# ECMWF update

## WGSIP 26

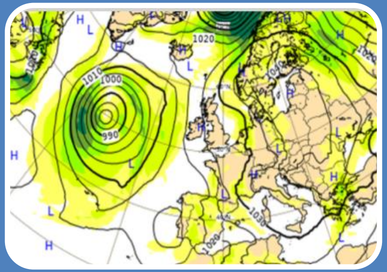
Stephanie Johnson representing the Earth System Predictability Section

Magdalena Alonso Balmaseda, Tim Stockdale, **Frederic Vitart**, Chris Roberts, Steffen Tietsche, Antje Weisheimer, Matthias Aengenheyster, Daniel Befort, Gokun Dai, Jonathan Day, Michael Mayer, Charles Pelletier, Jakob Schloer, Retish Senan, Kristian Strommen, Joshua Talib

Nov. 5, 2024

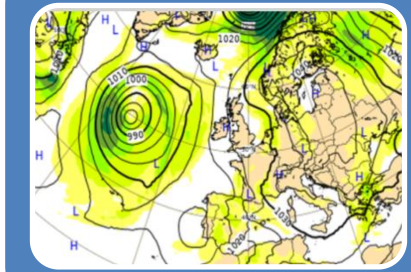
s.johnson@ecmwf.int

# Forecast systems at ECMWF



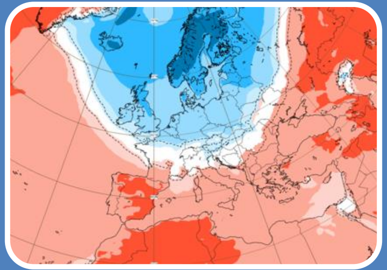
## Medium range

- 0-15 days
- 9 km resolution, 137 levels
- 51 ensemble members, run twice daily
- Upgraded approximately once a year



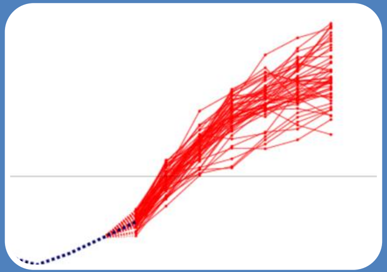
## AIFS: Medium range

- Experimental real-time system
- 0-15 days
- ~0.25 degree resolution
- Deterministic and ensemble versions
- Upgraded as developments are ready



## Extended range

- 0-46 days
- 36 km, 137 levels
- 101 ensemble members, run once daily
- Upgraded approximately once a year, with medium range



## Long range: SEAS5

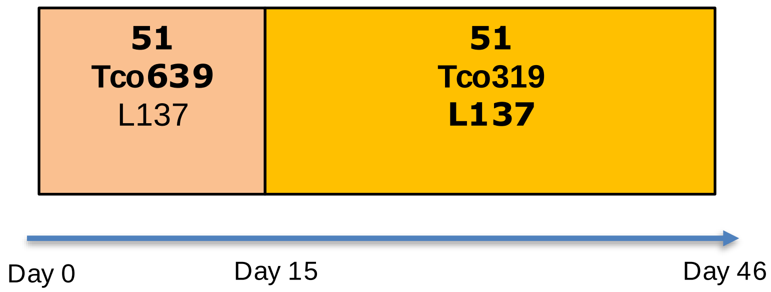
- 0-7 months
- 36 km, 137 levels
- 51 ensemble members, run once a month
- Four times a year, the forecast is run out to 13 months
- Last upgraded in 2017 (Cy43r1), next upgrade in 2025

A seamless forecast model underpins all dynamical systems (at implementation).

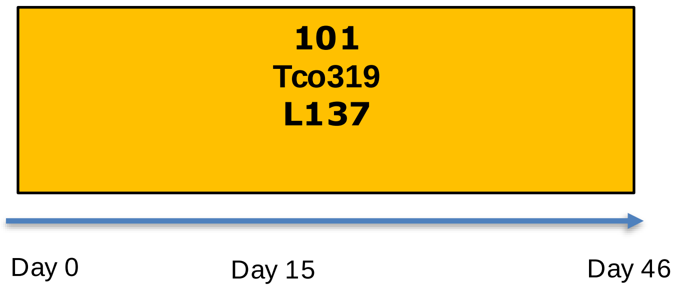
# Upgrades to operational systems: extended range configuration

- **Extended range forecast configuration has changed**
  - From June 2023 (Cy48r1), real-time forecast ensemble size increased to 101 members and issued daily.
  - To facilitate this, extended range initialized from day zero rather than as an extension of the ENS.
- **Reforecast configuration will also change**
  - From Nov 2024 (Cy49r1), ten perturbed and one control member will be run on every odd day (excluding 29 Feb).
  - This gives:
    - 75% more reforecasts, benefitting skill assessment and calibration.
    - Direct comparisons between re-forecasts produced in different years and with different systems.
    - Opportunities for combining medium and extended range ensembles to produce calibrated dual-resolution ensemble products.
- **Several notable upgrades in November implementation (Cy49r1)**
  - New land surface initialization, new stochastic physics scheme (SPP)

**47r2 configuration**




**48r1 configuration**



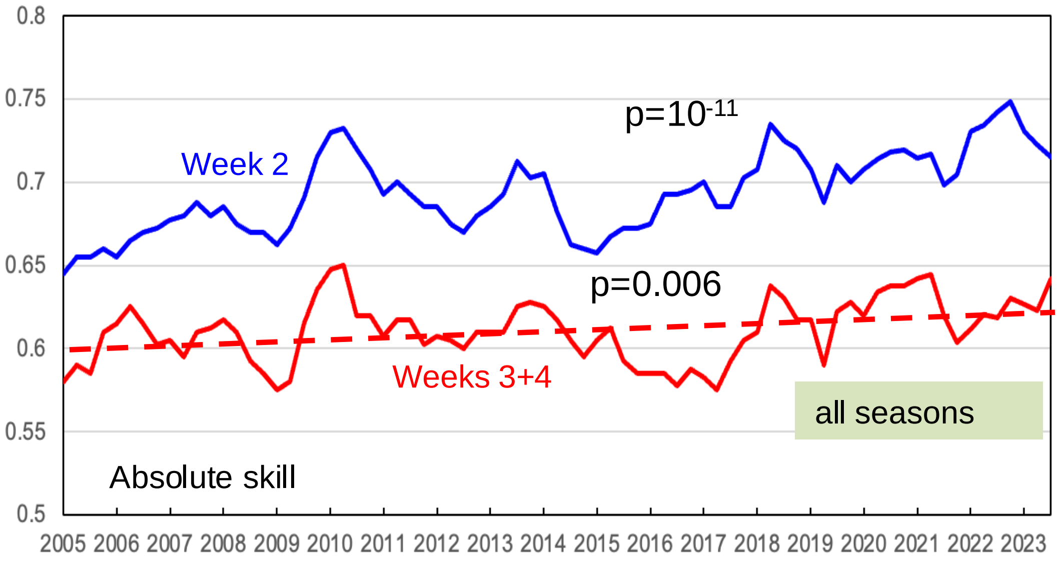

*F. Vitart*

# 20 Years of Sub-seasonal prediction:

## Are we improving?

## Real-time forecasts

ROC area (upper tercile) T2m NH extratropics



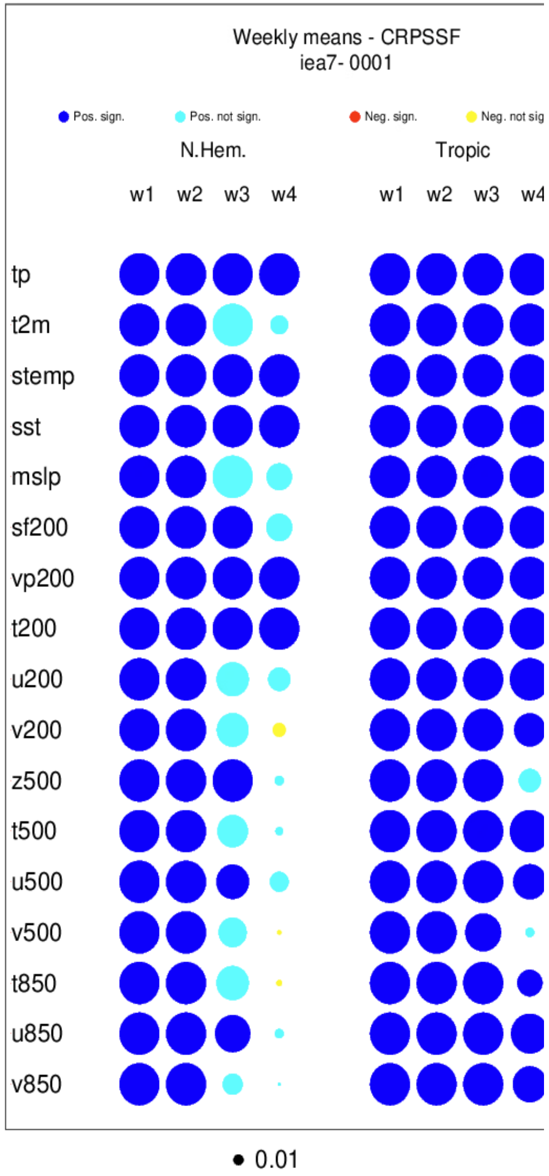

Significant improvement in week 2 (day12-18)
No significant improvement over persistence for Weeks 3+4 (day 19-32).

*F. Vitart*

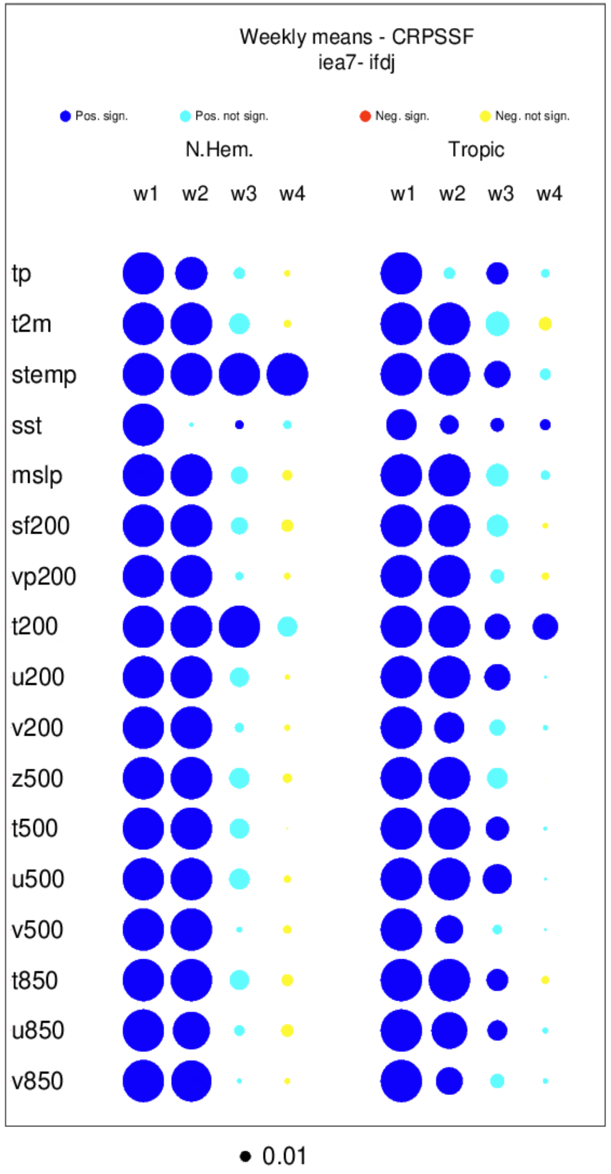
ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WE

