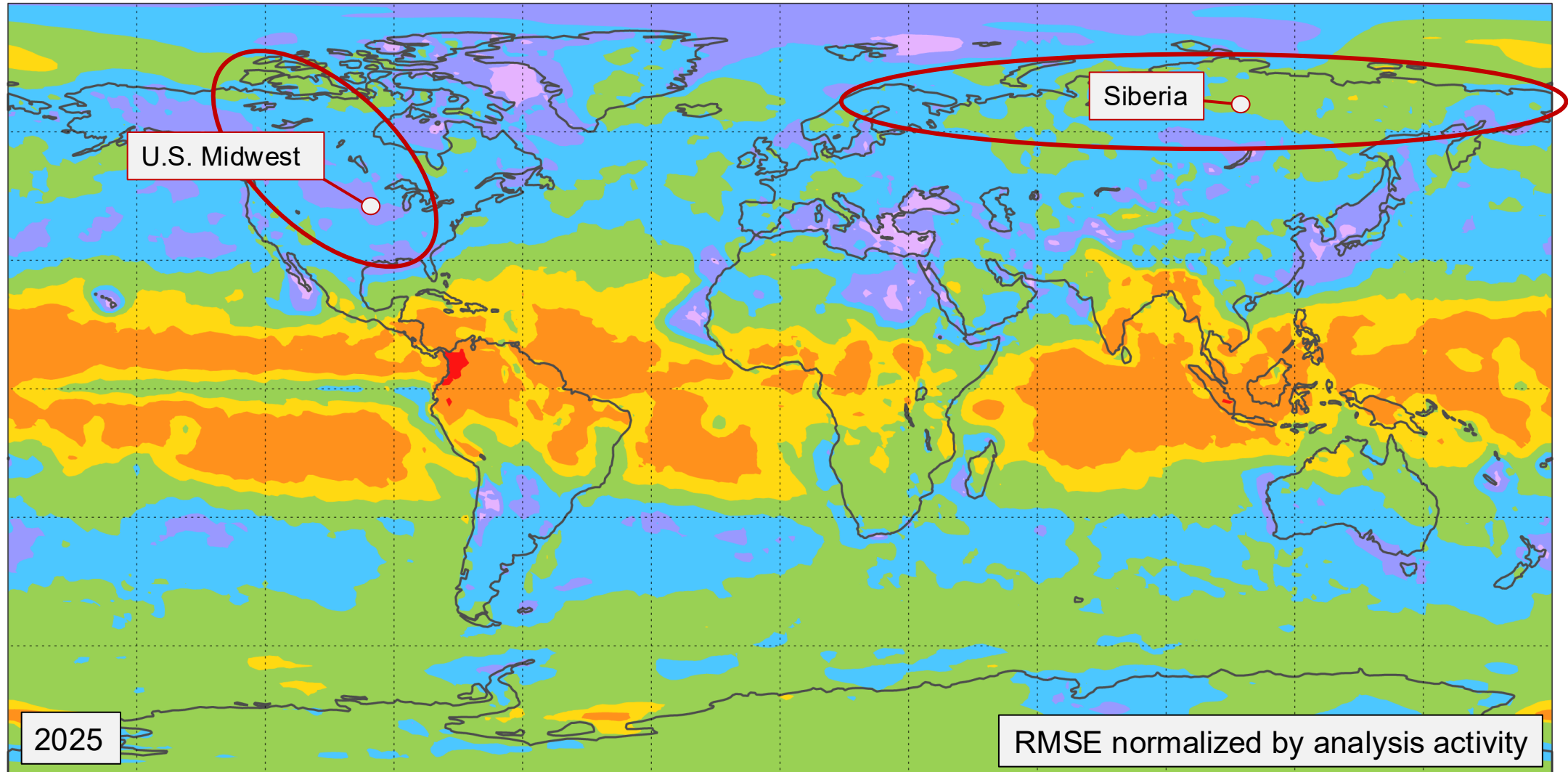
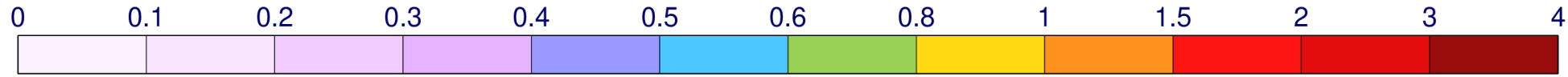


2m temperature forecast error (RMSE) at Day 5



- Largest errors in high latitude continental areas
- N. America has larger errors than Europe

2m temperature forecast error (normalized RMSE) at Day 5



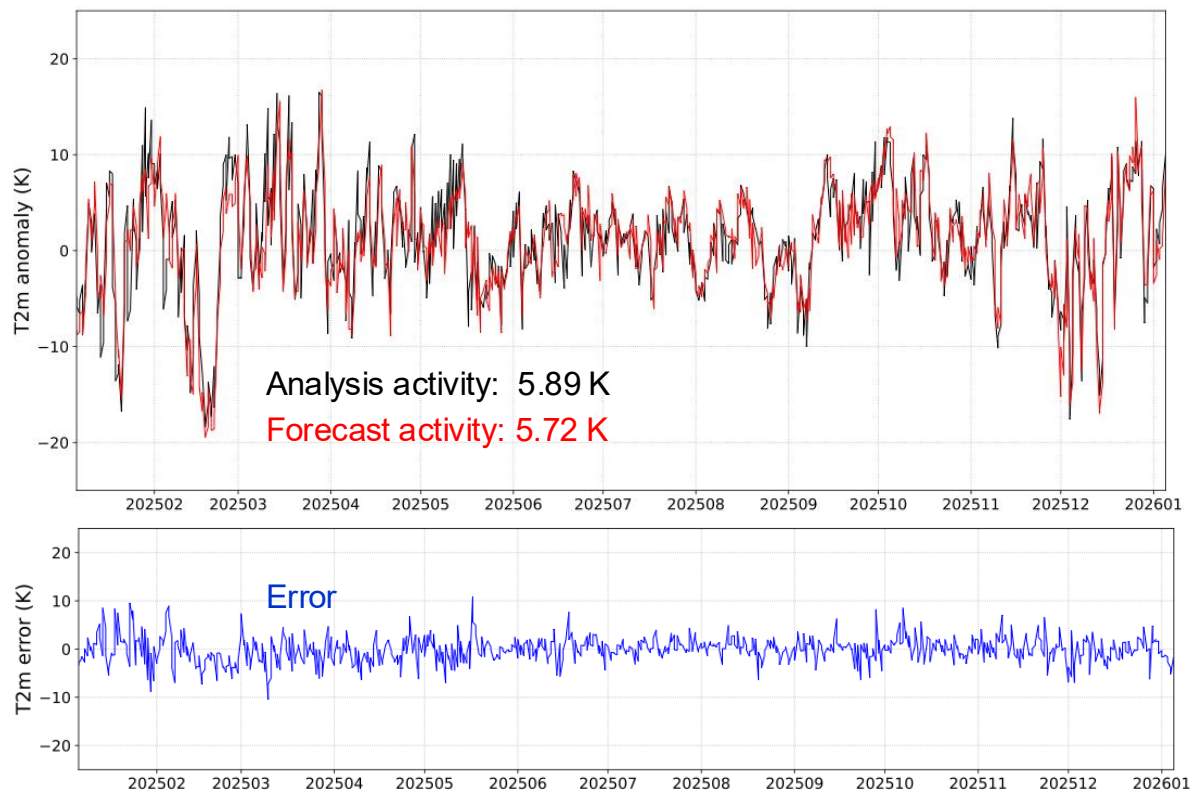
2025

RMSE normalized by analysis activity

Verification against own analysis

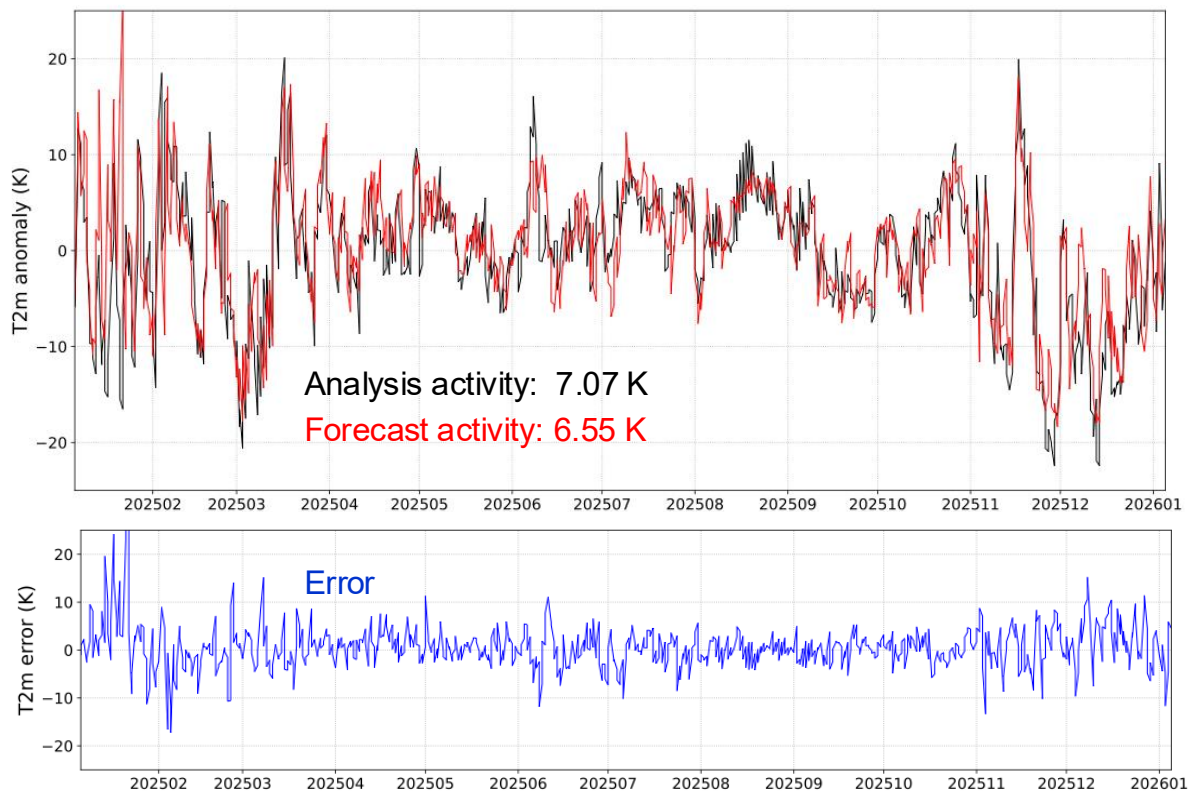
2m temperature anomalies at Day 5

U.S. Midwest



RMSE = 2.74 K
nRMSE = 0.47
ACC = 0.88

Siberia

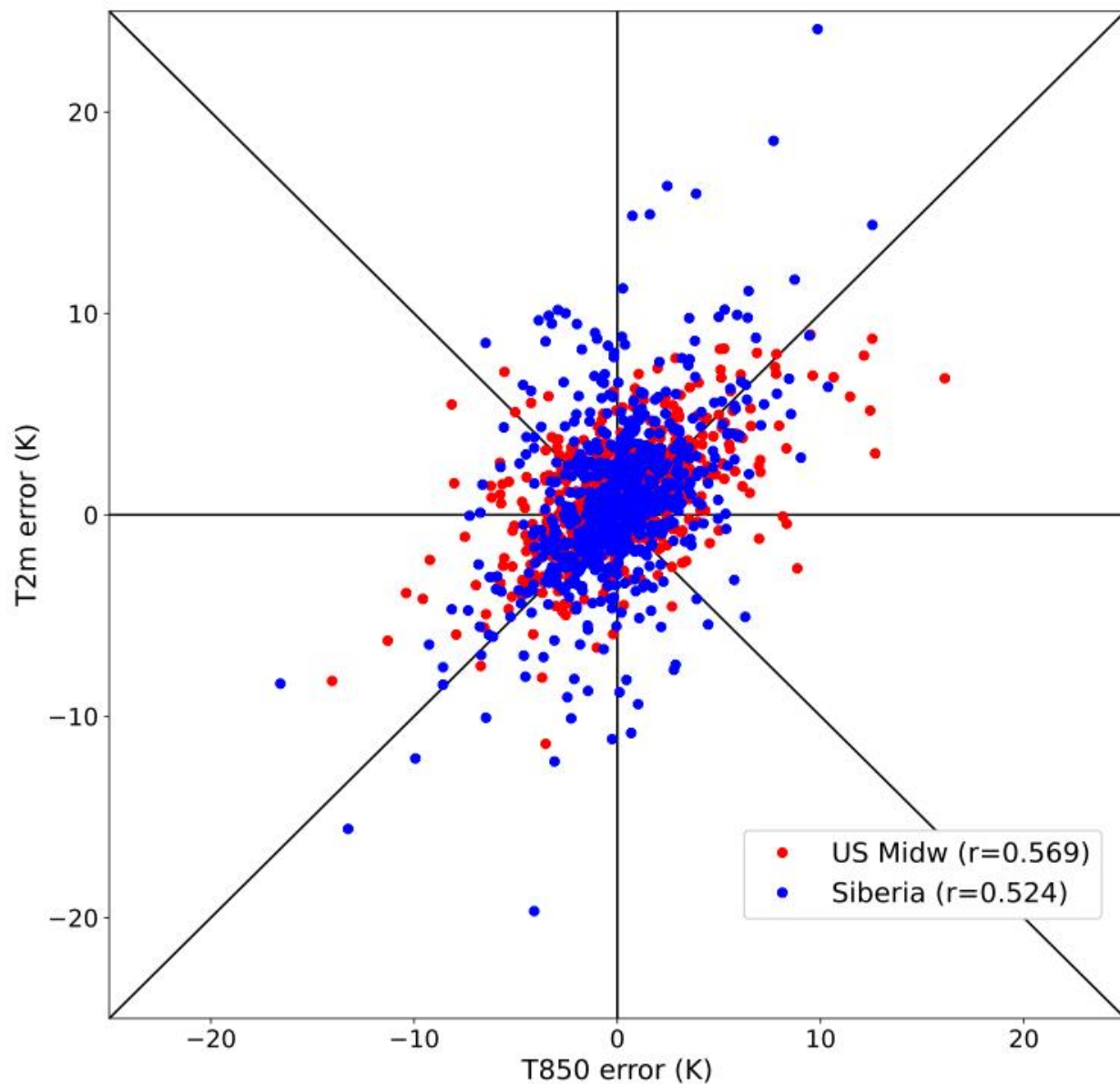


RMSE = 4.74 K
nRMSE = 0.67
ACC = 0.76

- Siberia has 20% higher activity but 70% larger RMSE
- Siberia has some very large errors, especially in winter

