Recent developments in km-scale ensembles and uncertainty representation at ECMWF

Aristofanis Tsiringakis

(with contributions from Martin Leutbecher, Simon Lang, Sarah-Jane Lock and Joffrey Dumont Le Brazidec)

ECMWF

Overview

Part 1. *Km-Scale Developments in Destination Earth*

- Km-scale ensembles for Uncertainty Quantification in the Global DT
- Downscaling for Destination Earth

Part 2. *General Developments in Uncertainty Representation at ECMWF*

- Recent updates in Uncertainty Representation coming to 50r1
- Training AIFS-ENS with multi-scale loss formulation
- Ongoing Work and Future Plan for Uncertainty Representation

Part 1

Km-Scale Developments in Destination Earth

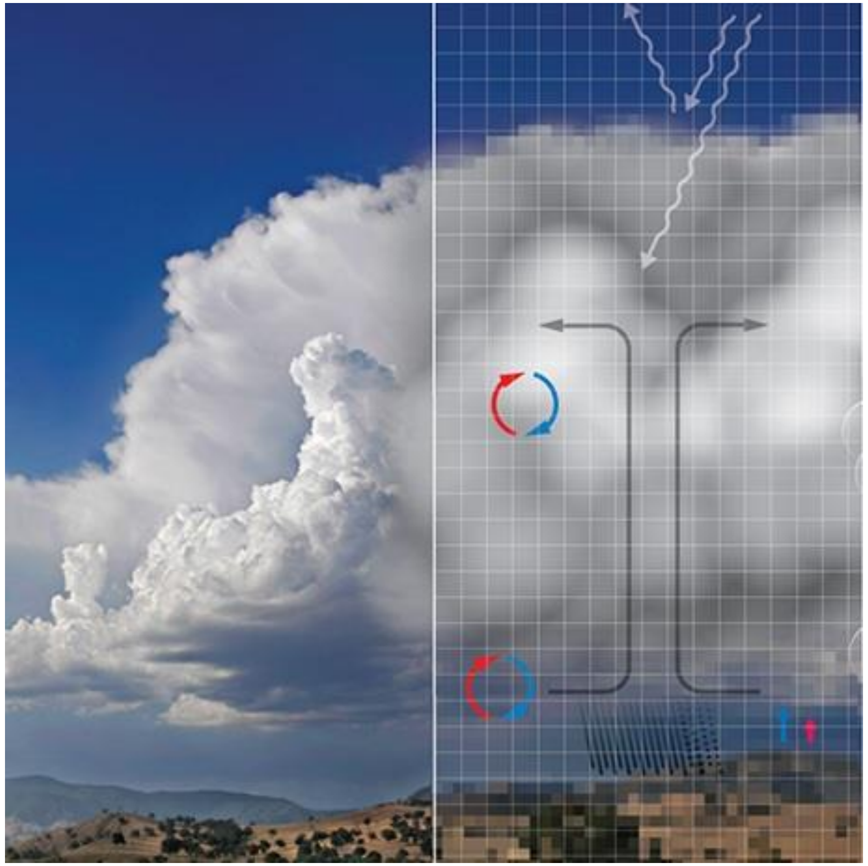
All models are imperfect

Earth System

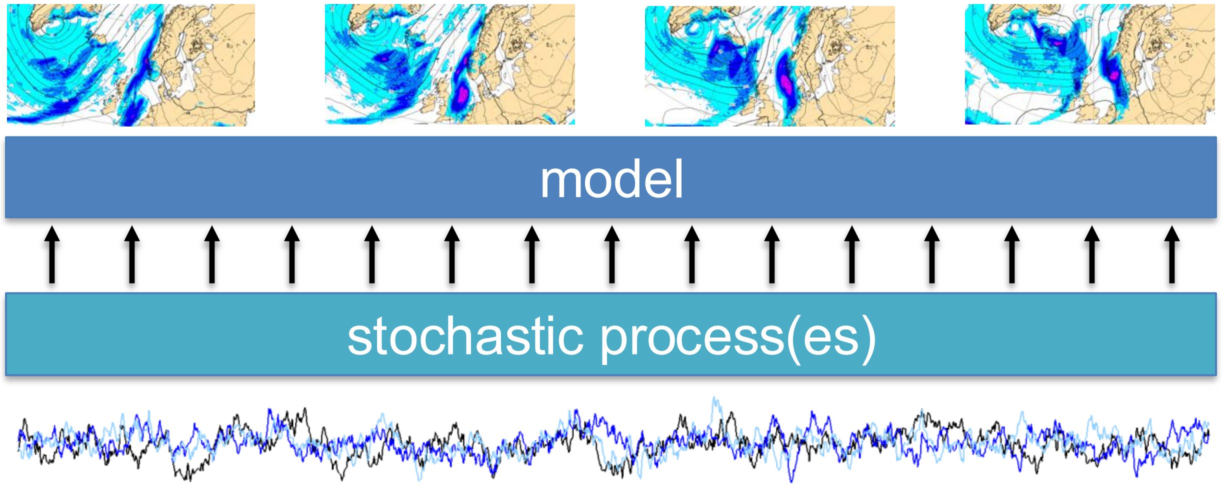
$$\mathbf{x}(t) \to \mathbf{x}_S(t + \Delta t)$$