## Reforecasts

2024 vs 2004 version




Same version (2024)
ERA5 vs ER40 Initialization



# Developing SEAS6

- **Ocean and atmosphere model upgrades:** IFS, NEMO4, SI3
    - Clear improvement in the stratosphere since SEAS5 from increase in vertical levels, hybrid linear ozone scheme…
    - New stochastic physics scheme (SPP, Lang et al. 2021, QJRMS)
    - NEMO4/SI3 – New version of the ocean model and an upgrade to the SI3 multi-category sea ice model

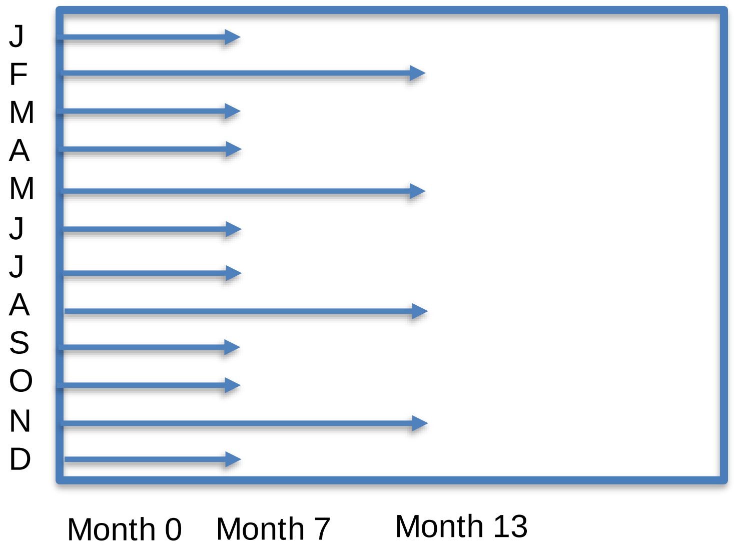
- **New initialization datasets:**
    - OCEAN6 - New version of the ocean reanalysis system including SST assimilation using an ensemble data assimilation approach. (H. Zuo)
    - New land surface initialization including data assimilation of soil moisture and snow – Cerise project (J. Day and coupled assimilation team)

- **Improved forcings:** CMIP6 GHG, volcanic aerosol, tropospheric aerosol
    - Generating time varying tropospheric aerosols for ERA6 and SEAS6 (CONFESS, R. Senan, T. Stockdale)
    - Volcanic aerosol parameterization for SEAS6 (CONFESS, T. Stockdale)

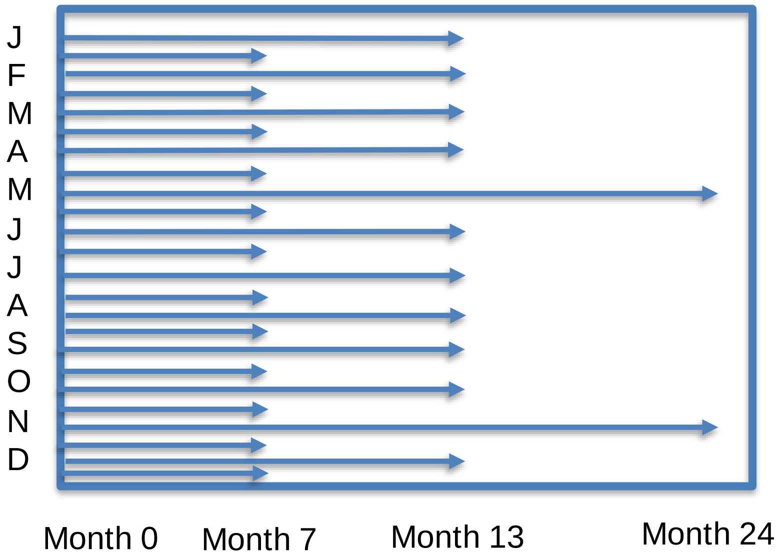
# SEAS6 configuration – Implementation 2025

- **Enhancement 1: Real-time 101 member ensemble**
  - Currently 51 members
- **Enhancement 2: Issue SEAS twice per month**
  - Initial date 1$^{st}$ and 16$^{th}$ of each month
- **Enhancement 3: Expand annual-range ENSO forecasts**
  - Issue forecast monthly not quarterly
  - Twice per year, increase range to 24 months
- **Enhancement 4: More comprehensive reforecasts**
  - Larger ensemble sizes and larger set of years
  - Some reforecasts will extend to the 1960s
  - Bias correction of products will use the recent period (1993-2022) for consistency with C3S and improved anomalies relative to a changing climate

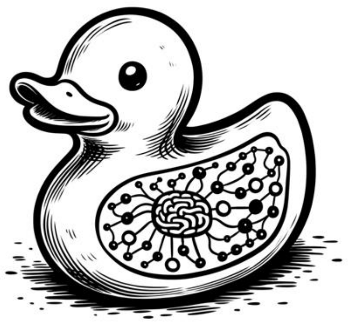
**SEAS5**




**SEAS6**



# AIFS - Artificial Intelligence Forecasting System



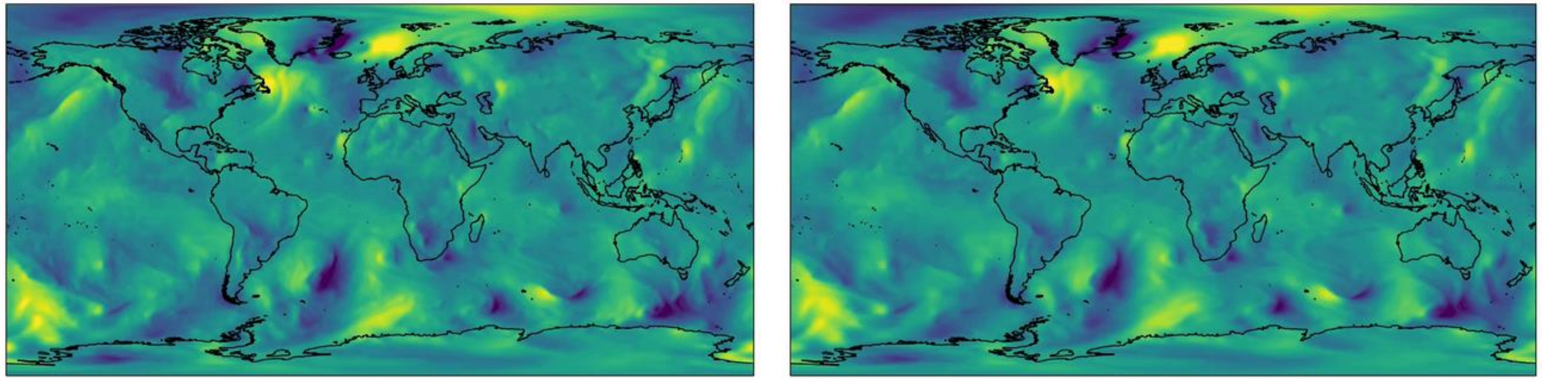
First implementation (~ 1deg resolution) in 2023, following Keisler 2022 and Lam et. al 2022:

- GNN architecture: Interaction Networks (Battaglia et. al 2016)
- Graph representation, hidden multi-scale mesh, edge features
- Scales to > 1000s of GPUs ; tensor parallel implementation, split model across multiple GPUs

Update beginning of 2024, update to ~ 0.25 deg:

- Attention / Transformer based GNN for encoder, decoder (Shi et al., 2021)
- Transformer backbone in processor (with sliding window, e.g. Child et al. 2019, Jiang et al. 2023)
- Trained on 64 GPUs ~ 1 Week, using ERA5 data (atmosphere only)

IFS



AIFS

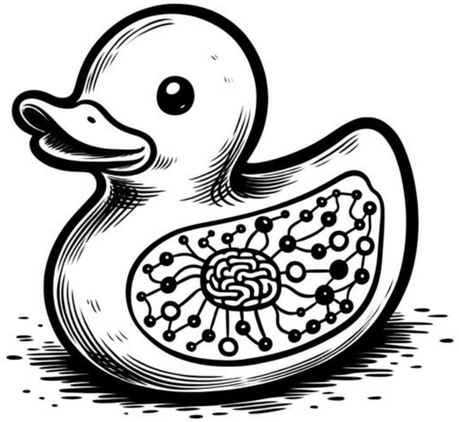
Why GNN: Encoder / Decoder: can handle arbitrary input / output grids, local and ad hoc grid refinement, changing grids etc. ; attractive for use in earth system science

See Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C. A., Lessig, C., Maier-Gerber, M., Magnusson, L., Bouallègue, Z. B., Nemesio, A. P., Dueben, P. D., Brown, A., Pappenberger, F., Rabier, F. (2024). AIFS - ECMWF's data-driven forecasting system. arXiv. https://doi.org/10.48550/ARXIV.2406.01465

*Simon Lang*

# Future directions in data-driven forecasts

- **Current machine learning activities at the sub-seasonal time-scale**
  - Assessment of AIFS developments at the extended range
  - Including other fields from other model components (e.g. SST and sea ice cover) in AIFS training
  - AI/ML post-processing methods (collaboration with S. Lerch)
  - Predictability studies using relaxation methods applied to data driven models
- **Activities at the seasonal timescale are at an early stage**
  - Ensemble AIFS systems may be more appropriate for seasonal timescales (either diffusion denoiser or CRPS trained)
  - Hybrid modelling options coupling data-driven model components to physical components
- **WeatherGenerator project**
  - Aims to develop a foundation model that could be used in many tasks, including sub-seasonal and seasonal forecasting



# Other activities

- **Future development of operational systems**
  - Sensitivity of the medium range to ocean initial condition resolution
  - Testing SPP in sea ice
  - Influence of ocean eddies, Tco1279 and Tco399 AMIP experiments with filtering applied to smooth ocean eddies (EERIE)
  - Development of land-atmosphere coupling diagnostics for evaluation of sub-seasonal to seasonal forecast systems (Cerise)
  - Bridging gaps in Seasonal to Decadal forecasts and Decadal forecasts to climate projections using 2-year predictions and the with the 30-year initialized climate outlooks (ASPECT)
  - Machine learning activities

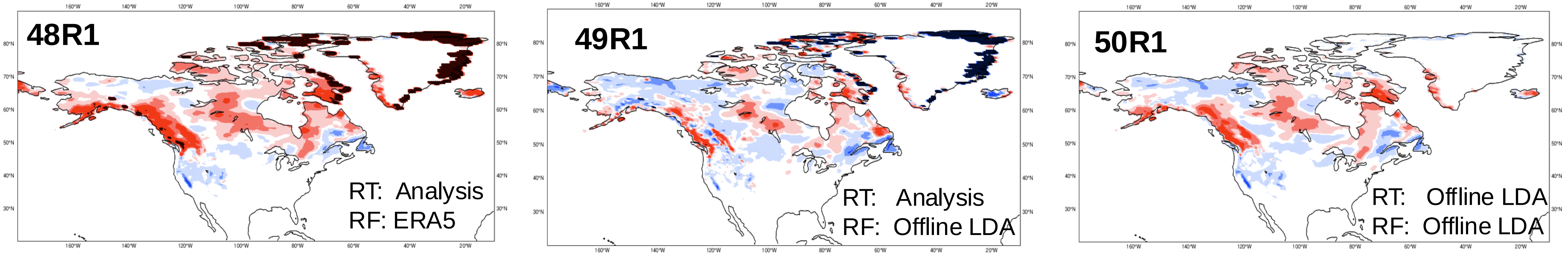
- **Investigating model performance**
  - Investigating influences on the record breaking global-mean temperatures of 2023
  - Investigating variability in Earth's energy imbalance, quantified via TOA net radiative flux
  - Investigating long term trends
  - Ocean Observations in S2S (coordinated experimentation as part of SynOBs program contributing to the Ocean Decade)
  - Tropical-extratropical teleconnections and their interaction

- **Development of end user products**
  - WMO lead centre for S2S reforecast evaluation and QED
  - Improve the quality of Arctic forecasts across Copernicus services (ACCIBERG)
  - End user NH sea ice products (ICECAP)

# Extra slides

# Upgrades to operational systems: extended range land initialization

**Snow depth anomaly 1st Jan 2021 – Week 4**




Offline land surface data assimilation system (LDAS) is now used to assimilate soil moisture (L1-3) and snow water equivalent for both reforecast and forecast initial conditions. More consistent initialization across reforecasts and forecasts gives more accurate anomalies early in the forecast.