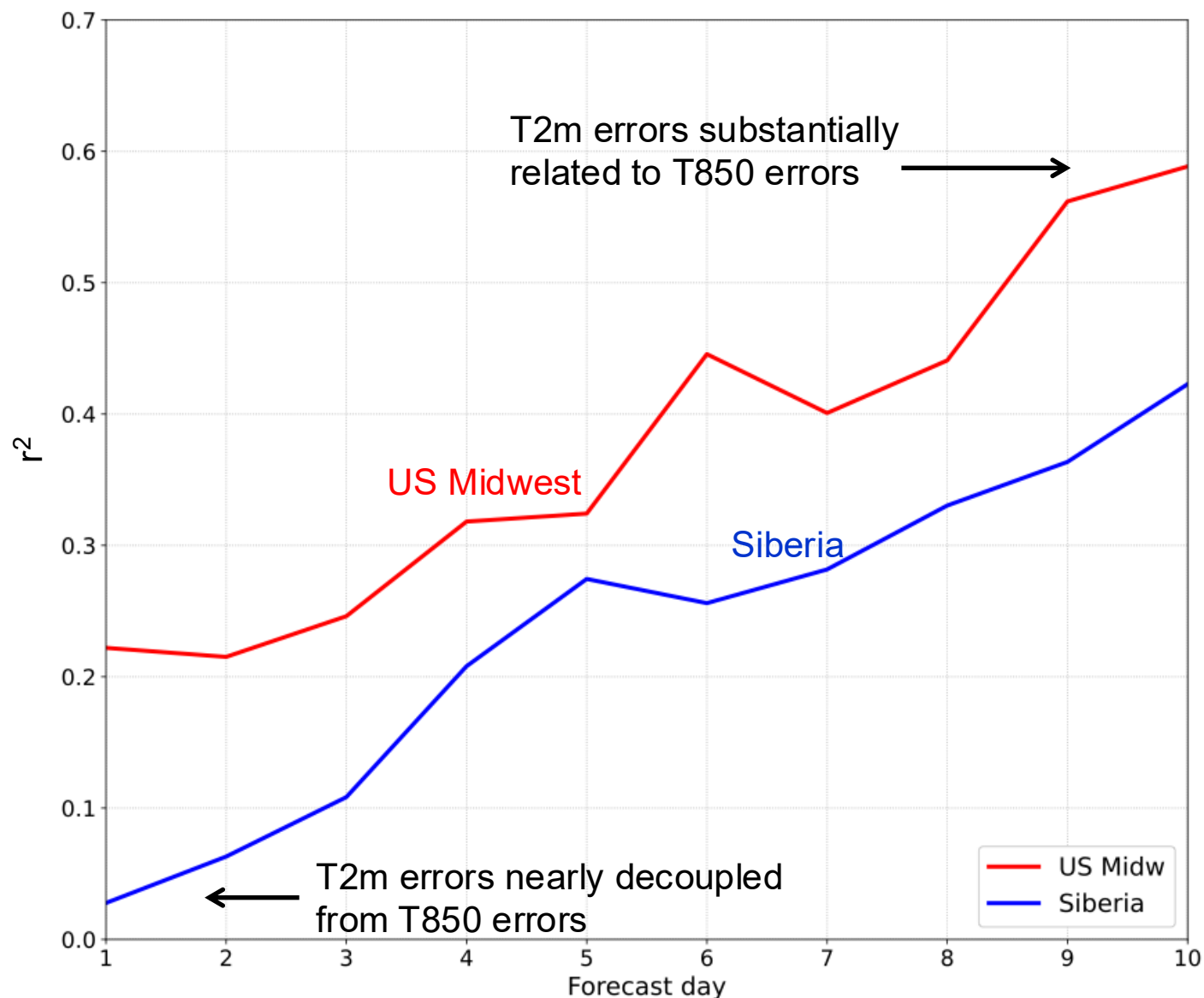
The role of 850 hPa temperature

Relationship between T2m and T850 errors (Day 5)



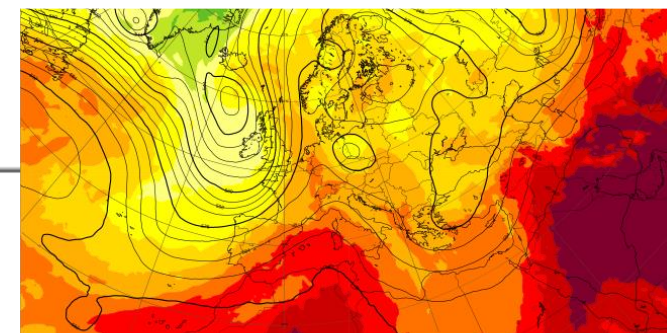
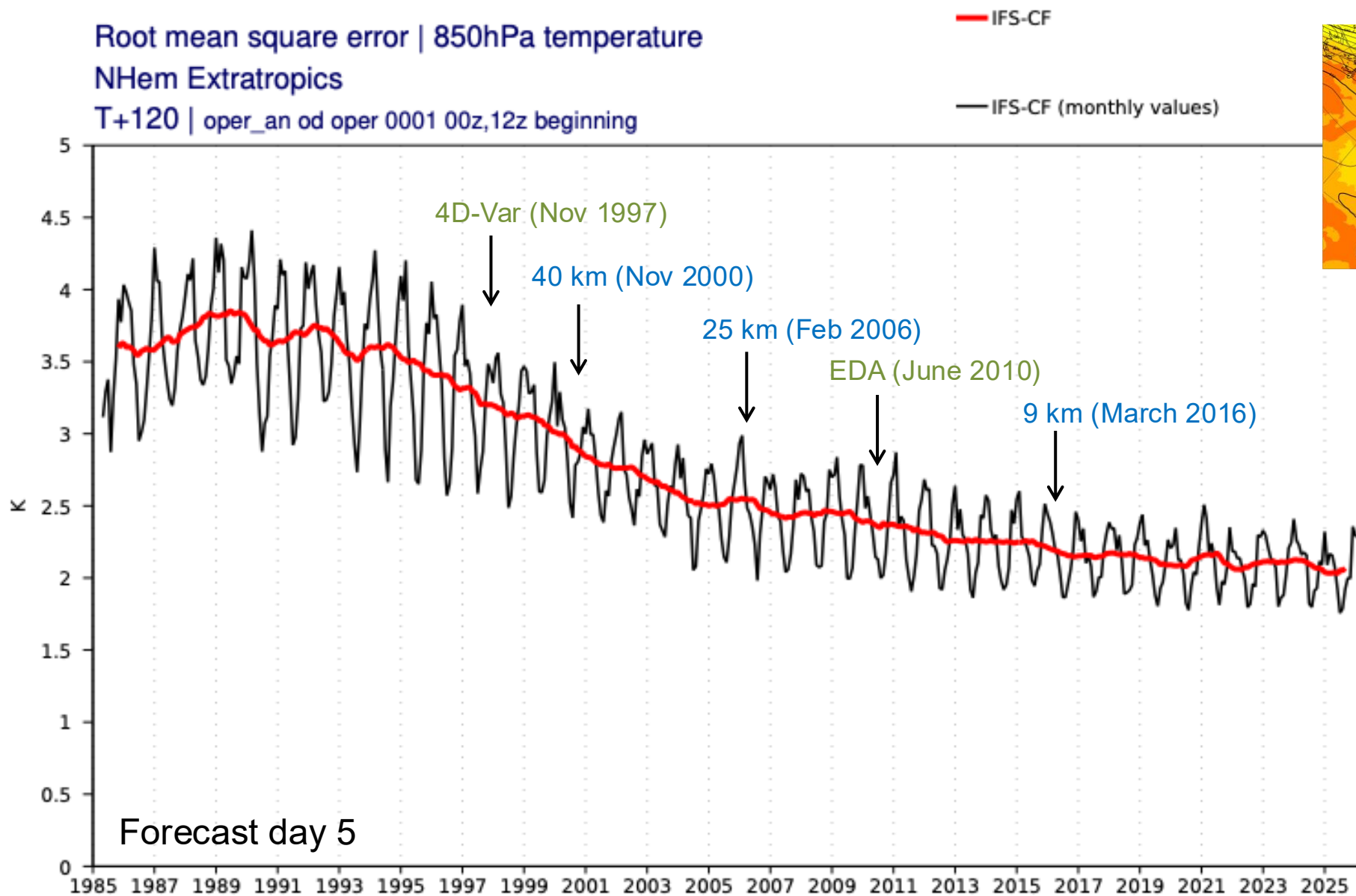
- Similar levels of correlation
- For correct T850, T2m errors in Siberia can be substantially larger than in US Midwest

To what extent are T2m errors related to T850 errors?



- Strength of T2m-T850 relationship strongly increases with lead time
- In Siberia, in the short range, T850 almost does not matter for T2m
- In the US Midwest at Day 10, T850 “explains” almost 60% of T2m errors

