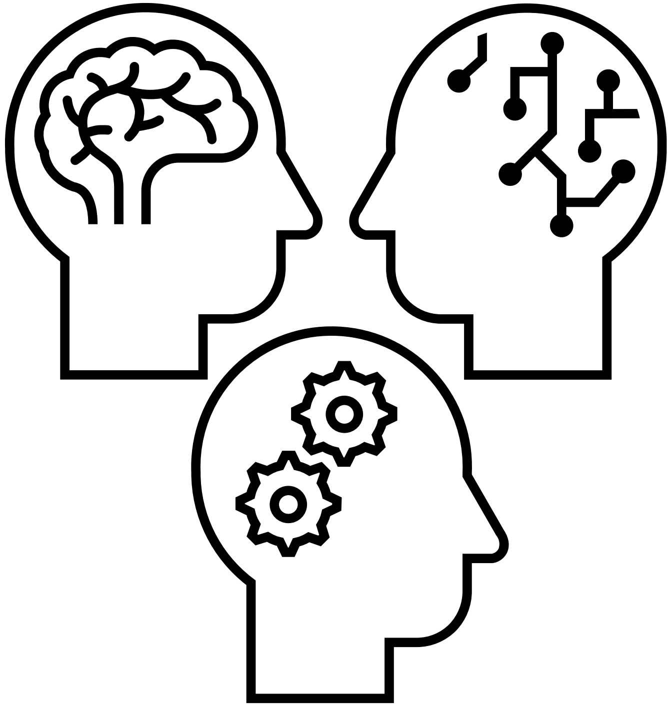
Machine Learning in NWP: Forecast evaluation

Linus Magnusson and colleagues at ECMWF



ECMWF

Think back a couple of decades:

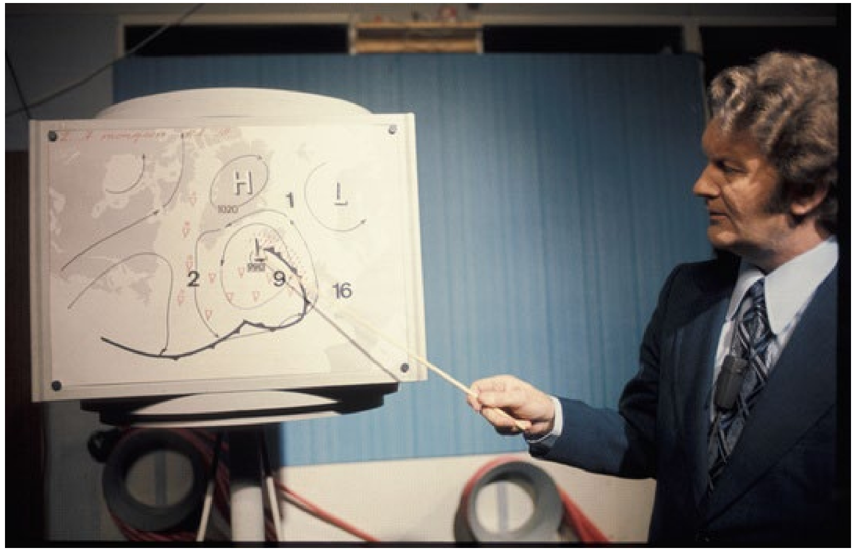
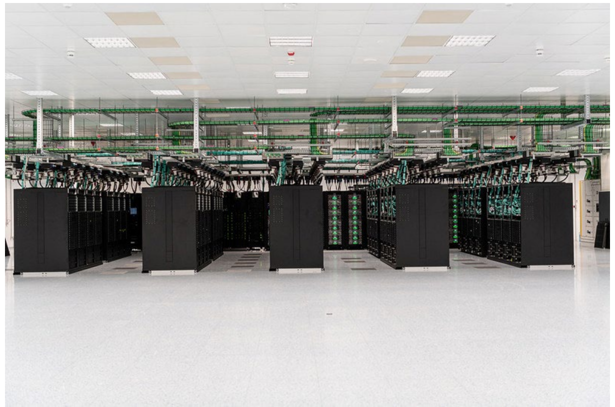
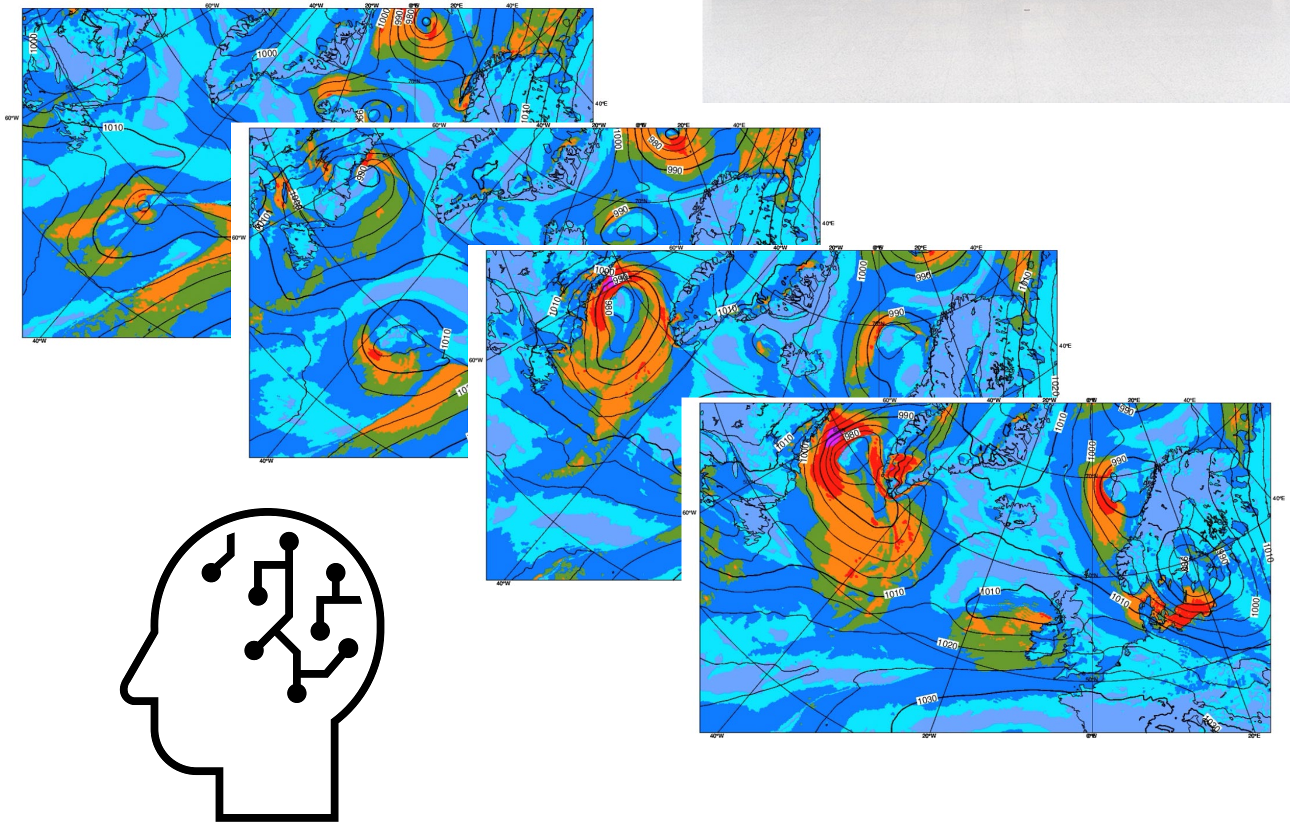


Photo credits: Icelandic Met service

With 40+ years of reanalysis, what if a computer could make forecasts based on the past experiences?





And what we are used to now:

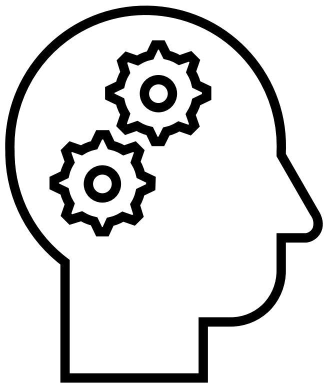
$$\frac{D\mathbf{V}}{Dt}+f\mathbf{k}\times\mathbf{V}=-\nabla\Phi$$

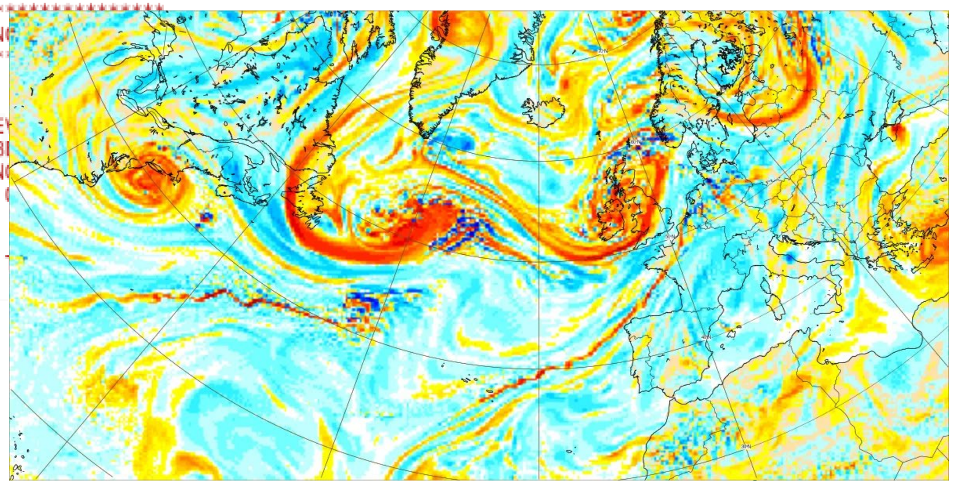
$$\frac{\partial\Phi}{\partial p}=-\alpha=-\frac{RT}{p}$$

$$\nabla\cdot\mathbf{V}+\frac{\partial\omega}{\partial p}=0$$

$$\left(\frac{\partial}{\partial t}+\mathbf{V}\cdot\nabla\right)T-S_p\omega=\frac{J}{c_p}$$

```
SUBROUTINE CALLPAR(YDGEOMETRY,YDVARS,YDSURF,YDMODEL,KDIM,&
 !-----------------------------------------------------------------
 & PAUX, PRAD, FLUX, PDIAG, PSURF, PCGPP, PCREC, PAG, PRECO, PDDHS, AUXL, SURFL, LLKEYS, PERTL, &
 ! - Model variables (t)
 & PGFL,&
 & PPERT, PSLPHY9,&
 ! - UPDATED TENDENCY
 & PTENGFL,&
 & PHYS_MWAVE,&
 ! Stored quantities
 & PSAVTEND, PGFLSLP,&
 & STATE_T0, TENDENCY_DYN, TENDENCY_CML, STATE_TMP, TENDENCY_TMP,&
 & TENDENCY_VDF, TENDENCY_SATADJ, TENDENCY_LOC, TENDENCY_PHY&
 & )

!**** *CALLPAR * - CALL ECMWF PHYSICS

!     PURPOSE.
!     --------
!     - CALL THE SUBROUTINES OF THE E.C.M.W.F. PHYSICS PACKAGE.

!     ***********************************************************************
!     ****** IDIOSYNCRASIES *** IDIOSYNCRASIES *** IDIOSYNC
!     ****************************************************
!     ***  HEALTH WARNING:
!     ***  ===============
!     ***  NOTE THAT WITHIN THE E.C.M.W.F. PHYSICS HALF-LEV
!     ***  ARE INDEXED FROM 1 TO NFLEVG+1 WHILE THEY ARE BE
!     ***  0 AND NFLEVG IN THE REST OF THE MODEL. THE CHANG
!     ***  CARE OF IN THE CALL TO THE VARIOUS SUBROUTINES O
!     ***  PHYSICS PACKAGE
!     ***
!     ***    THIS IS SUPPOSED TO BE A "TEMPORARY" FEATURE T
!     ***    STRAIGHTENED OUT IN THE "NEAR" FUTURE
```





Physics-based forecast systems using IFS model

Medium-range

- Atmosphere: 9 km
- 15 days lead time
- 50 ensemble members + 1 unperturbed control forecasts
- Distributed around 8 hours after initialisation time*

Sub-seasonal

- Atmosphere: 36 km
- 6 weeks lead time
- 101 ensemble members
- Initialised every day at 00UTC
- Distributed in the evening (UTC time)

Seasonal (SEAS5)

- Atmosphere: 36 km
- 7 months lead time
- 51 ensemble members
- Initialised 1st every month
- Distributed 5th every month

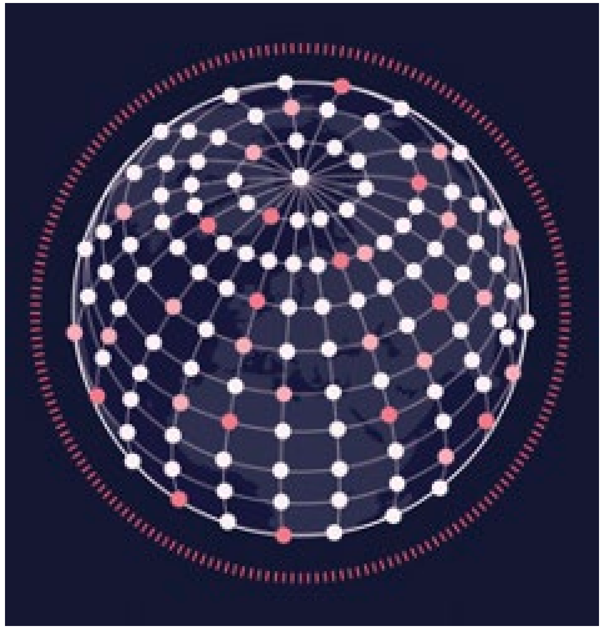
ECMWF

* In the assimilation system, it has used observations +3h after the labelled initialisation time

Machine-learning based forecast systems using AIFS model

Medium-range deterministic AIFS-single v1.1

- Atmosphere: AIFS Model v1.0 using **MAE loss**, N320 (0.25°, ~25 km)
- 15 days lead time
- Initialised from physics-based analysis



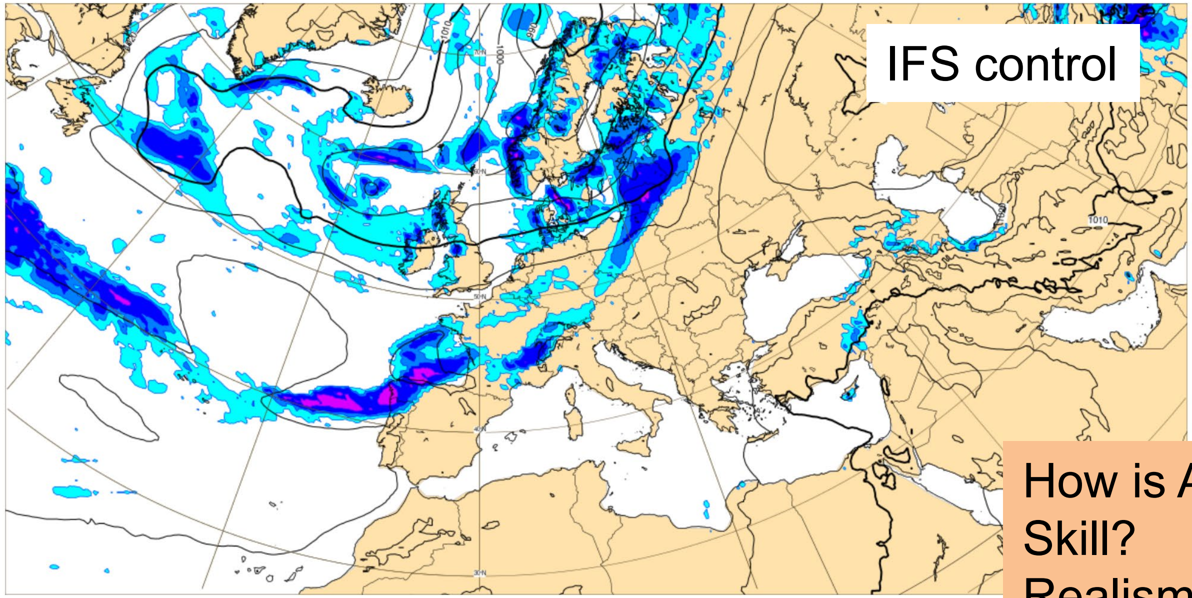
Medium-range ensemble AIFS-ENS v1.0

- Atmosphere: AIFS Model v1.0 using **CRPS loss**, N320 (0.25°, ~25 km), 13 vertical levels
- 15 days lead time
- 50 ensemble members + 1 unperturbed** control forecasts
- Same initial conditions including perturbations as physics-based ensemble

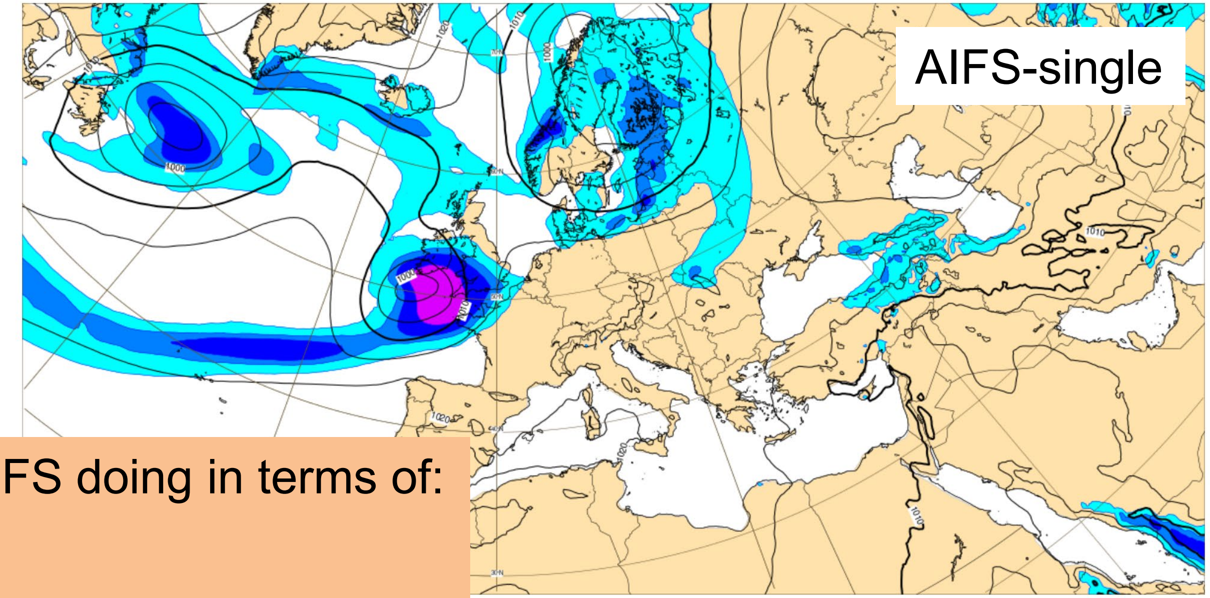
* In the assimilation system, it has used observations +3h after the labelled initialisation time
** With the use of CRPS-loss, also the control forecast include the “noise” to simulate model uncertainty

So here are forecast (MSLP and precipitation) from a week ago:

Base time: Wed 10 Sep 2025 00 UTC Valid time: Wed 17 Sep 2025 12 UTC (+180h) Area : Europe Interval (hr) : 6

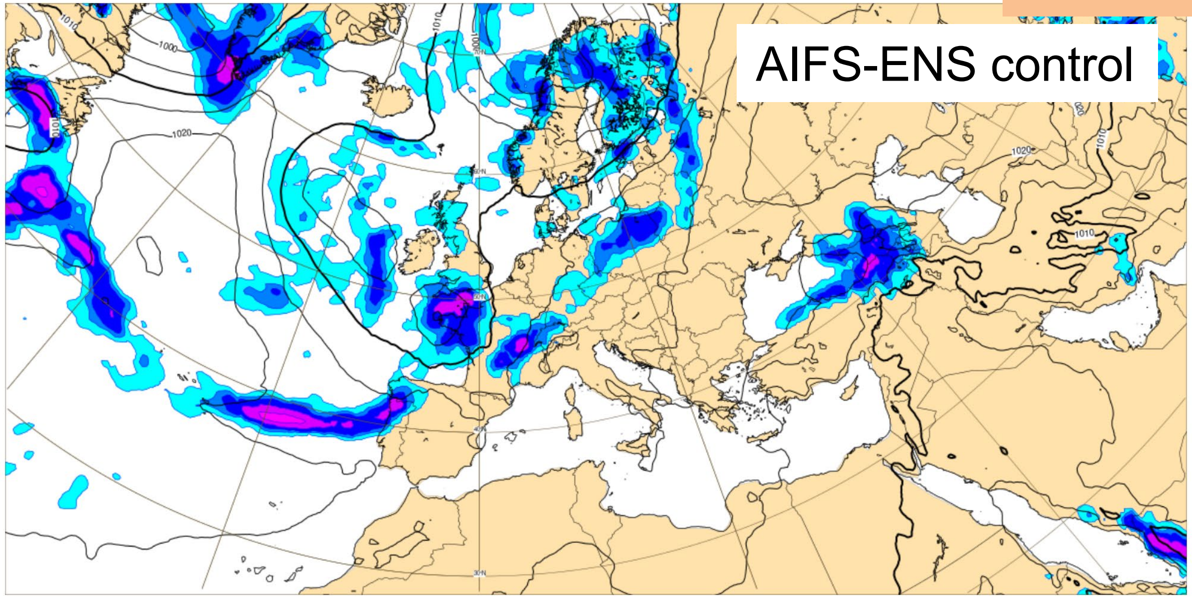


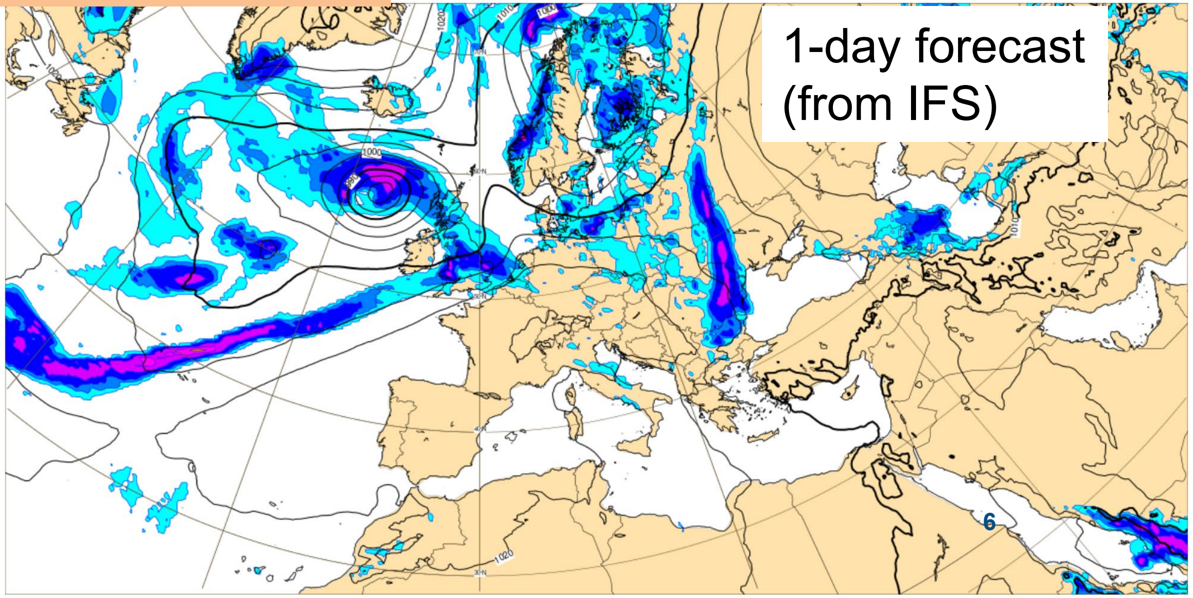
Base time: Wed 10 Sep 2025 00 UTC Valid time: Wed 17 Sep 2025 12 UTC (+180h) Area : Europe Interval (hr) : 6



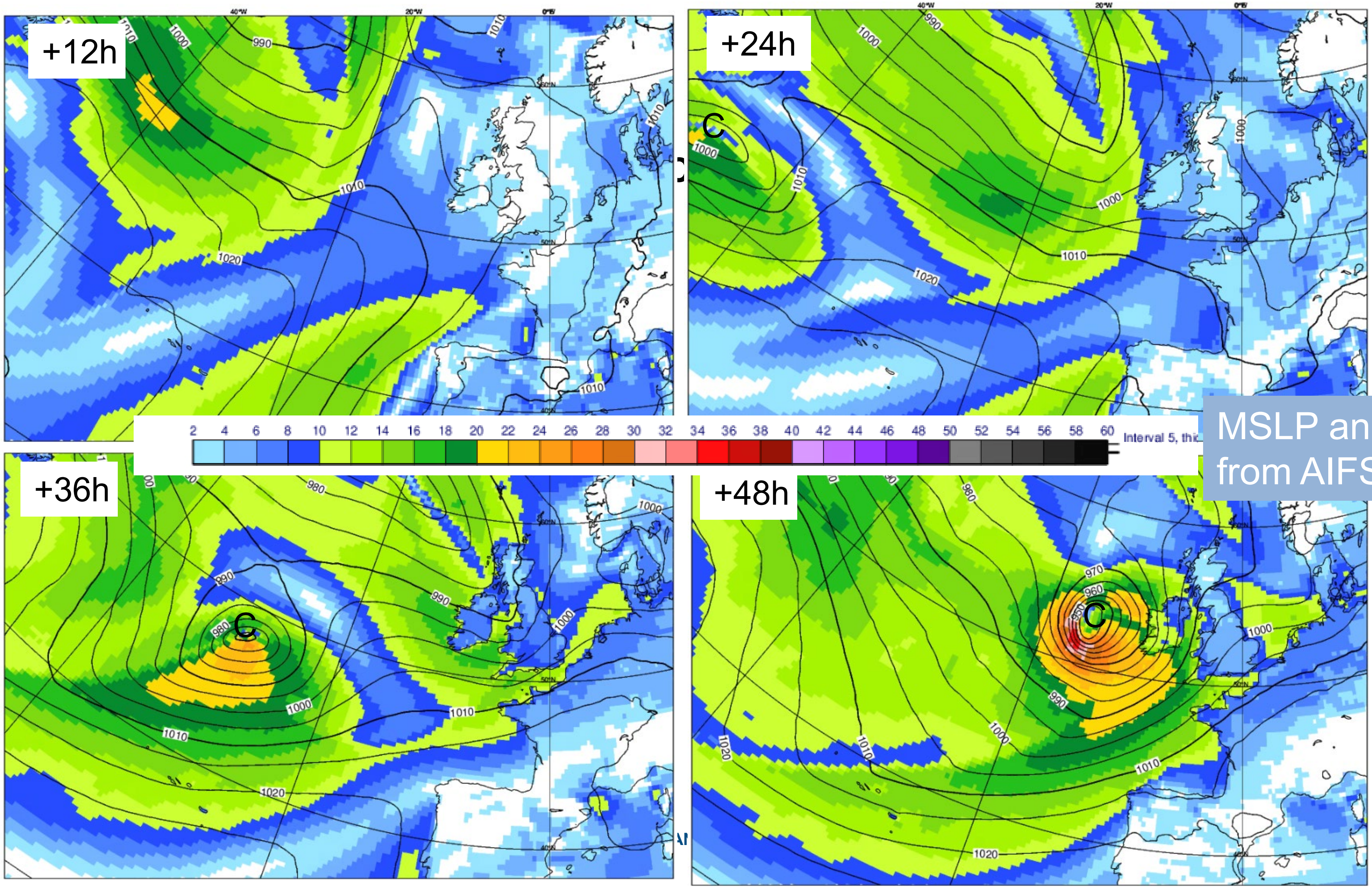
How is AIFS doing in terms of:
Skill?
Realism?
Extremes?