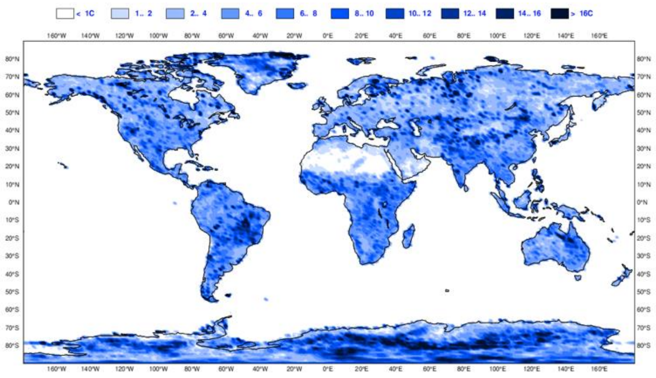
*F. Vitart*

# New land initialization for sub-seasonal to seasonal timescales

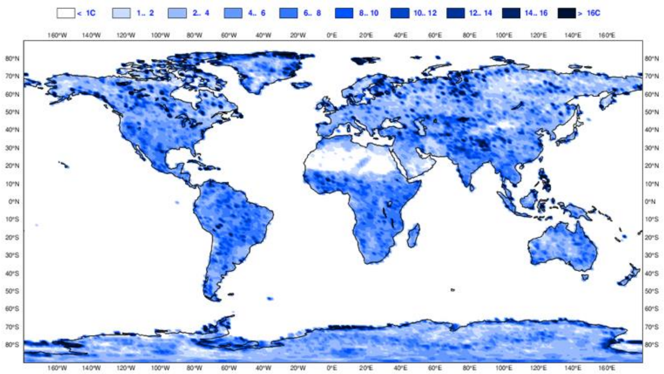
Offline land surface data assimilation system (LDAS) is now used to assimilate soil moisture (L1-3) and snow water equivalent for both reforecast and forecast initial conditions. More consistent initialization across reforecasts and forecasts gives more accurate anomalies early in the forecast.

Soil Water Level 1

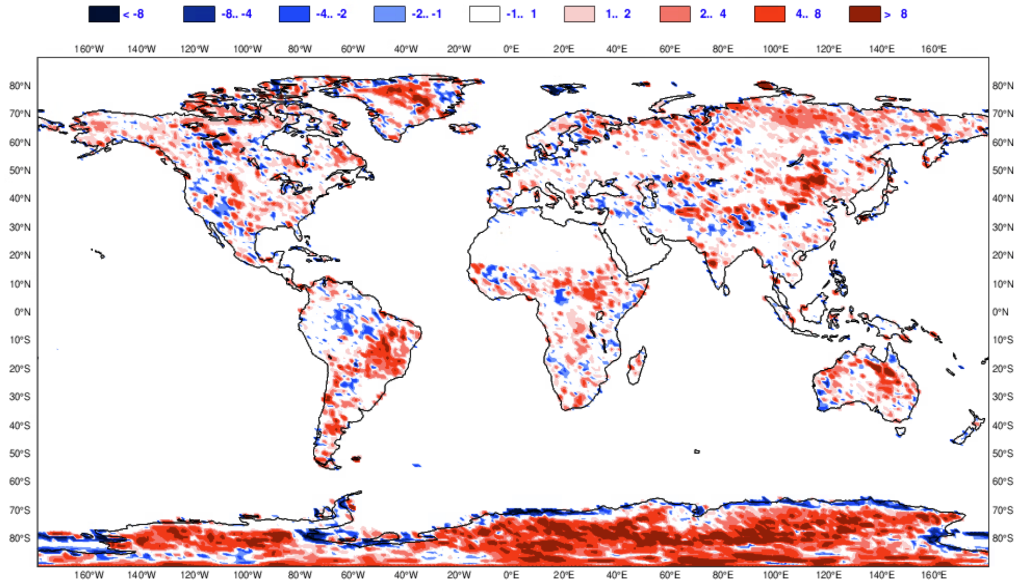
RMSE ERA5 relative to 49r1 e-suite



RMSE offline LDAS relative to 49r1 e-suite analysis



RMSE ERA5 – RMSE offline LDAS



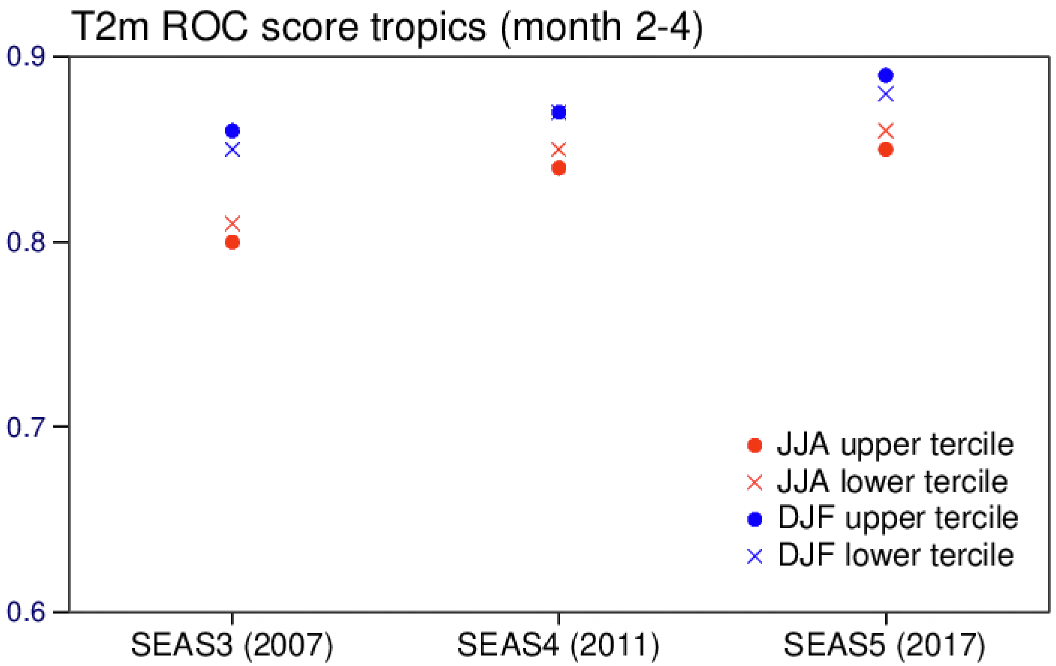
More consistency in LDAS

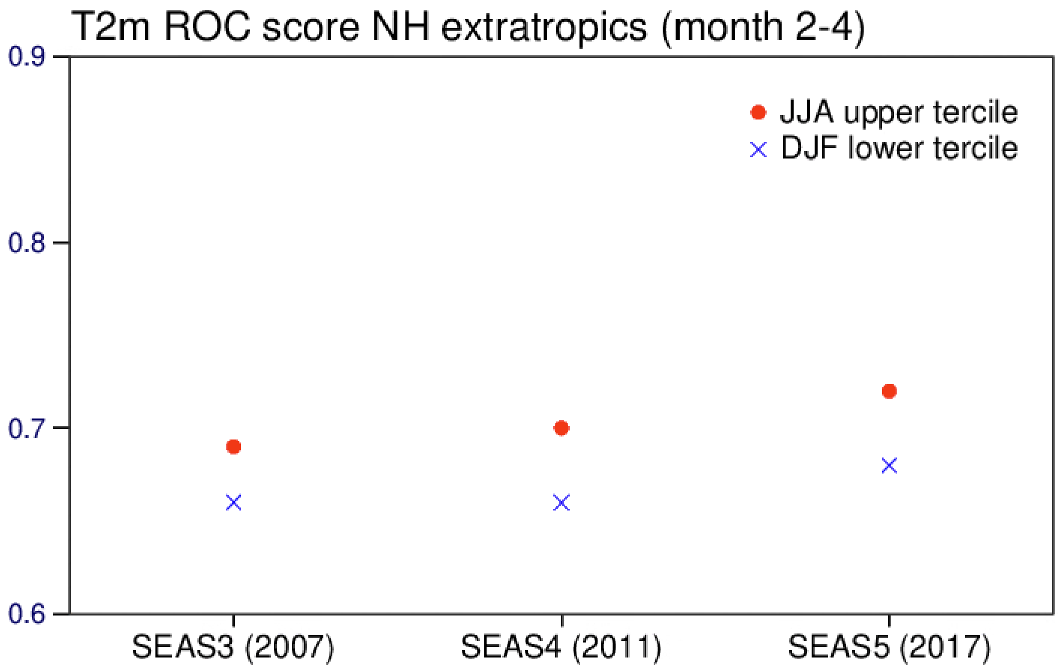
Less consistency in LDAS

Ratio = 0.9952

# Seasonal prediction:
## Are we improving?

*May 1, Nov 1 reforecasts over 1981-2011*
*SEAS3: 41 members SEAS4,5: 51 members*



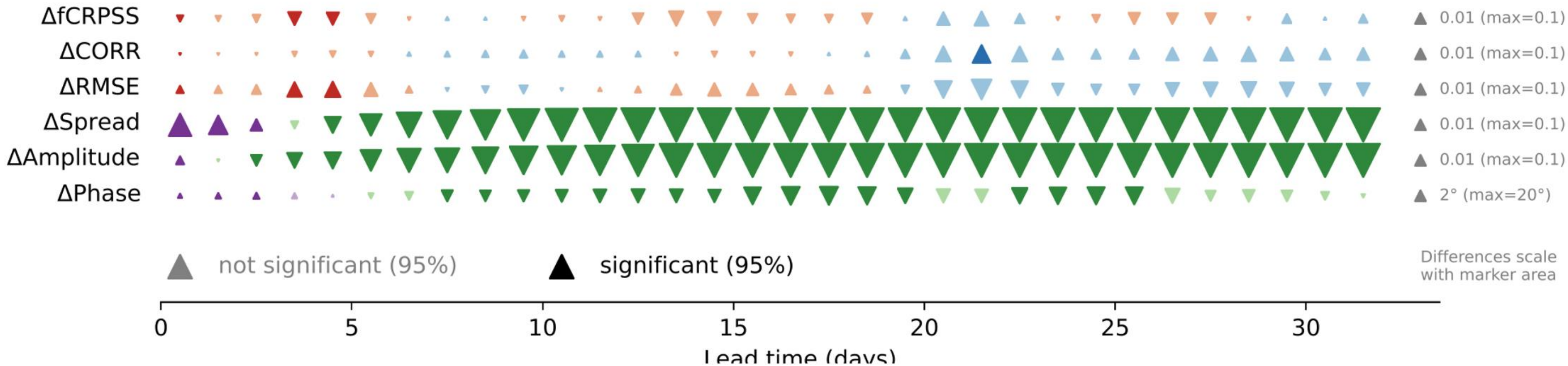



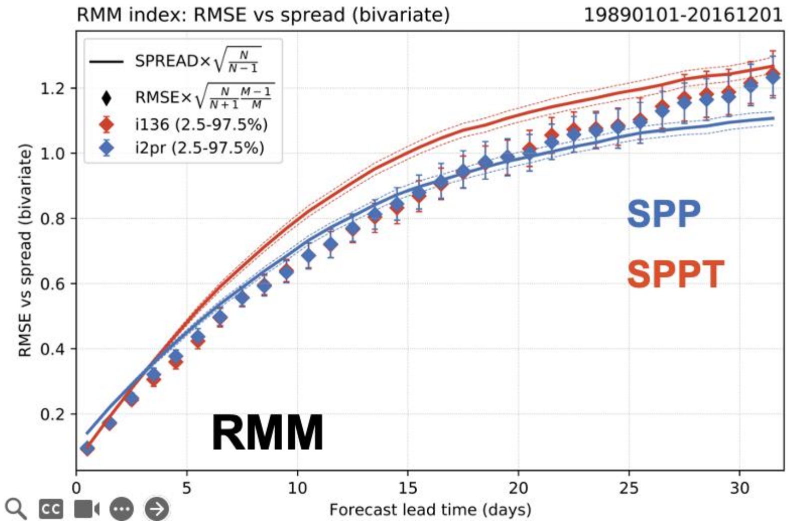

Steady, small improvements over last three systems… SEAS6 due in 2025.