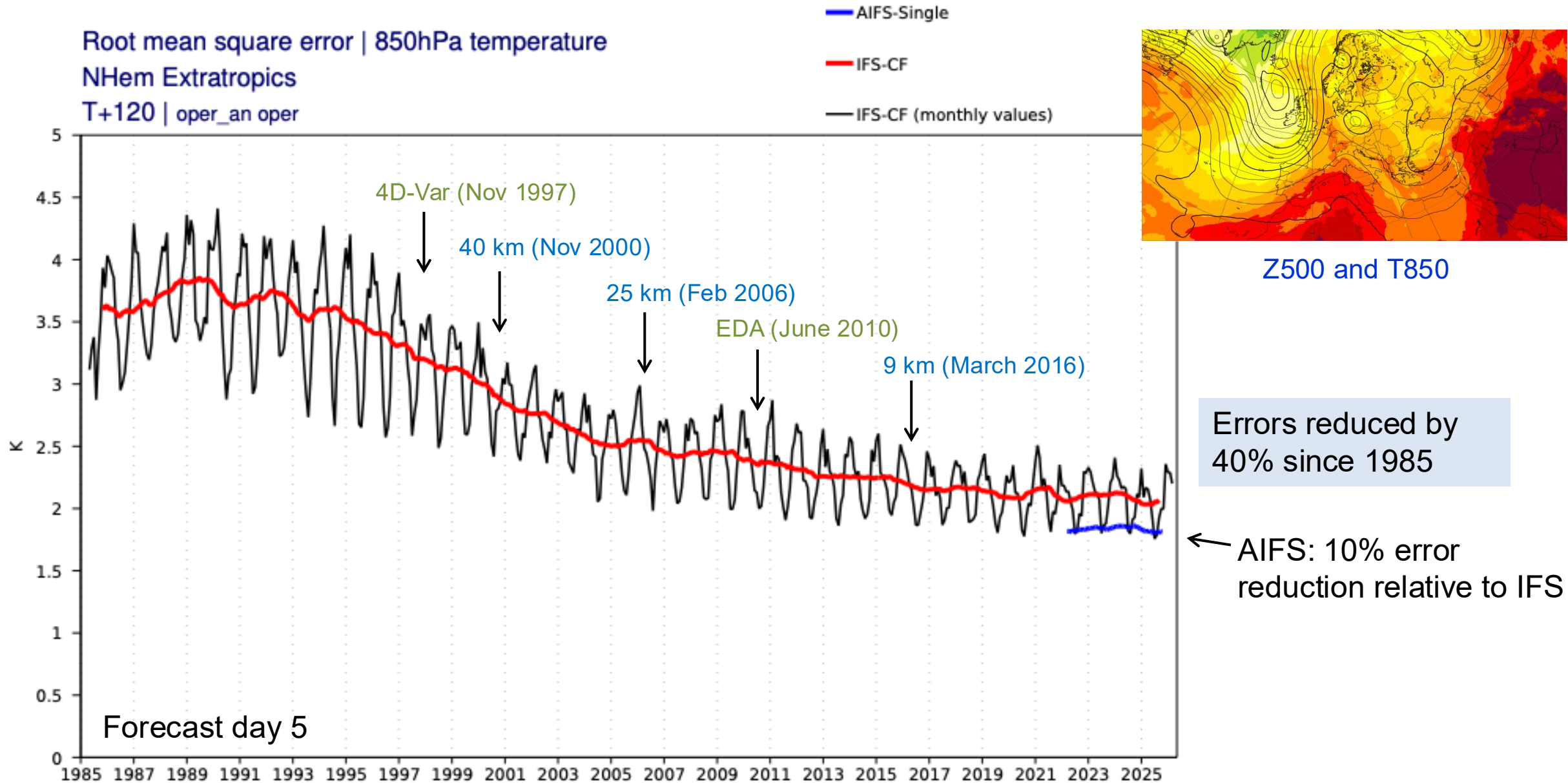
RMSE of 850 hPa temperature



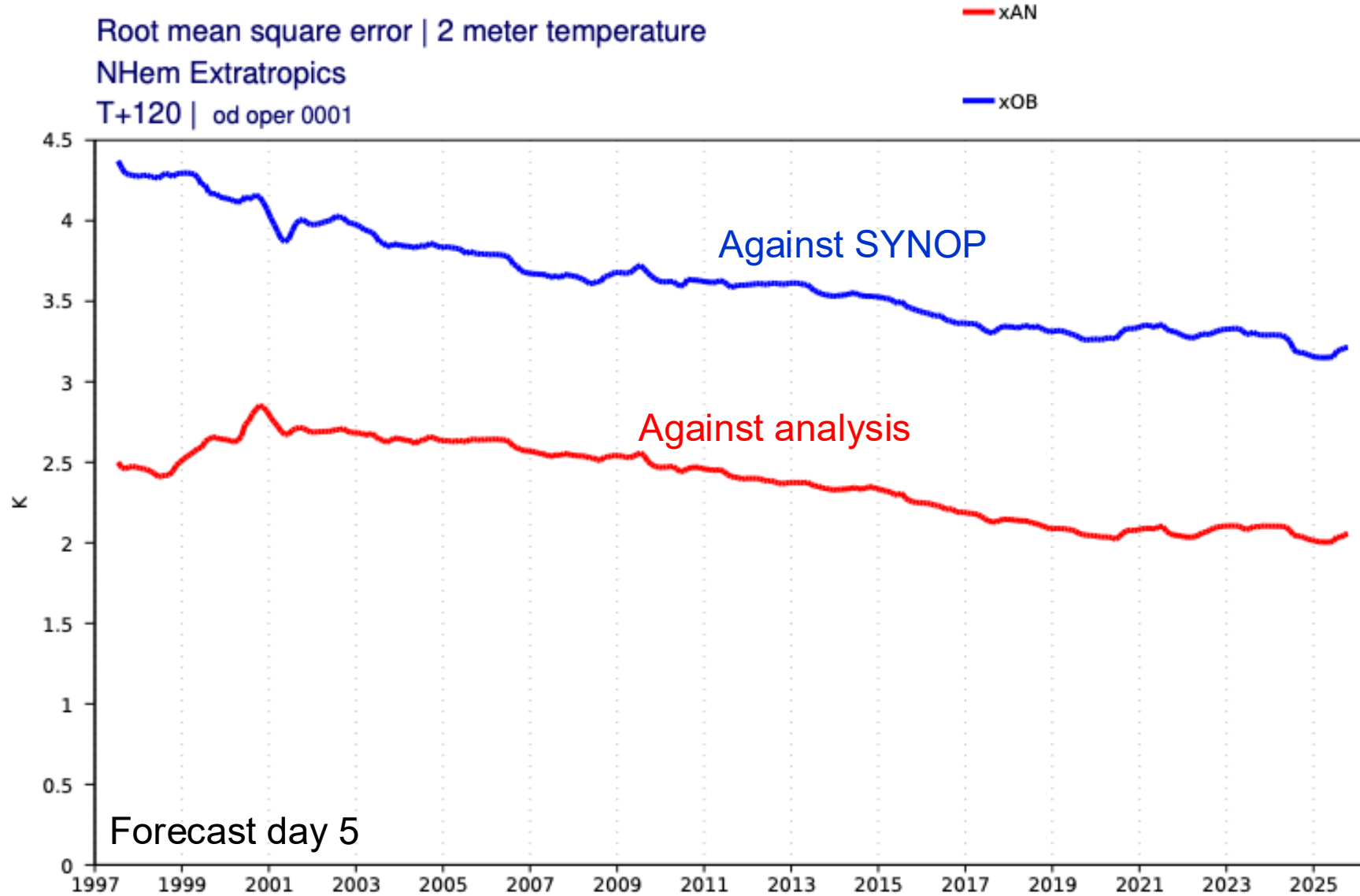
Z500 and T850

Errors reduced by 40% since 1985

RMSE of 850 hPa temperature

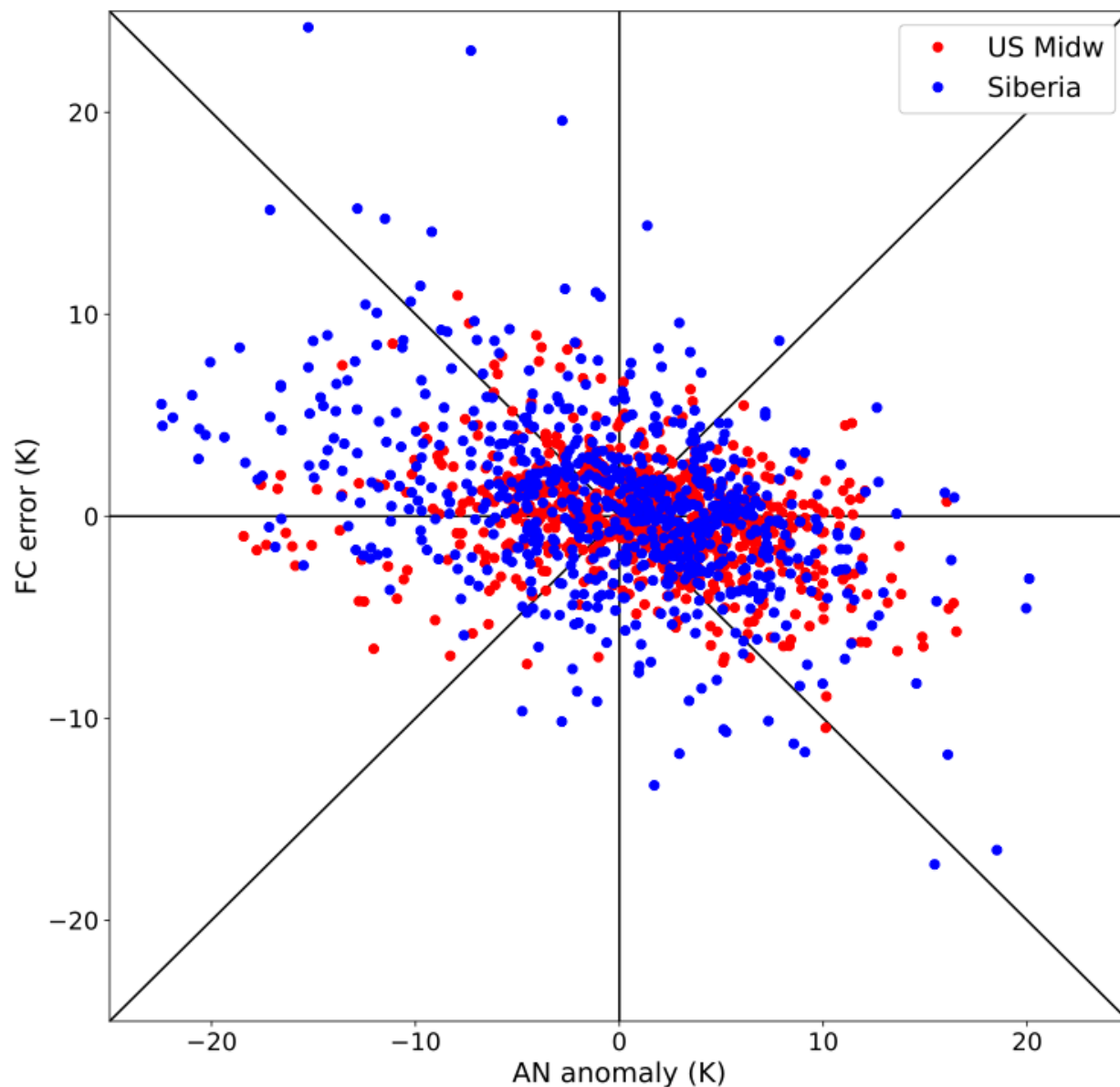


RMSE of 2m temperature



25% error reduction
over 30 years

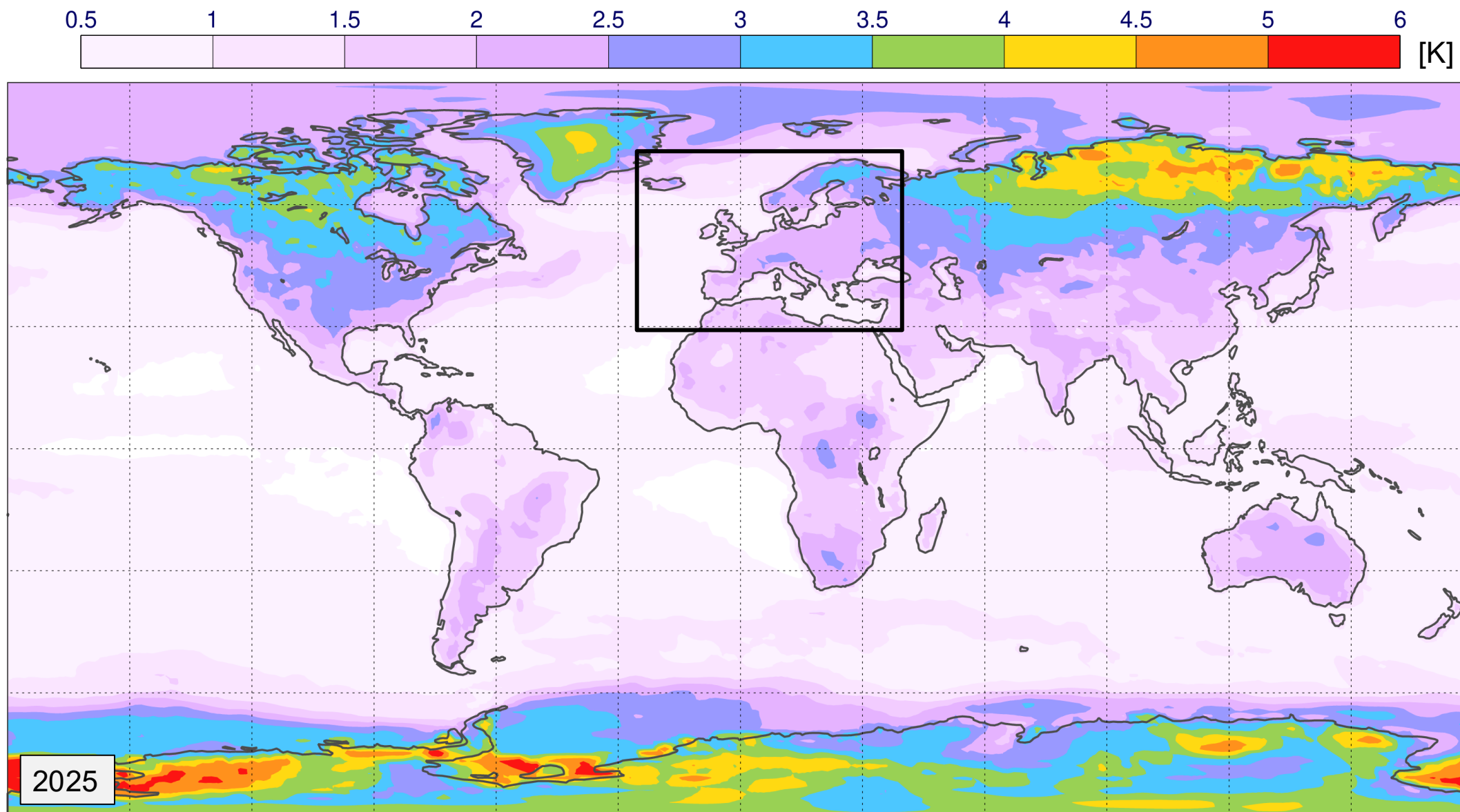
Larger T2m errors for larger T2m anomalies?



- US Midwest: errors grow slightly with increasing anomaly
- Siberia: some very large errors for large cold anomalies
- Siberia: more systematic warm bias for cold anomalies than in US Midwest

Verification against observations (SYNOP)

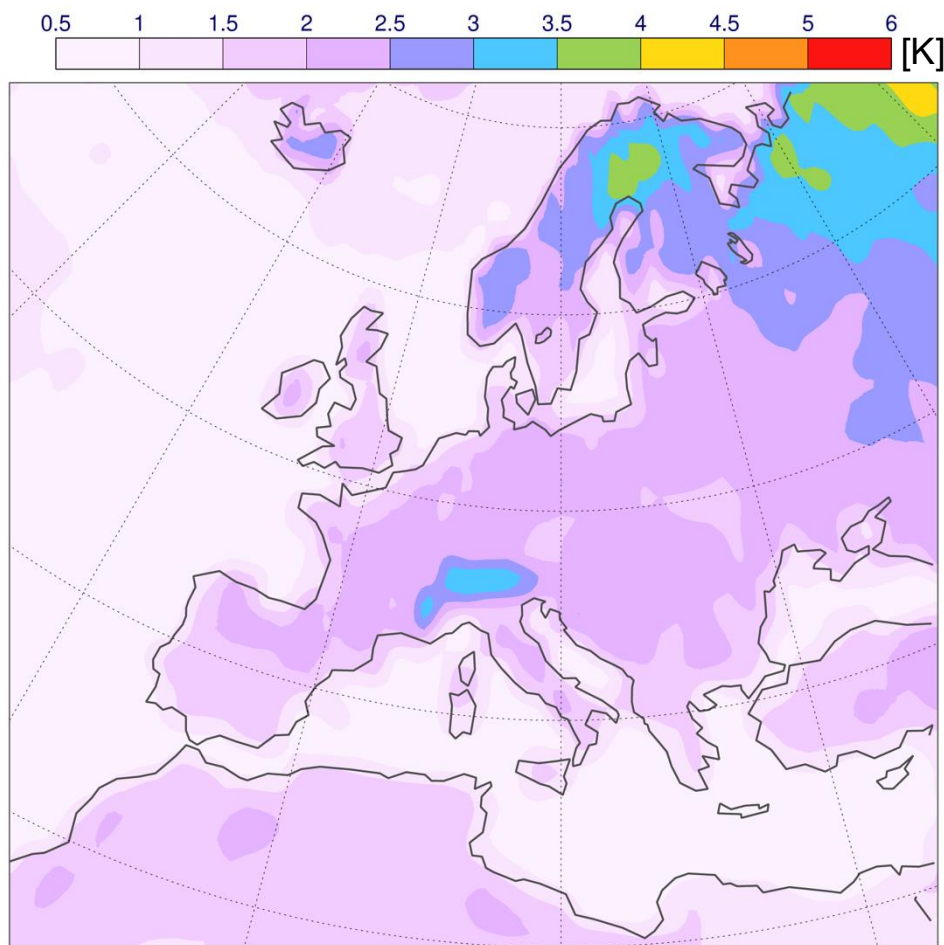
2m temperature forecast error (RMSE) at Day 5



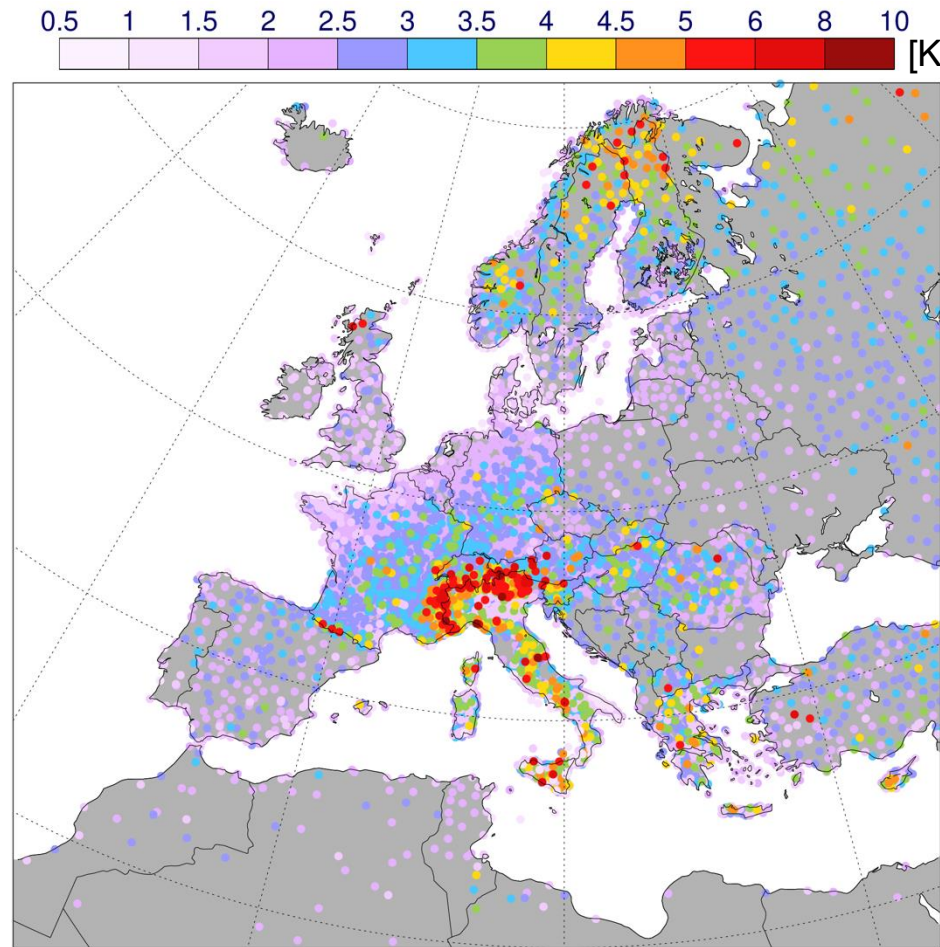
Verification against own analysis

Nighttime 2m temperature RMSE at Day 5 (00 UTC)

2025



Verification against analysis

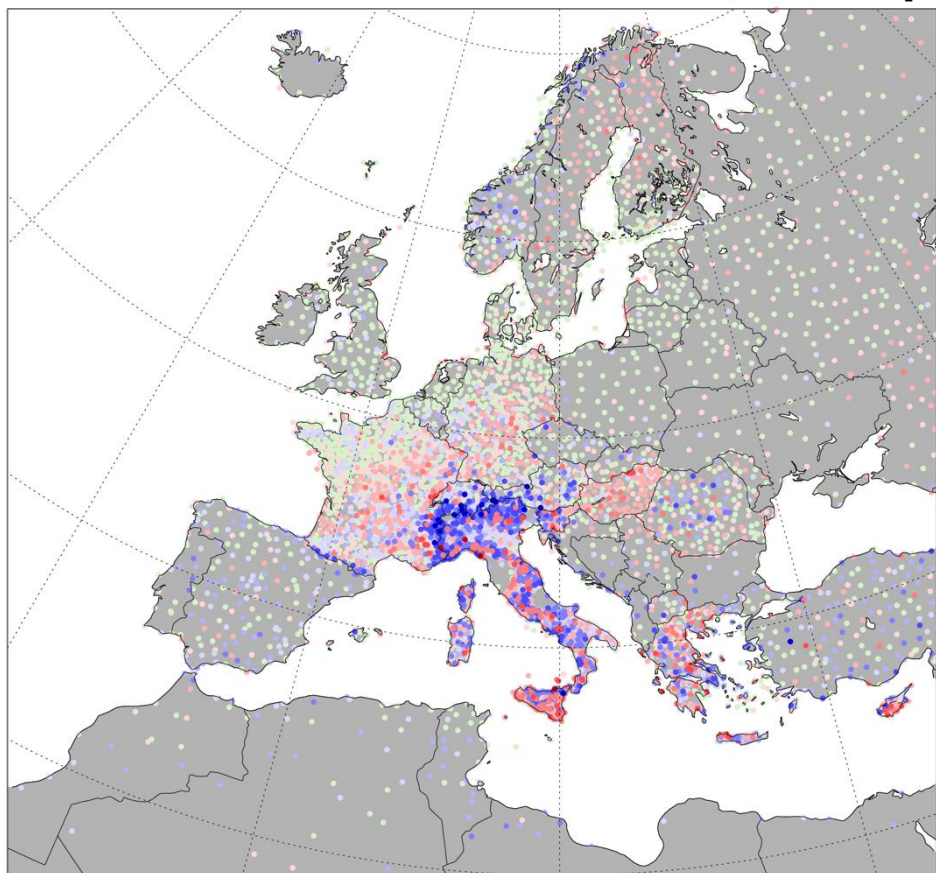
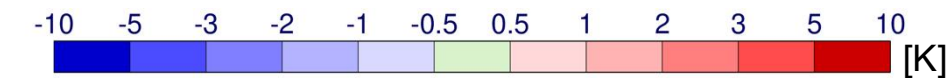


Verification against SYNOP

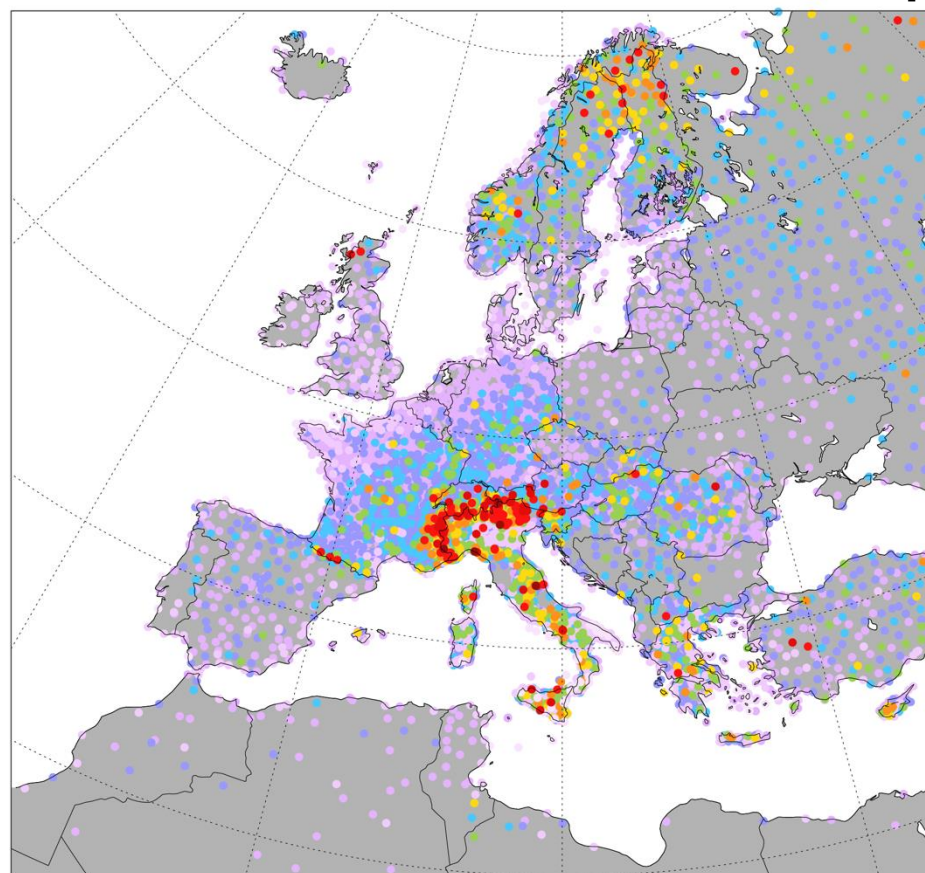
- Errors increase with:
- latitude
 - continentality
 - orographic complexity