Base time: Wed 10 Sep 2025 00 UTC Valid time: Wed 17 Sep 2025 12 UTC (+180h) Area : Europe Interval (hr) : 6





AIFS-single: Storm Eowyn (forecast from 22 Jan 2025 00UTC)



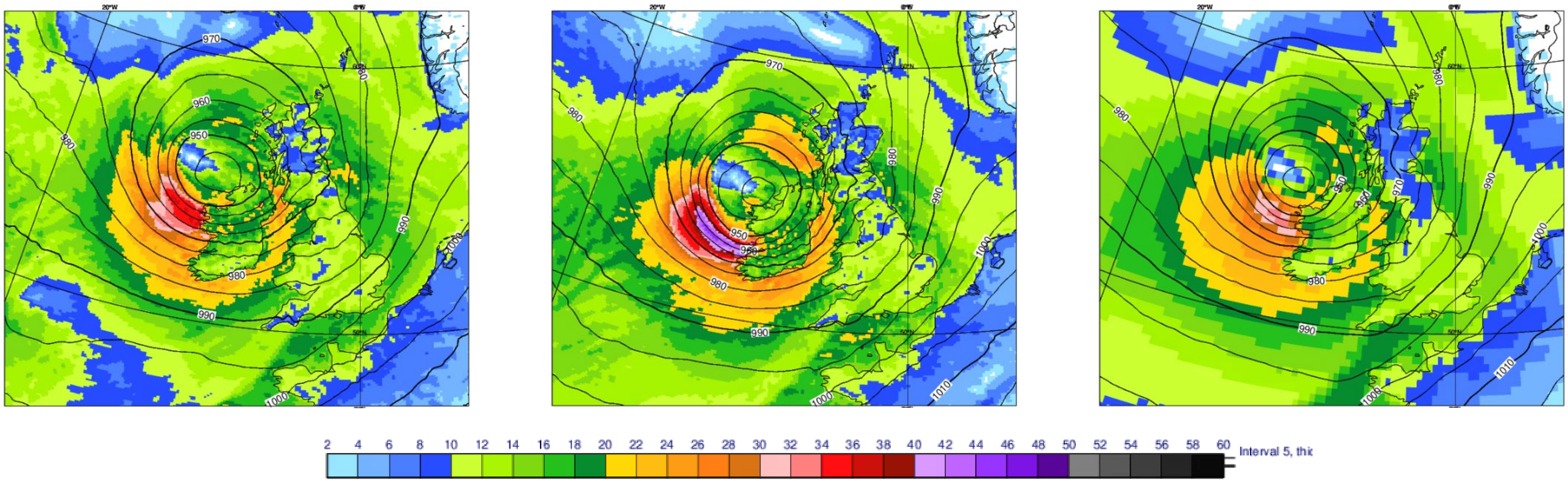
MSLP and wind speed from AIFS model

Storm Eowyn (2-day forecasts valid 24 Jan 2025 06UTC)

Analysis

IFS CF

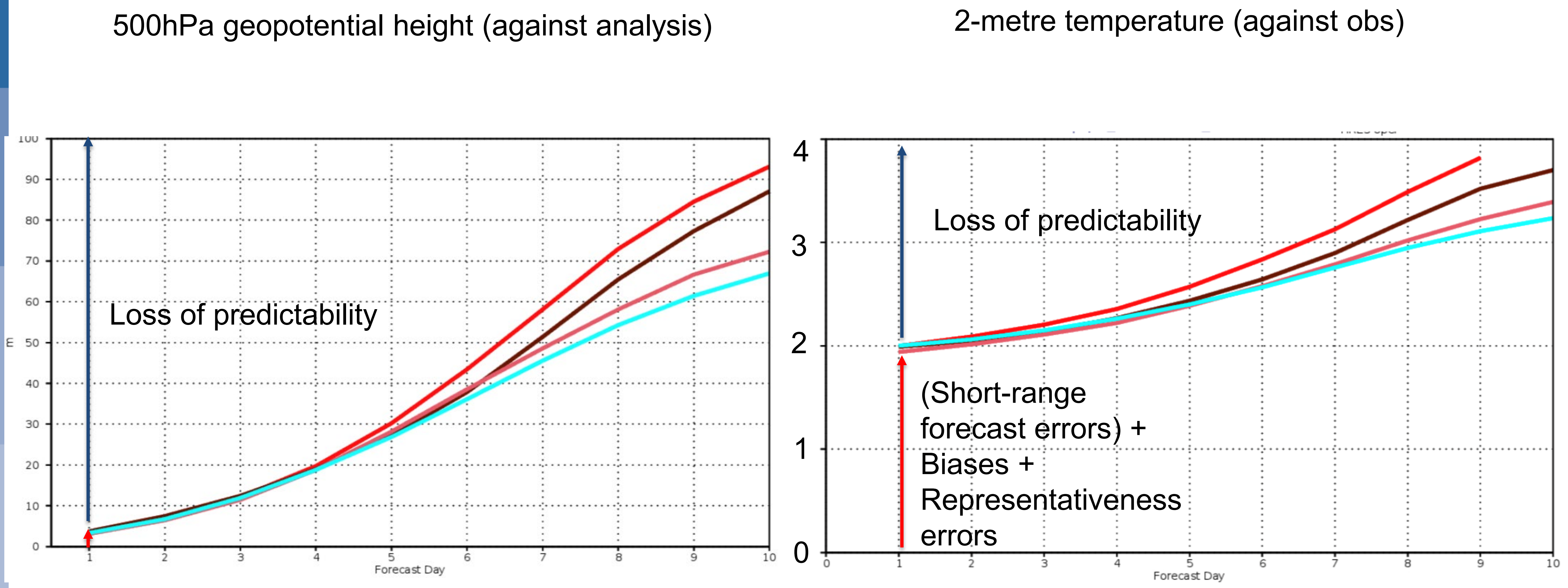
AIFS-single v1.0



- Better position forecast in of the maximum wind in AIFS
- Similar minimum pressure 935-940hPa
- Less extreme wind speed in ML models 46 m/s vs. 33 m/s

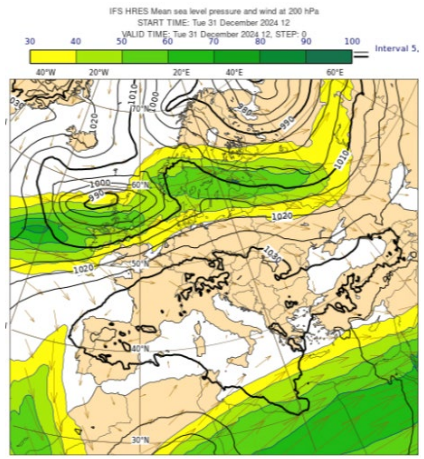
See ECMWF Newsletter 183

RMSE over Europe - 1 July to 1 September

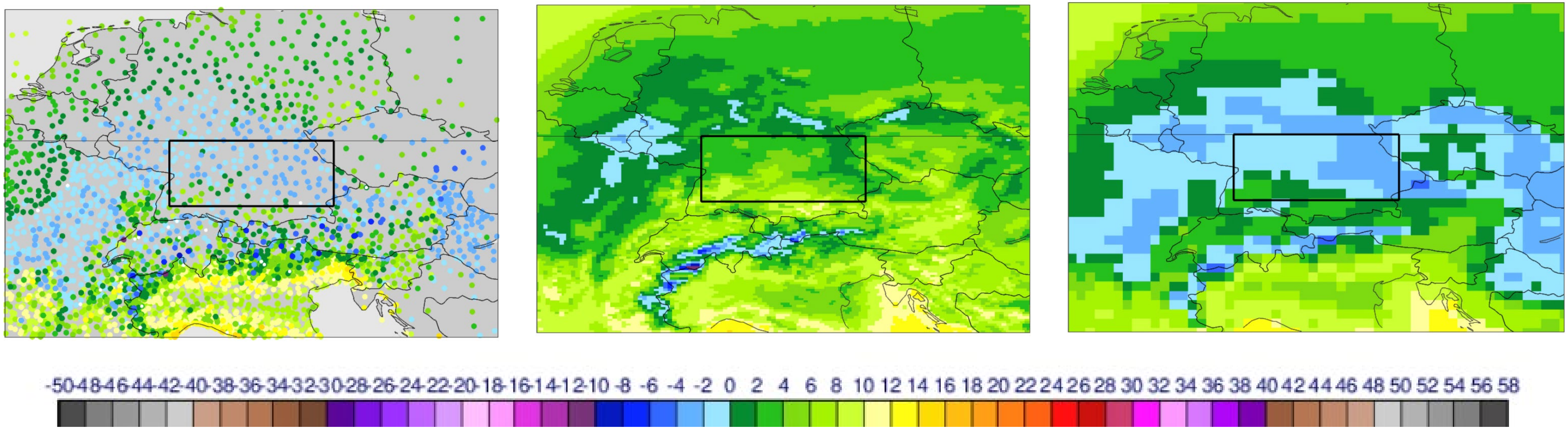


IFS control – red
AIFS-single - brown
IFS-ENS ensemble mean - pink
AIFS-ENS ensemble mean - cyan

Conditional systematic errors:
Example valid 31 December 12UTC: 2-metre temperature



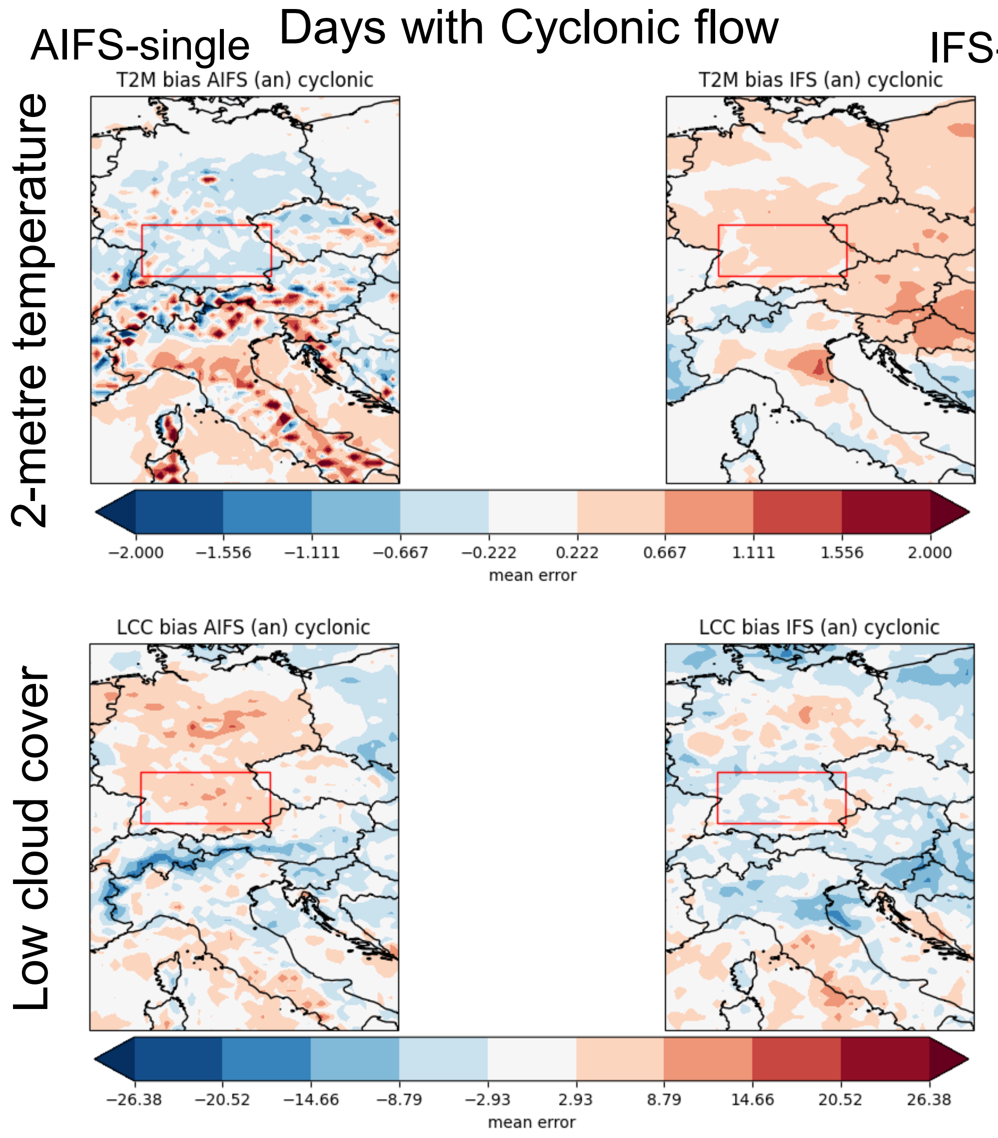
Observations | 48-hour IFS Control | 48-hour AIFS-single



Karmouche et al. (2025):
AIFS blog: https://www.ecmwf.int/en/about/media-centre/aifs-blog/2025/verifying-2-m-temperature-forecasts-wintertime-anticyclonic

Conditional verification based on flow pattern, Winter 2024/2025

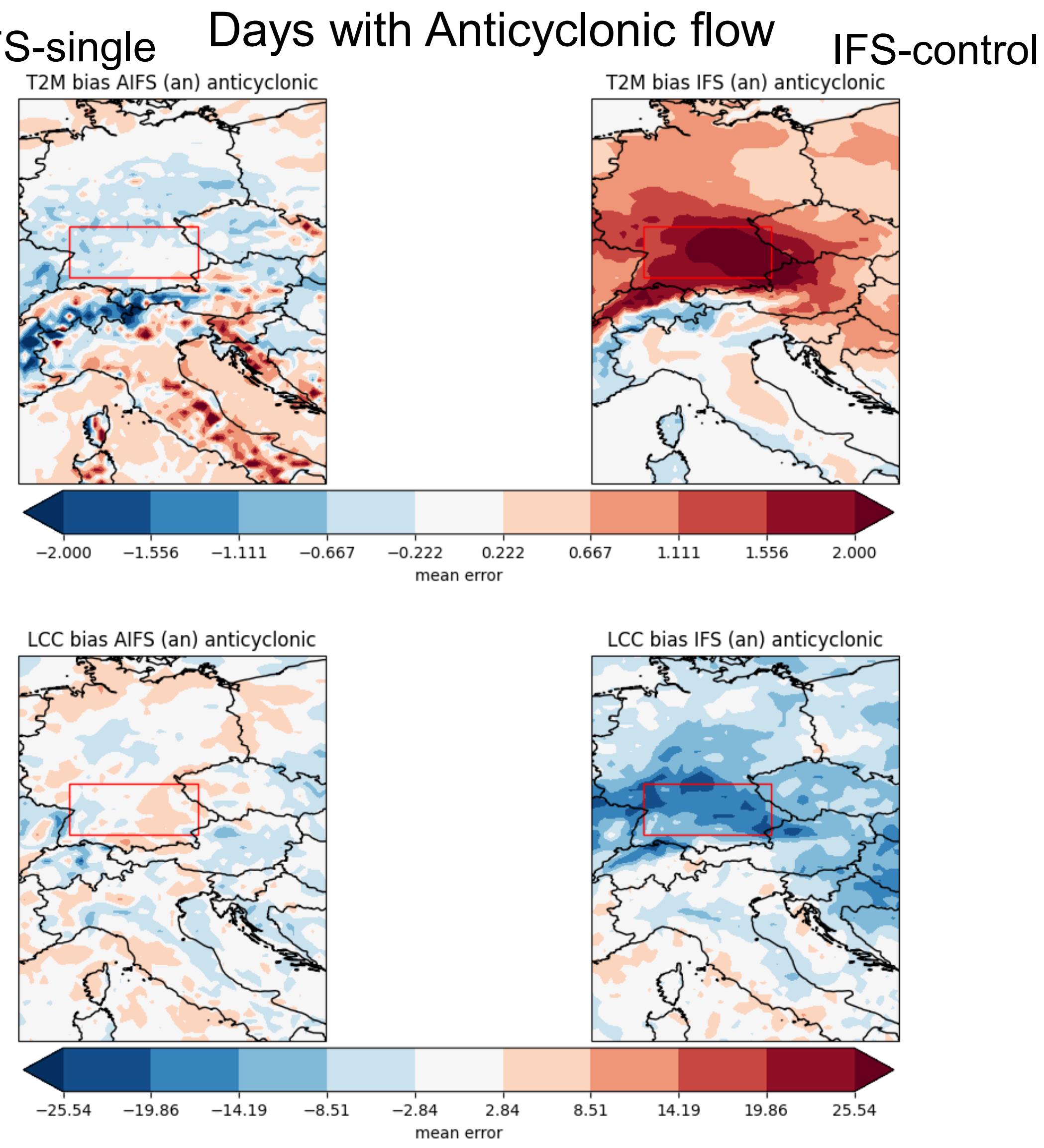
(verified against ECMWF analysis)



Bias, step +36h, validity dates (2024-10-02 12UTC - 2025-02-25 12UTC)

Flow conditions based on CURV index developed by Nigel Roberts, ECMWF

Thanks to Soufiane Karmouche, ECMWF

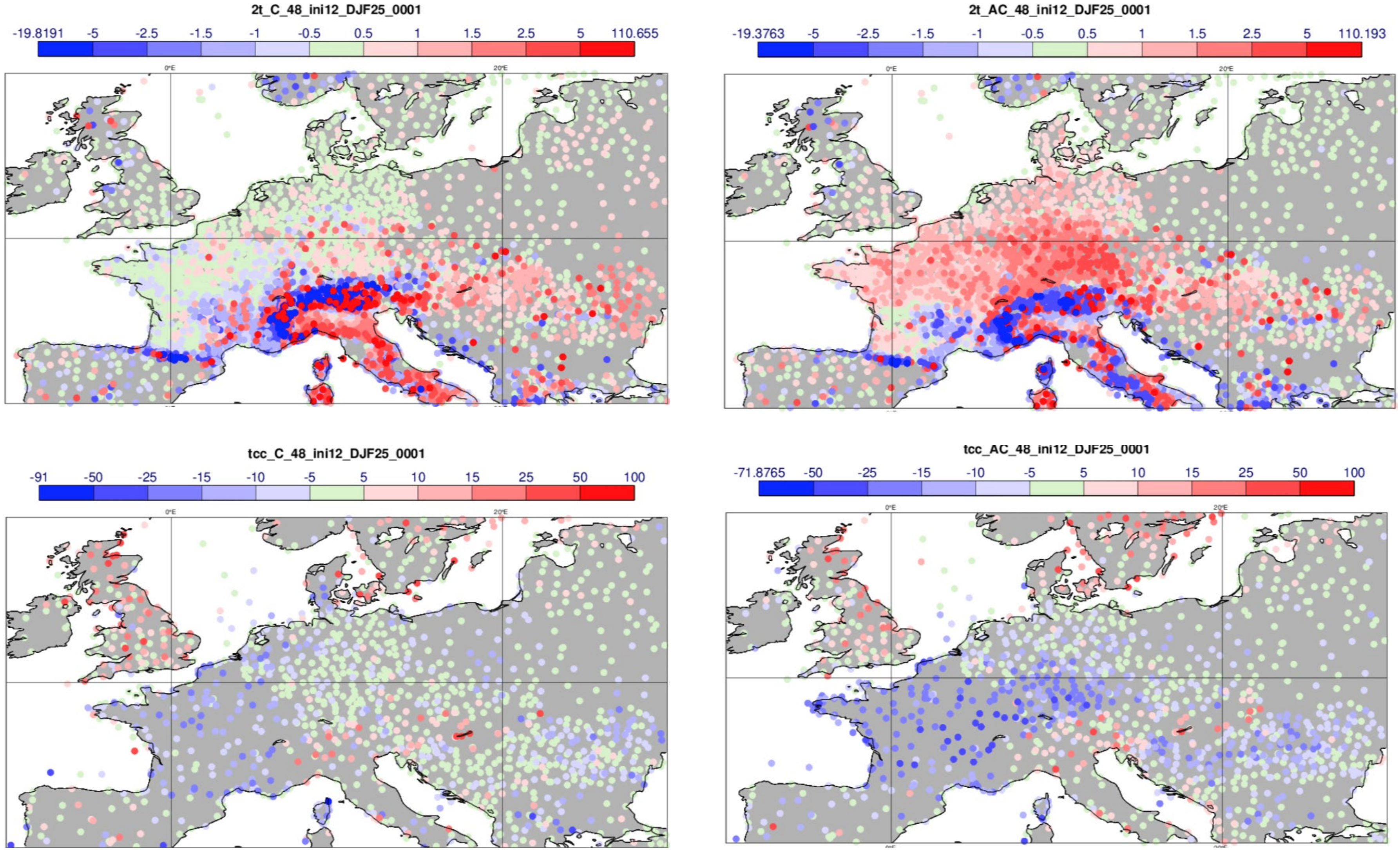


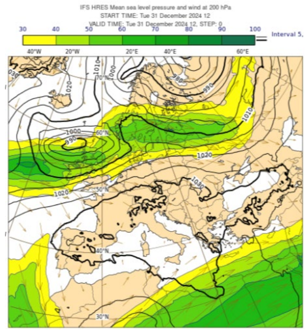
IFS Bias based on SYNOP observations winter 2024/25 (12UTC + 48h)