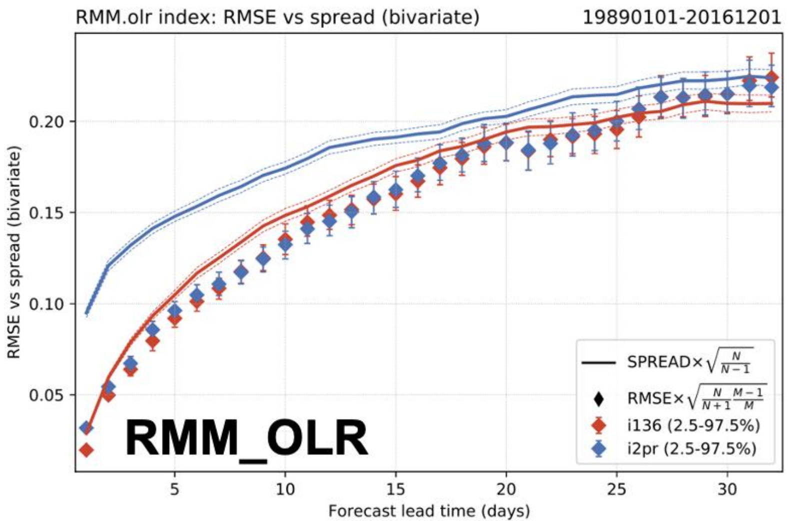
# SPP vs SPPT – impact on MJO spread/error

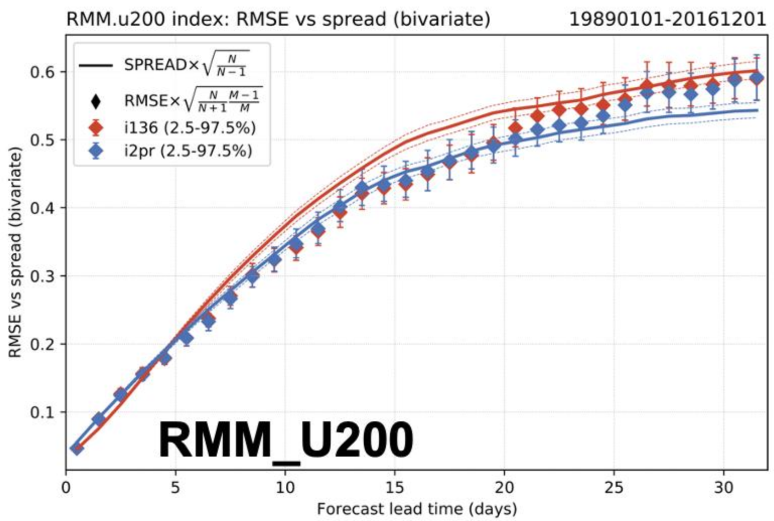
- SPP reduces spread (and thus average member amplitude) in bivariate RMM index but scores ~neutral.
- Spread/error improved overall, especially for u200/u850 components that dominate RMM index.
- However, small degradation during days 1-5 associated with over-dispersion of OLR component.

**SPP vs SPPT (49R1_v8 vs 49R1_v7) - bivariate RMM scorecard (PF only)**



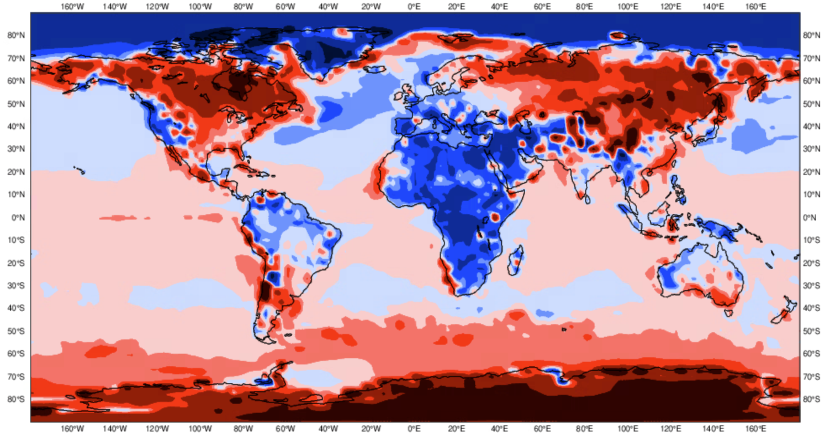
# Biases relative to ERA5

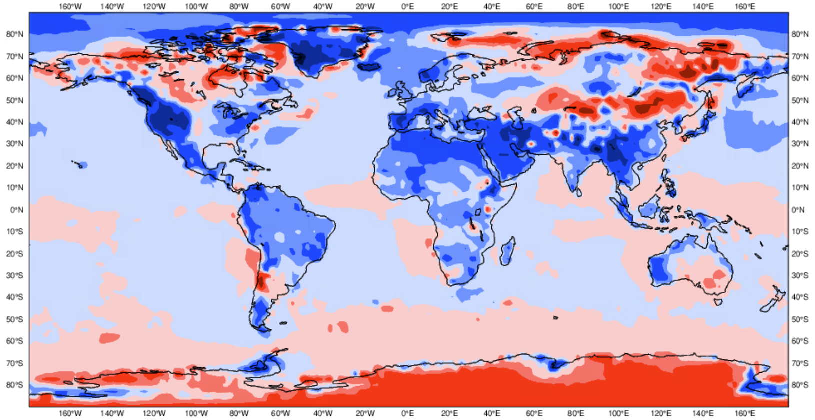
## 2-metre temperature

### Day 5-11

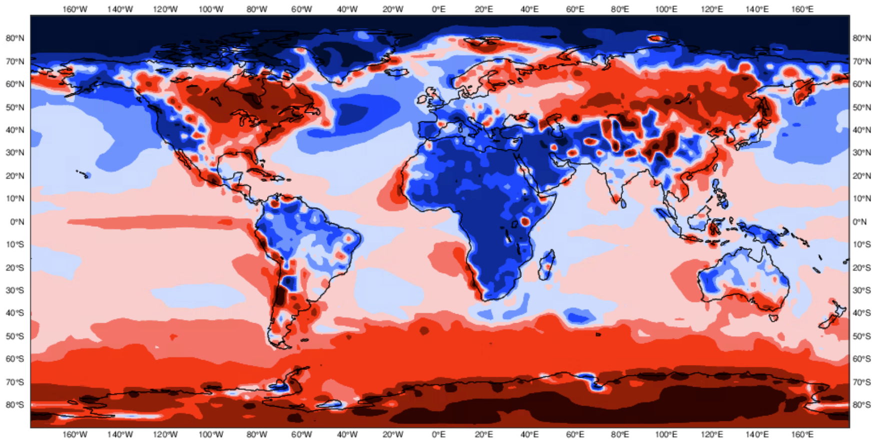
2004 version (28R1)

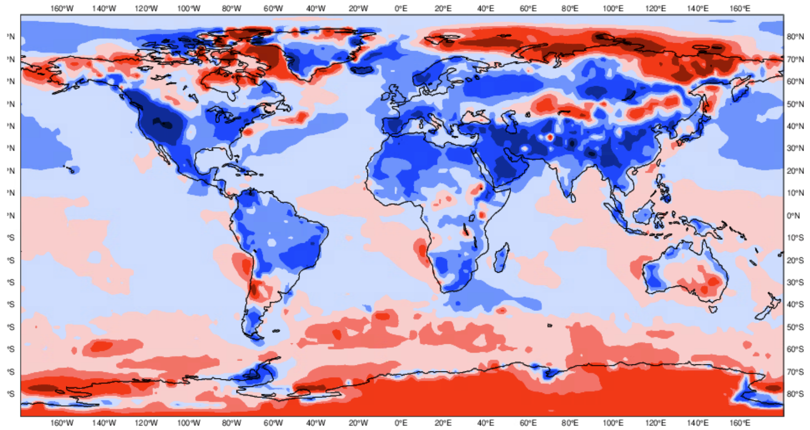
2024 version (48R1)





### Day 26-32






48r1 Reforecasts
Same re-forecast period and start dates as 2004 re-forecasts.

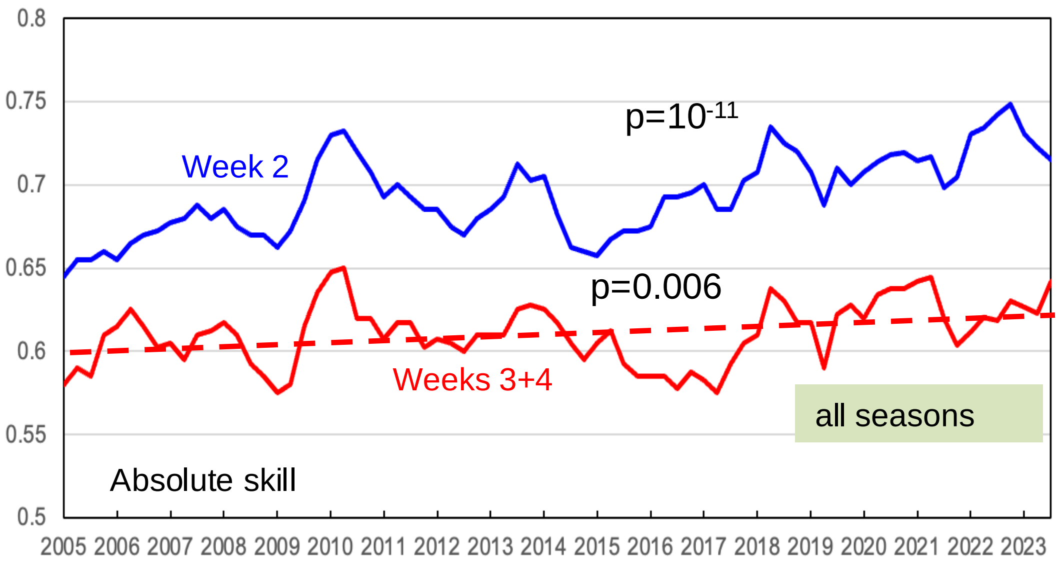
*F. Vitart*

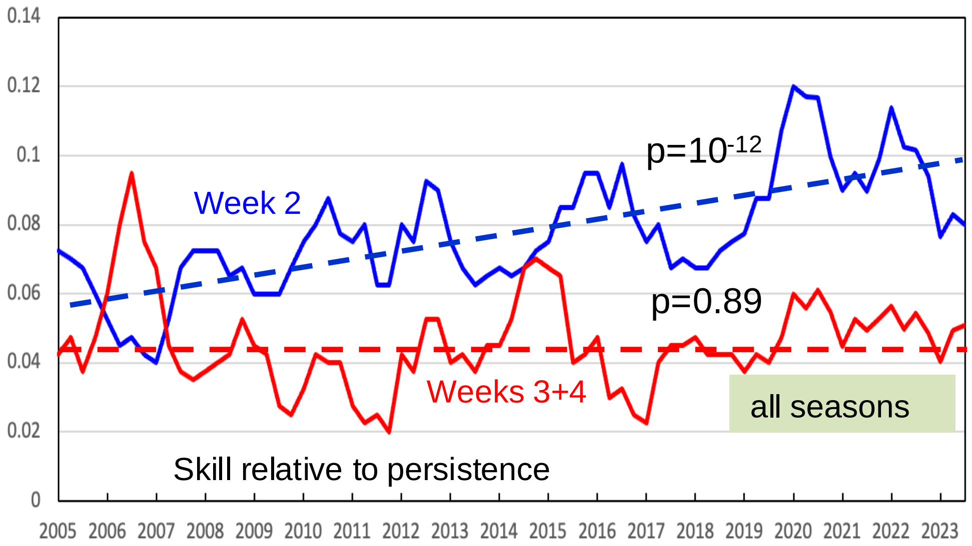
# 20 Years of Sub-seasonal prediction

## Forecast skill. Are we improving?

### Extratropics

T2m ROC area (upper tercile)




Significant improvement in week 2 (day12-18)
No significant improvement for Weeks 3+4 (day 19-32) !