Nighttime 2m temperature at Day 5 (00 UTC)

2025



BIAS

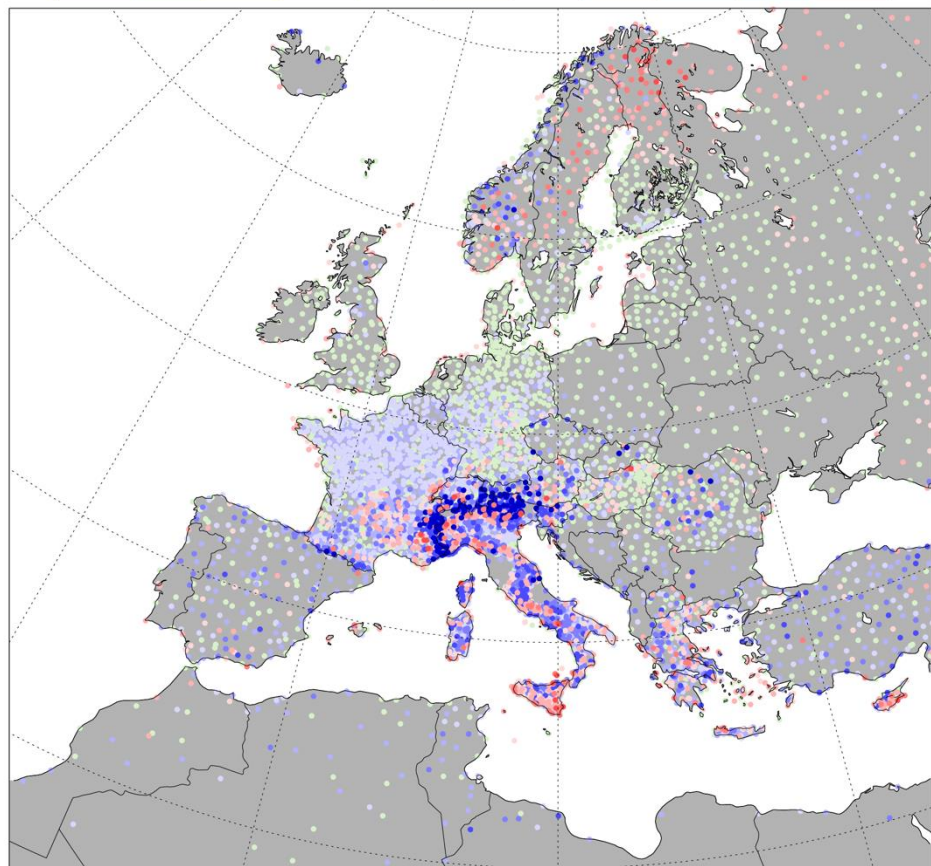
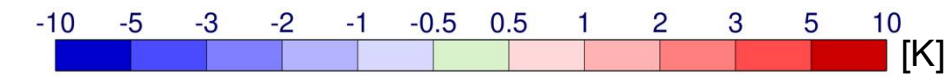


RMSE

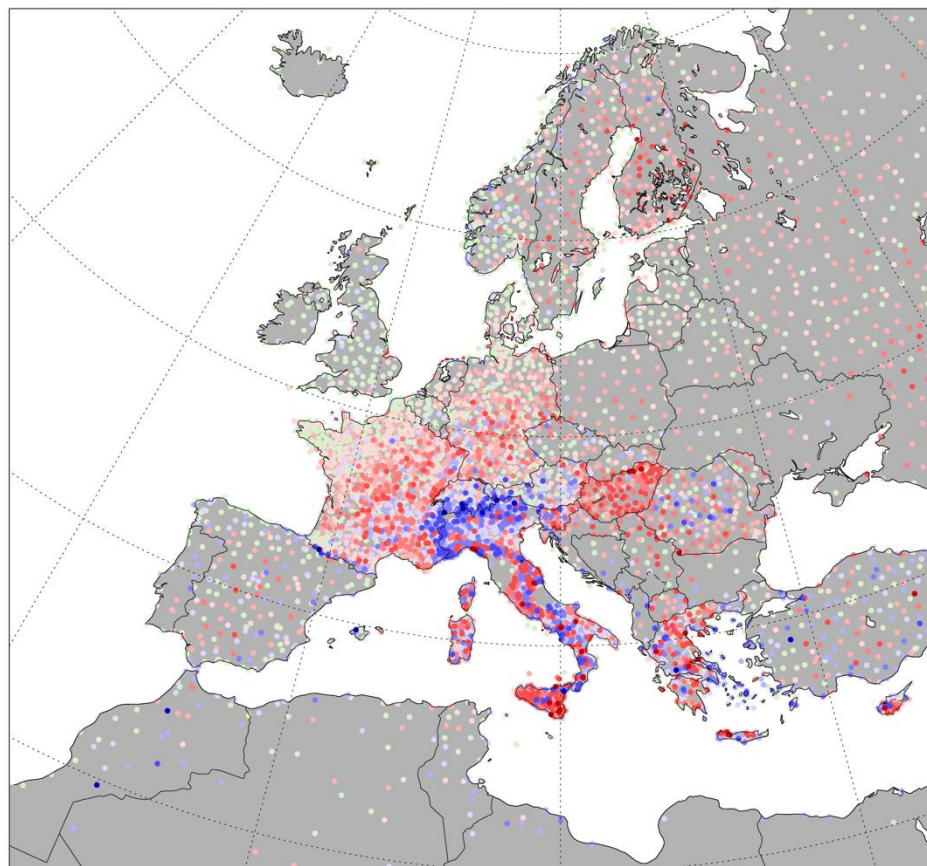
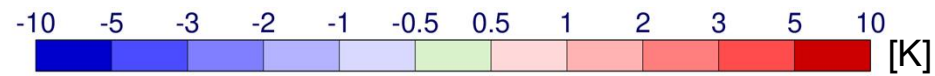
- Complex bias pattern
- Bias contributes substantially to RMSE in some areas

Nighttime 2m temperature bias at Day 5 (00 UTC)

2025



Winter
(DJF 2025)



Summer
(JJA 2025)

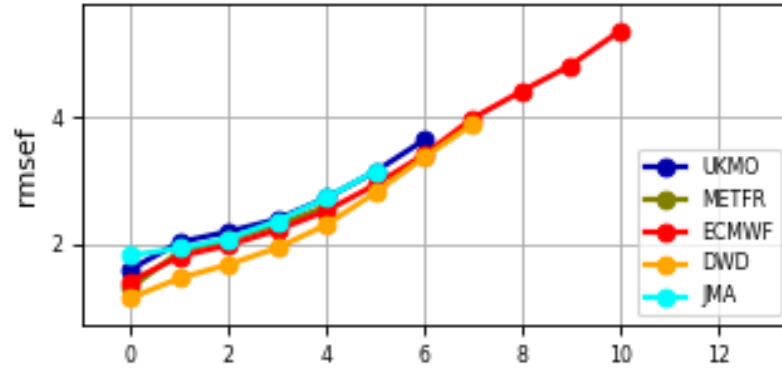
Biases strongly dependent on season

2m temperature model intercomparison (WMO scores)

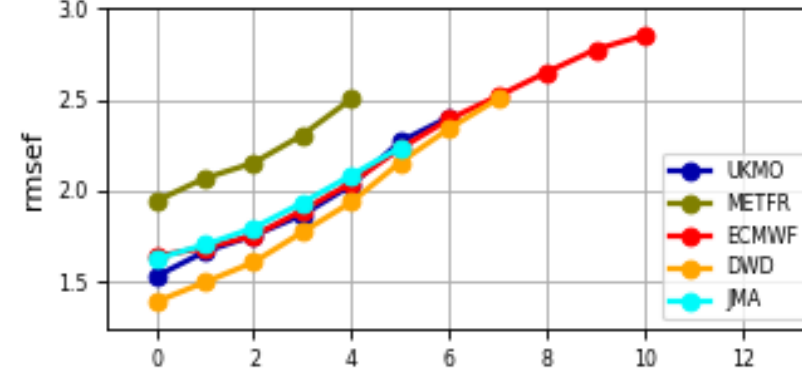
DJF 2026, 00 UTC

RMSE

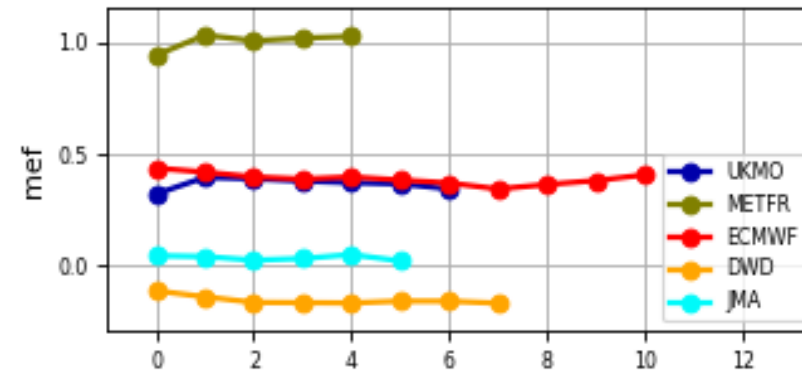
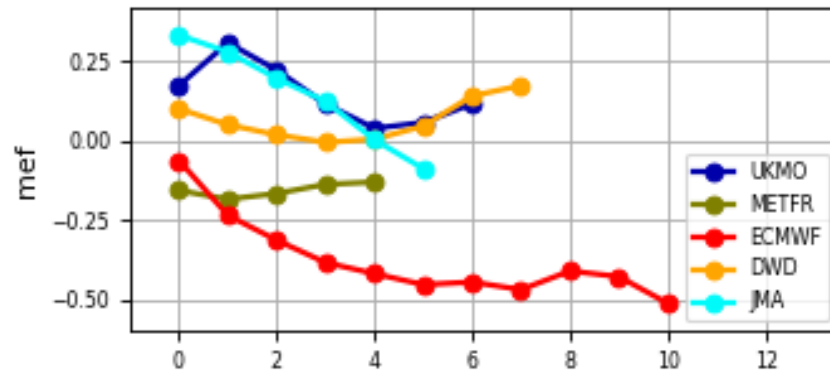
2t – DJF 2026 – Europe [75, -12, 35, 42]



10ff – DJF 2026 – Europe [75, -12, 35, 42]



BIAS

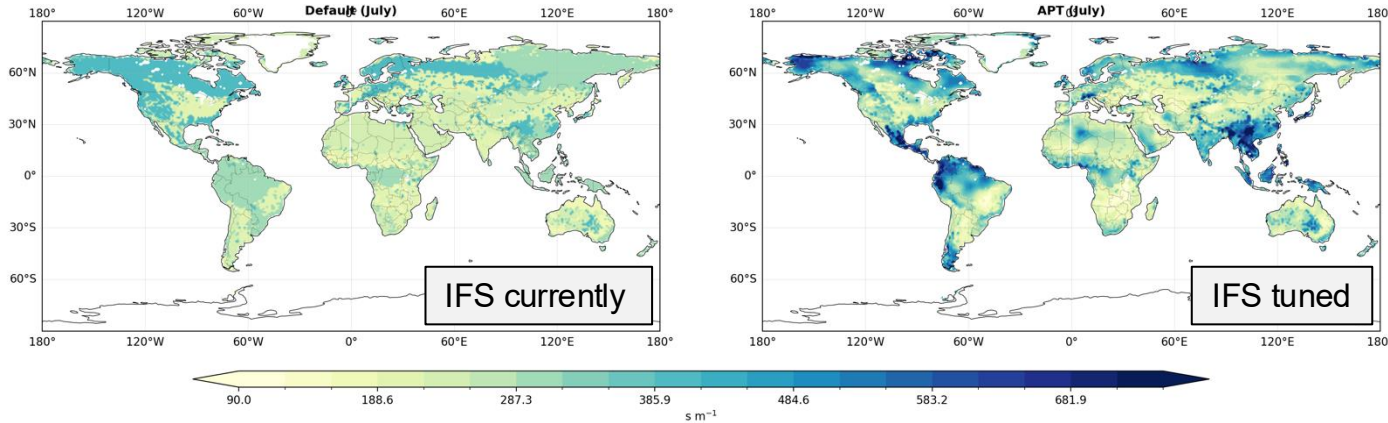


2m temperature

10m wind speed

- DWD generally leading due to adaptive parameter tuning
- ECMWF in the middle ground

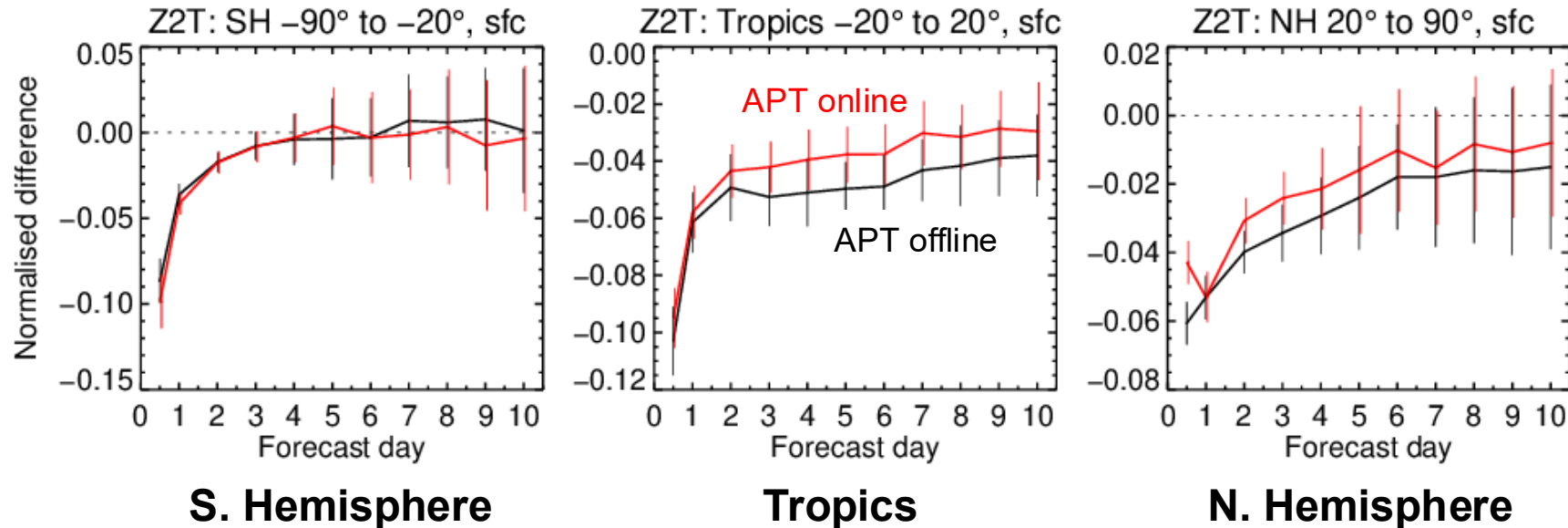
Offline parameter tuning (ECMWF approach)



Default minimum stomatal resistance for high vegetation (evaporation control)

3-Jun-2024 to 21-Jul-2024 from 78 to 97 samples. Verified against own-analysis.

Confidence range 95% with AR(2) inflation and Sidak correction for 8 independent tests.