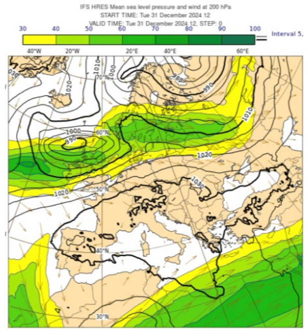
Cyclonic | Anticyclonic

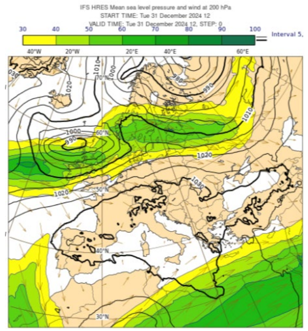
2-metre temperature

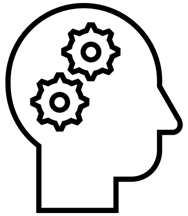
Total cloud cover

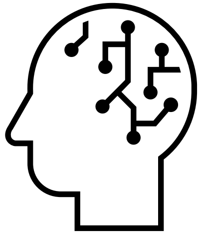




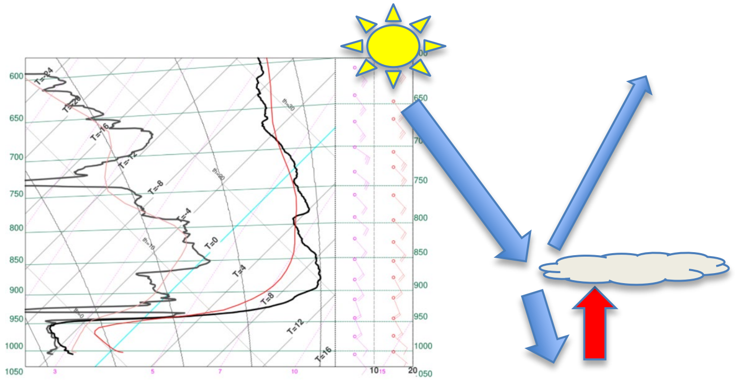


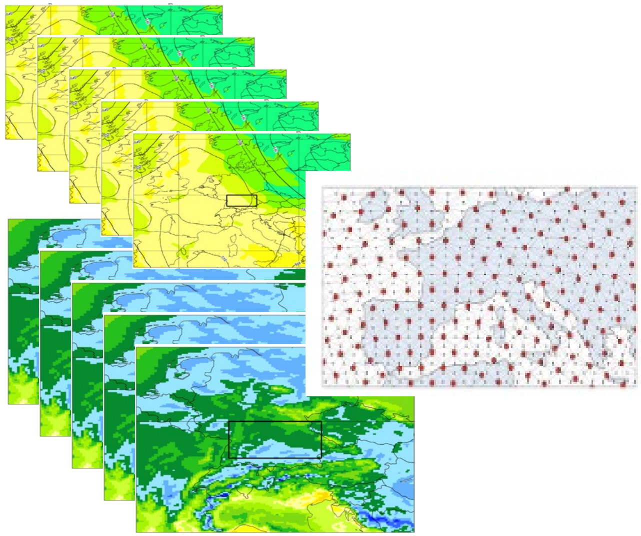


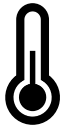


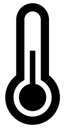


Ω + winter







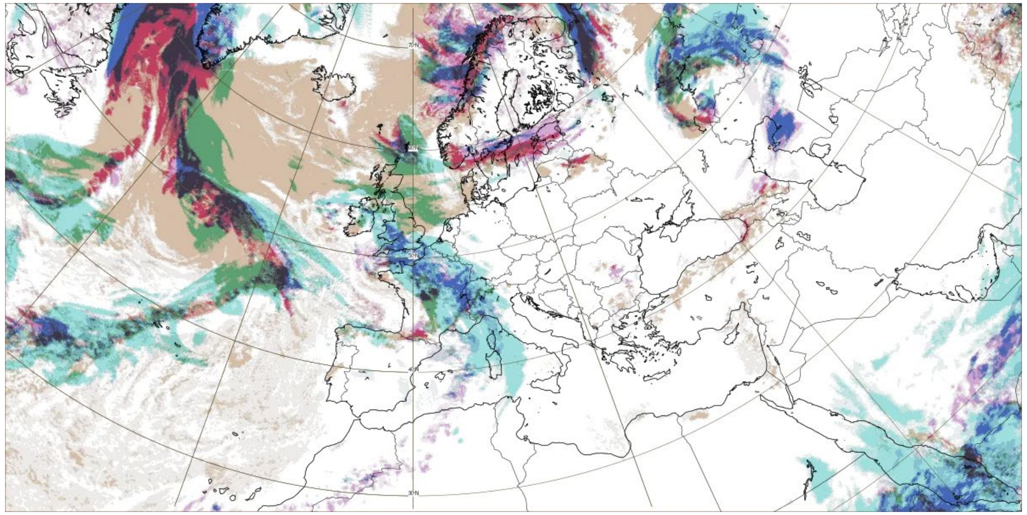


But not everything is perfect (yet) in the AI model world…

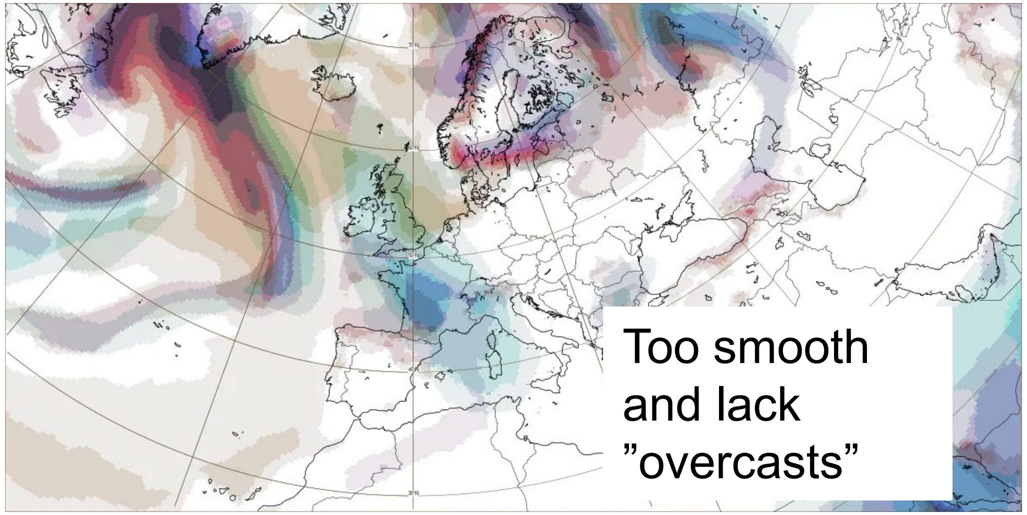
Remember that AIFS is still a very new model…

Example of cloud forecasts, 14 August 2025 00UTC+144h

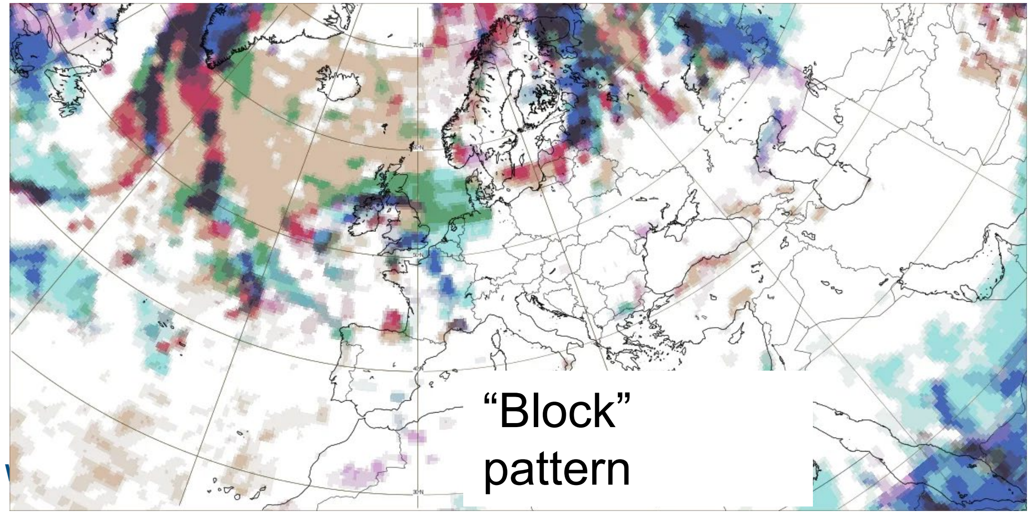
IFS CF



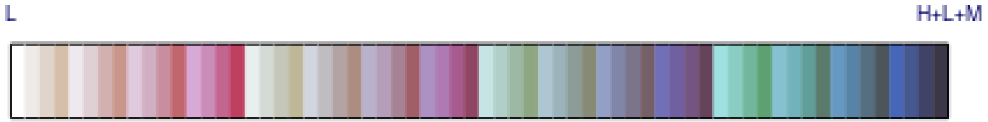
AIFS-single



AIFS-ENS control



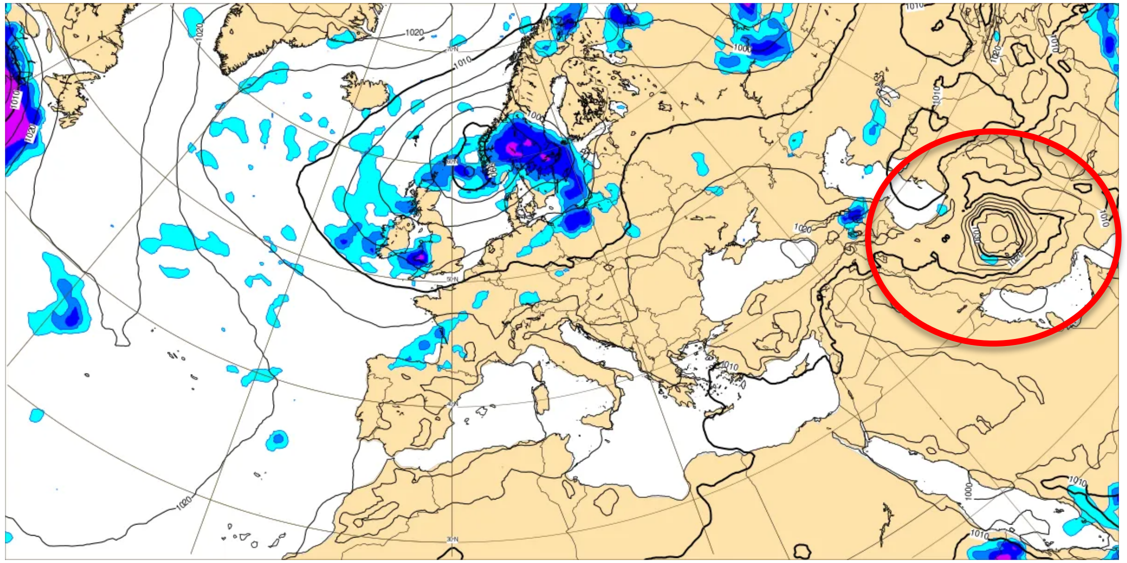
Total cloud cover



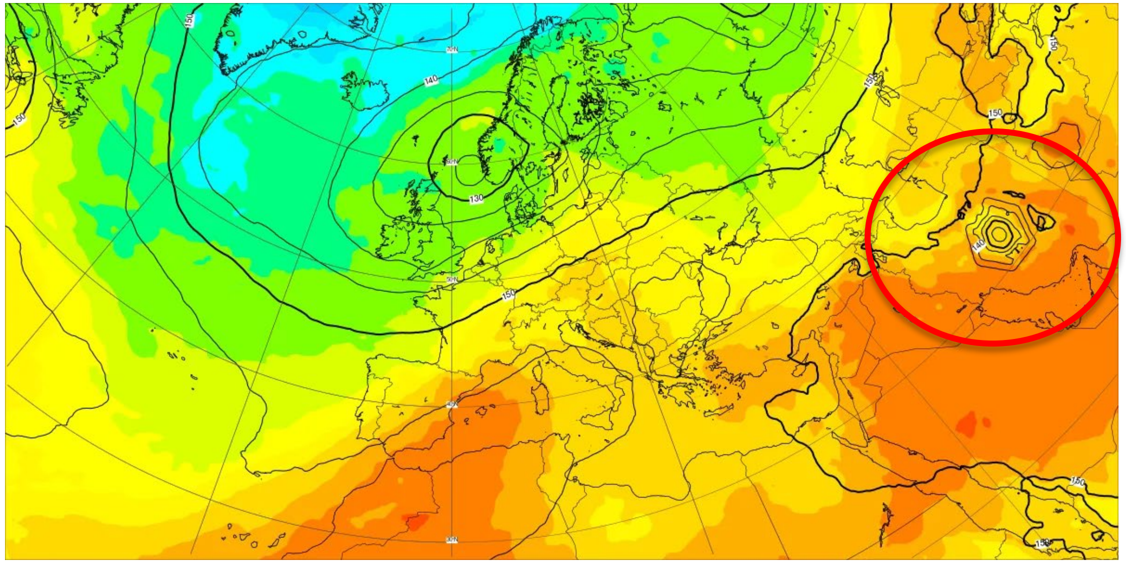
Problem over "warm" orography in AIFS-ENS

Day 15 forecast from AIFS-ENS control

MSLP and precipitation



850hPa temperature and geopotential



Seems to develop after day 10

Spurious snowfall in AIFS-ENS

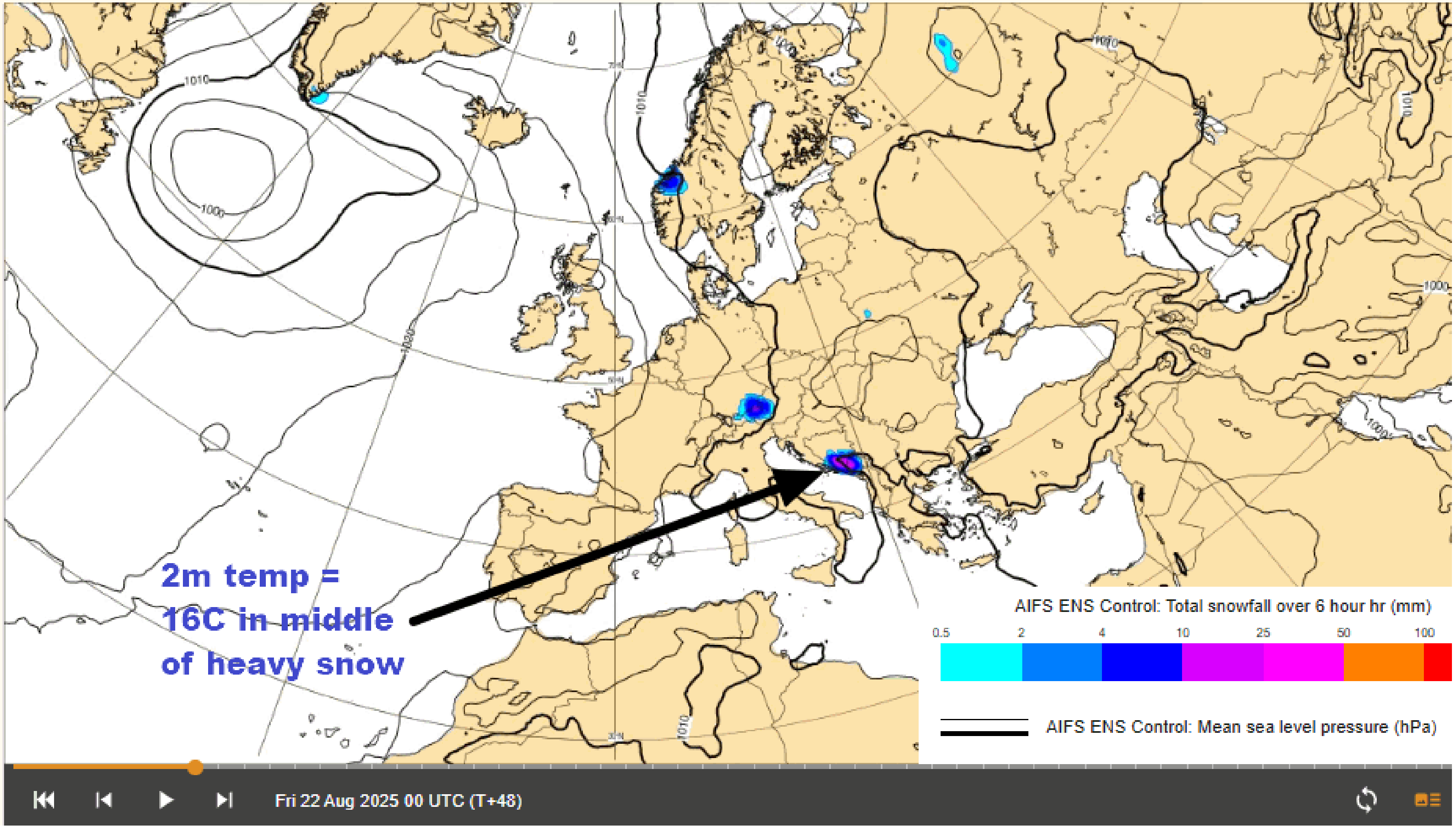
AIFS ENS Control: Total snowfall during last 6 hours

AIFS ENS

Base time
Wed 20 Aug 2025 00 UTC

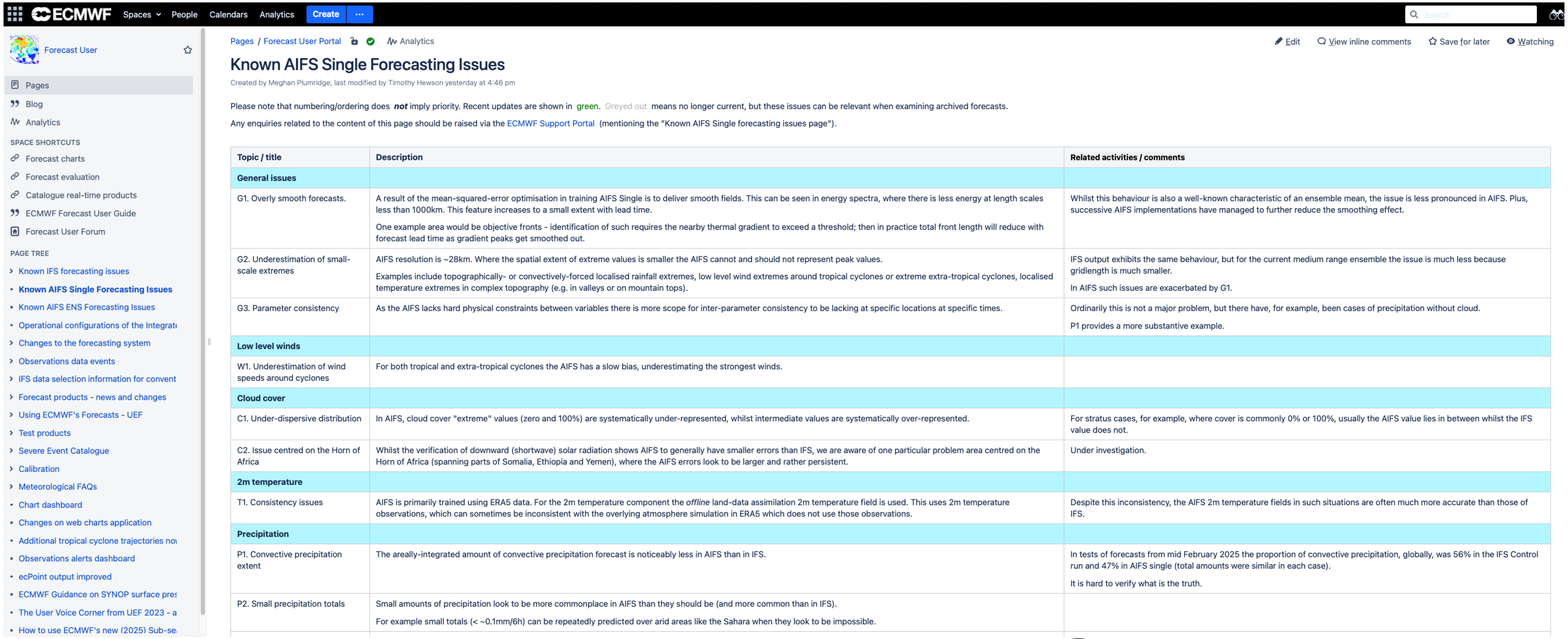
Valid time
Fri 22 Aug 2025 00 UTC (T+48)

Area
Europe



Known forecast issues page for AIFS-single / AIFS-ENS

(in a similar way as we document issues in IFS)


ECMWF Spaces People Calendars Analytics Create ···

Forecast User

Pages
Blog
Analytics

SPACE SHORTCUTS
Forecast charts
Forecast evaluation
Catalogue real-time products
ECMWF Forecast User Guide
Forecast User Forum

PAGE TREE
- Known IFS forecasting issues
- Known AIFS Single Forecasting Issues
- Known AIFS ENS Forecasting Issues
- Operational configurations of the Integrat
- Changes to the forecasting system
- Observations data events
- IFS data selection information for convent
- Forecast products - news and changes
- Using ECMWF's Forecasts - UEF
- Test products
- Severe Event Catalogue
- Calibration
- Meteorological FAQs
- Chart dashboard
- Changes on web charts application
- Additional tropical cyclone trajectories no
- Observations alerts dashboard
- ecPoint output improved
- ECMWF Guidance on SYNOP surface pres
- The User Voice Corner from UEF 2023 - a
- How to use ECMWF's new (2025) Sub-se

Pages / Forecast User Portal Analytics

Edit View inline comments Save for later Watching

Known AIFS Single Forecasting Issues

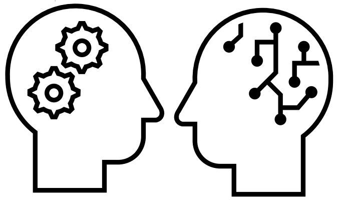
Created by Meghan Plumridge, last modified by Timothy Hewson yesterday at 4:46 pm

Please note that numbering/ordering does *not* imply priority. Recent updates are shown in green. Greyed out means no longer current, but these issues can be relevant when examining archived forecasts.

Any enquiries related to the content of this page should be raised via the ECMWF Support Portal (mentioning the "Known AIFS Single forecasting issues page").

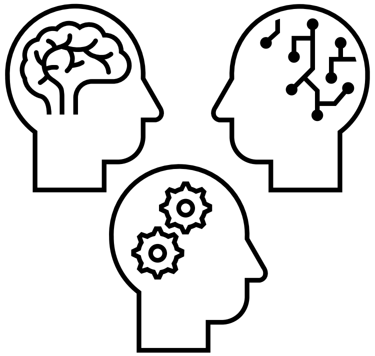
Topic / title	Description	Related activities / comments
General issues		
G1. Overly smooth forecasts.	A result of the mean-squared-error optimisation in training AIFS Single is to deliver smooth fields. This can be seen in energy spectra, where there is less energy at length scales less than 1000km. This feature increases to a small extent with lead time. One example area would be objective fronts - identification of such requires the nearby thermal gradient to exceed a threshold; then in practice total front length will reduce with forecast lead time as gradient peaks get smoothed out.	Whilst this behaviour is also a well-known characteristic of an ensemble mean, the issue is less pronounced in AIFS. Plus, successive AIFS implementations have managed to further reduce the smoothing effect.
G2. Underestimation of small-scale extremes	AIFS resolution is ~28km. Where the spatial extent of extreme values is smaller the AIFS cannot and should not represent peak values. Examples include topographically- or convectively-forced localised rainfall extremes, low level wind extremes around tropical cyclones or extreme extra-tropical cyclones, localised temperature extremes in complex topography (e.g. in valleys or on mountain tops).	IFS output exhibits the same behaviour, but for the current medium range ensemble the issue is much less because gridlength is much smaller. In AIFS such issues are exacerbated by G1.
G3. Parameter consistency	As the AIFS lacks hard physical constraints between variables there is more scope for inter-parameter consistency to be lacking at specific locations at specific times.	Ordinarily this is not a major problem, but there have, for example, been cases of precipitation without cloud. P1 provides a more substantive example.
Low level winds		
W1. Underestimation of wind speeds around cyclones	For both tropical and extra-tropical cyclones the AIFS has a slow bias, underestimating the strongest winds.	
Cloud cover		
C1. Under-dispersive distribution	In AIFS, cloud cover "extreme" values (zero and 100%) are systematically under-represented, whilst intermediate values are systematically over-represented.	For stratus cases, for example, where cover is commonly 0% or 100%, usually the AIFS value lies in between whilst the IFS value does not.
C2. Issue centred on the Horn of Africa	Whilst the verification of downward (shortwave) solar radiation shows AIFS to generally have smaller errors than IFS, we are aware of one particular problem area centred on the Horn of Africa (spanning parts of Somalia, Ethiopia and Yemen), where the AIFS errors look to be larger and rather persistent.	Under investigation.
2m temperature		
T1. Consistency issues	AIFS is primarily trained using ERA5 data. For the 2m temperature component the *offline* land-data assimilation 2m temperature field is used. This uses 2m temperature observations, which can sometimes be inconsistent with the overlying atmosphere simulation in ERA5 which does not use those observations.	Despite this inconsistency, the AIFS 2m temperature fields in such situations are often much more accurate than those of IFS.
Precipitation		
P1. Convective precipitation extent	The areally-integrated amount of convective precipitation forecast is noticeably less in AIFS than in IFS.	In tests of forecasts from mid February 2025 the proportion of convective precipitation, globally, was 56% in the IFS Control run and 47% in AIFS single (total amounts were similar in each case). It is hard to verify what is the truth.
P2. Small precipitation totals	Small amounts of precipitation look to be more commonplace in AIFS than they should be (and more common than in IFS). For example small totals (< ~0.1mm/6h) can be repeatedly predicted over arid areas like the Sahara when they look to be impossible.	

However, solving the issues require different strategies for an AI model compared to NWP



Summary

- Different characteristics in AIFS-single and AIFS-ENS, but both:
  - Able to make very skilful prediction of the large-scale flow
  - Capture extremes, but missing the magnitude for some types of extreme weather
  - Improves on (conditional) systematic error in IFS
- Continued efforts in the evaluation team to find strange behaviours in AIFS
- AIFS limited by the skill of the training data (ERA5 + operational analysis)
  - Efforts to add observation data into training
- Under development:
  - Including more earth-system components (e.g waves)
  - Higher resolution
  - Reforecast dataset
  - Sub-seasonal system
- How are we going to use the information from the different models in practice?
  - Come and discuss this afternoon!