# AIFS training

## 3 Training

AIFS is trained to produce 6-hour forecasts. It receives as input a representation of the atmospheric states (ERA5 or ECMWF's operational analysis) at $t_{-6h}$, $t_0$, and then forecasts the state at time $t_{+6h}$. The full list of input and output fields of AIFS is shown in Table 1.

| Field | Level type | Input/Output |
|---|---|---|
| Geopotential, horizontal and vertical wind components, specific humidity, temperature | Pressure level: 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, 1000 | Both |
| Surface pressure, mean sea-level pressure, skin temperature, 2 m temperature, 2 m dewpoint temperature, 10 m horizontal wind components, total column water | Surface | Both |
| Total precipitation, convective precipitation | Surface | Output |
| Land-sea mask, orography, standard deviation of sub-grid orography, slope of sub-scale orography, insolation, latitude/longitude, time of day/day of year | Surface | Input |

Table 1: Input and output variables of AIFS.

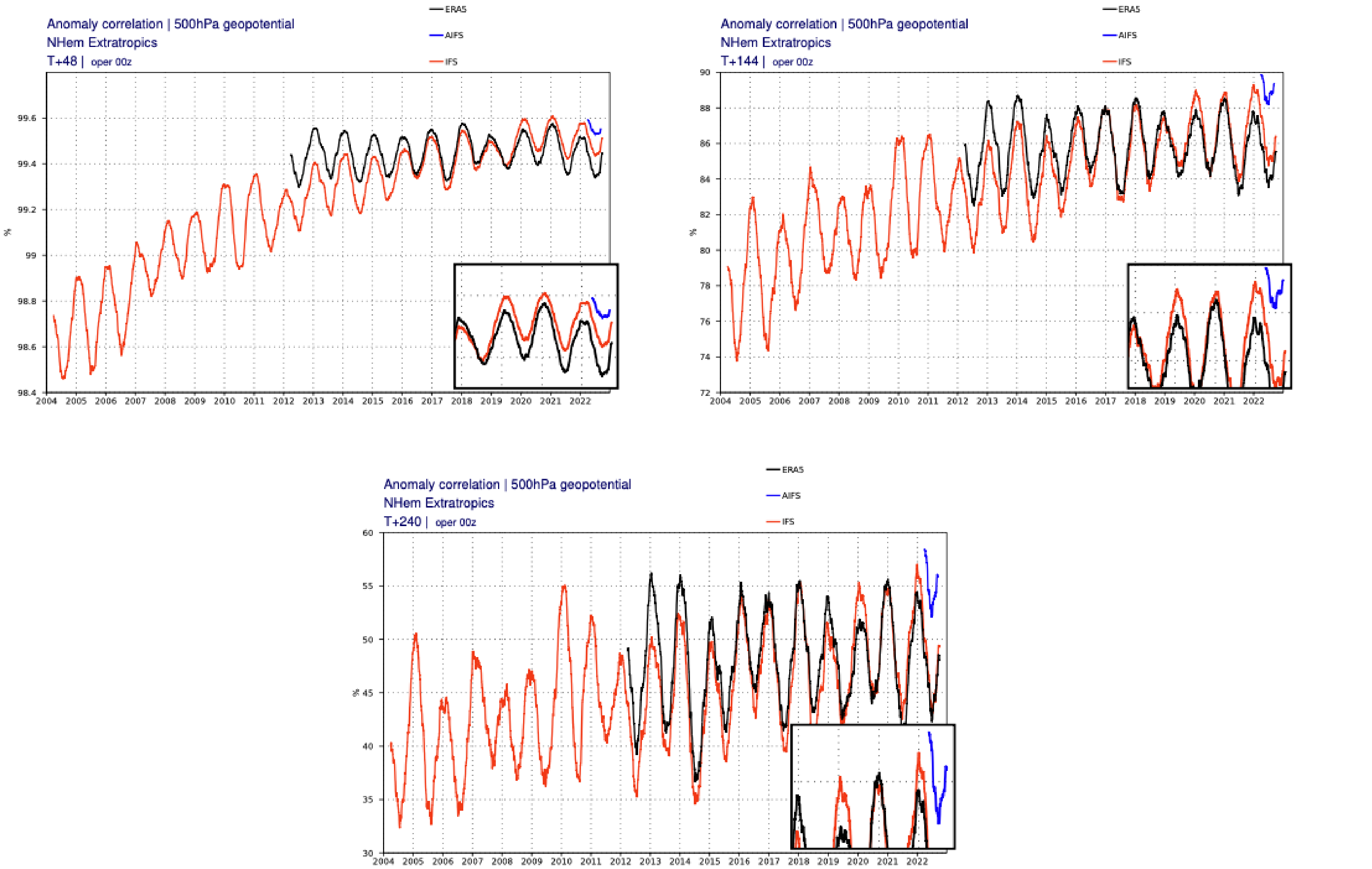
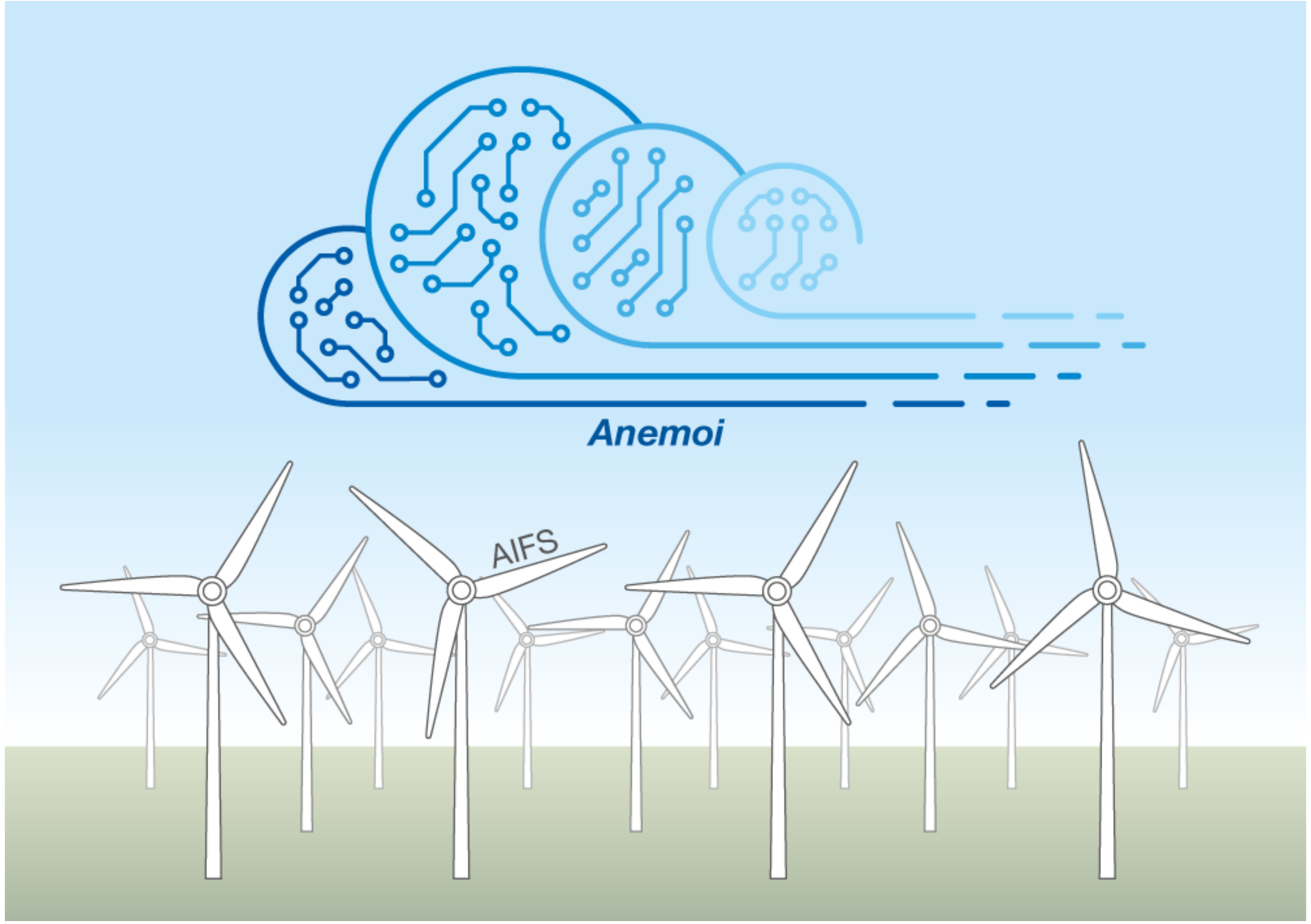



Figure 4: Northern Hemisphere ACC (anomaly correlation) of geopotential at 500 hPa for AIFS (blue) and IFS (red) and ERA5 (black) forecasts for different years: 2 day forecasts (top left), 6 day forecasts (top right) and 10 day forecasts (bottom). Forecasts are initialised on 00 UTC each day and shown is a 30 day running mean. Insets show a zoomed-in view.

# Anemoi and AIFS

**Aspiration**
Set of tools, shared/co-developed across Europe, and beyond, for building data driven forecasting systems.
Users can bring their data and pick a suitable architecture and training method.
More advanced users can add new architectures and training methods.

Anemoi will be open source, with many pieces already being open.