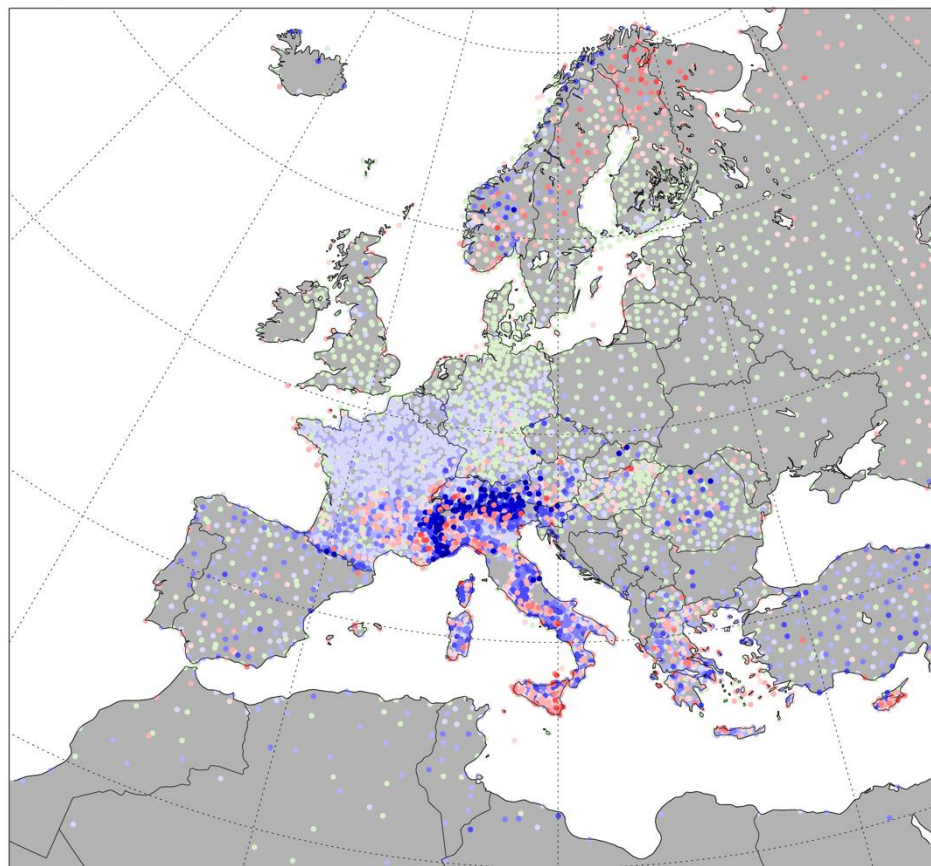
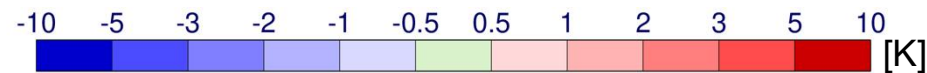
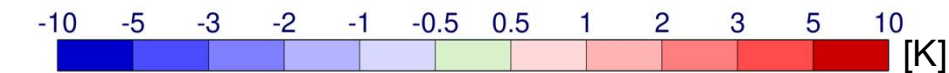


- Substantial improvement, even larger in ENS (10%)
- To be implemented in 51r1

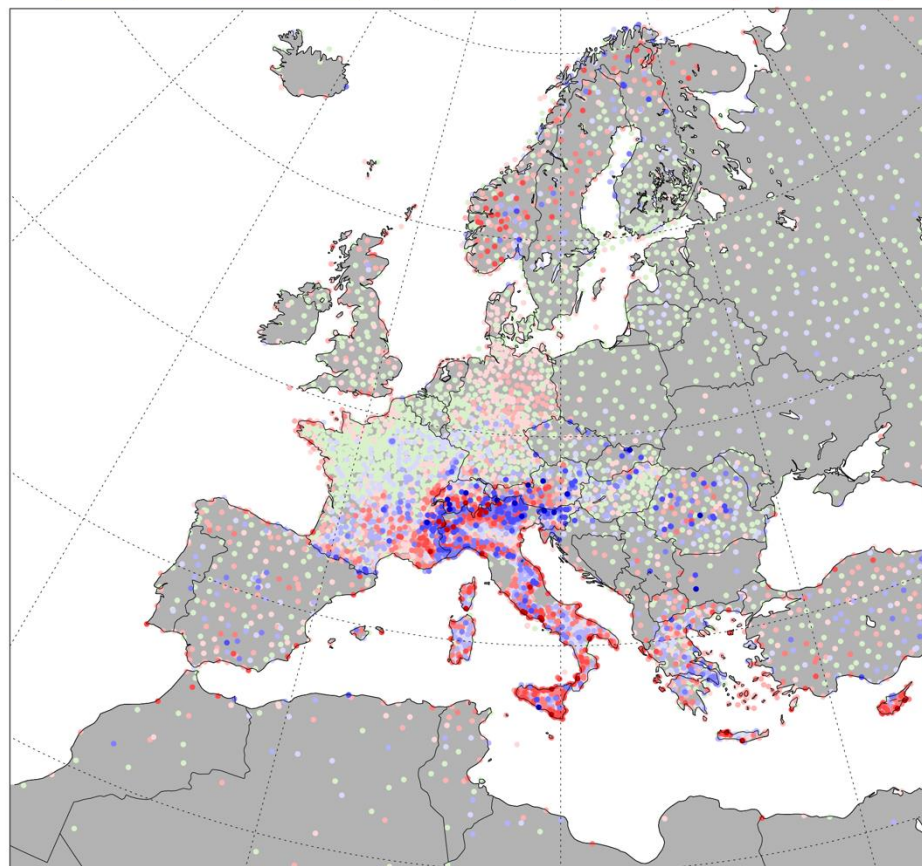
AIFS forecasts of 2m temperature

Nighttime 2m temperature bias at Day 5 (00 UTC)

DJF 2025



IFS

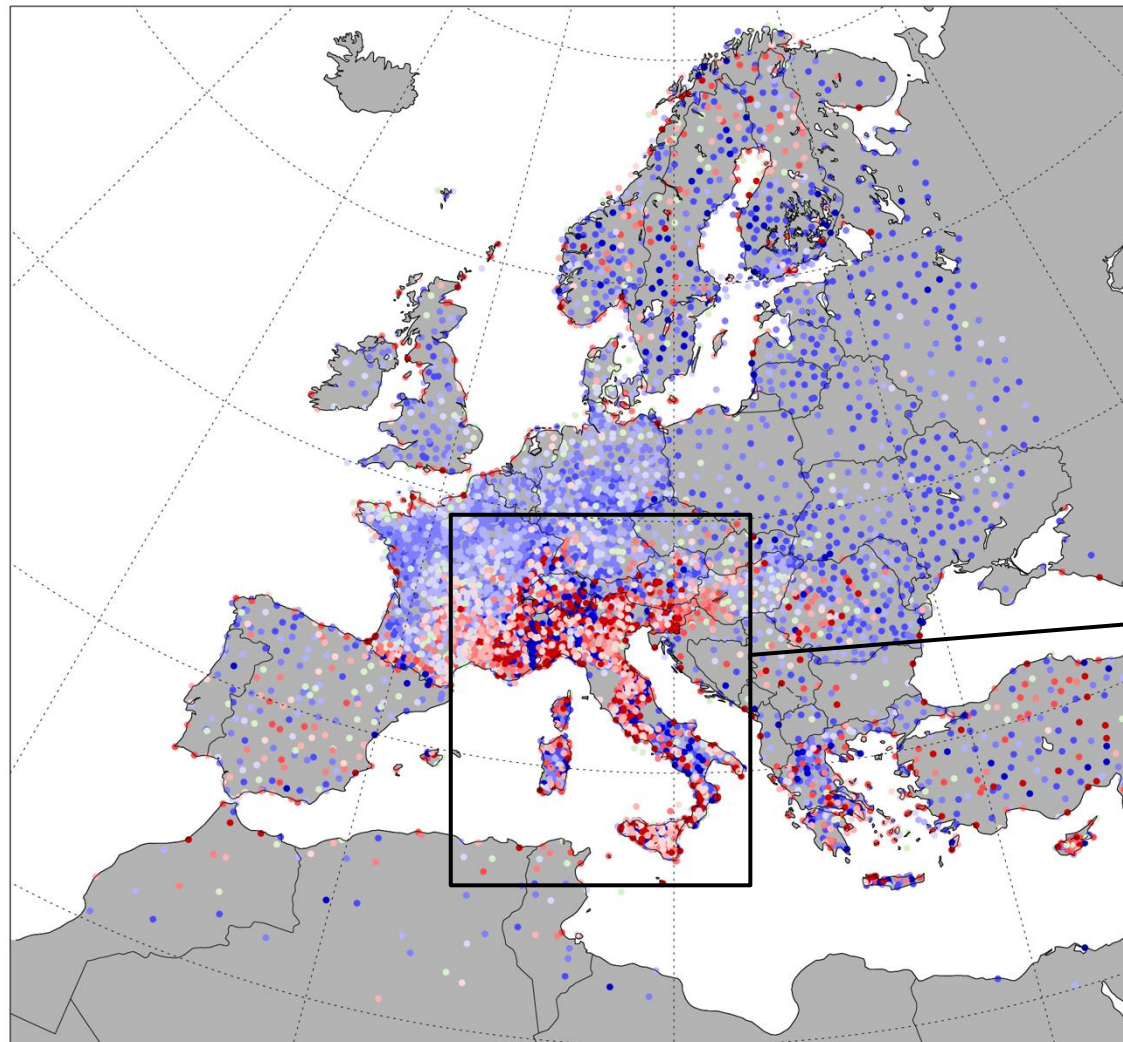
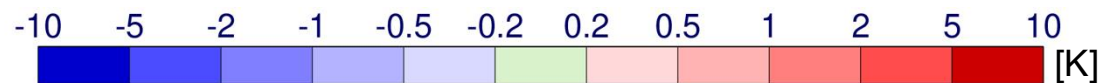


AIFS

AIFS bias smaller in some areas but not everywhere

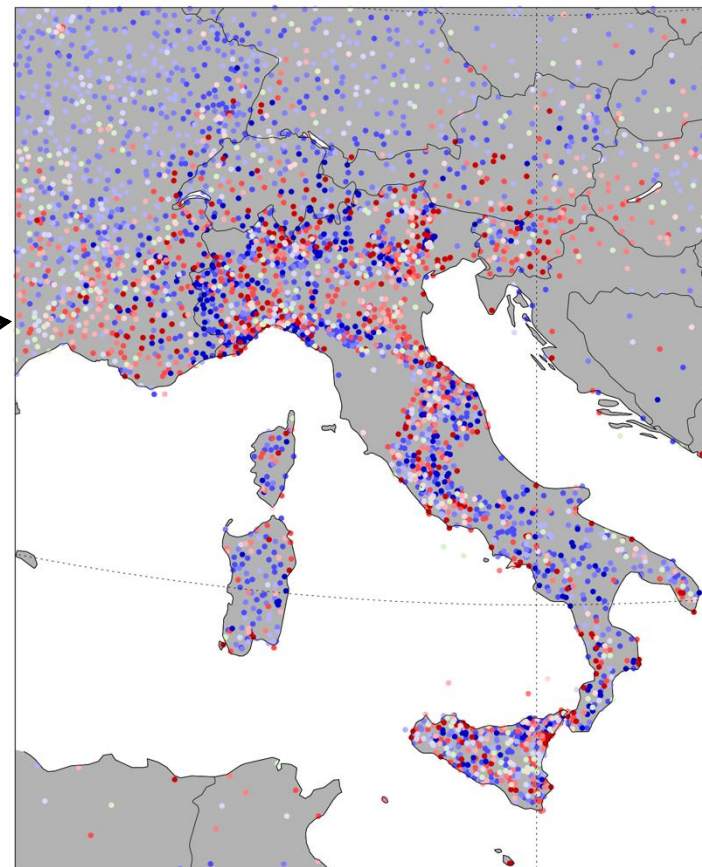
Where does AIFS improve on IFS?

DJF 2025



RMSE(AIFS) – RMSE(IFS)

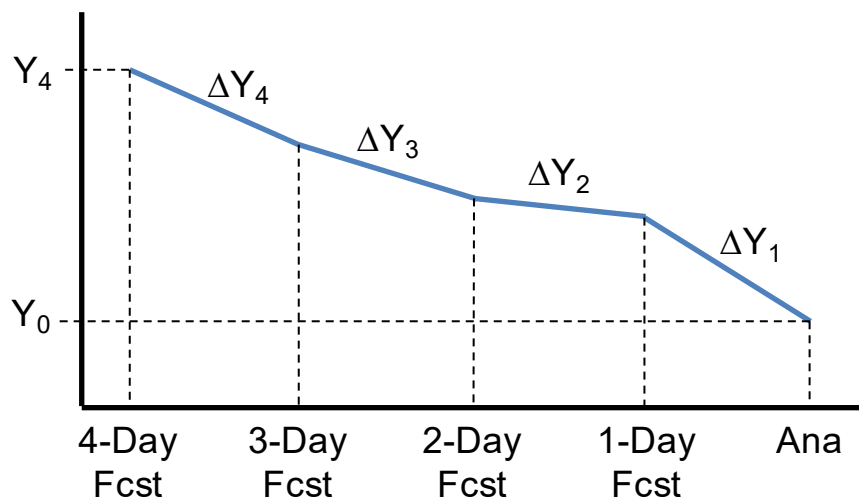
- AIFS improvement not just due to smaller bias
- Scandinavia improved, but still problematic
- Orographically complex areas worse in AIFS



Day 5, 00 UTC,
Verification against
SYNOP

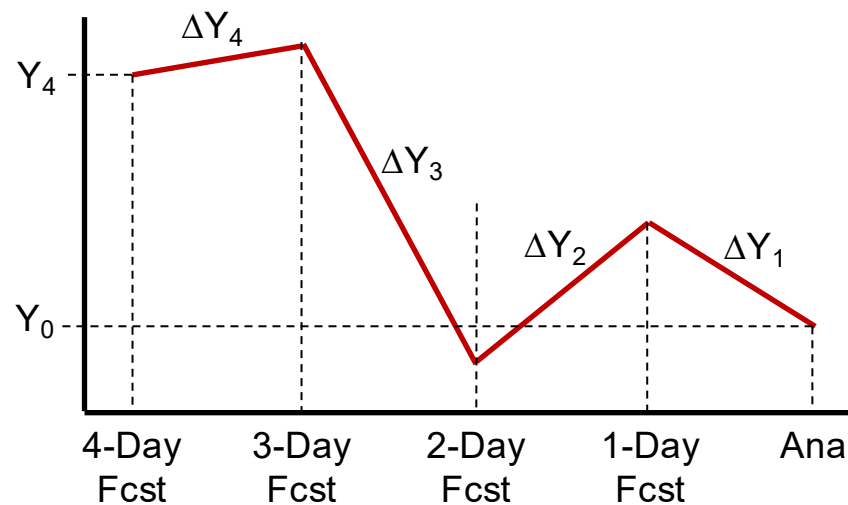
Jumpiness

Measuring jumpiness ('flip-flop' behaviour)