Further reading:

ECMWF newsletter articles over the past year about extreme events like Storm Eowyn, Storm Boris, …

AIFS: Lang et al., AIFS -- ECMWF's data-driven forecasting system https://arxiv.org/abs/2406.01465

AIFS-ENS: Lang et al., AIFS-CRPS: Ensemble forecasting using a model trained with a loss function based on the Continuous Ranked Probability Score
https://arxiv.org/html/2412.15832v1

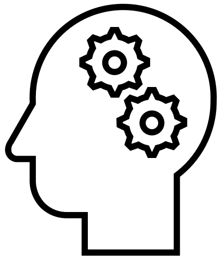
AI-Model assessments: Ben Bouallègue et al., The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning–Based Weather Forecasts in an Operational-Like Context: https://doi.org/10.1175/BAMS-D-23-0162.1

A growing list of AIFS output:

Variable name	Short name	Level type Pressure level (50-1000 hPa) or Surface	Variable type: Prognostic, Diagnostic, Forcing
Geopotential	z	Pl	P
Horizontal wind components	u, v	Pl	P
Specific humidity	q	Pl	P
Temperature	t	Pl	P
Surface pressure	sp	S	P
Mean sea-level pressure	msl	S	P
Skin temperature	skt	S	P
2 m temperature	2t	S	P
2 m dewpoint temperature	2d	S	P
10 m horizontal wind components	10u, 10v	S	P
Total column water	tcw	S	P
Volumetric soil water level 1 and 2*	swvl1, swvl2	S	P
Soil temperature level 1 and 2*	stl1, stl2	S	P
Total precipitation	tp	S	D
Convective precipitation	cp	S	D
Snowfall*	sf	S	D
Total cloud cover*	tcc	S	D
High cloud cover*	hcc	S	D
Medium cloud cover*	mcc	S	D
Low cloud cover*	lcc	S	D
Runoff*	ro	S	D
Surface solar radiation downwards*	ssrd	S	D
Surface thermal radiation downwards*	strd	S	D
100 m horizontal wind components*	100u, 100v	S	D
Land-sea mask	lsm	S	F
Orography	z	S	F
Standard deviation of sub-grid orography	sdor	S	F
Slope of sub-scale orography	slor	S	F
Insolation	insolation	S	F
Latitude/longitude (cos/sin)	lat/lon	S	F
Time of day/day of year	local time, julian day	S	F

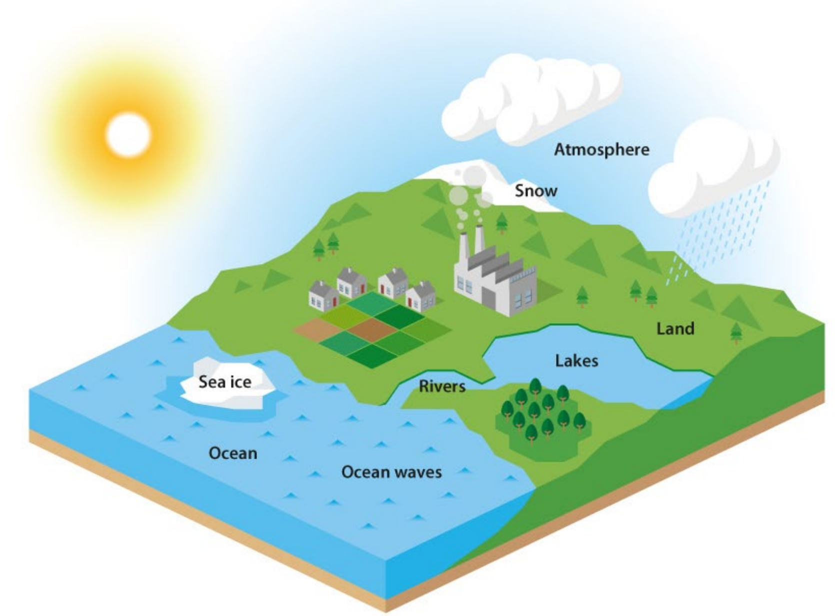
Pressure levels: 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925 and 1000hPa

Physics-based forecast systems using IFS model



Medium-range

- Atmosphere: IFS Model 48r1, TCo1279 (**9 km**), 137 vertical levels
- **15 (00 and 12UTC) and 6 (06 and 18UTC) days lead time**
- 50 ensemble members + 1 unperturbed control forecasts
- Ocean: NEMO (0.25°)
- Reforecasts: 11 members, past 20 years
- Initialised every day at 00/06/12/18 UTC
- Distributed around 8 hours after initialisation time*



Sub-seasonal

- Atmosphere: IFS Model 49r1, TCo319 (36 km), 137 vertical levels
- **6 weeks lead time**
- **101 ensemble members**
- Ocean: NEMO (0.25°)
- Reforecasts: 11 members, past 20 years
- **Initialised every day at 00UTC**
- Distributed in the evening (UTC time)

Seasonal (SEAS5)

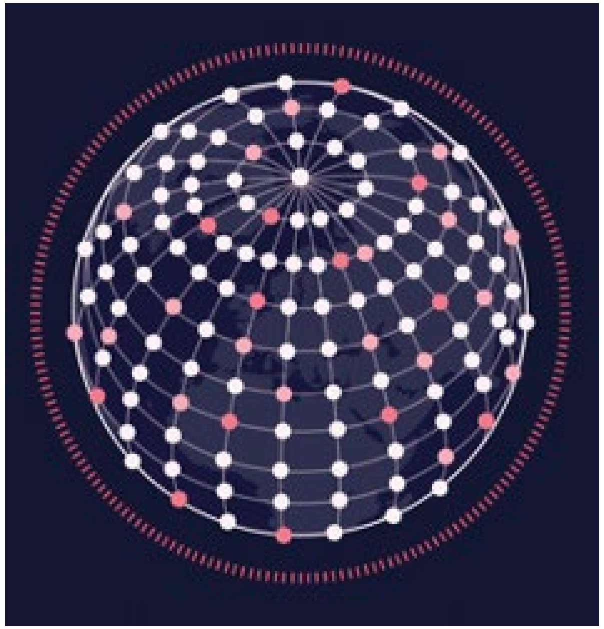
- **Operational since November 2017**
- Atmosphere: IFS Model 43r1, TCo319 (**36 km**), 91 vertical levels
- **7 months lead time**
- 51 ensemble members
- Ocean: NEMO (0.25°)
- Reforecasts: 25 members, 1981-2016 (the full reforecast period is only used for skill estimates)
- **Initialised 1st every month**
- Distributed 5th every month

* In the assimilation system, it has used observations +3h after the labelled initialisation time

Machine-learning based forecast systems using AIFS model

Medium-range deterministic AIFS-single v1.1

- Atmosphere: AIFS Model v1.0 using **MAE loss**, N320 (0.25°, ~25 km), 13 vertical levels
- 15 days lead time
- Initialised every day at 00/06/12/18 UTC
- Initialised from physics-based analysis
- Distributed around XX hours after initialisation time*

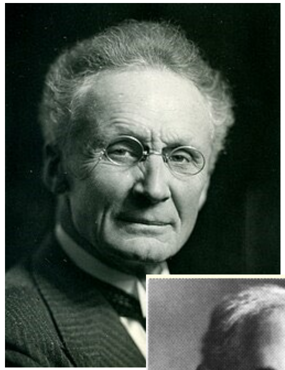


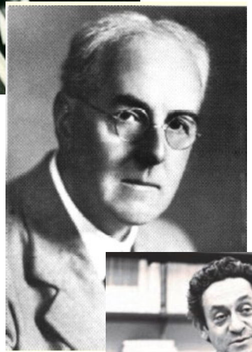
Medium-range ensemble AIFS-ENS v1.0

- Atmosphere: AIFS Model v1.0 using **CRPS loss**, N320 (0.25°, ~25 km), 13 vertical levels
- 15 days lead time
- 50 ensemble members + 1 unperturbed** control forecasts
- Same initial conditions including perturbations as physics-based ensemble
- Distributed around XX hours after initialisation time*

* In the assimilation system, it has used observations +3h after the labelled initialisation time
** With the use of CRPS-loss, also the control forecast include the “noise” to simulate model uncertainty

And what we are used to now:



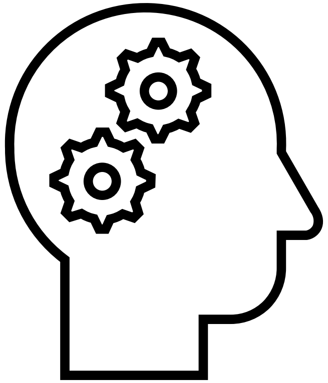


$$\frac{D\mathbf{V}}{Dt} + f\mathbf{k}\times\mathbf{V} = -\nabla\Phi$$

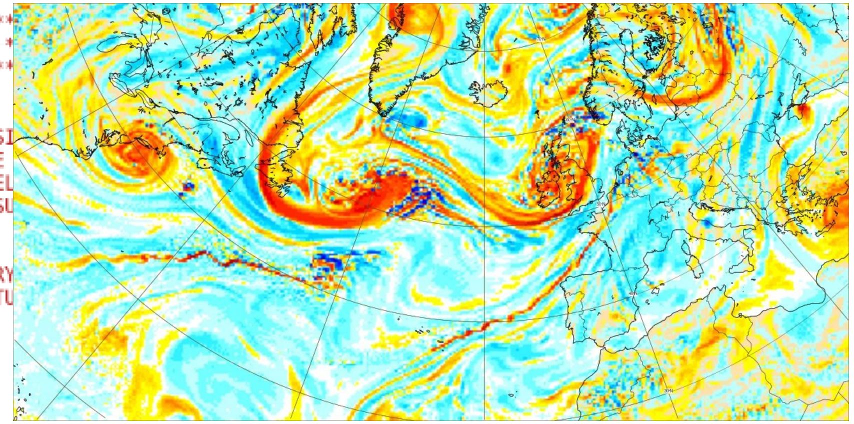
$$\frac{\partial\Phi}{\partial p} = -\alpha = -\frac{RT}{p}$$

$$\nabla\cdot\mathbf{V} + \frac{\partial\omega}{\partial p} = 0$$

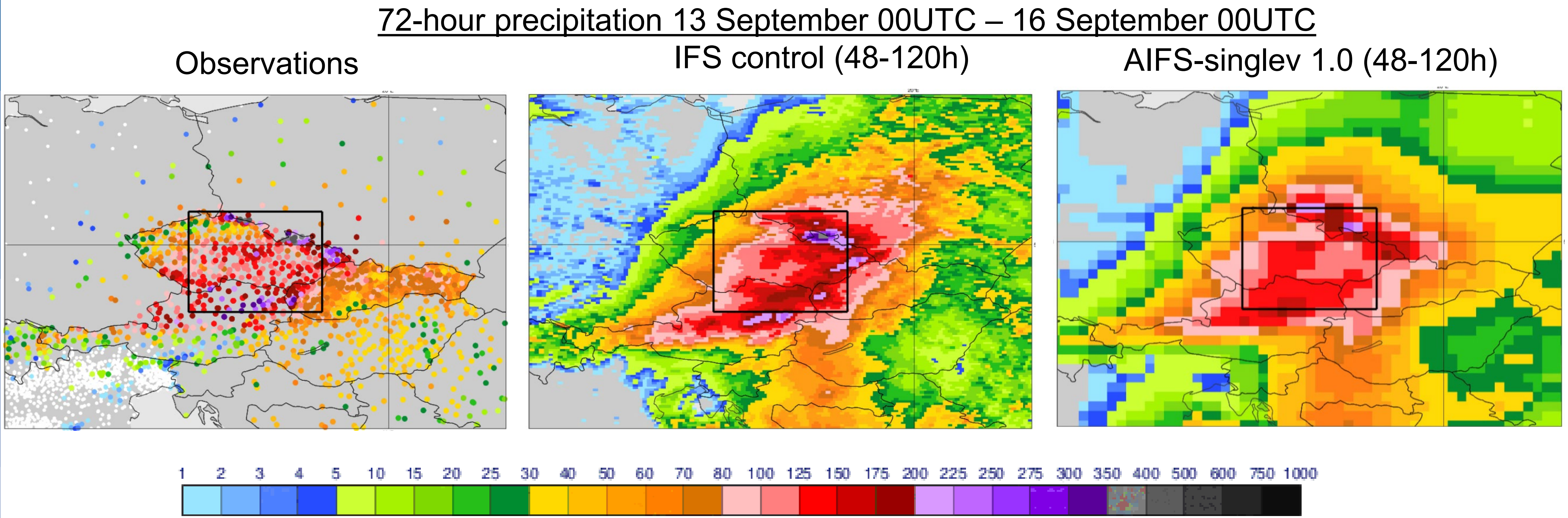
$$\left(\frac{\partial}{\partial t} + \mathbf{V}\cdot\nabla\right)T - S_p\omega = \frac{J}{c_p}$$



```
SUBROUTINE CALLPAR(YDGEOMETRY,YDVARS,YDSURF,YDMODEL,KDIM,&
 !-----------------------------------------------------------------
 & PAUX, PRAD, FLUX, PDIAG, PSURF, PCGPP, PCREC, PAG, PRECO, PDDHS, AUXL, SURFL, LLKEYS, PERTL, &
 ! - Model variables (t)
 & PGFL,&
 & PPERT, PSLPHY9,&
 ! - UPDATED TENDENCY
 & PTENGFL,&
 & PHYS_MWAVE,&
 ! Stored quantities
 & PSAVTEND, PGFLSLP,&
 & STATE_T0, TENDENCY_DYN, TENDENCY_CML, STATE_TMP, TENDENCY_TMP,&
 & TENDENCY_VDF, TENDENCY_SATADJ, TENDENCY_LOC, TENDENCY_PHY&
 & )

!**** *CALLPAR * - CALL ECMWF PHYSICS

!     PURPOSE.
!     --------
!     - CALL THE SUBROUTINES OF THE E.C.M.W.F. PHYSICS PACKAGE.

!     ****************************************
!     ****** IDIOSYNCRASIES *** IDIOSYNCRASIES *
!     ****************************************
!     ***  HEALTH WARNING:
!     ***  ===============
!     ***  NOTE THAT WITHIN THE E.C.M.W.F. PHYSI
!     ***  ARE INDEXED FROM 1 TO NFLEVG+1 WHILE
!     ***  0 AND NFLEVG IN THE REST OF THE MODEL
!     ***  CARE OF IN THE CALL TO THE VARIOUS SU
!     ***  PHYSICS PACKAGE
!     ***
!     ***    THIS IS SUPPOSED TO BE A "TEMPORARY
!     ***    STRAIGHTENED OUT IN THE "NEAR" FUTU
```

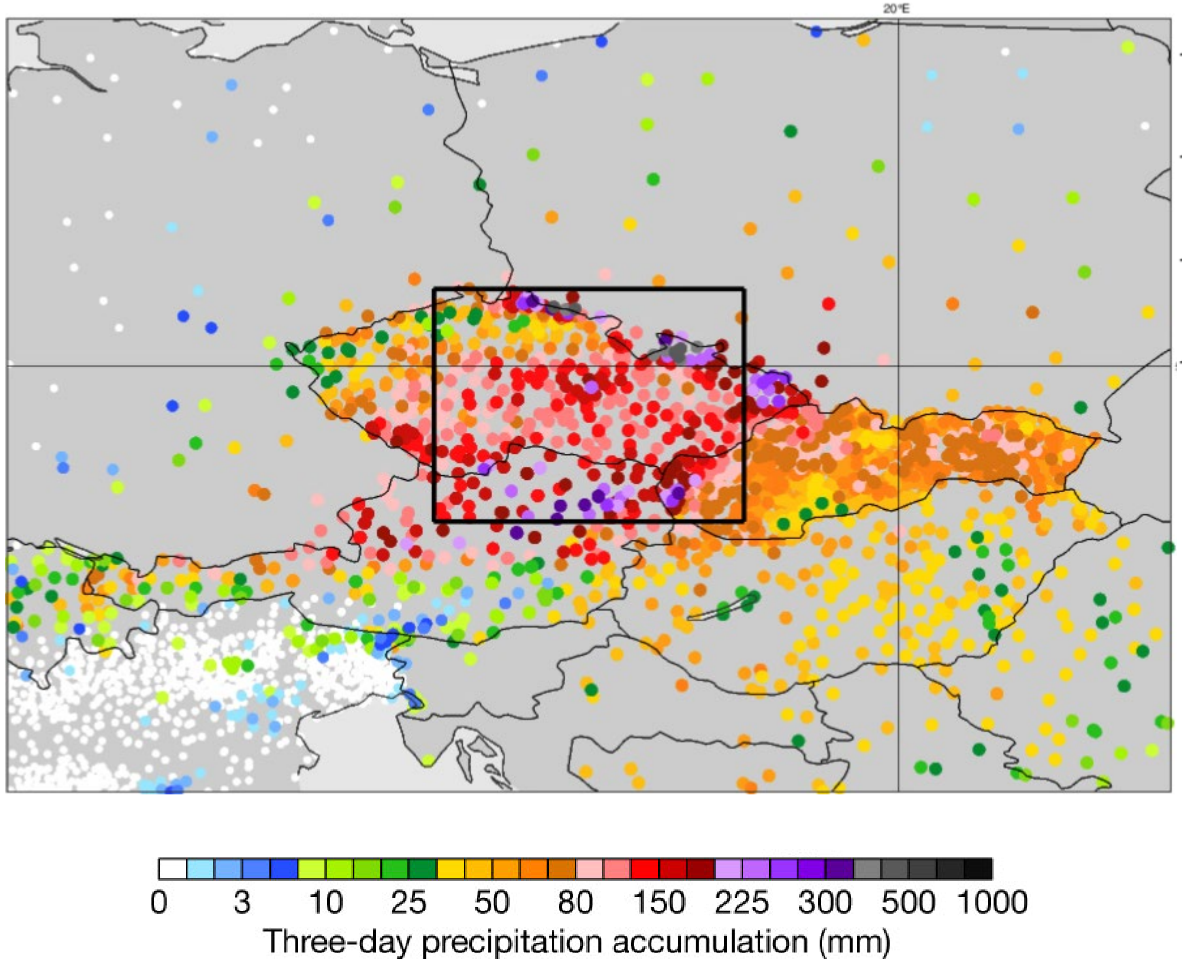


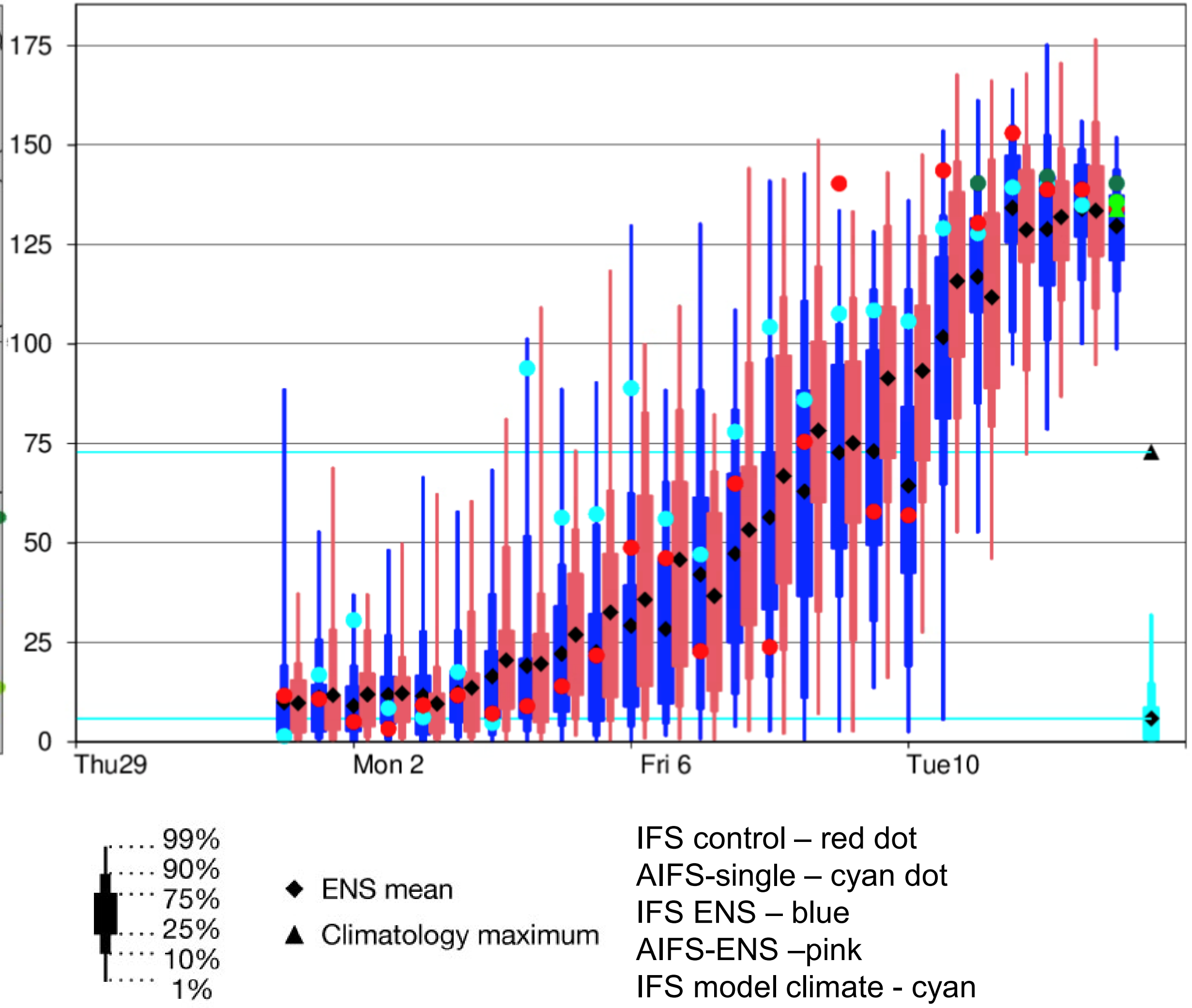
Extreme precipitation during Storm Boris in central Europe, September 2024



- Smooth precipitation field from AIFS (do not capture local structures)
- AIFS predicted very extreme values for this region

3-day Precipitation during Storm Boris 13-16 September 2024





IFS control – red dot
AIFS-single – cyan dot
IFS ENS – blue
AIFS-ENS –pink
IFS model climate - cyan

Extreme weather cases

Pages / Forecast User Portal   Analytics   Edit   View inline comments   Save for later   Watching   Share   ···

Severe Event Catalogue

Created by Florian Pappenberger, last modified by Timothy Hewson on Nov 09, 2022

r gui

On this space we collect material for evaluation of severe/extreme weather events. The focus is on the meteorological conditions and the forecast performance. The amount of material differs from case to case, and we are not claiming to give the full picture of the cases here. Users are welcome to contribute with material for the cases by using the comment function in the bottom of each page. To suggest a new case to evaluate, please contact us at the email address given below. If you have any initial comments and material, please include them in the mail.

Contact email address	servicedesk@ecmwf.int

(Please note that some of the links on the pages are only accessible from ECMWF.)