ECMWF

anemoi.datasets

- Expanding functionality to meet new users demands. e.g. interfacing with Member States datasets.
- **Open source, a read the docs in progress** https://anemoi-datasets.readthedocs.io/

anemoi.inference

ai-models

- Modular framework for inference with data-driven models.
- Utilising multIO to remove IO bottleneck from inference.
- Supporting multiple input streams for LAM & stretched-grid inference.
- Plugins can handle “bespoke” complexity (i.e. ai-models-pangu, ai-models-graphcast)
- **Open source, a read the docs in progress** https://anemoi-inference.readthedocs.io/

anemoi.models

anemoi.graphs

- Models for data-driven modelling.
- GPU parallel training,
- Tools to help build & visualise graphs for graph neural networks.
- **Open source, a read the docs in progress** https://anemoi-models.readthedocs.io/

anemoi.training

- Ensembles, supporting score optimisation & diffusion training
- Dynamic graphs for training towards observations
- **Open source, a read the docs in progress** https://anemoi-training.readthedocs.io/

ECMWF



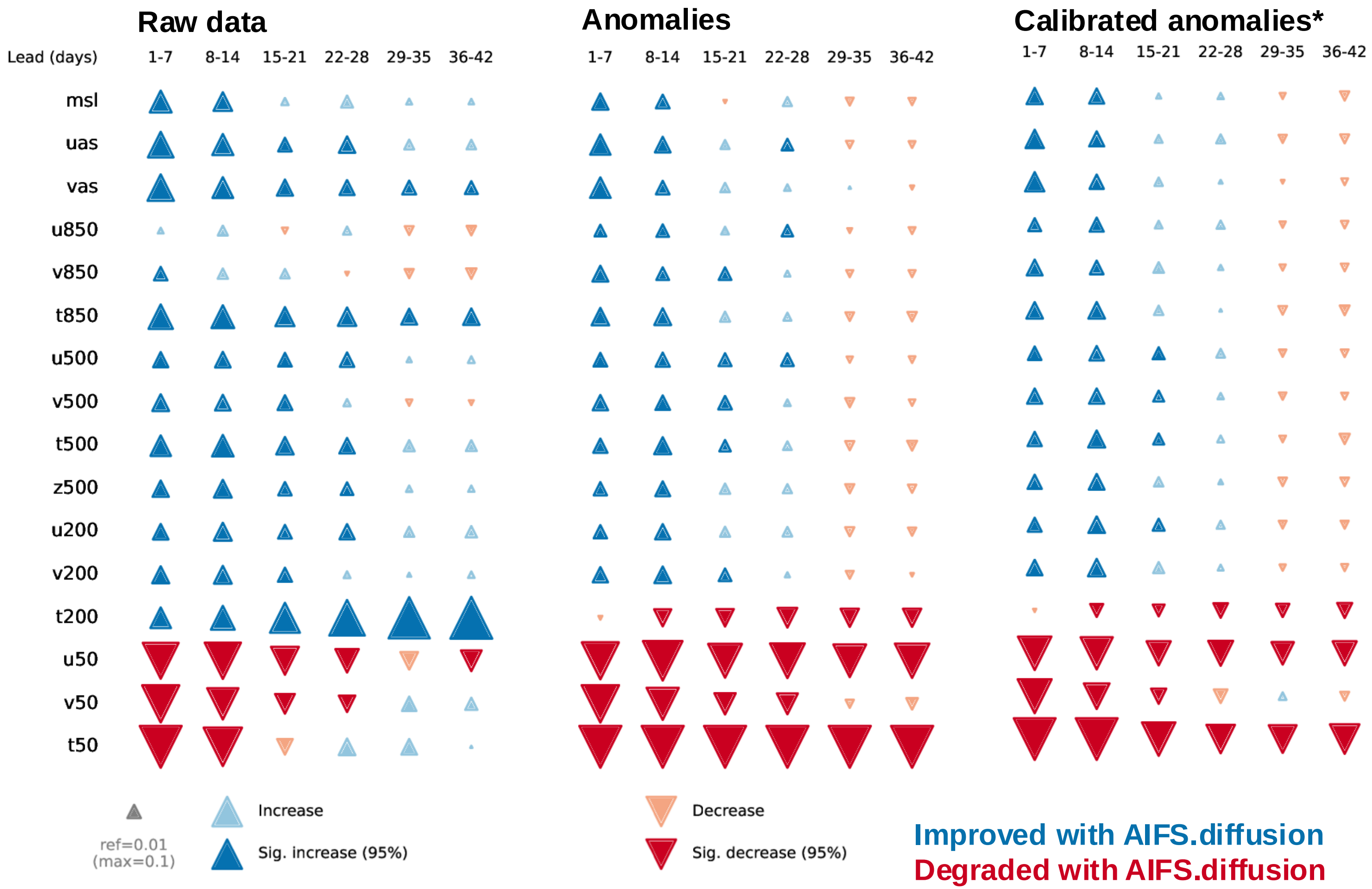
Health warning: work in progress, some functionality still being implemented …

# Preliminary evaluation of AIFS.diffusion for S2S timescales

**Reforecast configuration**
- AIFS.diffusion vs IFS (2023 47r2/48r1)
- 8 perturbed members.
- 46-days, twice per week.
- 2018-2022.

Reforecasts from Simon Lang using initial version of AIFS.diffusion (no IniPert) trained on ERA5 for medium-range and applied to S2S timescales without modification.

## ΔfCRPSS: NHEM




**Improved with AIFS.diffusion**
**Degraded with AIFS.diffusion**

**Troposphere**
AIFS.diffusion mean state is very good, sufficient to improve fCRPSS of raw forecasts relative to IFS (2023).

Skill of anomalies (or calibrated anomalies) is very similar at S2S lead times.

**Stratosphere**
AIFS.diffusion mean state is good but anomaly forecasts are significantly worse than IFS.

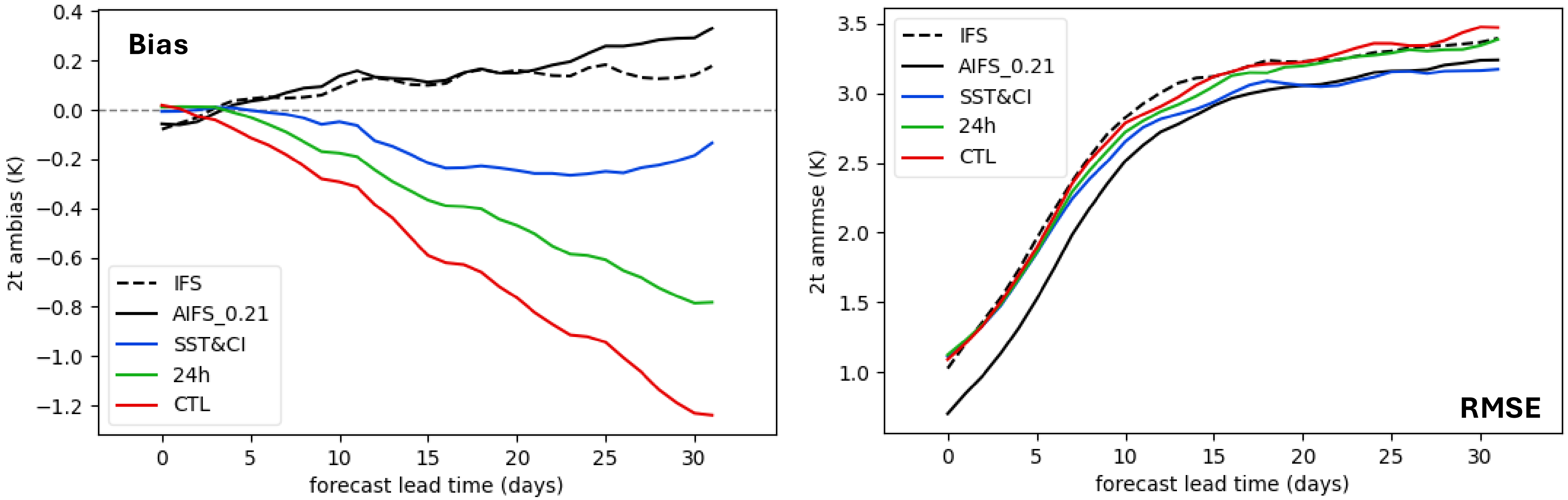
**MJO (not shown)**
AIFS.diffusion MJO correlation skill slightly better than IFS.2023 during days 1-15 and similar afterwards.

**After in-sample statistical calibration that enforces perfect reliability in terms of total variance and spread-error ratio.*

*Chris Roberts*

# Training AIFS on additional surface fields to stabilize bias and RMSE

*Bias (left) and RMSE (right) for JJA t2m in NH Extratropics 2018-2022*




**CTL**: ~AIFS 0.21 but smaller model and no fine-tuning on oper. ana. (~100x cheaper to train)
**24h**: as CTL but 24 h model time step
**SST&CI**: as 24h but including prognostic ERA5 SST and sea ice cover in the training

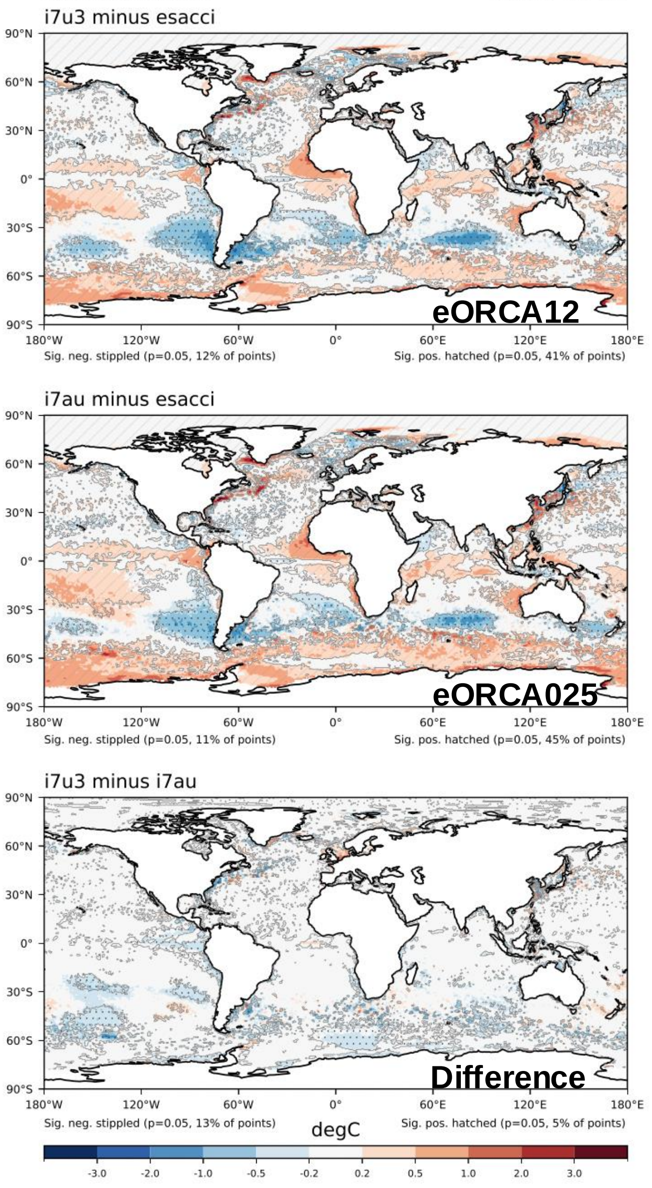
*Steffen Tietsche*

ECMWF

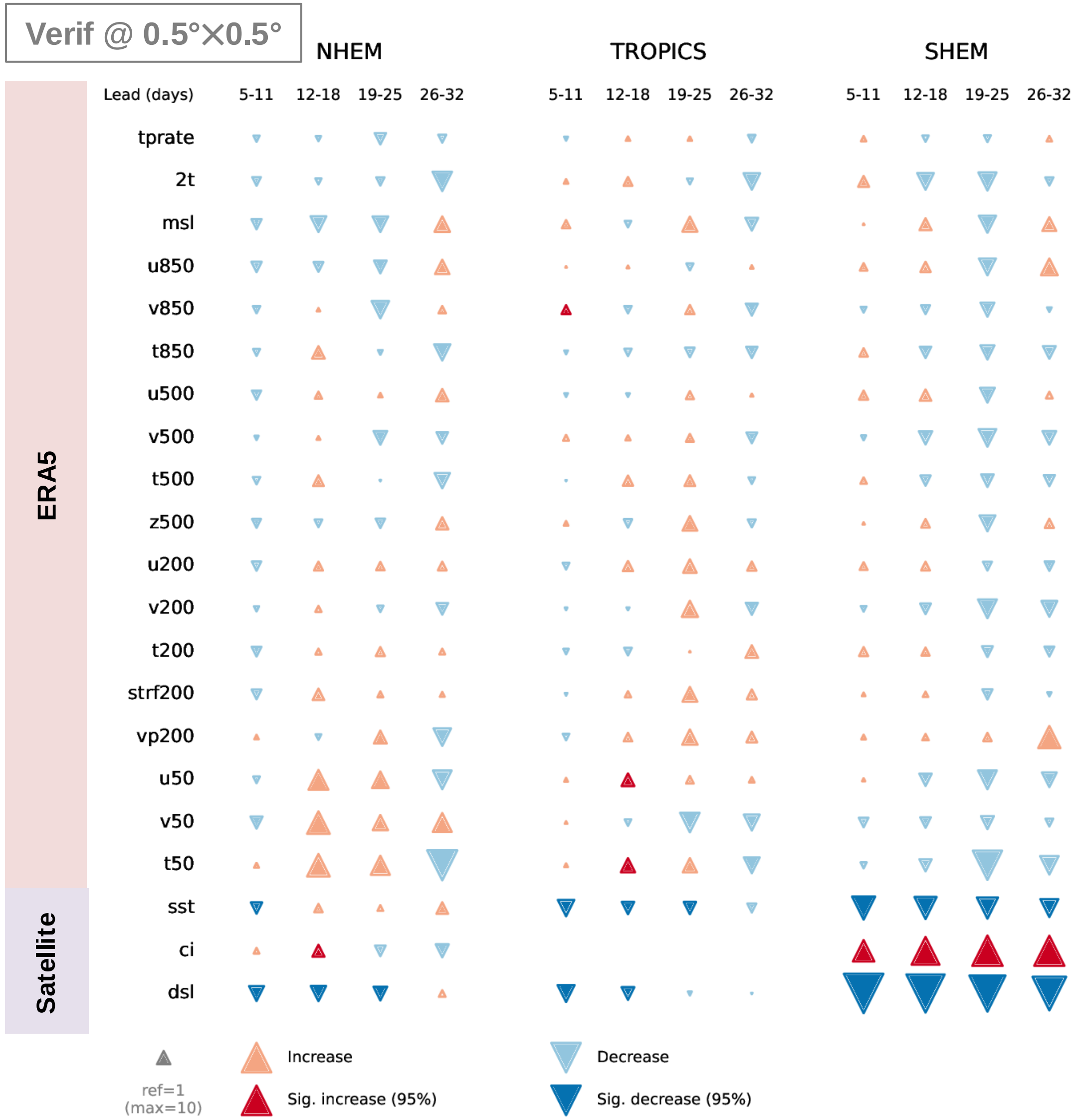
# Impact of “eddy-rich” NEMO4/SI3 in deterministic forecasts