Monotonous approach towards analysis

$$\sum |\Delta Y_i| = |Y_4 - Y_0|$$



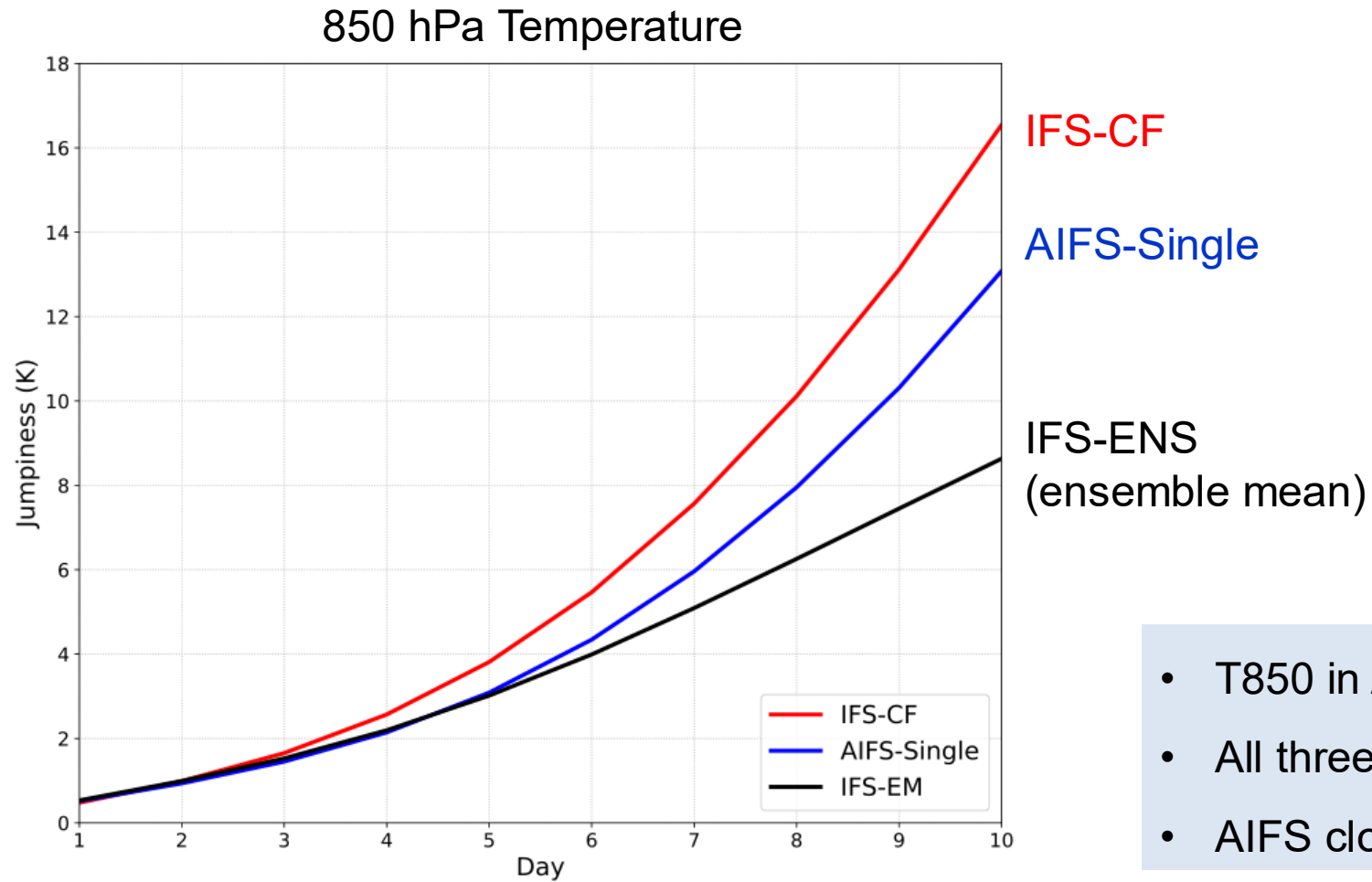
Jumpy approach towards analysis

$$\sum |\Delta Y_i| > |Y_4 - Y_0|$$

- Quantifies the smoothness of approach towards the observed value



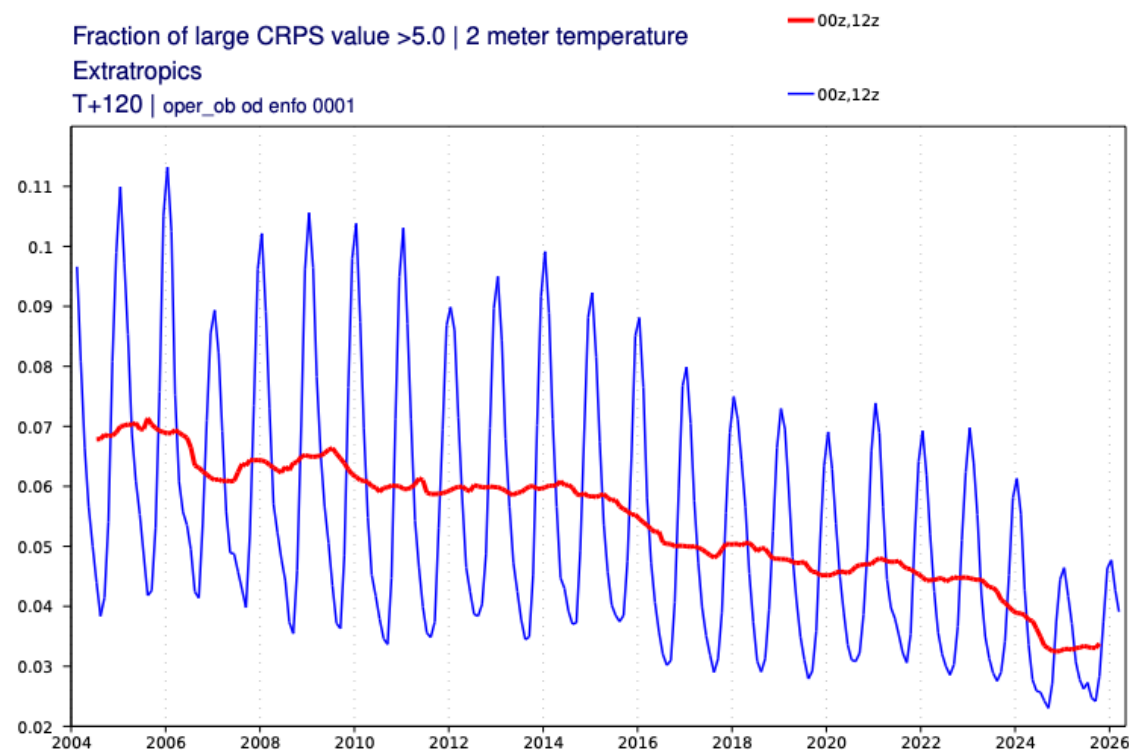
Jumpiness: AIFS vs IFS (DJF 2025-26)



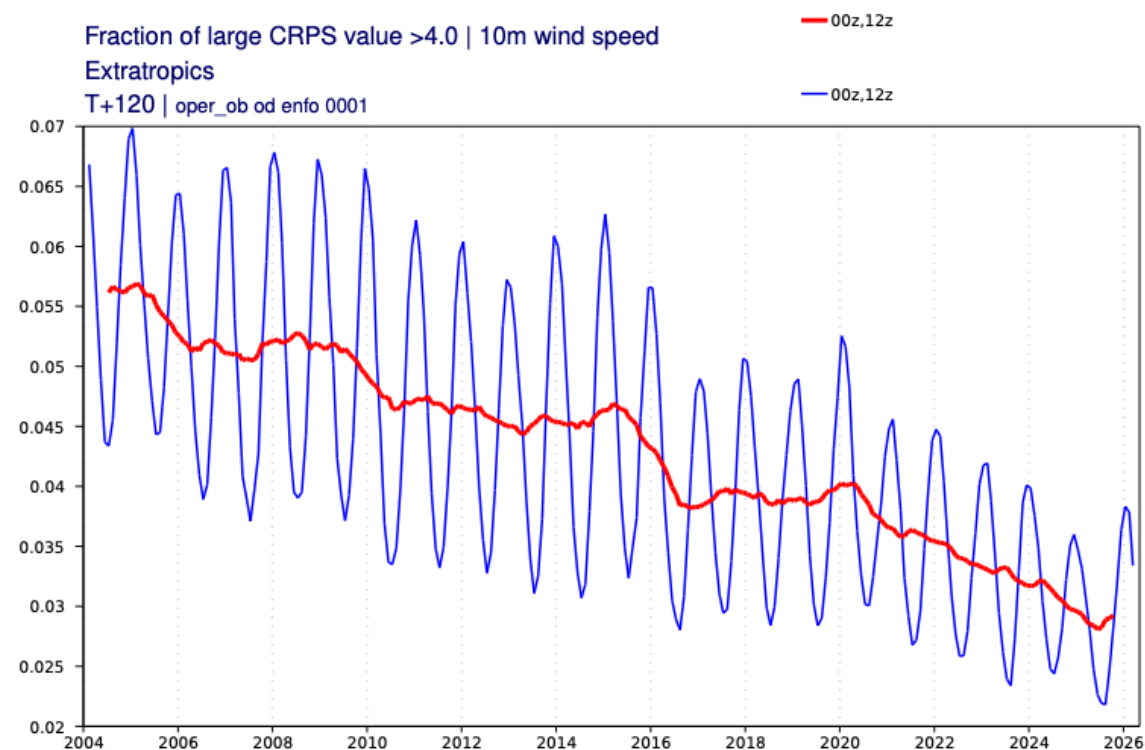
- T850 in AIFS about 20% less jumpy than in IFS
- All three very similar up to Day 3
- AIFS close to IFS ensemble mean up to Day 5

Ensemble forecasts

Number of large ENS errors



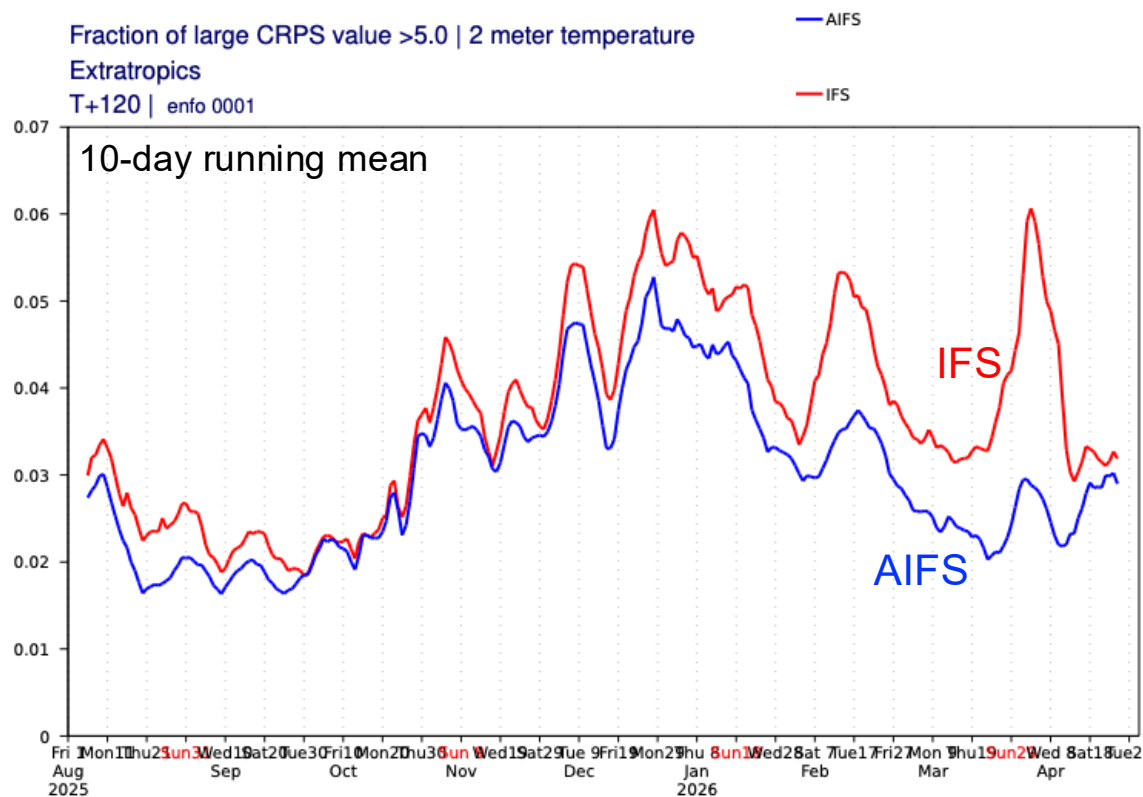
2 m temperature



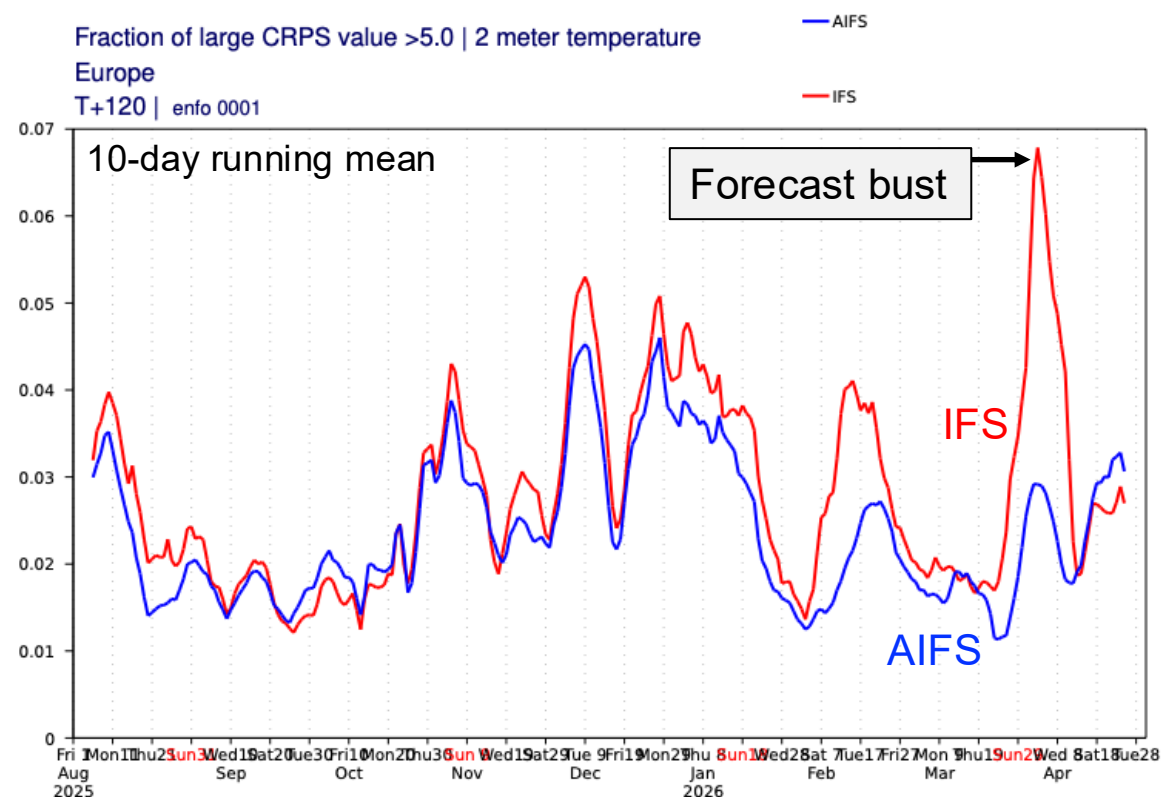
10 m wind speed

- Cycle 49r1 (SPP, T2m assimilation over land) has brought substantial improvement in 2024
- Wintertime maximum in number of large T2m errors strongly reduced
- Cycle 50r1 (May 2026) not expected to bring major T2m improvement

Fraction of large T2m errors: AIFS-ENS vs IFS-ENS (Day 5)



Extratropics



Europe

- Extratropics: typically 10-20% fewer large errors in DJF in AIFS but occasionally 50%!
- Europe: AIFS not always better than IFS but some peaks strongly reduced

Ensemble forecasts of T2m: DWD vs ECMWF

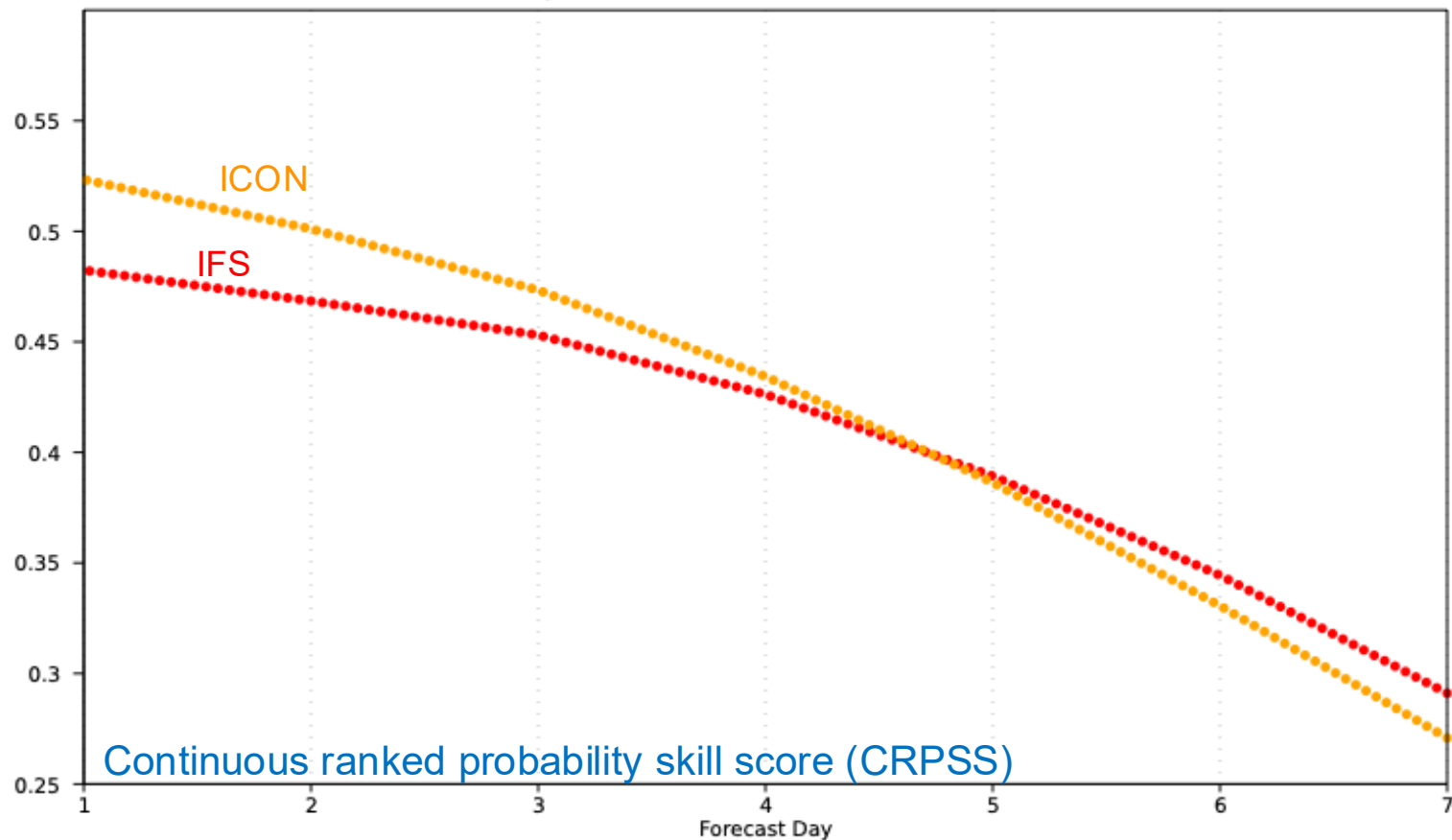
Continuous ranked probability skill score | 2 meter temperature