Navigation

List of (recent) cases

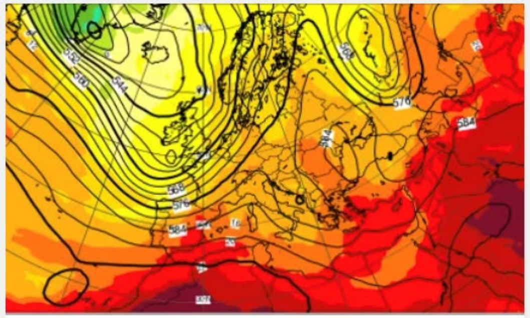
- 202404 - Snowfall / Cold - Sweden / Finland
- 202404 - Rainfall - UAE
- 202404 - Rainfall - Brazil
- 202404 - Cold -Europe

Search (for old cases enter the year and month of the event, as yyyymm)

Screenshot

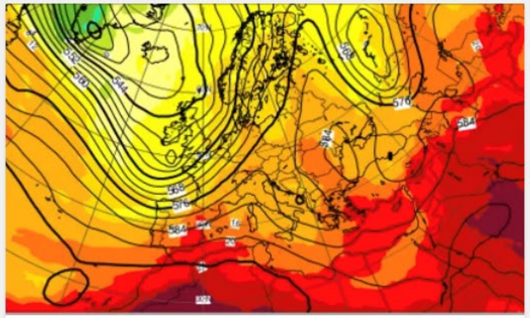
AIFS cases usually included in the ECMWF severe event catalogue https://confluence.ecmwf.int/display/FCST/Severe+Event+Catalogue

Real-time forecasts available on OpenCharts



Latest forecast

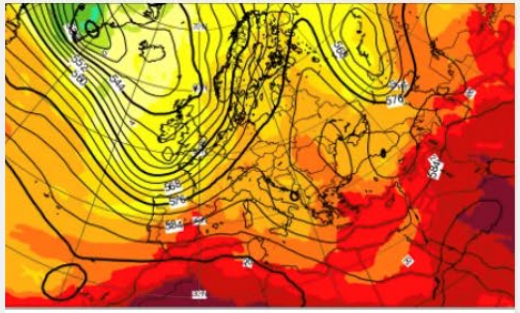
AIFS Single: 500 hPa geopotential height and 850 hPa temperature



Latest forecast

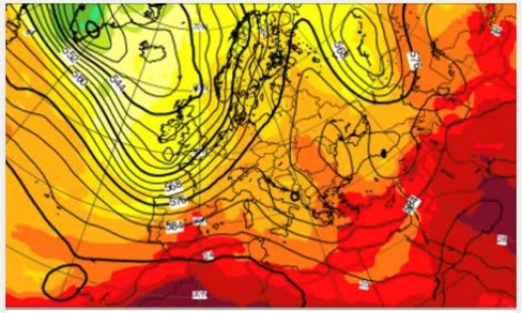
AIFS ENS Control: 500 hPa geopotential height and 850 hPa temperature

..together with other ML models experimentally run at ECMWF:



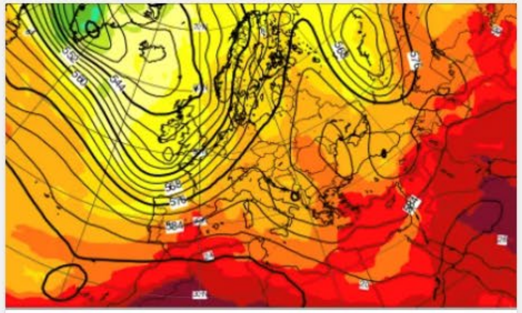
Latest forecast

Experimental: Aurora ML model: 500 hPa geopotential height and 850 hPa temperature



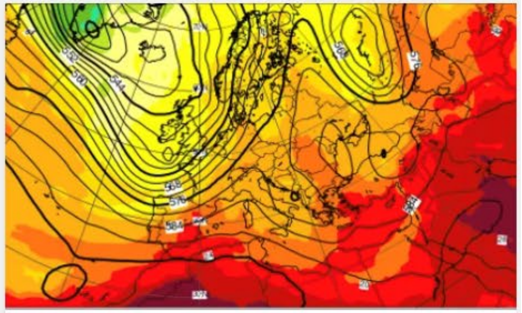
Latest forecast

Experimental: FourCastNet ML model: 500 hPa geopotential height and 850 hPa temperature



Latest forecast

Experimental: GraphCast ML model: 500 hPa geopotential height and 850 hPa temperature



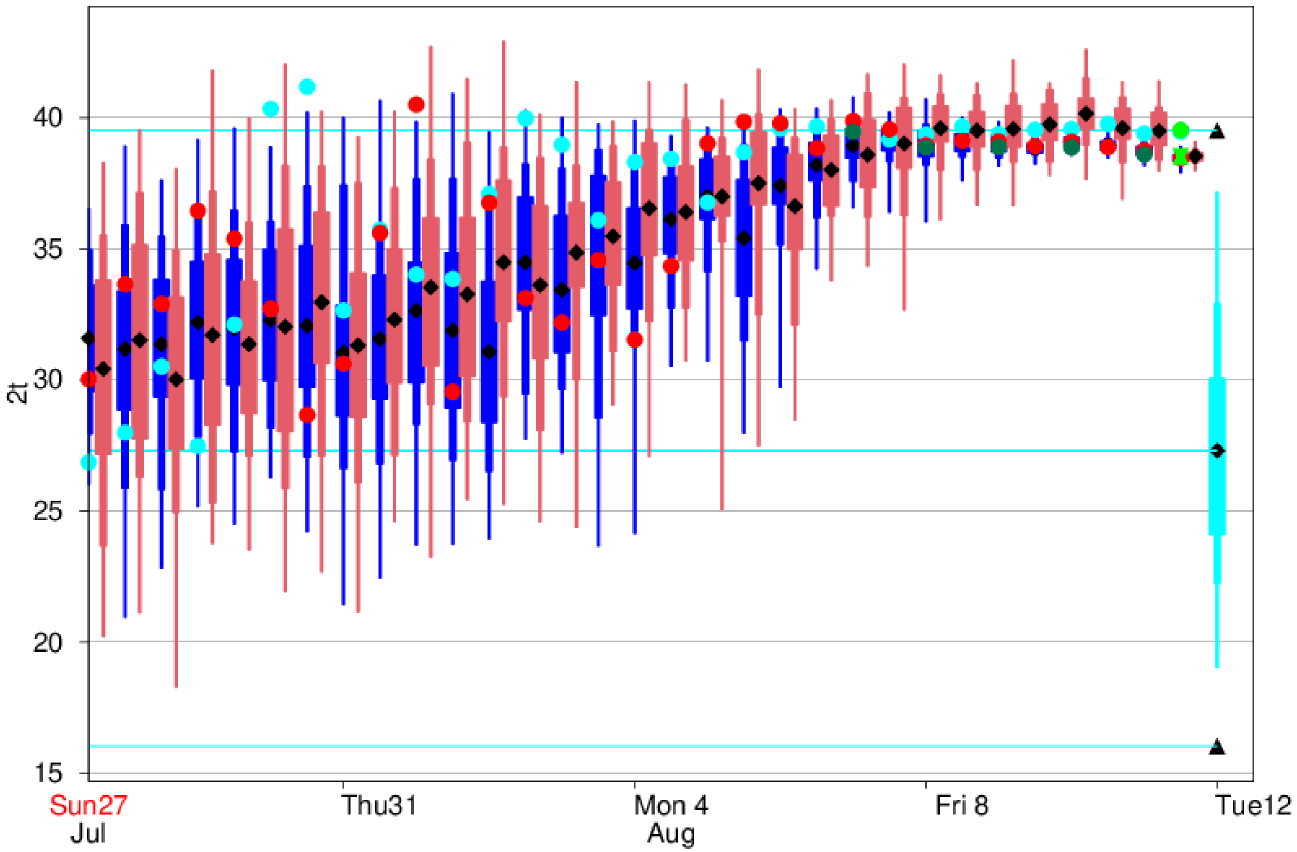
Latest forecast

Experimental: Pangu-Weather ML model: 500 hPa geopotential height and 850 hPa temperature

All models are trained on ERA5 reanalysis (~0.25 degree resolution), but some fine-tuned on HRES analysis

In all experiments below, we have initialised all ML models from ECMWF initial conditions.

2-metre temperature 31 August 12UTC around Toulouse



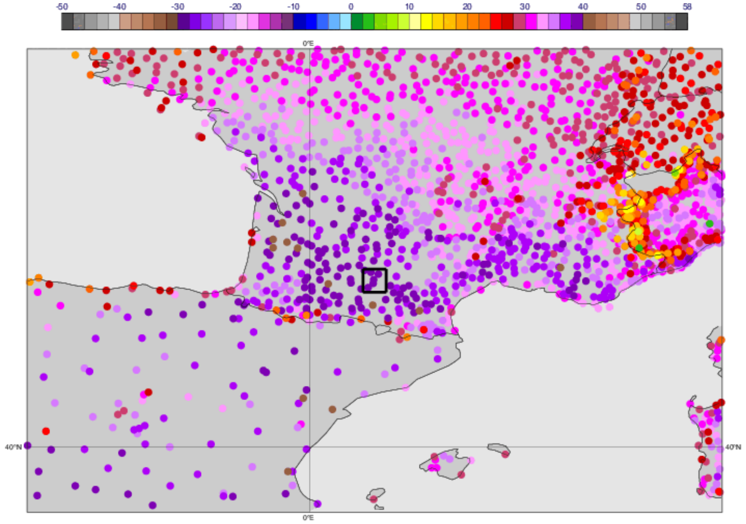
Observation – green hourglass
Analysis – green dot
IFS-CF – red
AIFS-single – cyan
IFS-ENS – blue
AIFS-ENS – pink

M-climate – cyan

2-metre temperature 31 August 12UTC

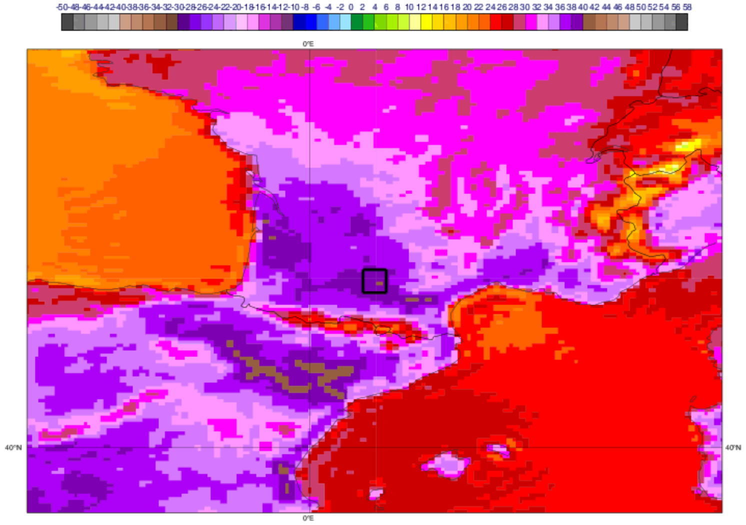
Observations

2t_Toulouse0.0_2025081112_od_oper_ecmf_0001_obs
Observations



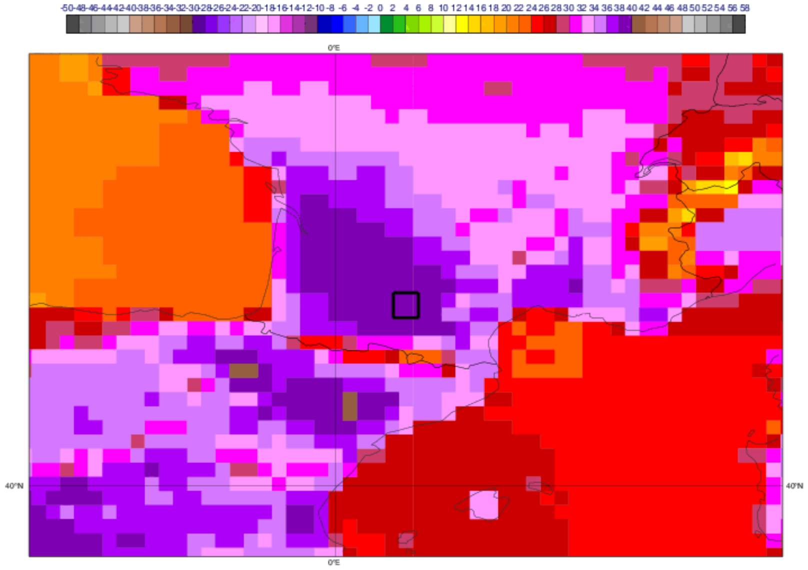
IFS ENS control +36h

2t_Toulouse0.0_2025081112_od_enfo_ecmf_0001_cf
Member: 0, Step: 36



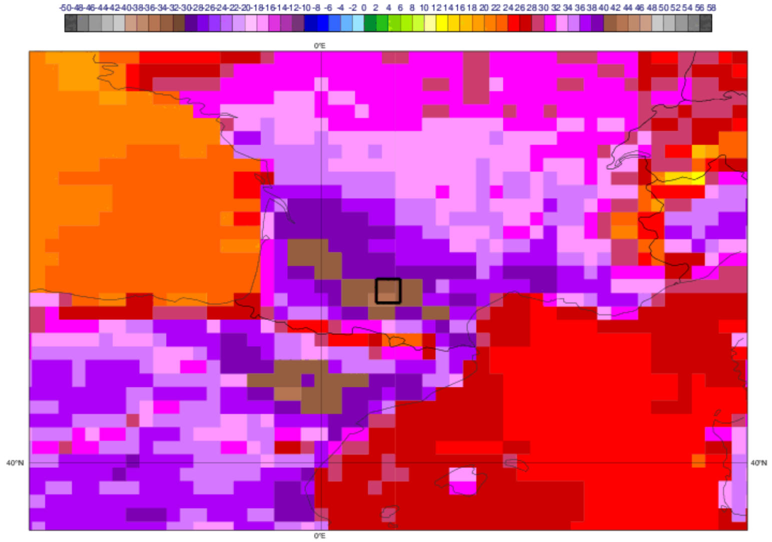
AIFS-single +36h

2t_Toulouse0.0_2025081112_ai_oper_ecmf_0001_fc
Member: 0, Step: 36



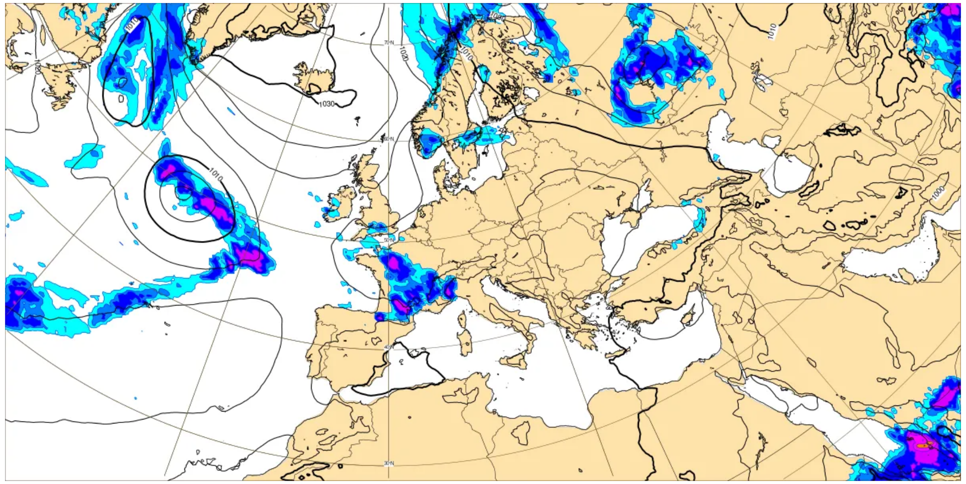
AIFS-ENS mem 1 +36h

2t_Toulouse0.0_2025081112_ai_enfo_ecmf_0001_pf
Member: 1, Step: 36

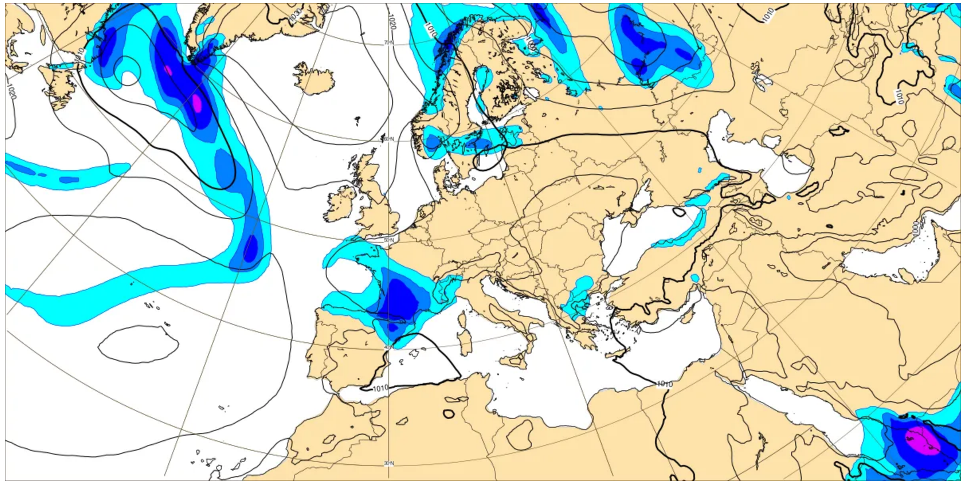


Smoothness in fields in AIFS-single

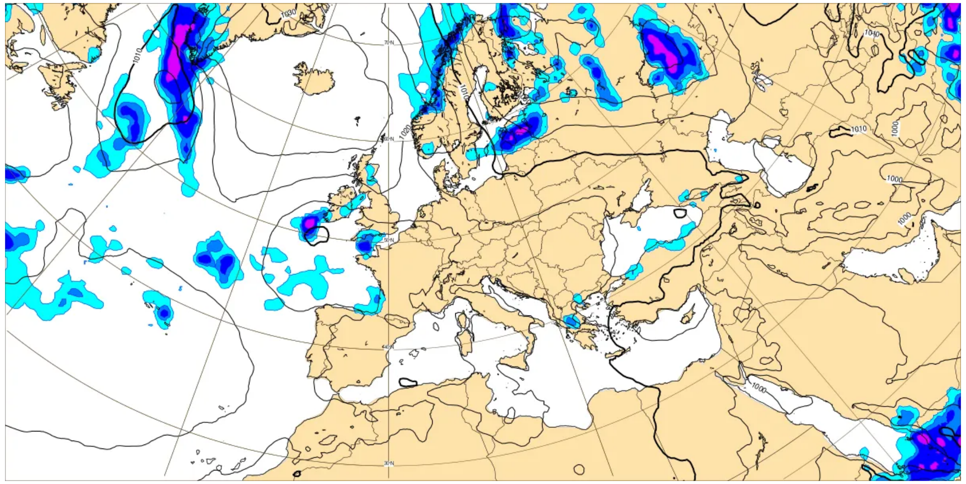
IFS control



AIFS-single



AIFS-ENS control

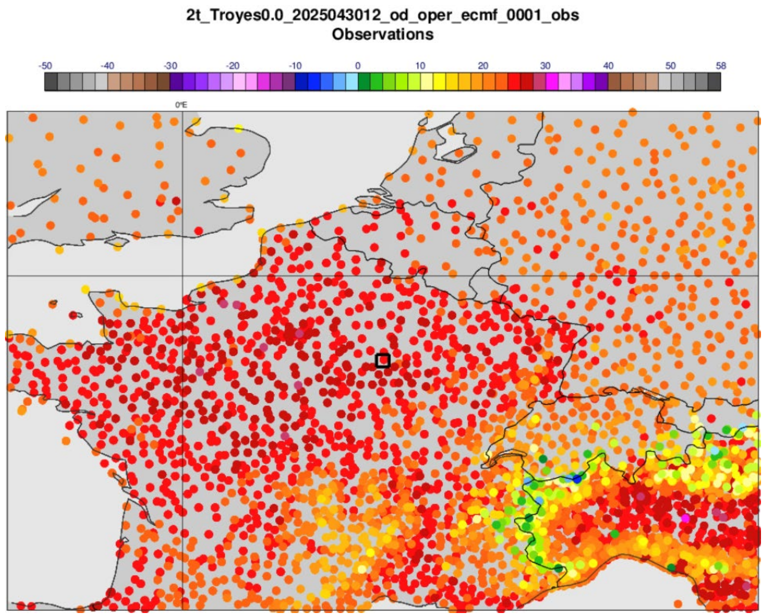


ECMWF

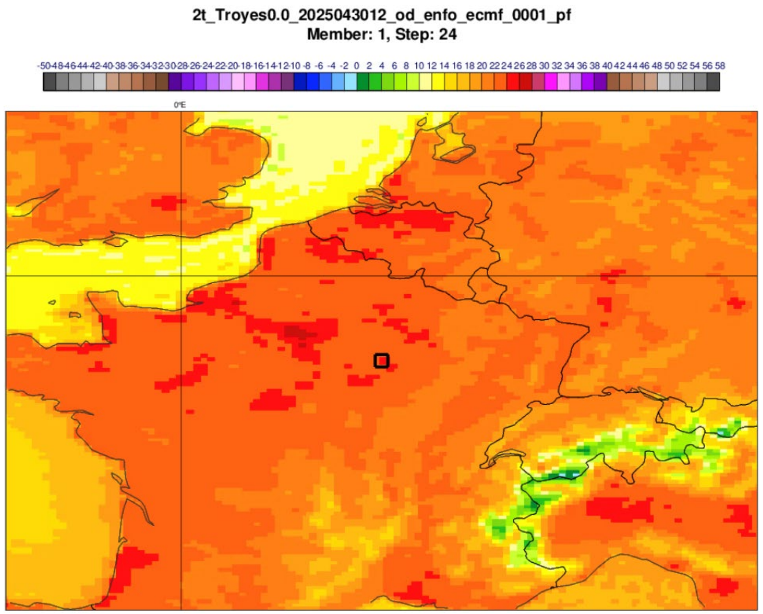
Understanding the enhanced large-scale skill in AIFS

Early heatwave in April 2025: 2-metre temperature 30 April 12UTC

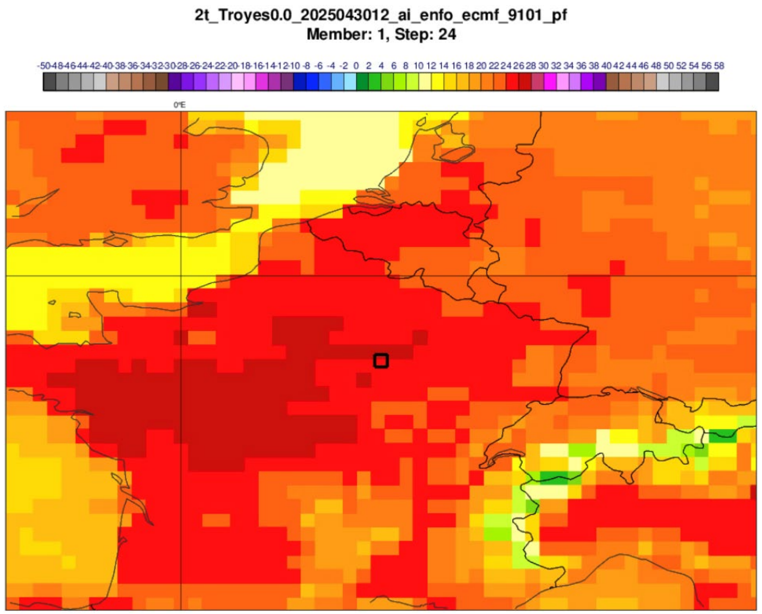
Observations



24h forecast from IFS-ENS Mem 1

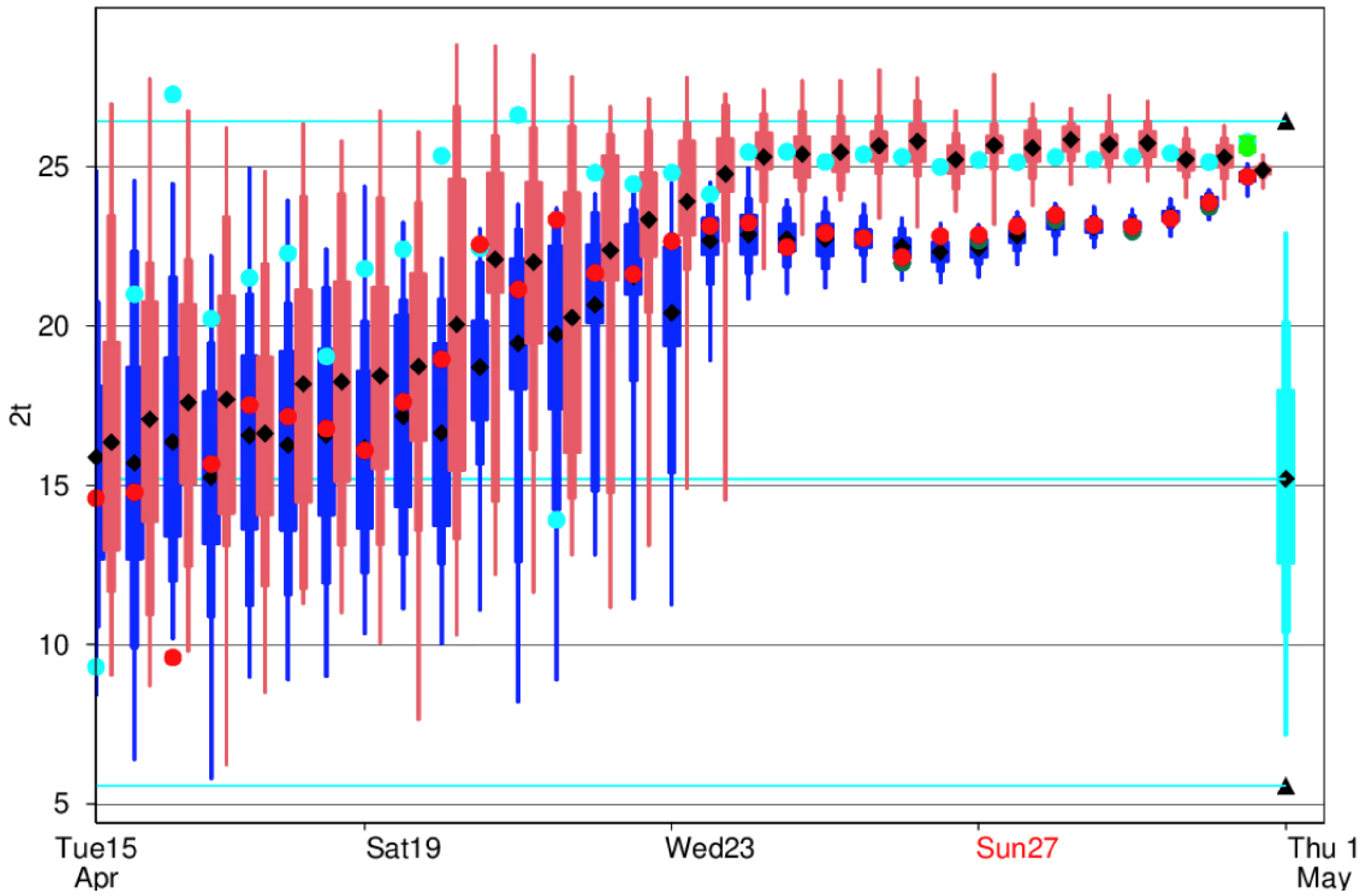


24h forecast from AIFS-ENS Mem 1



ECMWF

Evolution of forecasts for 2-metre temperature 30 April 12UTC in Troyes, France

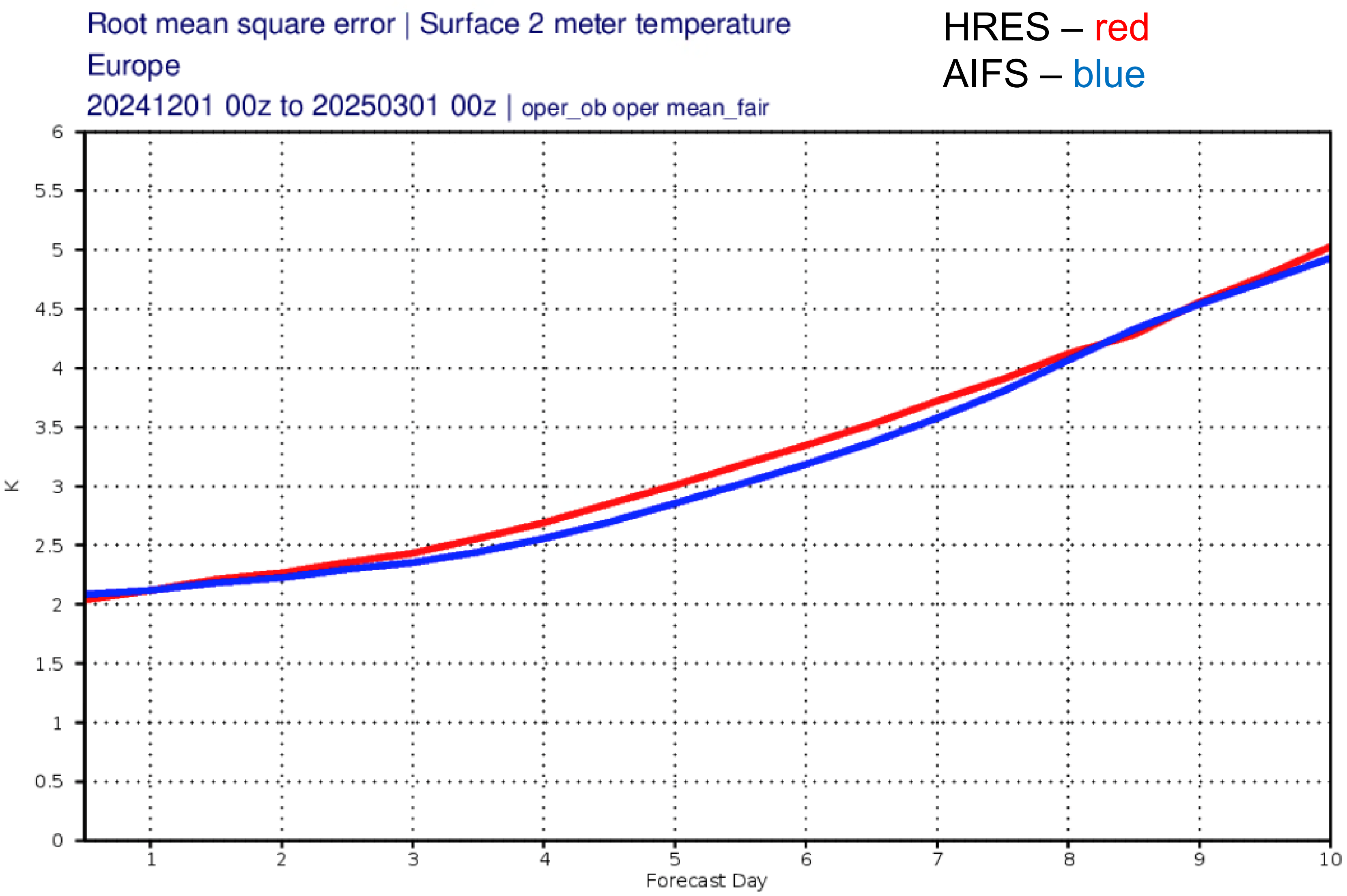


Observation – green hourglass
Analysis – green dot
IFS-CF – red
AIFS-single – cyan
IFS-ENS – blue
AIFS-ENS – pink