Model

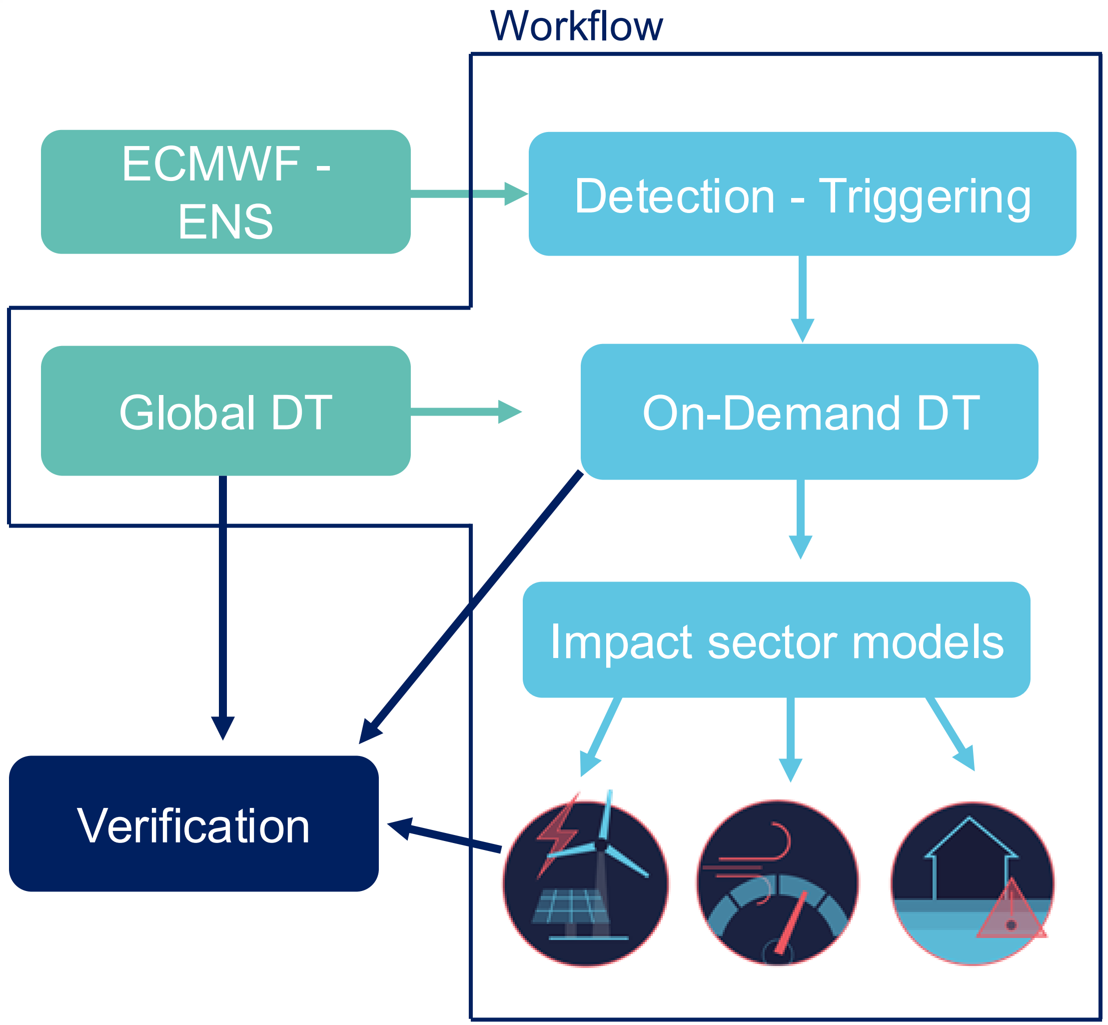
$$\mathbf{x}(t) \to \mathbf{x}_M(t + \Delta t)$$



- representing random errors of model improves reliability of ensemble
- Stochastic representation of model uncertainties



THE EXTREMES DT



What is the role of Uncertainty Quantification in the Extremes DT?

Uncertainty information is/can be used in almost every component of the extremes DT as:

- Input (e.g., EFI index for triggering the On-Demand DT, boundary condition uncertainty from Global to On-Demand)
- Output (e.g., probability of exceeding specific precipitation thresholds from Global DT)

Depending on the use case, uncertainty information can take many different forms :

- **Ensemble spread** → Verification for Global or On-Demand DT
- **Probabilities** → Predicting likelihood of extreme events
- **Uncertainty at boundaries** → Forcing for the sub-km scale ensembles, impact sector models etc.