Europe

20250701 00z to 20260430 12z | enfo mean_fair

●●● DWD (TIGGE)

●●● IFS 0.5deg



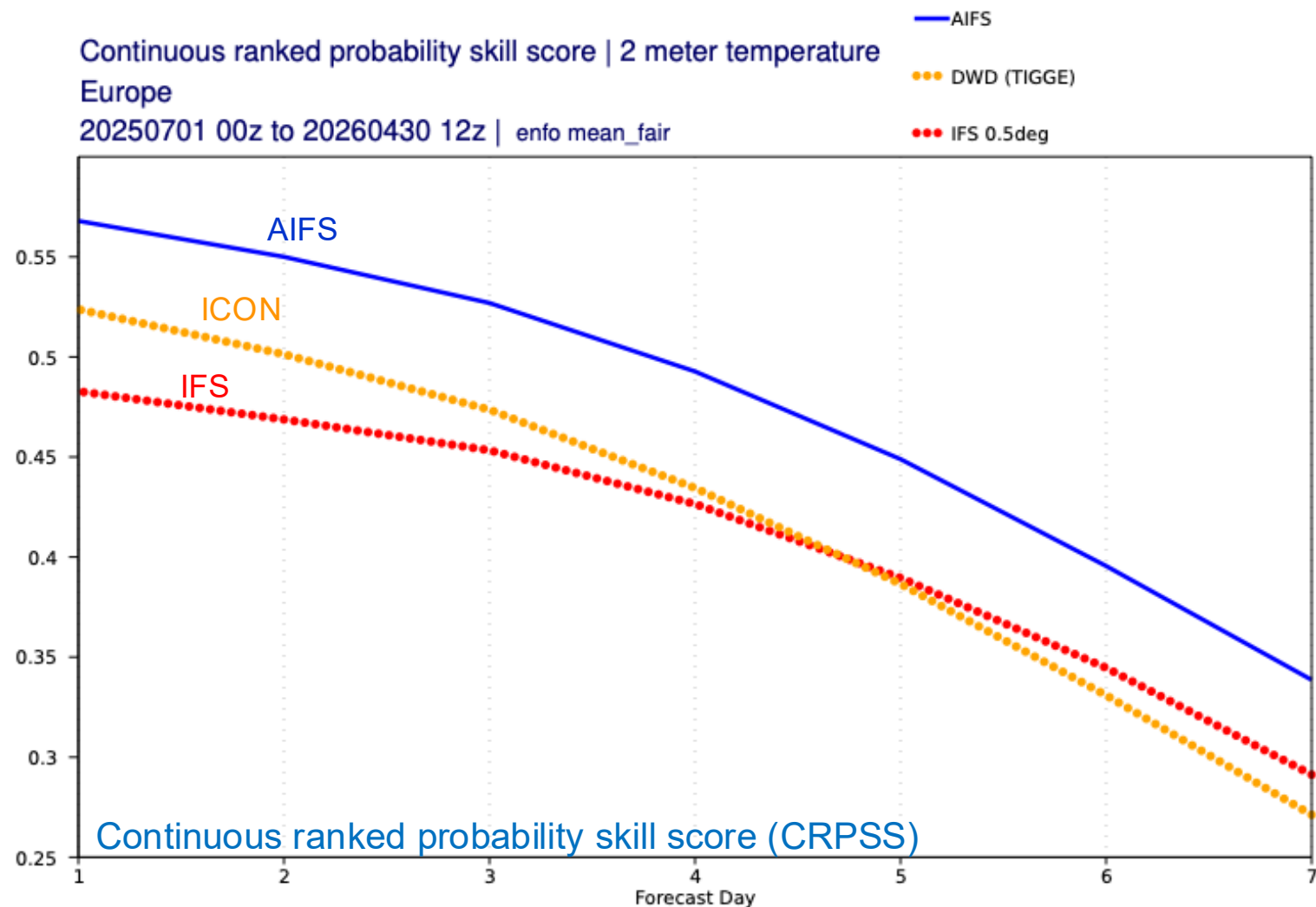
- Comparison based on TIGGE, i.e. 0.5 deg (not on original model grid)
- DWD leading up to Day 4
- Advantage mostly due to 'adaptive parameter tuning'

Ensemble forecasts of T2m: DWD vs ECMWF

Continuous ranked probability skill score | 2 meter temperature

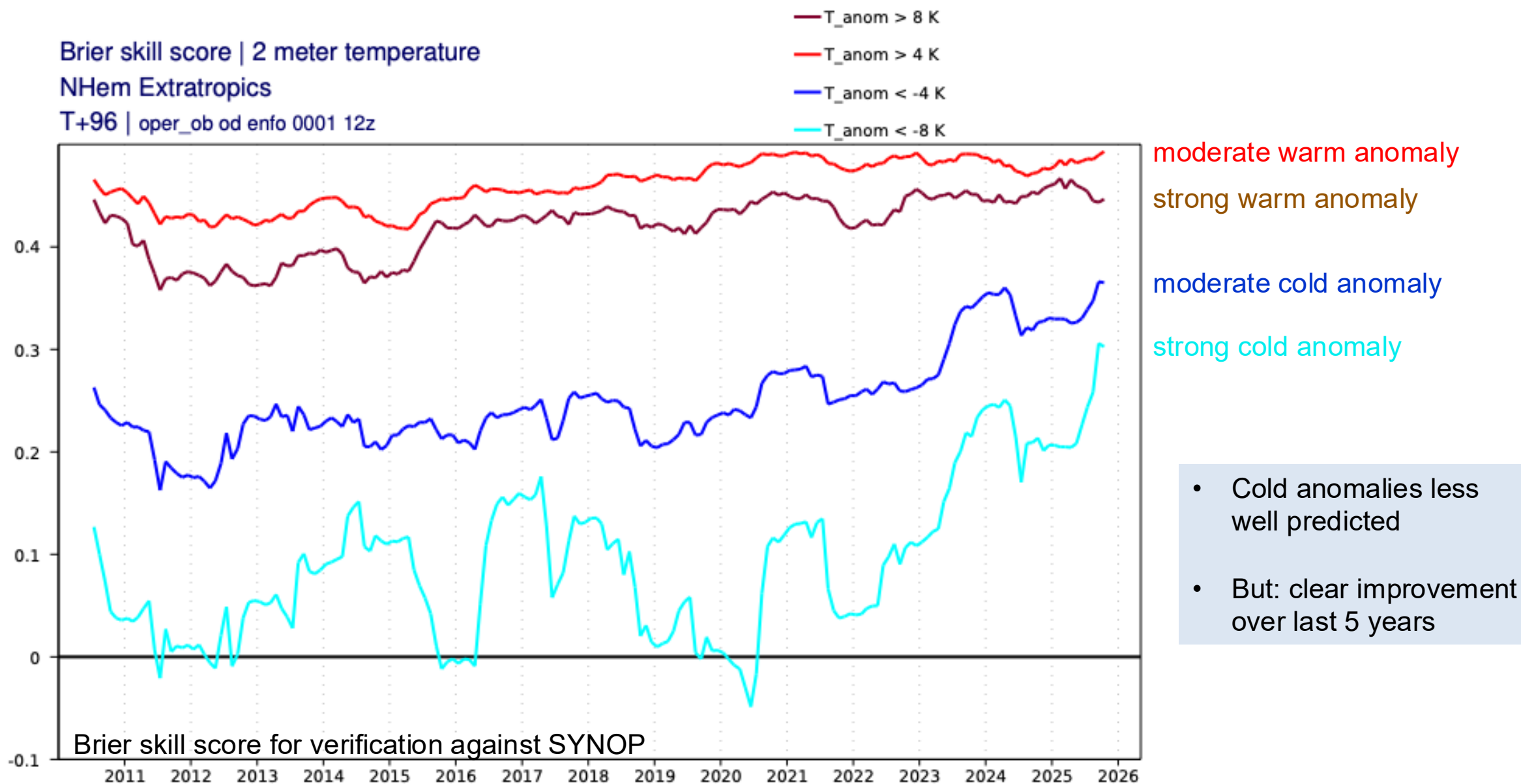
Europe

20250701 00z to 20260430 12z | enfo mean_fair

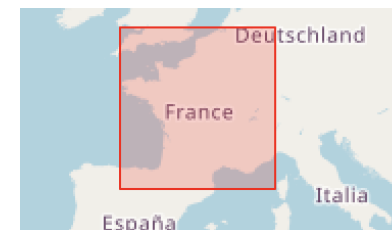
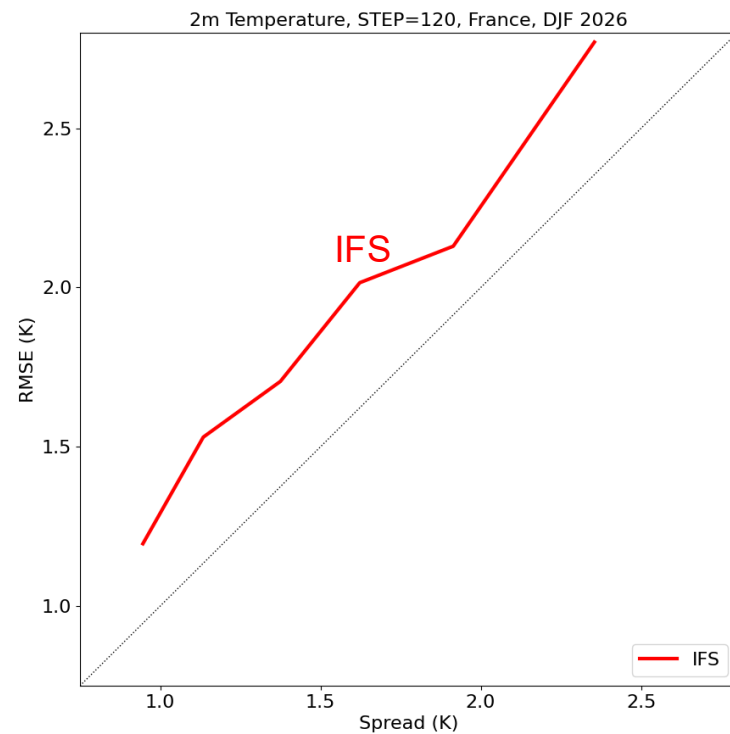
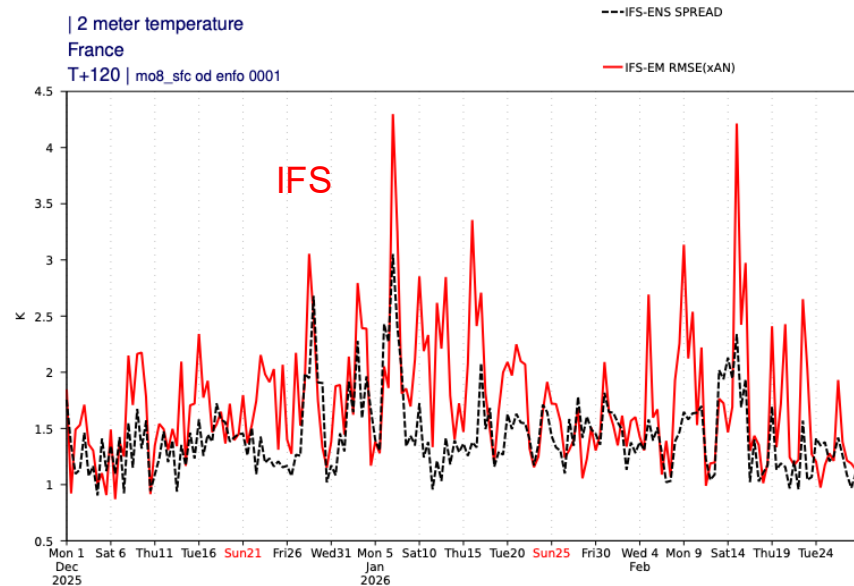


- Comparison based on TIGGE, i.e. 0.5 deg (not on original model grid)
- DWD leading up to Day 4
- Advantage mostly due to 'adaptive parameter tuning'

ENS skill in predicting temperature anomalies (Day 4)

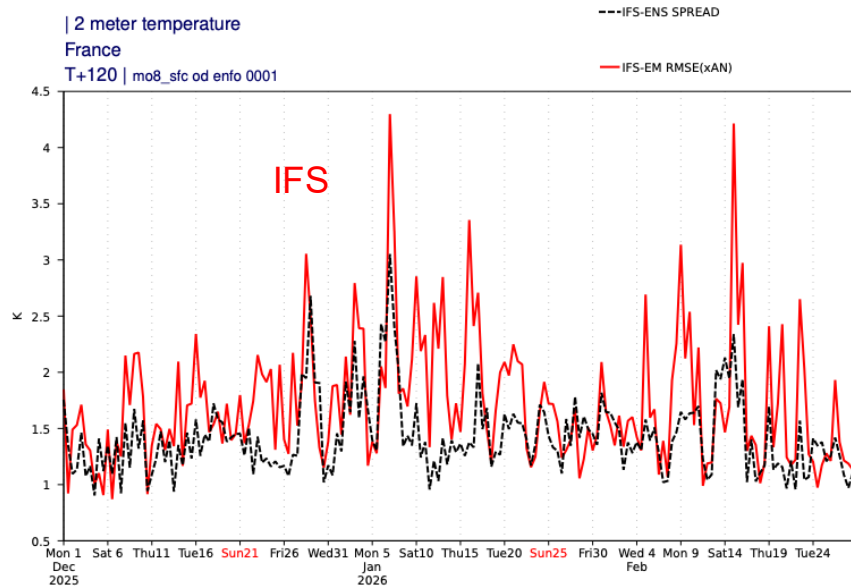


T2m spread reliability (Day 5)