M-climate – cyan

ECMWF

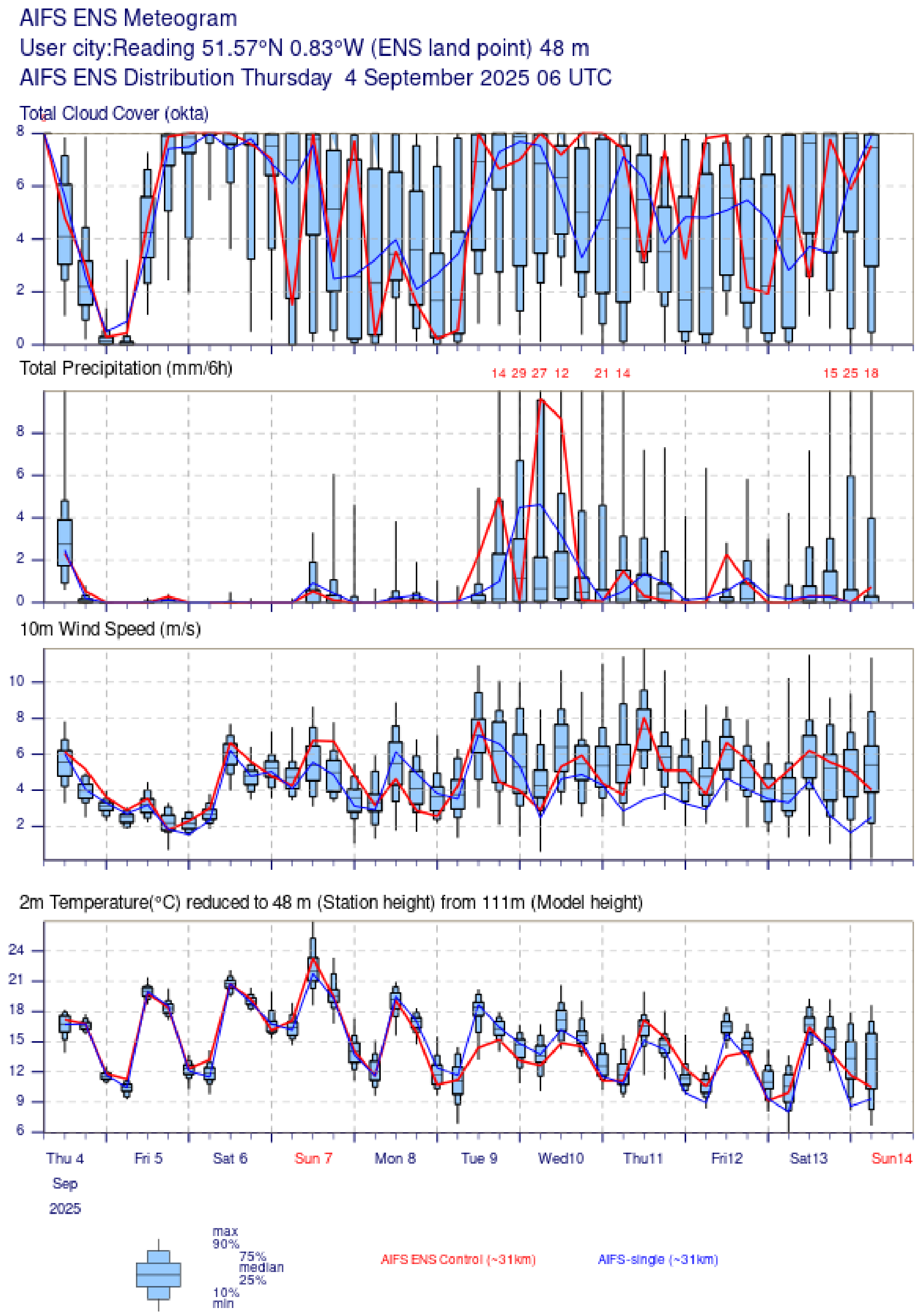
2-metre temperature RMSE against observations – DJF 2024-25



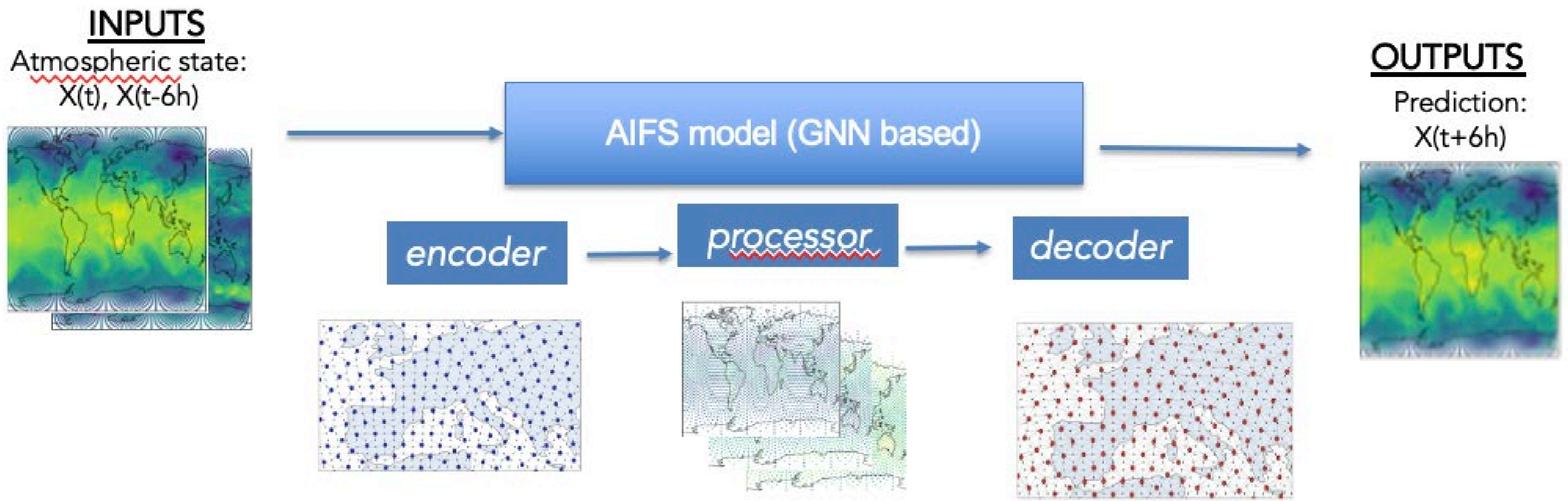
HRES – red
AIFS – blue

Difference due to:
Biases?
Conditional systematic errors?
Better large-scale predictions?

Metgrams also available:



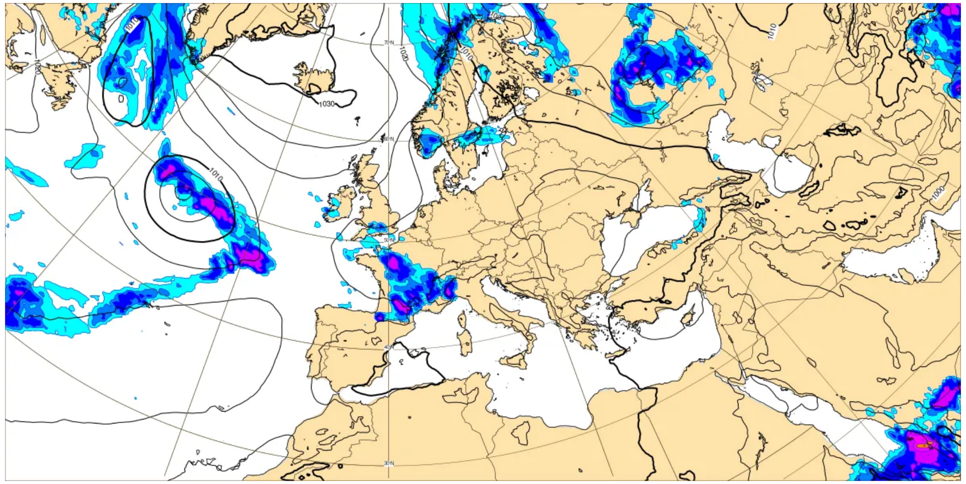
AIFS Single v1



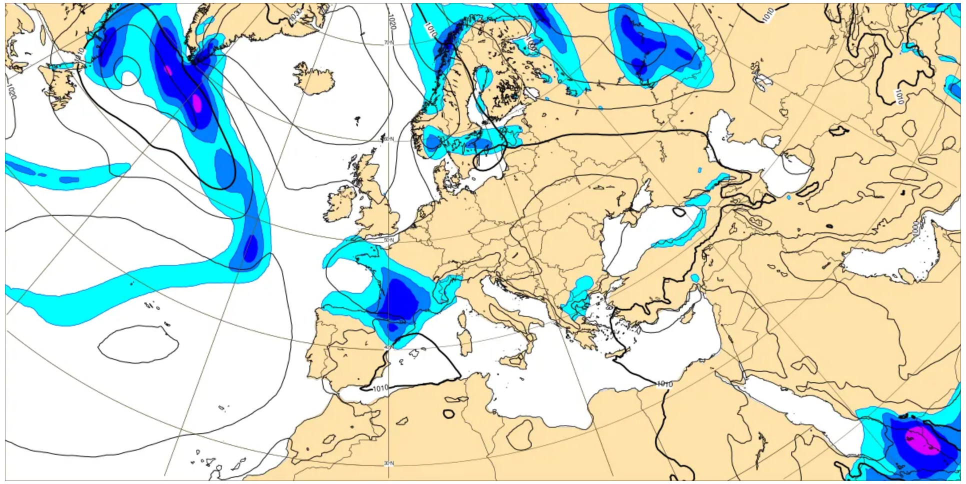
Lang et al 2024a
Operational system from 25/2/25

Smoothness in fields in AIFS-single

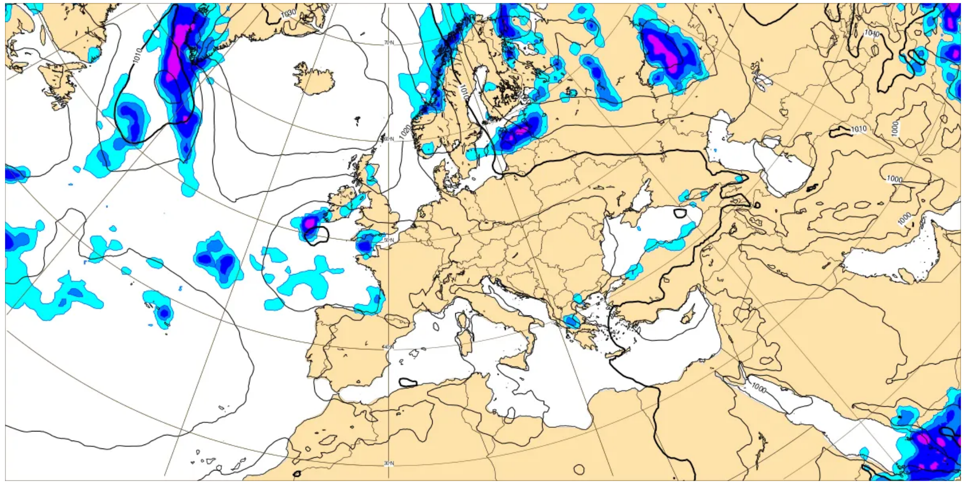
IFS control



AIFS-single



AIFS-ENS control

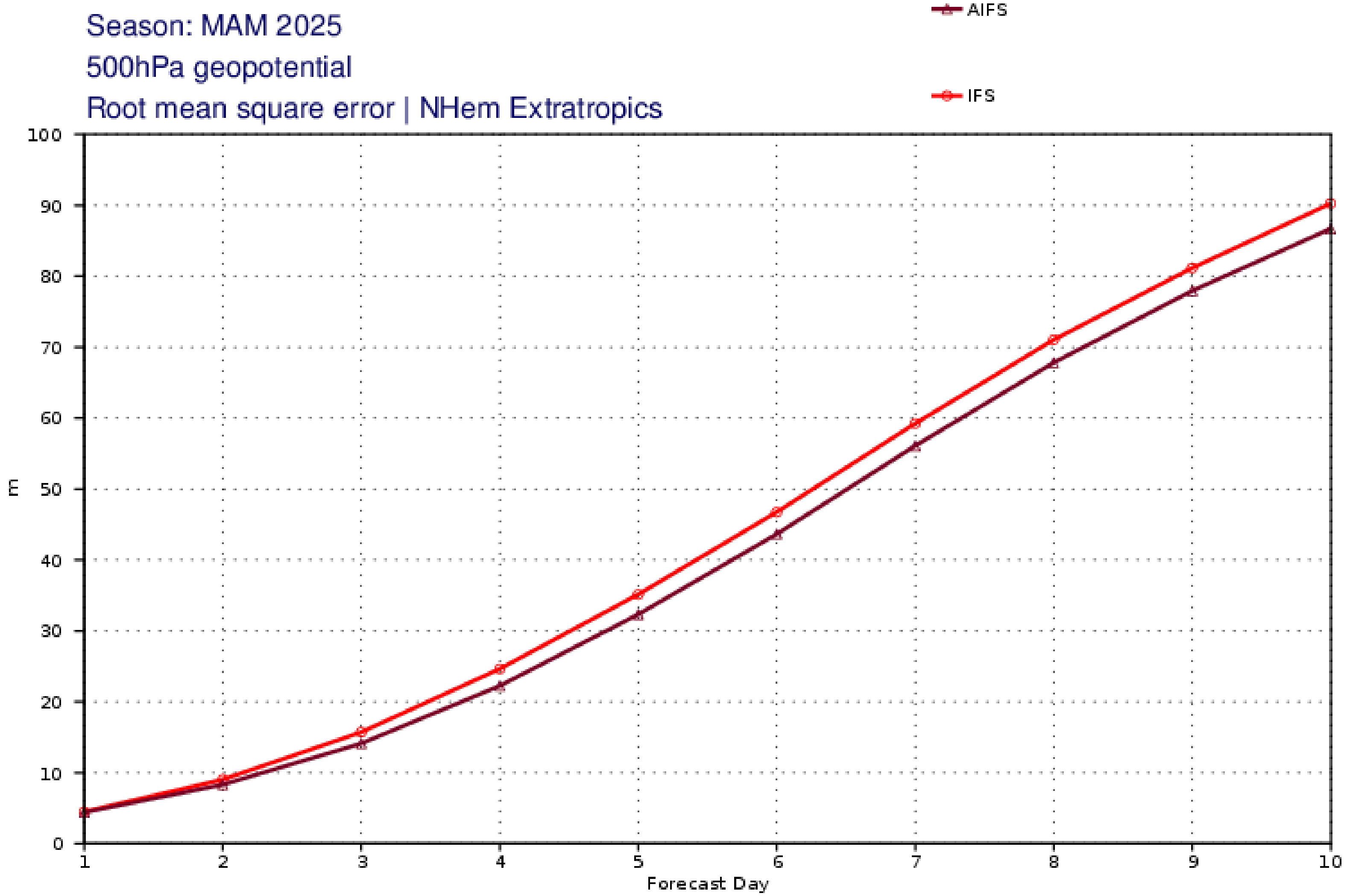


ECMWF

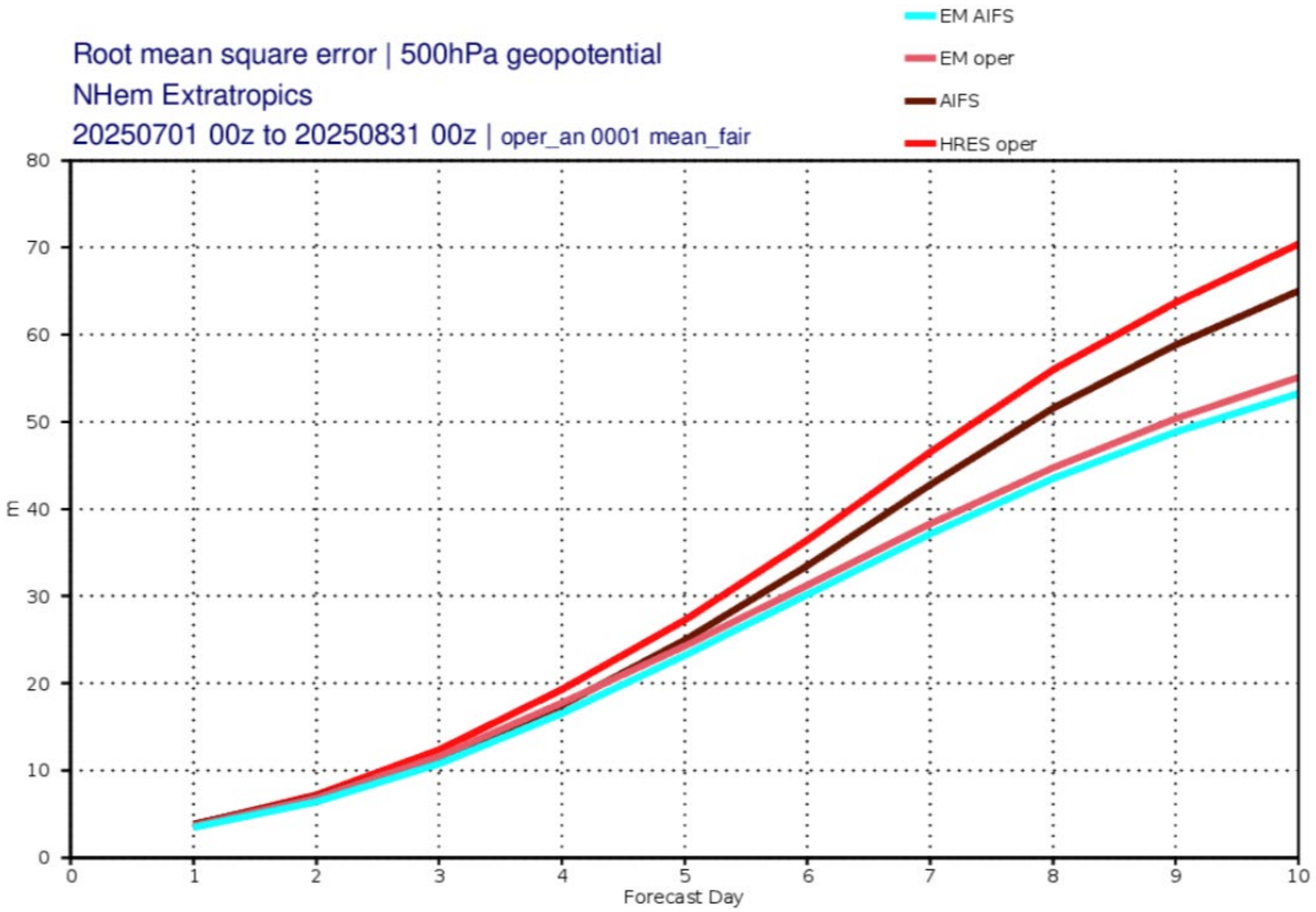
Reminder:

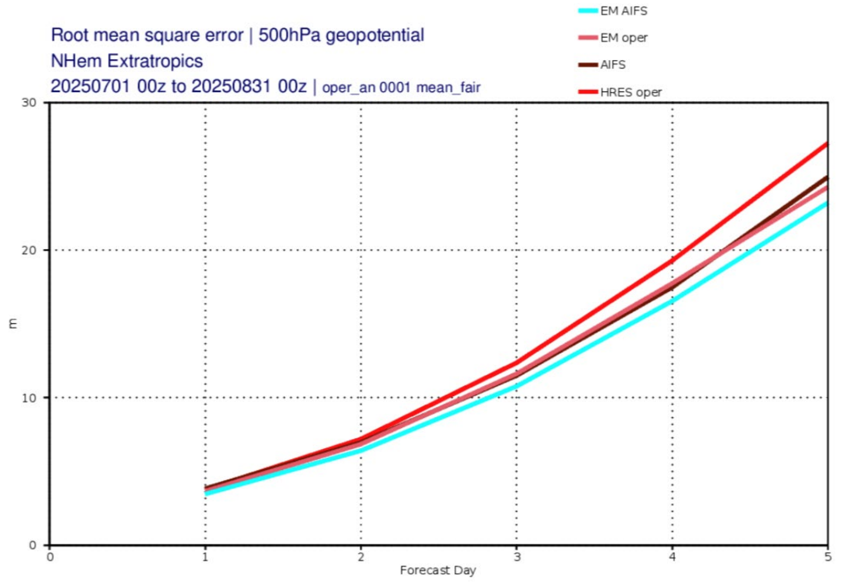
- AIFS (and all other AI models on OpenCharts):
  - Trained on NWP-based reanalysis (ERA5 + operational analysis)
  - Initialised from NWP analysis generated by IFS