**Reforecast configuration**
- 48r1Tco319L137 + NEMO4/SI3
- Control member only
- Initialized 1st of each month.
- 1995-2013.
- Ocean ini from GLORYS12.

**SST bias (week 4)** **Jan 1st starts**




## Impact on weekly mean anomalies (ΔRMSE)




Preliminary work with Charles Pelletier and Kristian Mogensen to evaluate “eddy-rich” (i.e. 1/12th degree) NEMO4/SI3 for S2S.

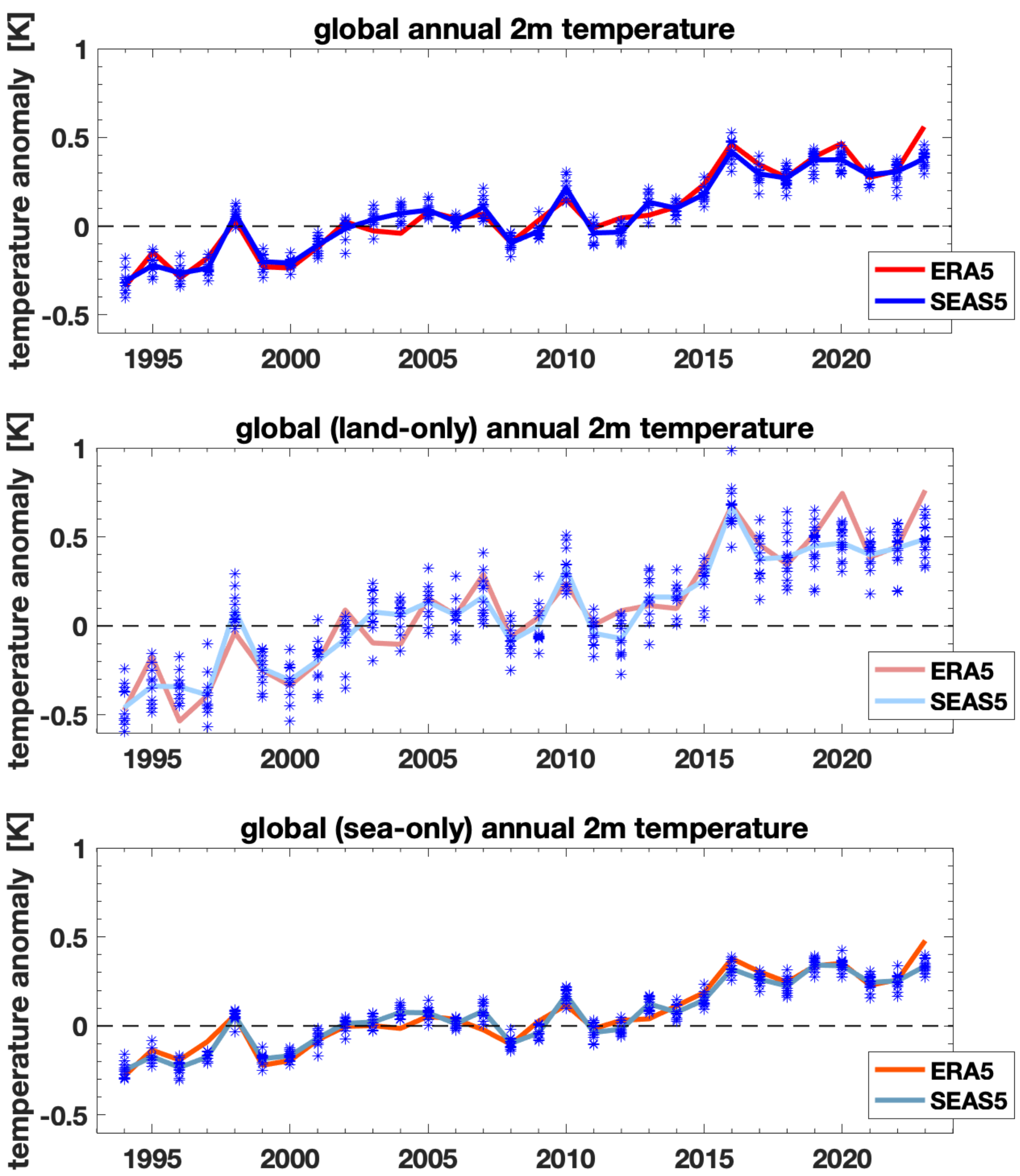
**Caveats**: single member, Tco319 L137 atmosphere.

**Key results:**

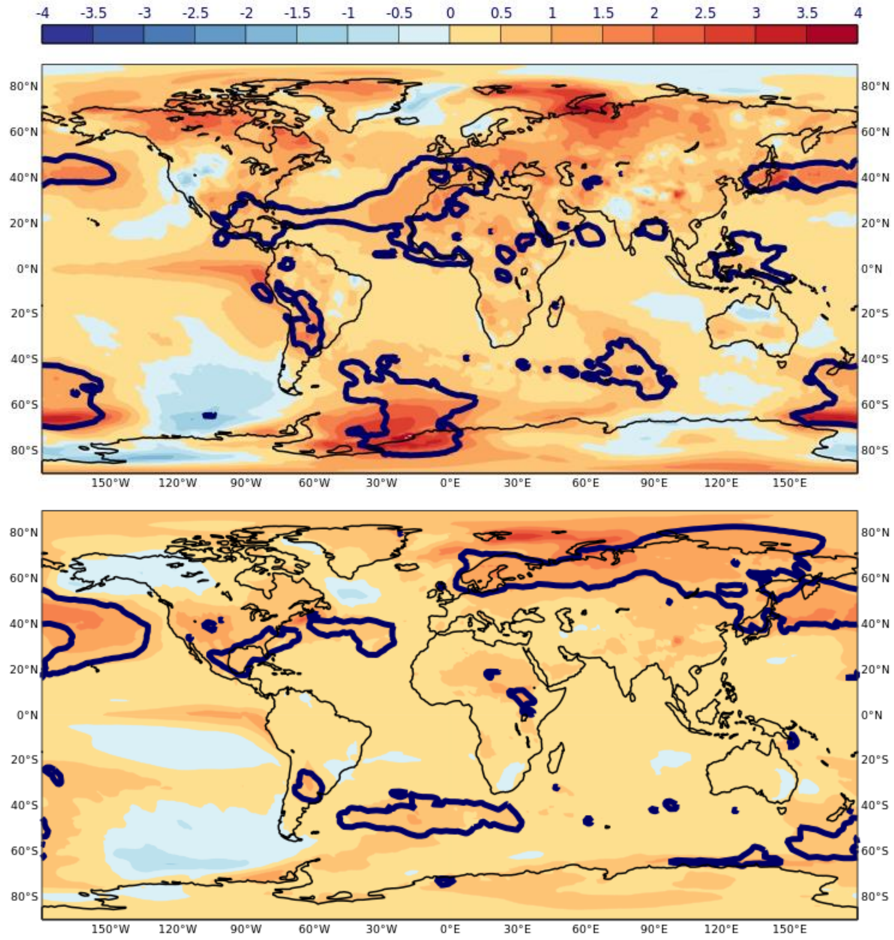
- S2S ocean mean state and anomaly scores are slightly improved with eddy-rich eORCA12, with some exceptions in SHEM.
- Differences are minimized when initial conditions for both systems derived from same source (i.e. Glorys12 reanalysis).
- Limited impact on atmospheric fields at S2S timescales.

*Chris Roberts*

# The record-breaking global mean temperatures of 2023: SEAS5 forecast




- SEAS5 underestimated the warming over land and sea
- Highly skillful in hindcast period, also after detrending
- Good signal to noise ratio