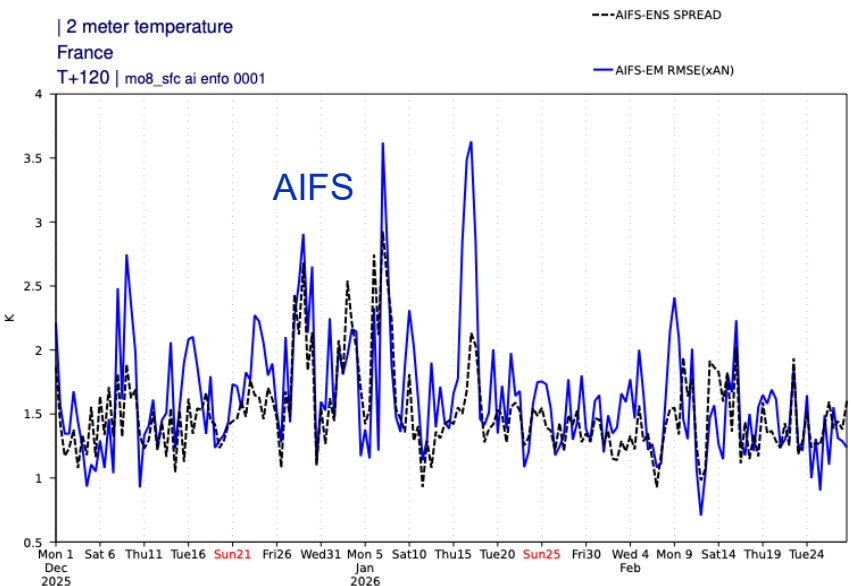
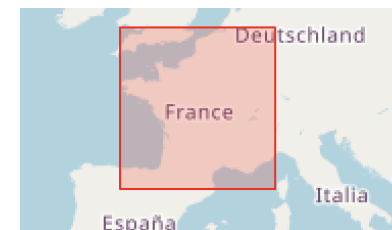
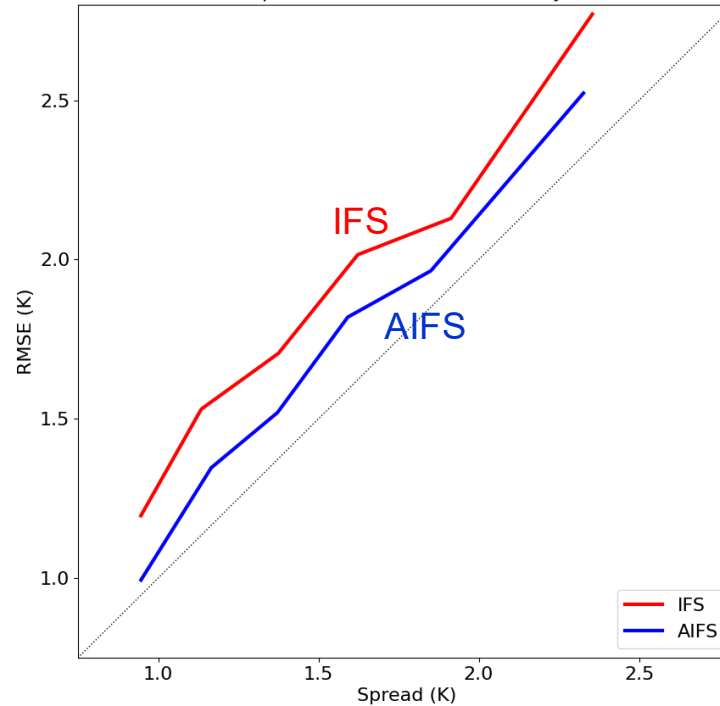


- Spread useful predictor of day-to-day RMSE changes averaged over large area (against analysis)

T2m spread reliability (Day 5)



2m Temperature, STEP=120, France, DJF 2026



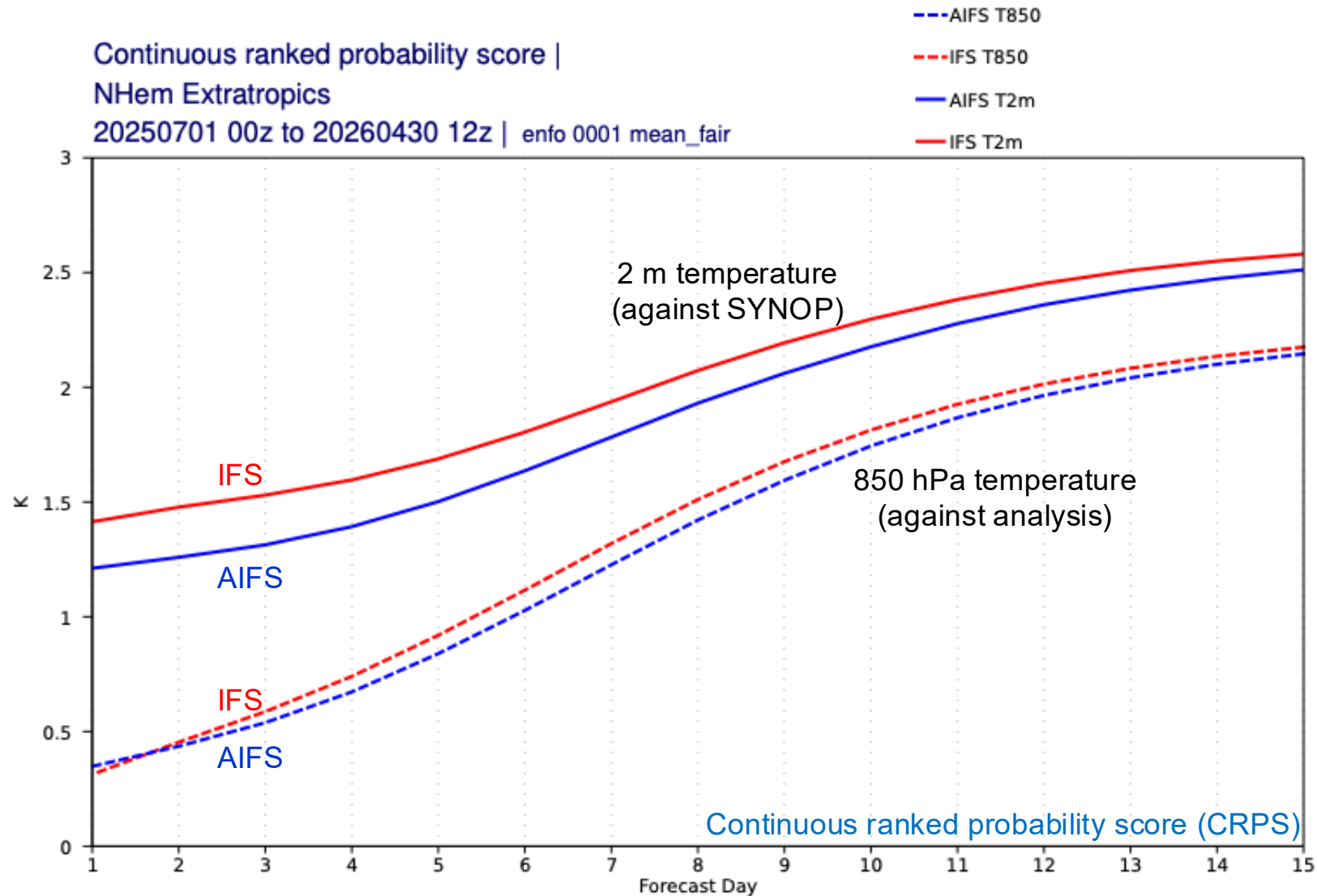
- Spread useful predictor of day-to-day RMSE changes averaged over large area (against analysis)
- AIFS similar, with closer match of spread and RMSE

Summary

- IFS-CF T2m forecast has **improved about 1% per year** on average
- **ECMWF currently not leading** in T2m in early medium-range
- **AIFS** provides additional **10-15% skill** relative to IFS
- **Stable boundary layer** has large errors (e.g. northern Scandinavia in winter)
- With 49r1 (2024), the number of **large T2m errors** in IFS-ENS has further **decreased**
- Prediction of **cold anomalies** by IFS-ENS has **improved**
- Major improvement expected from **51r1 (2027)**

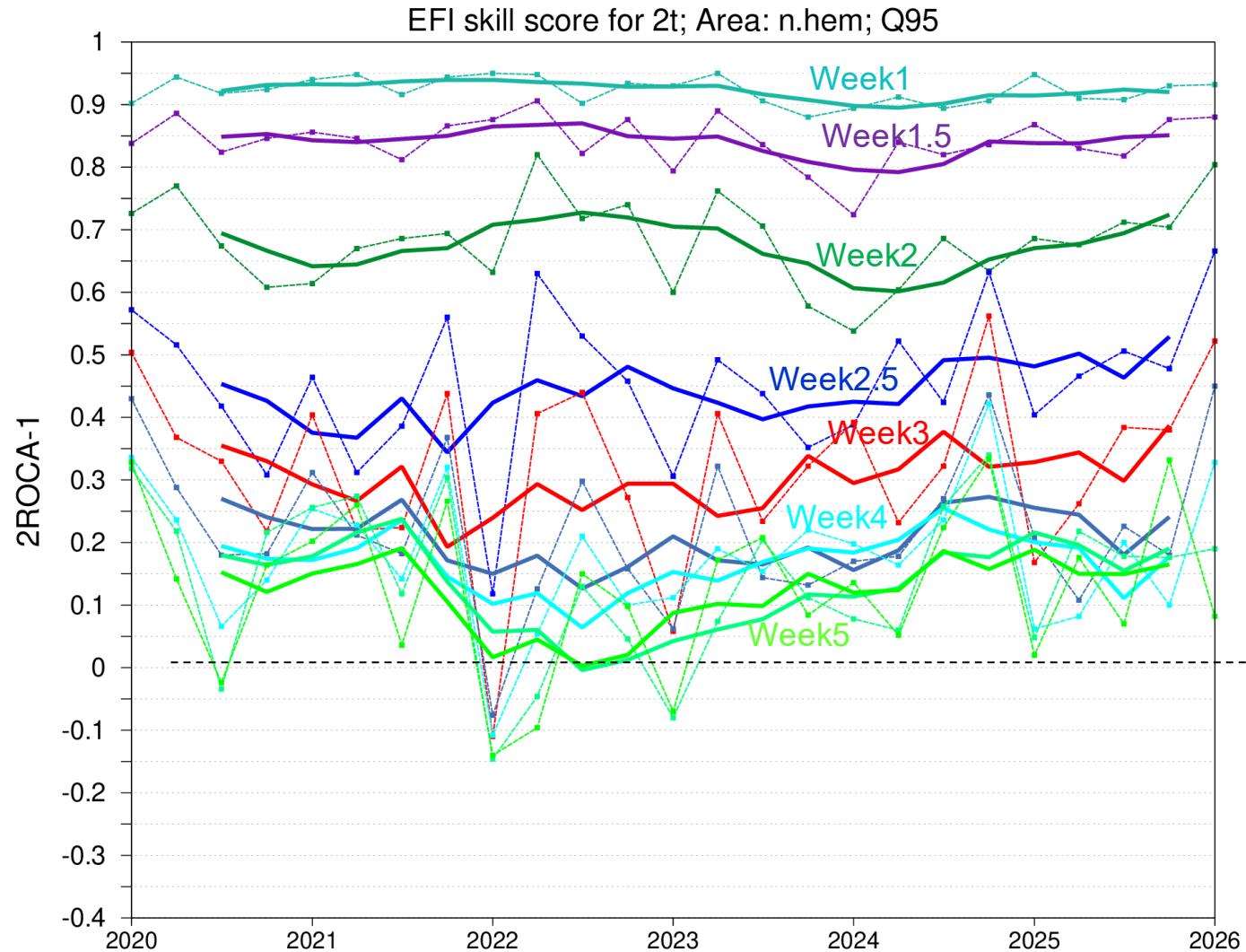
Thank you

AIFS-ENS (operational since 1 July 2025) vs IFS-ENS



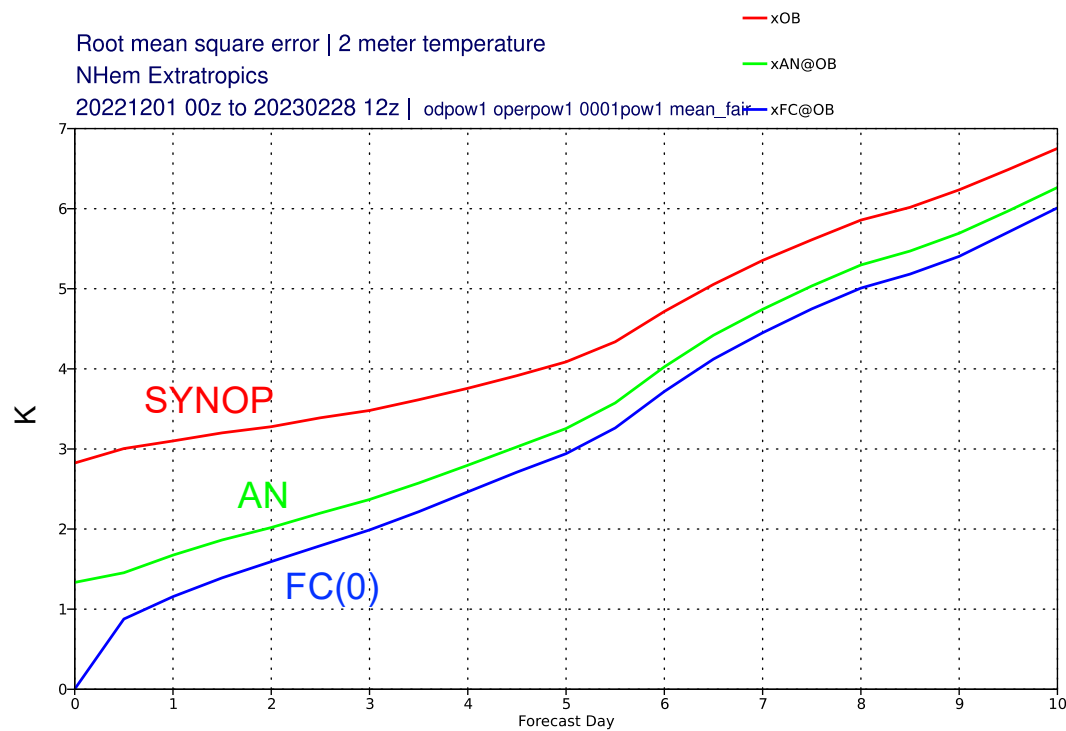
- T2m errors reduced up to 15%
- T850 errors reduced up to 10%
- In the short range, T2m improved without T850 being improved
- AIFS has seen the IFS land surface analysis in the training

EFI skill for T2m in the sub-seasonal range (Q95)

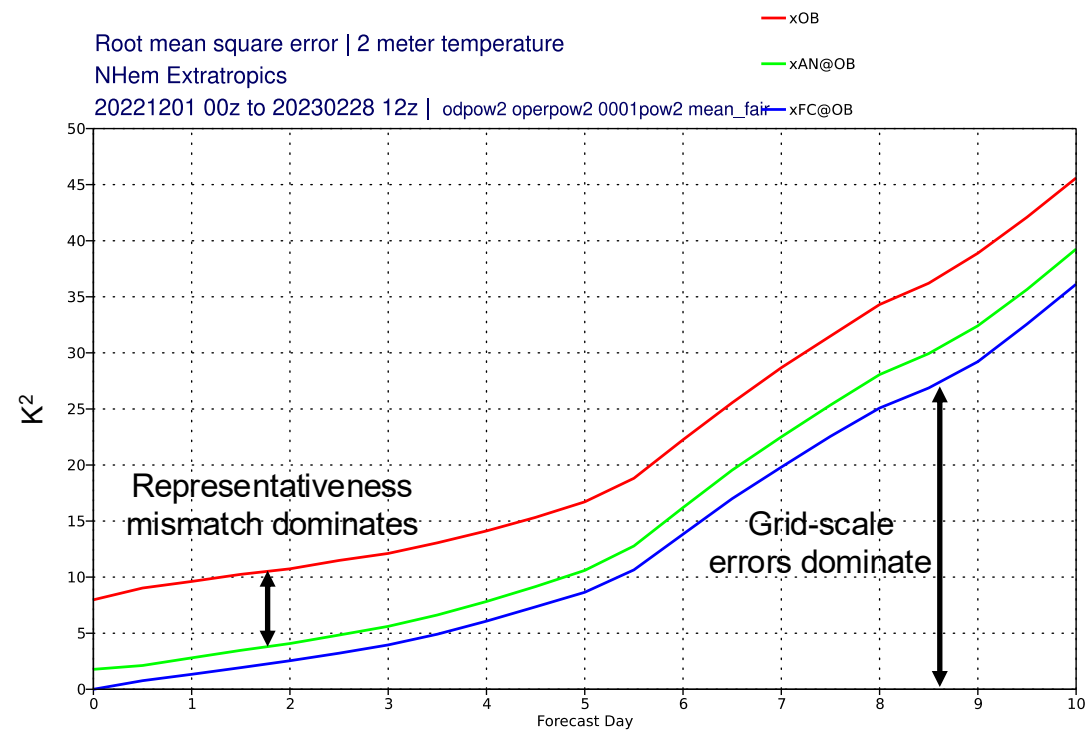


- Positive skill up to Week 5
- No clear trend since 2020
- Largest skill drop from Week 2 to 2.5

2m temperature RMSE verified against OBS, AN, and FC(0)



RMSE



MSE

Red = verification against SYNOP stations

Green = verification against T2m analysis (at the same stations)

Blue = verification against T2m forecast at STEP=0 (at the same stations)