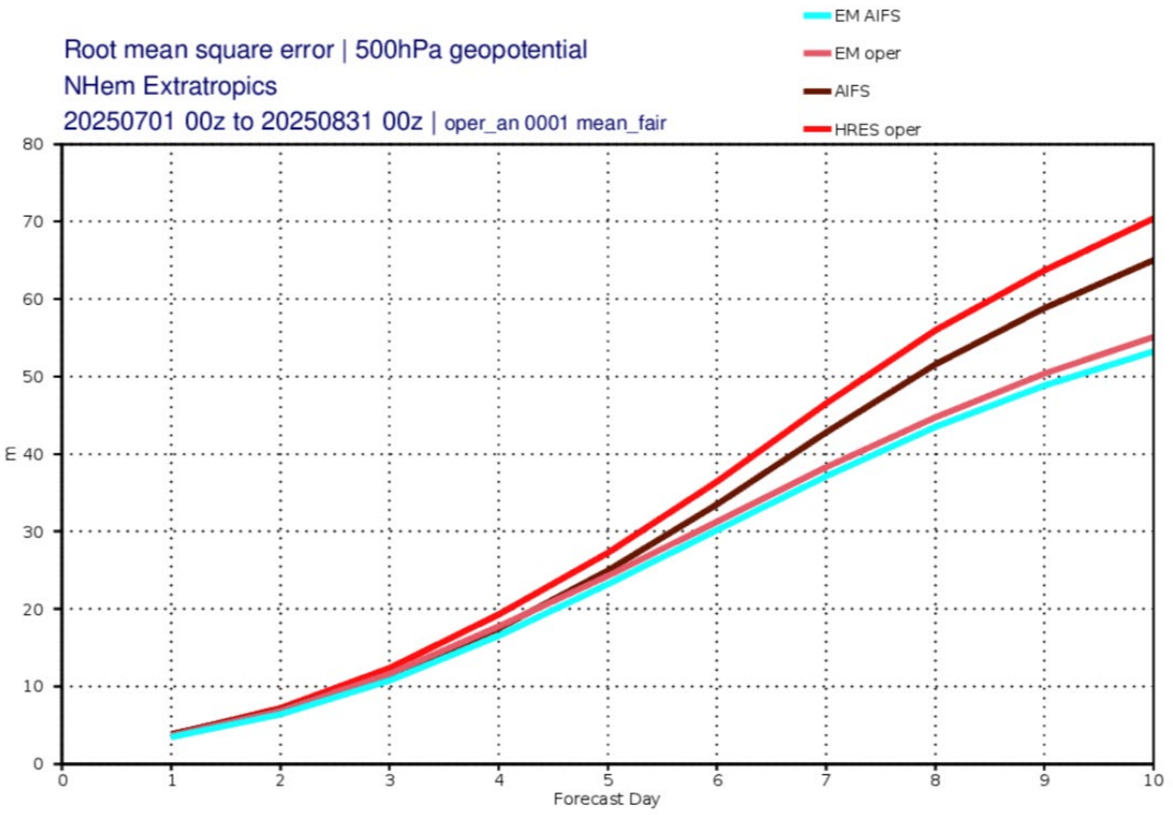
Operational verification

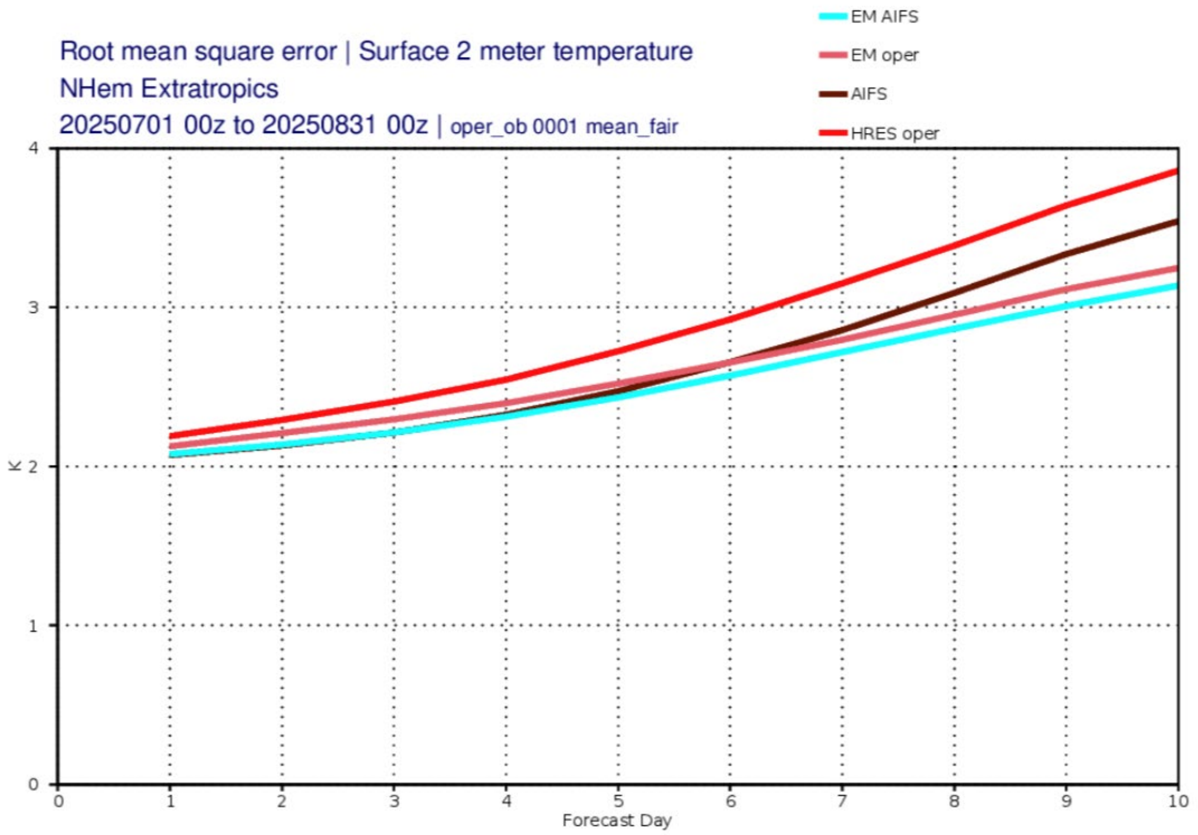


But forecast performance is much more than 500hPa geopotential height…


EM AIFS
EM oper
AIFS
HRES oper
Root mean square error | 500hPa geopotential
NHem Extratropics
20250701 00z to 20250831 00z | oper_an 0001 mean_fair
80
70
60
50
E 40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
Forecast Day

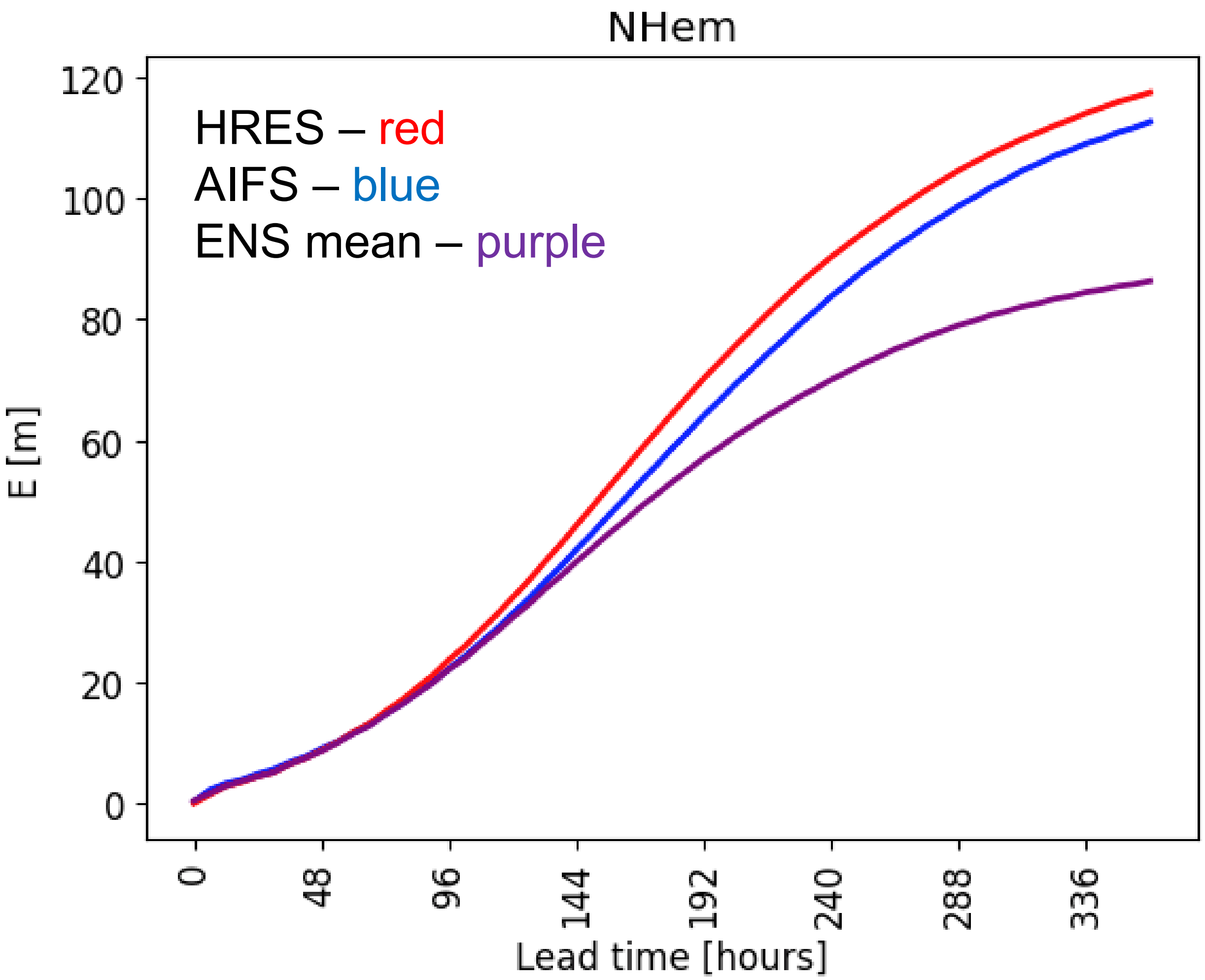

EM AIFS
EM oper
AIFS
HRES oper
Root mean square error | 500hPa geopotential
NHem Extratropics
20250701 00z to 20250831 00z | oper_an 0001 mean_fair
30
20
E
10
0
0
1
2
3
4
5
Forecast Day


EM AIFS
EM oper
AIFS
HRES oper
Root mean square error | 500hPa geopotential
NHem Extratropics
20250701 00z to 20250831 00z | oper_an 0001 mean_fair
m
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
Forecast Day


EM AIFS
EM oper
AIFS
HRES oper
Root mean square error | Surface 2 meter temperature
NHem Extratropics
20250701 00z to 20250831 00z | oper_ob 0001 mean_fair
K
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
10
Forecast Day

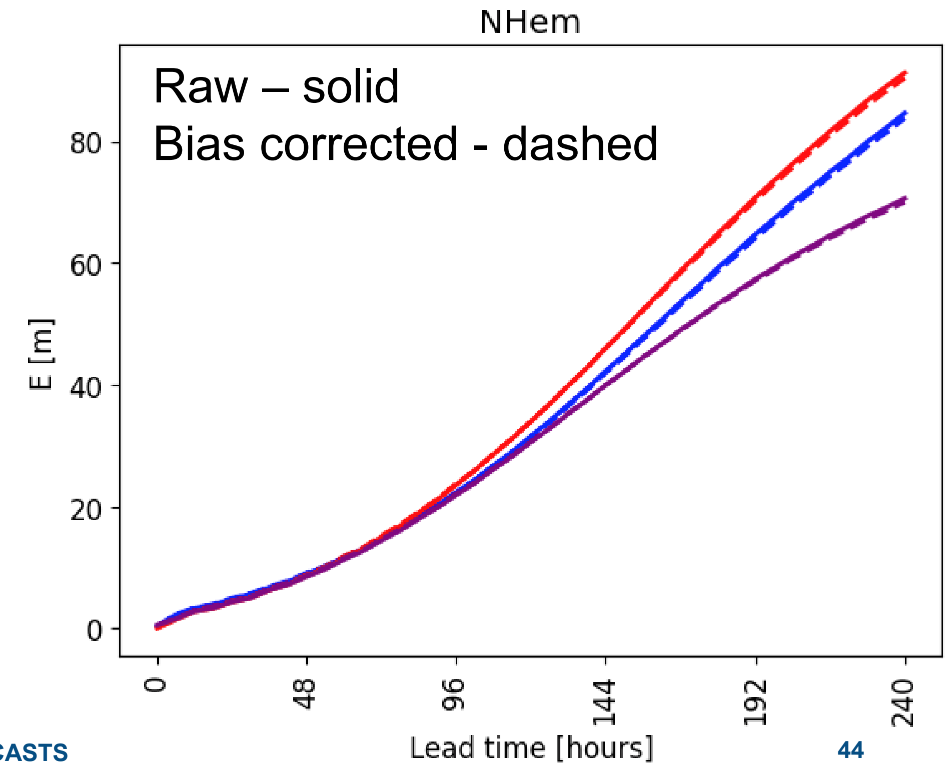
Understanding large-scale score differences between IFS and ML models
RMSE for z500, N.Hem March 2024 – mid-Feb 2025

- Chaotic error growth
- Initial condition error
- Real / Model activity
- "Random" model error



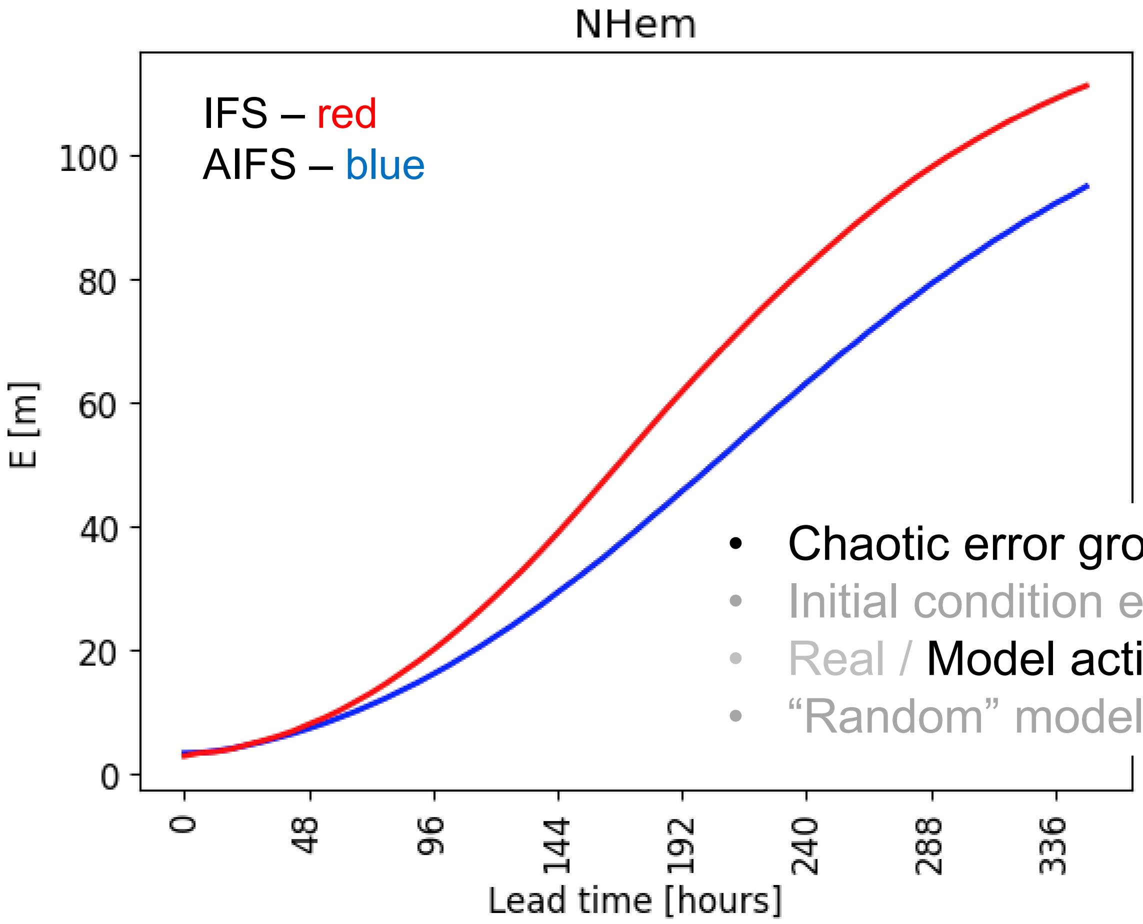
- AIFS initialised from ECMWF analysis

Effect of bias-correction



Role of chaotic ”error” growth:
Jumpiness - difference between consecutive (12h) forecasts verifying the same time

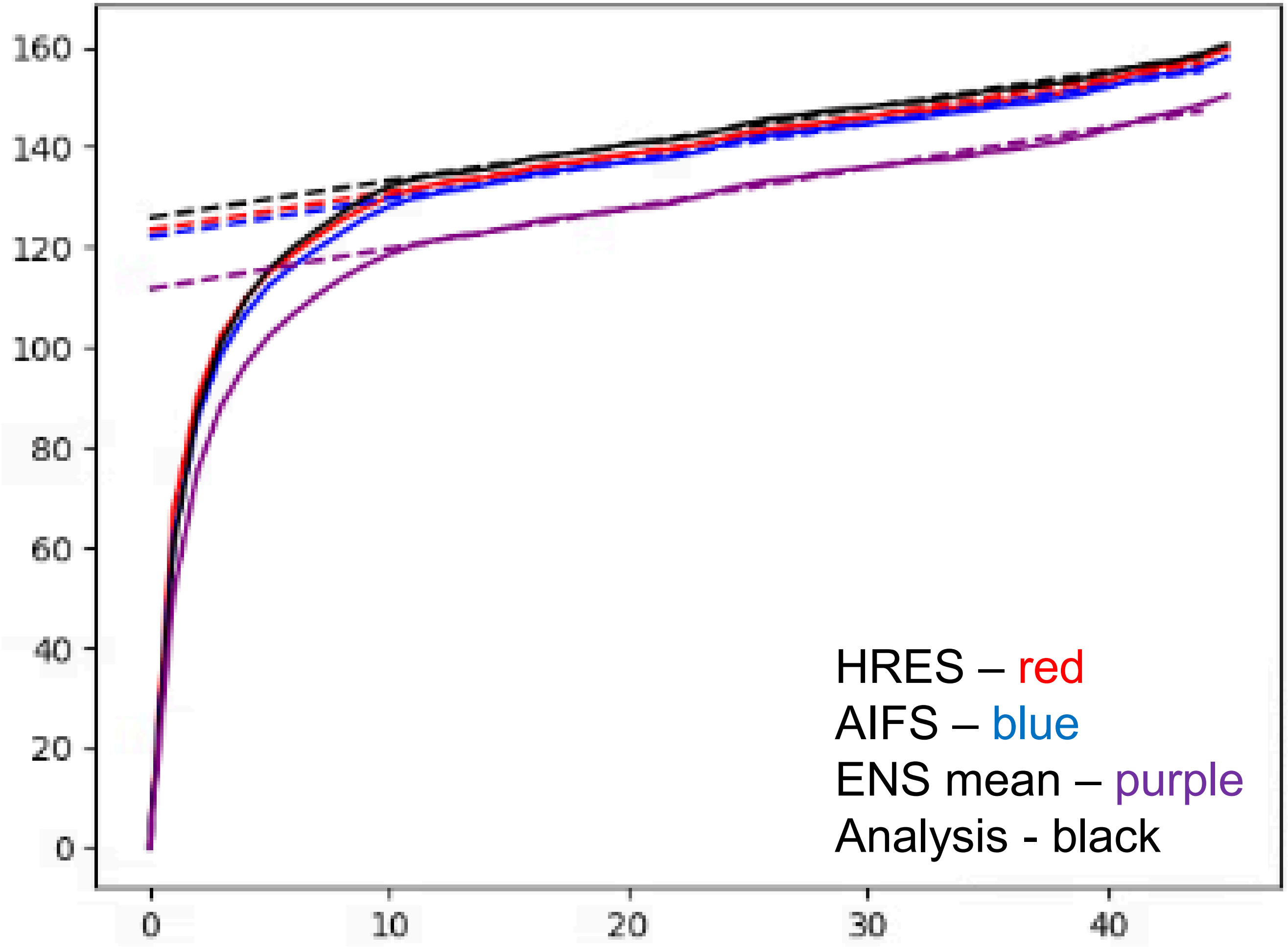
1 March 2024 - 20 Feb 2025



- Chaotic error growth
- Initial condition error
- Real / Model activity
- “Random” model error

Forecast activity day 6

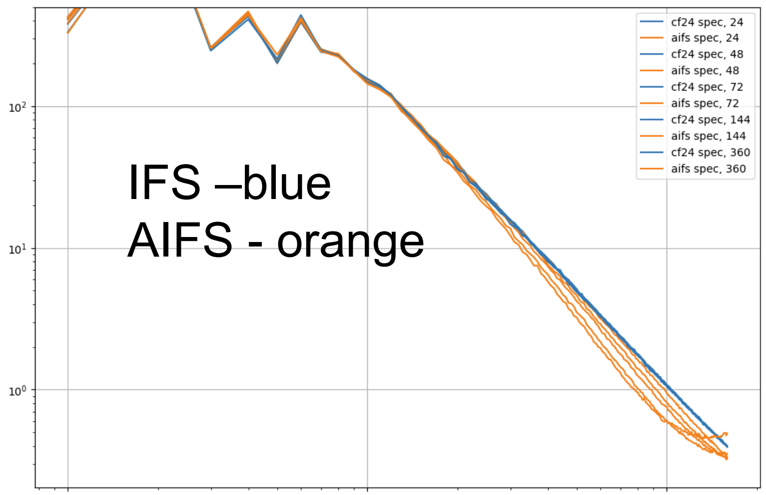
Difference between 6-day forecasts with x days apart



(Method from Bengtsson et al., 2008)

HRES - 121m
AIFS – 120m
ENS mean – 110m
AN – 124m

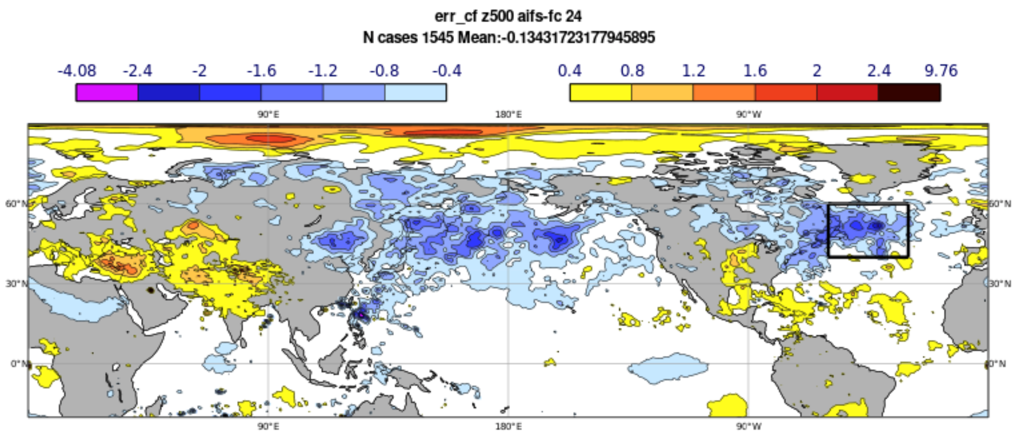
Power spectra for z500



Difference in error growth between 48h-24h, z500

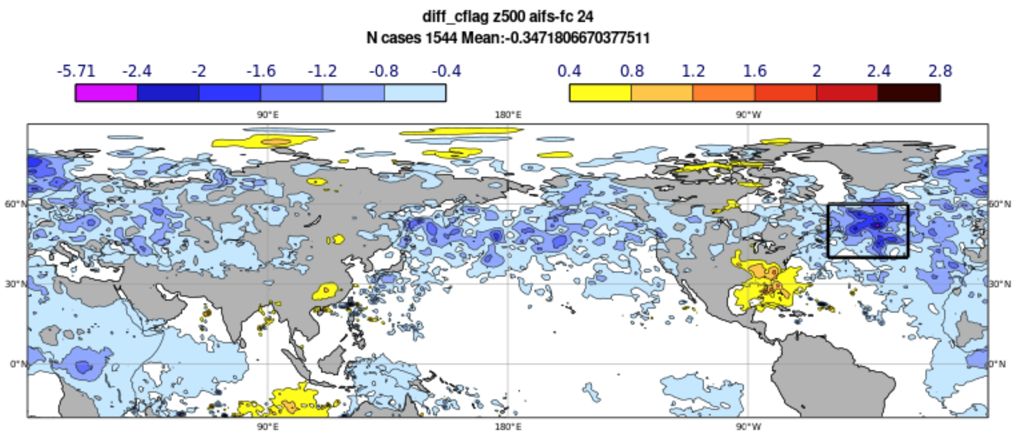
(based on ~2 years of forecasts)

RMS error



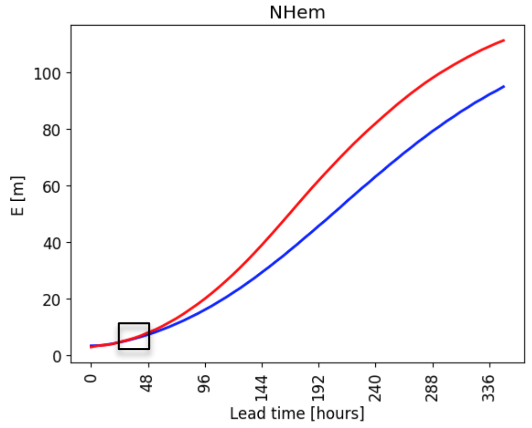
[RMS(AIFS-AN)$_{48}$ - RMS(AIFS-AN)$_{24}$] -
[RMS(HRES-AN)$_{48}$ - RMS(HRES-AN)$_{24}$]

RMS difference between lagged forecasts



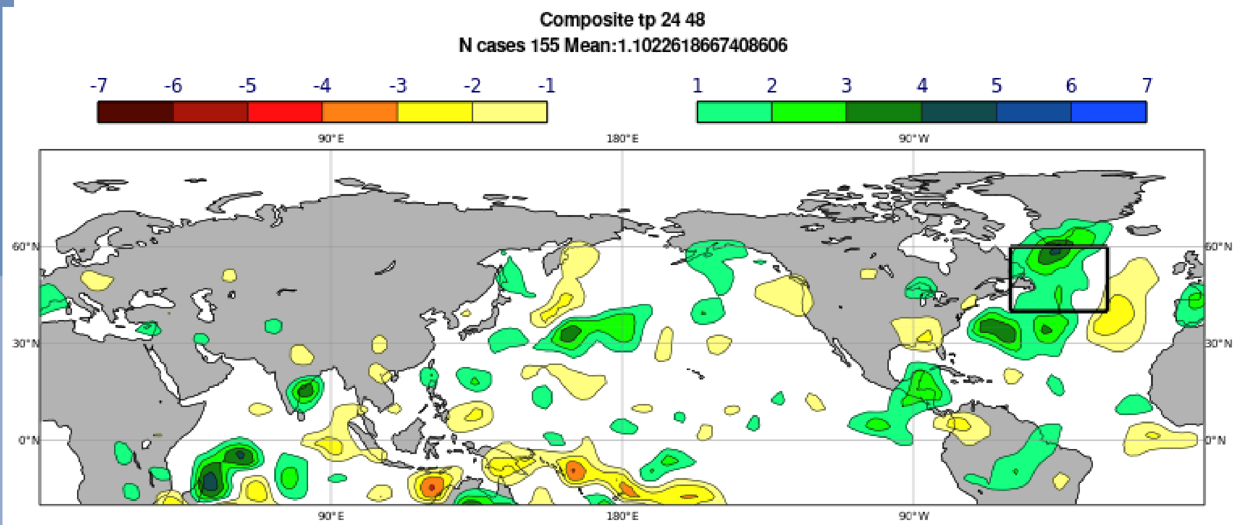
[RMS(AIFS-AIFS12)$_{48}$ - RMS(AIFS-AIFS12)$_{24}$] -
[RMS(HRES-HRES12)$_{48}$ - RMS(HRES-HRES12)$_{24}$]