ERA5

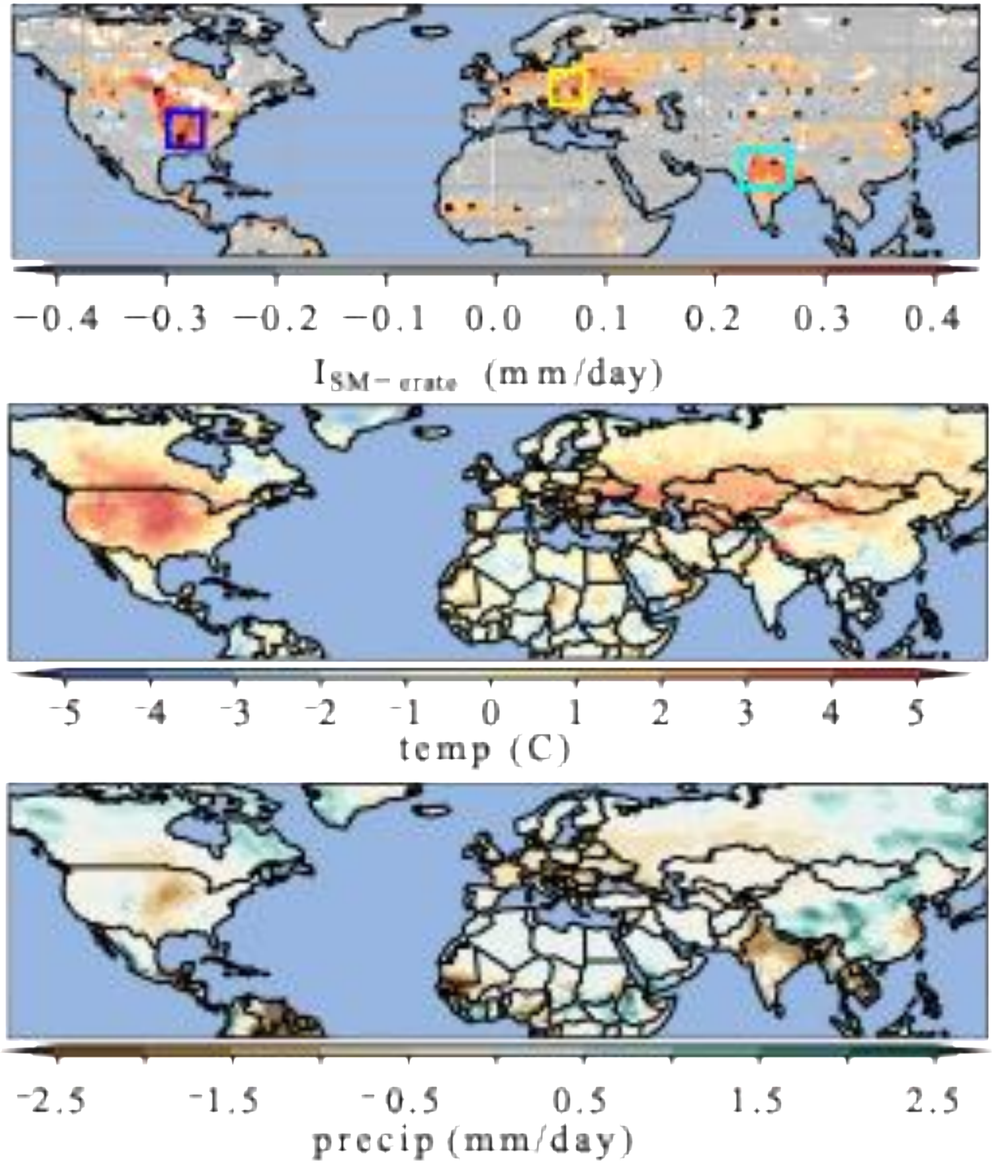
SEAS5

*A. Weisheimer*

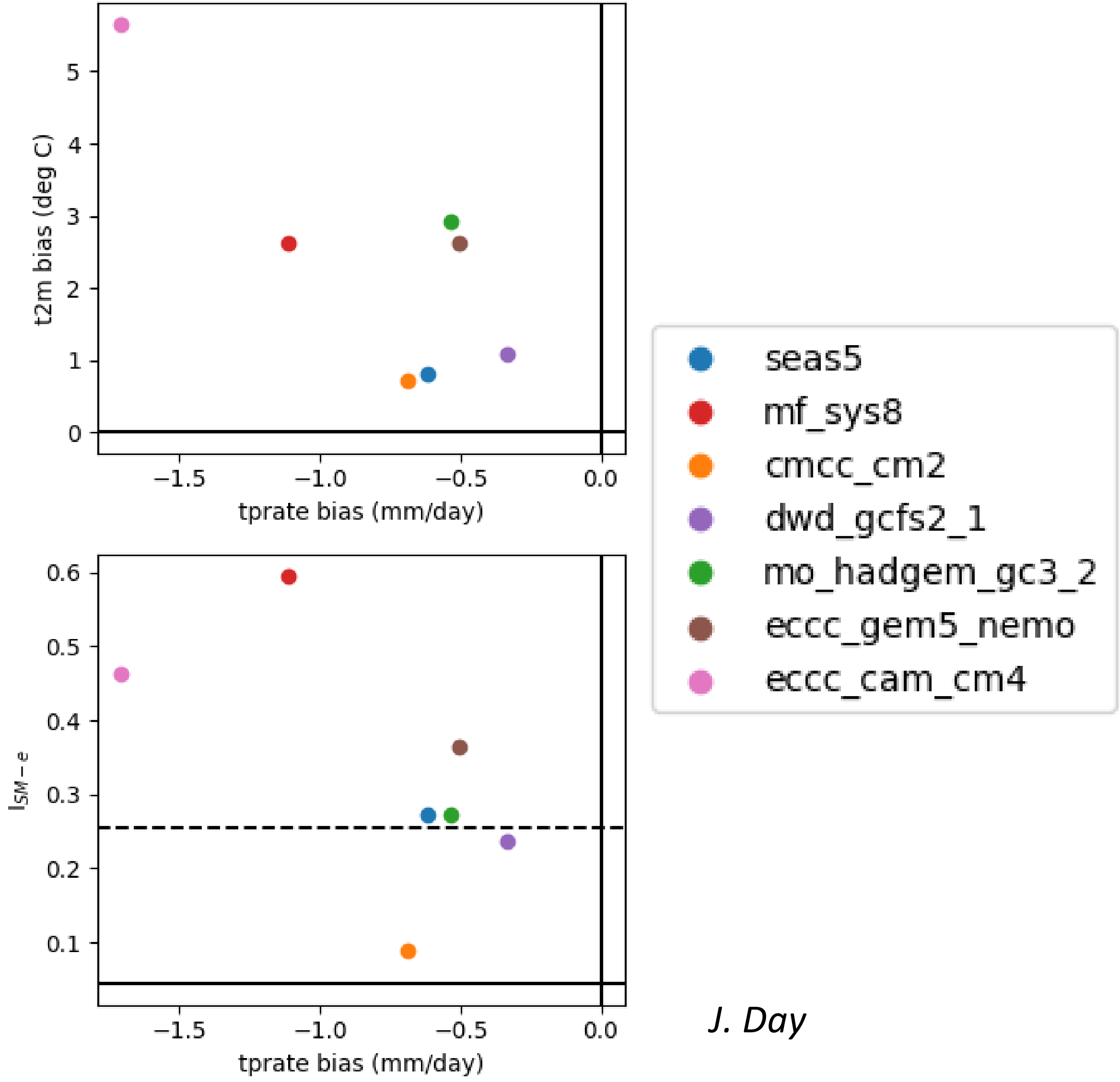
Funded by the European Union

# Coupling errors linked to model bias

## C3S multi-model mean bias




## East-Central US



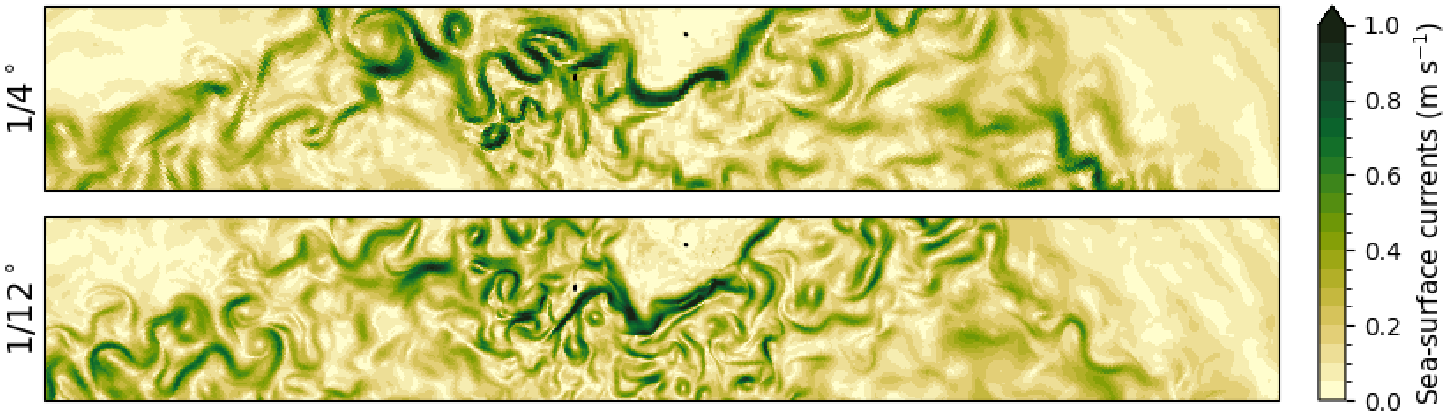

*J. Day*

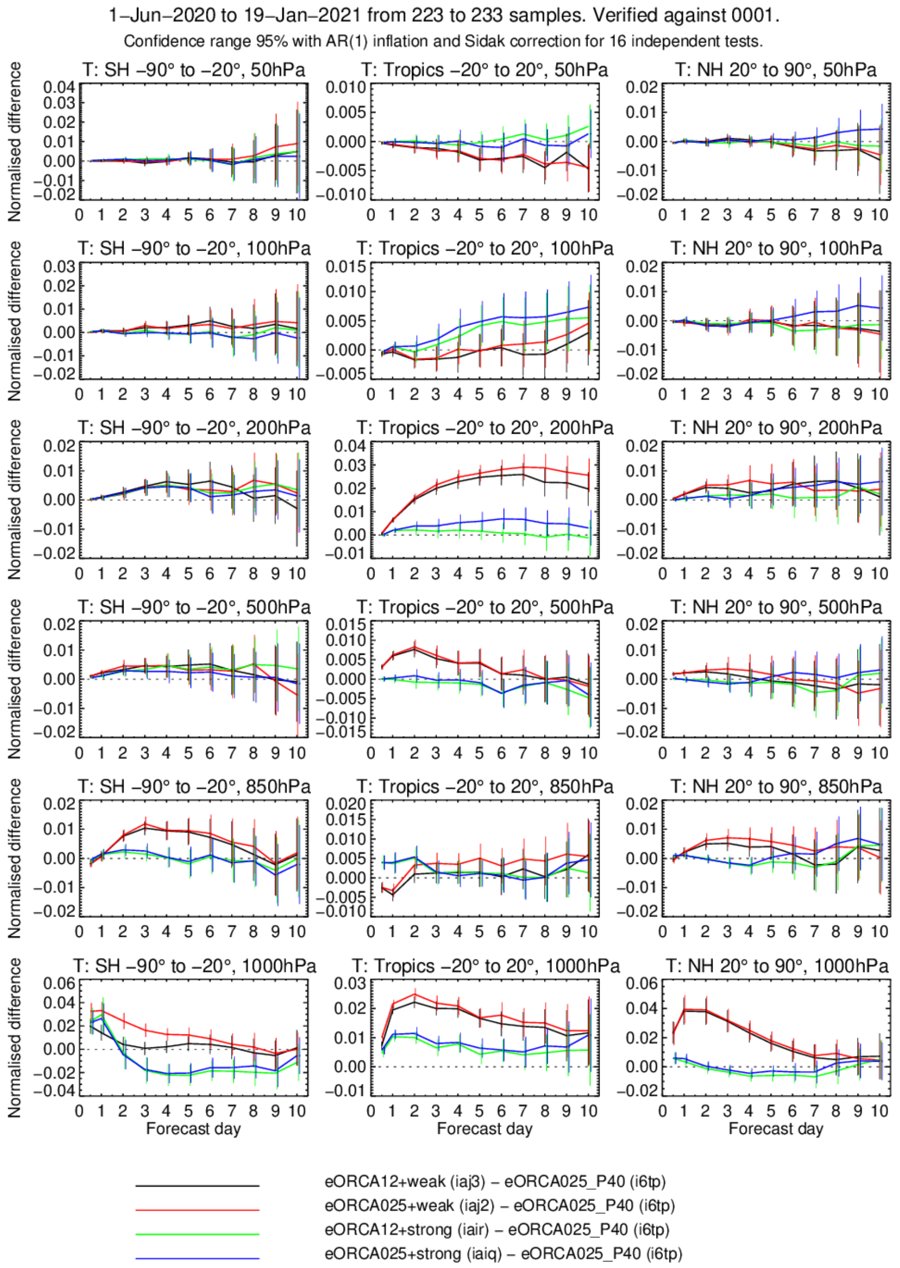
See also Ardilouze et al., (2019; WAF) and Lin et al., (2017; Nat Comms.)

# Medium-range testing of higher-resolution ocean initial conditions

- Sensitivity of IFS (tco1279) – NEMO4 coupled forecasts to ocean model resolution (1/4º vs. 1/12º) and initial conditions.
- Testing was carried using GLORYS12v1 (CMEMS ocean reanalysis) due to product then-availability.
- Quality of the (re)-analysis used as constraint on ocean ICs dominates the skill over the model resolution.
- Initialising from (pre-)ORAS6 provides best performance.
- Sanity check for 1/12º ocean : no clear ”bursts” (nor improvement).

*Below: snapshot of sea-surface velocities from ocean initial conditions at 1/4º vs 1/12º.*
*Right: drmse iver score for atmospheric temperatures using weakly vs. strongly constrained ocean ICS, and 1/4º vs 1/12º ocean.*






Chris Roberts

***K. Mogensen, C. Pelletier, C. Roberts***