Blue = slower "error" growth in AIFS

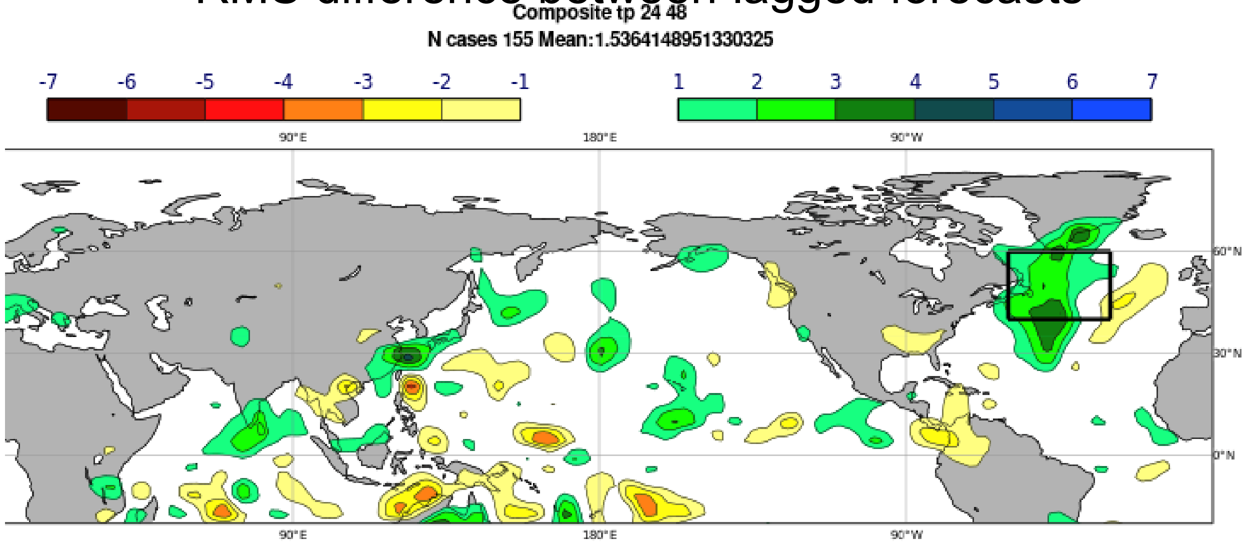


24h-48 h precipitation anomaly composites of cases with large difference in error growth (based on IFS precipitation)

RMS error



RMS difference between lagged forecasts



Reference composite based on dates shifted 10 days

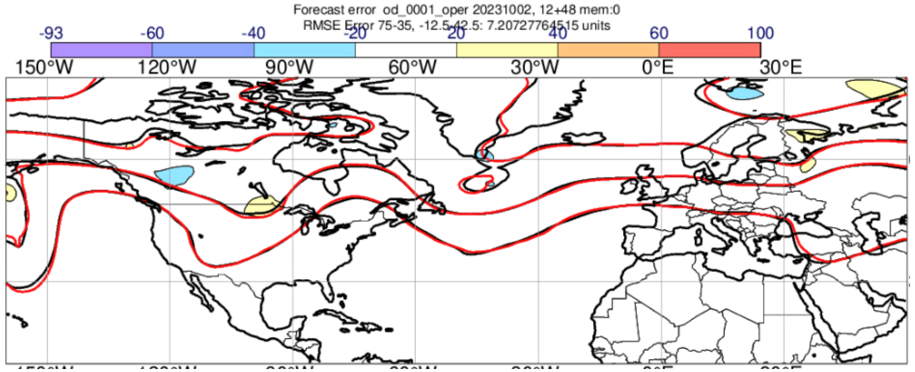
Example of bust in IFS – z500 from 20231005 12UTC

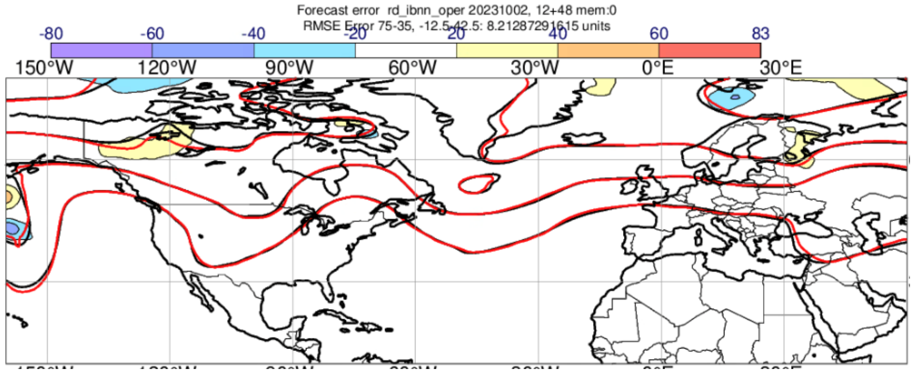
IFS

AIFS

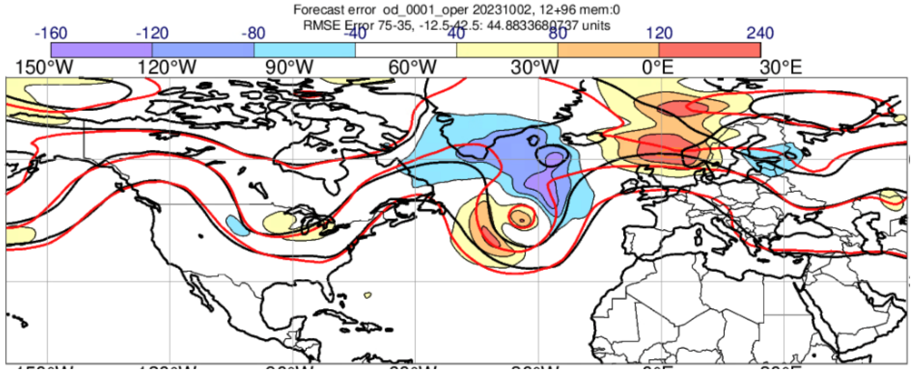
Forecast –black
Analysis –red
Error - shading

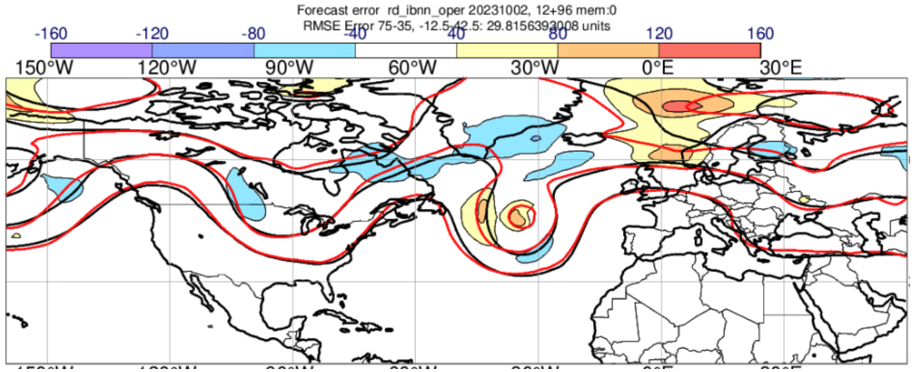
+48h



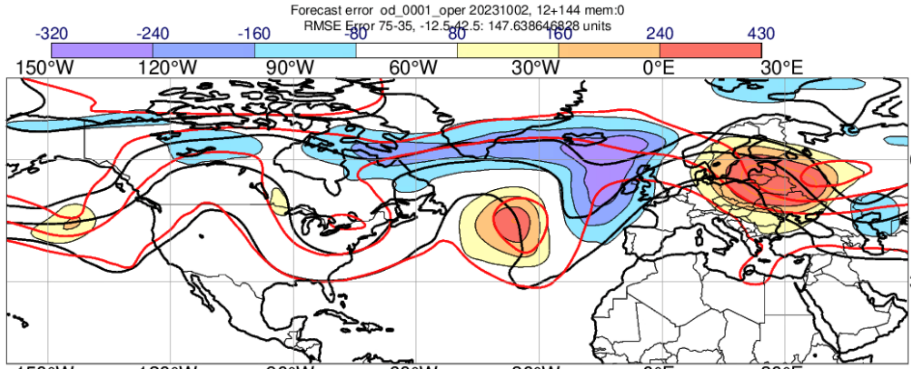


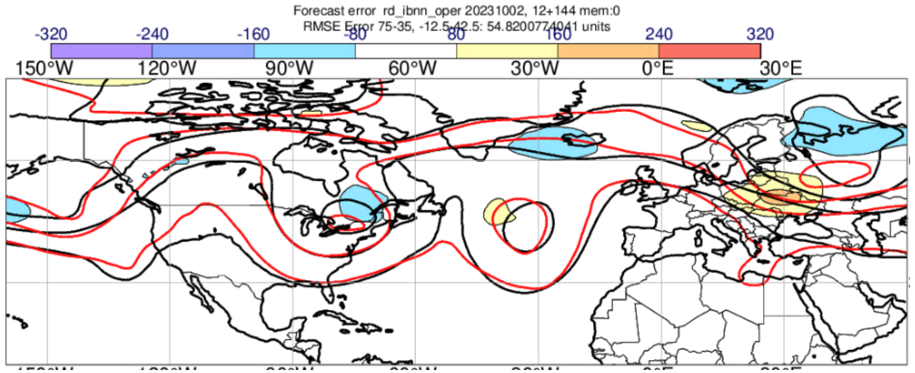
+96h



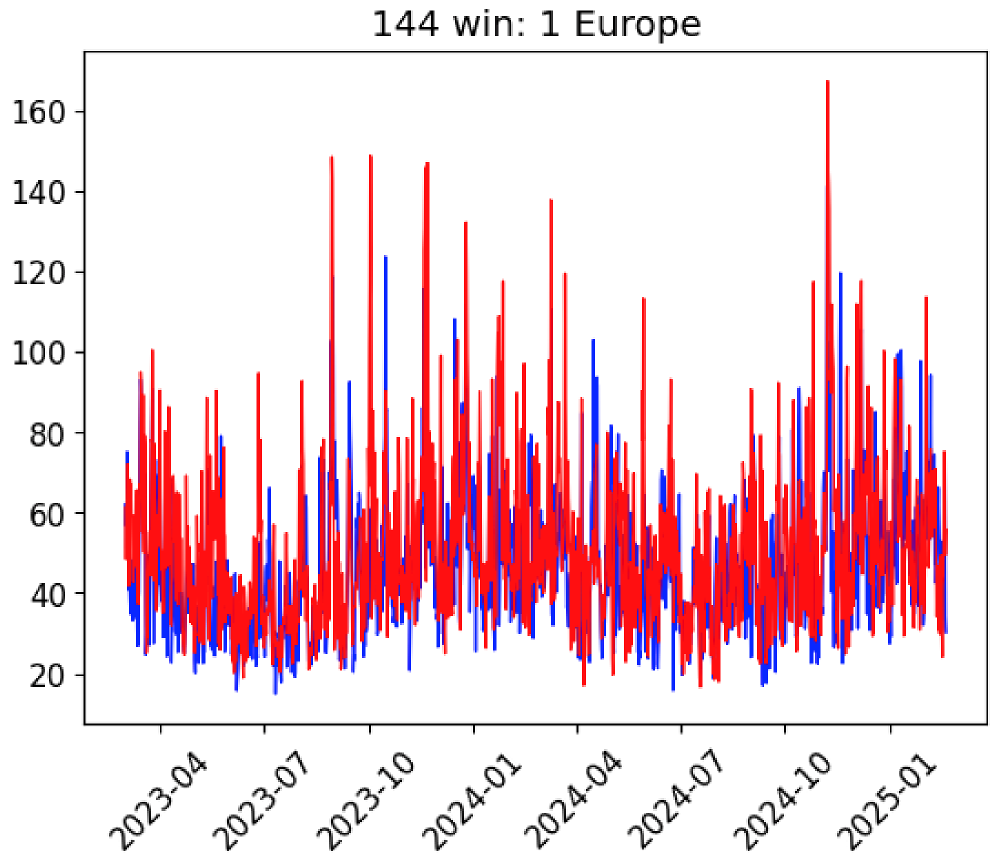


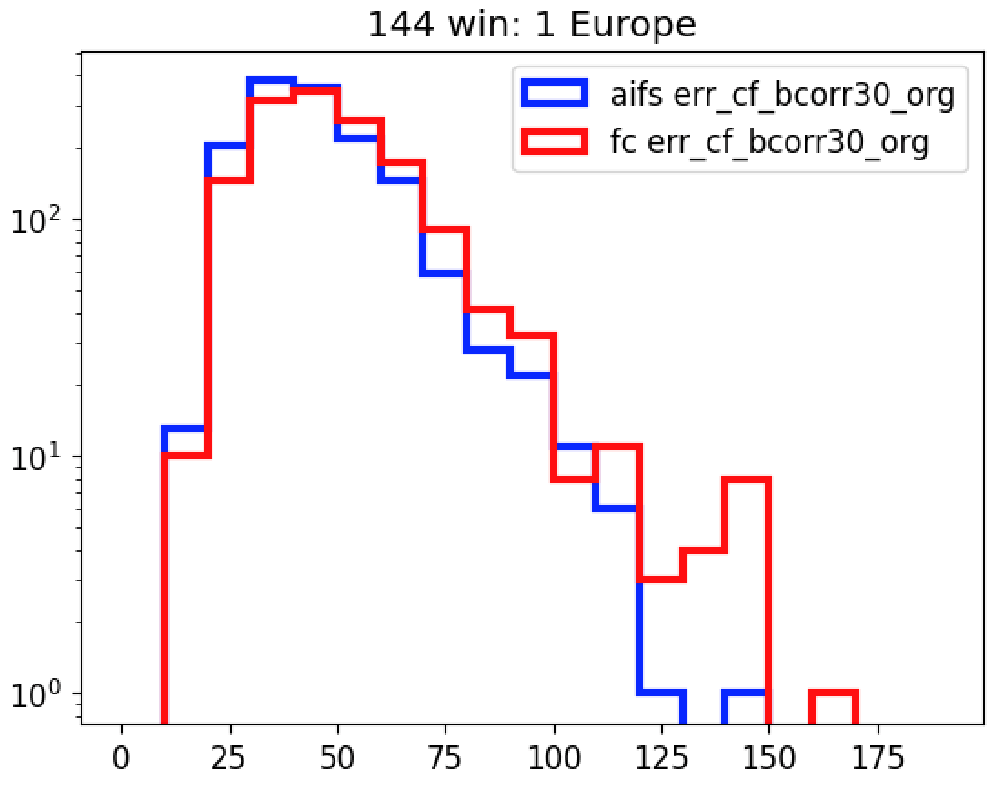
+144h





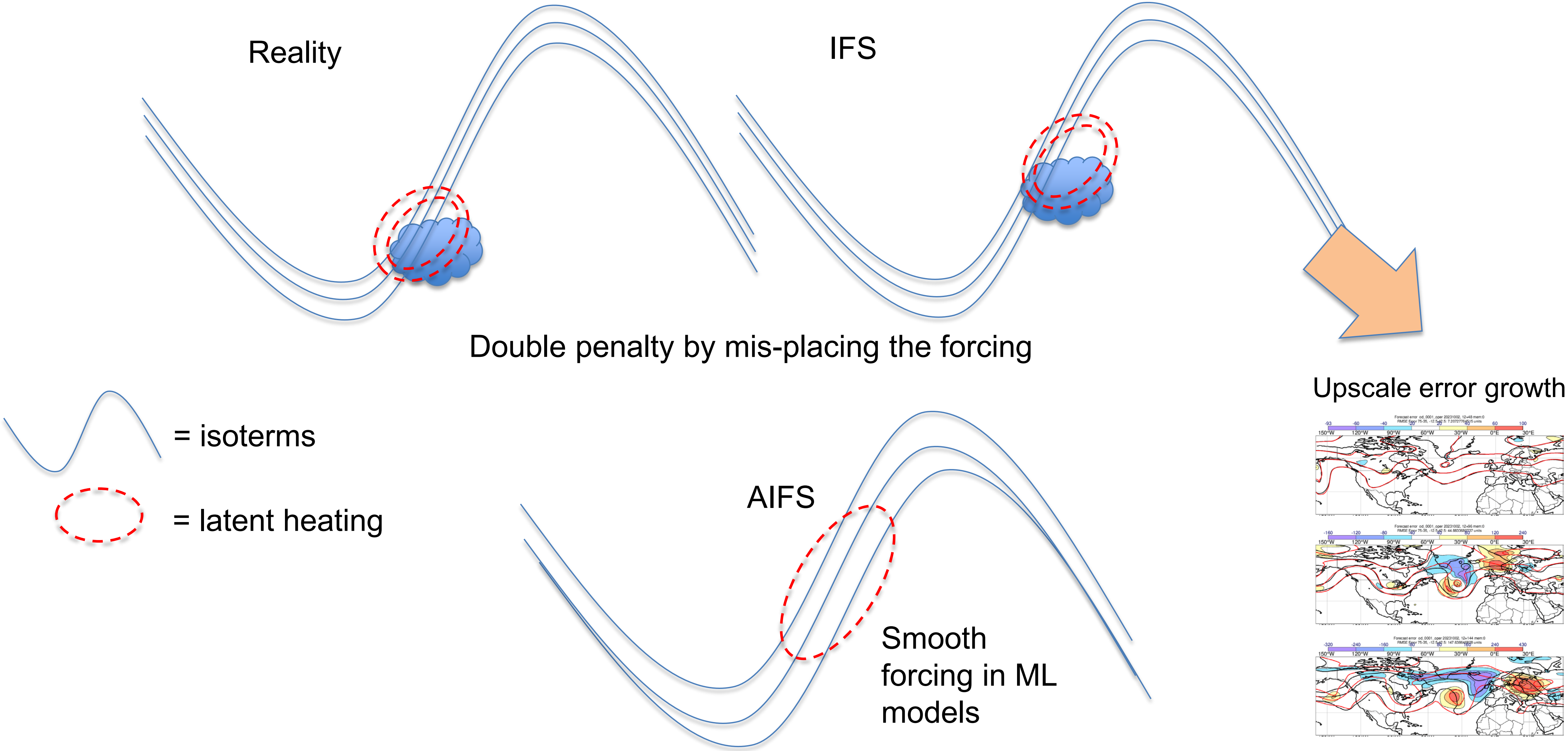
Frequency of large errors over Europe day 6





I hope you watched the poster from Tobias Selz!

Conceptual idea of reduced upscale error growth in AIFS

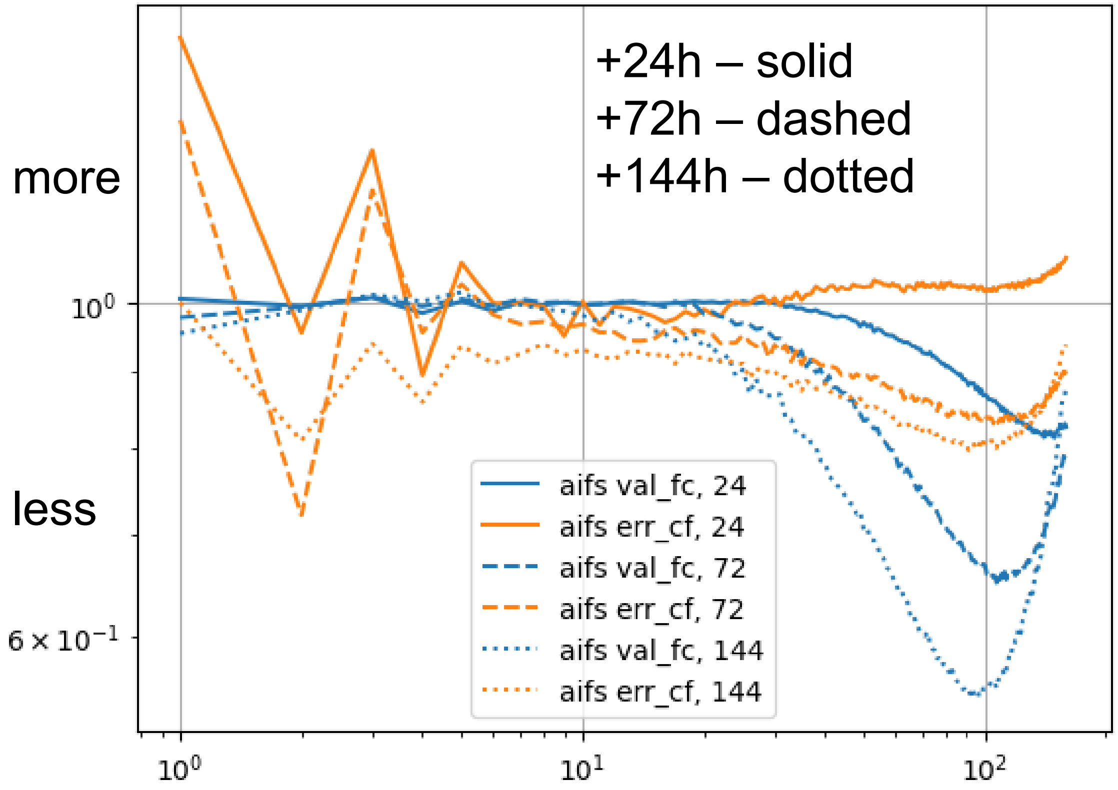


Spectral view on forecast activity, error and lagged difference (z500, global)

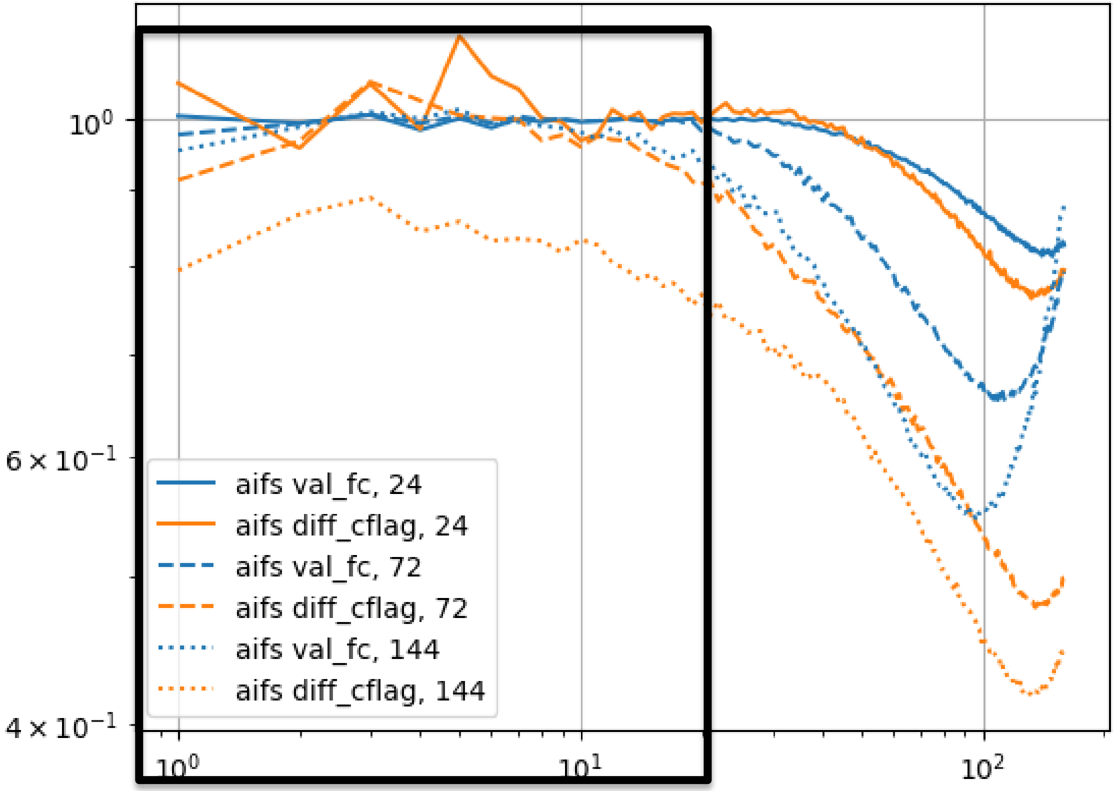
Difference in error spectra (orange)
Difference in forecast activity (blue)

AIFS more

AIFS less



Difference in lagged fc diff spectra (orange)
Difference in forecast activity (blue)



Tiny damping in AIFS vs. IFS at 144h in synoptic scales,
But large difference in lagged-fc difference

Summary error growth (ongoing work…)

- AIFS is less chaotic than IFS (as seen in forecast jumpiness)
- A slower chaotic growth in AIFS seems to be linked to precipitation events
- Less forecasts busts over Europe in AIFS
- AIFS mainly damped activity on wave number >20, but the lagged forecast difference is much less for wave number <20 at step 144h

- Does slower upscale error growth from e.g warm-conveyor belts explain the lower large-scale errors in AIFS? (See e.g Grams et al., 2018 for mechanism)
- Due to smoothing in unpredictable scales OR due to lower systematic errors during these conditions?