Uncertainty Quantification in the Global DT

Aim : Improve Uncertainty Quantification in the Global DT via the use of km-scale ensembles

Improved representation of extreme events

Improved UQ information provided to the On-Demand DT

Better UQ provided for impact sector models

Starting Point

IFS operational ensemble

- 50 ensemble members
- 15-day forecasts
- EDA+SV+SPP/SPPT
- 9 km resolution

End of Phase 1

Km-scale ensemble Global DT v1 (first setup)

- 10 ensemble members
- 5-day forecasts
- EDA+SV+SPP
- 4.4 km resolution

End of Phase 2

Km-scale ensemble Global DT v2

- Better initial conditions
- Km-scale appropriate physics
- Improved model uncertainty representation for km-scale

Verification 4.4 km vs 9 km

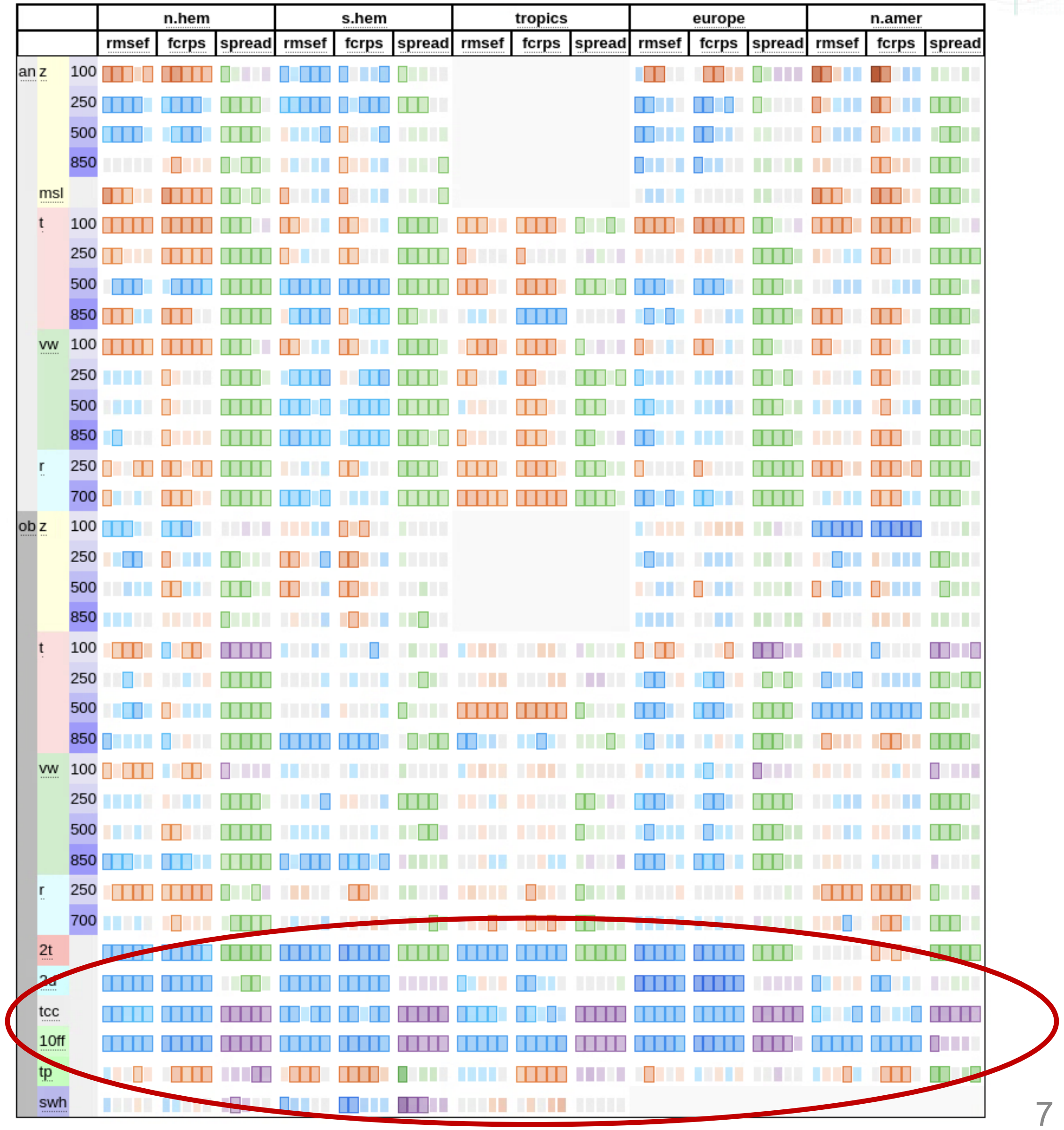
- Substantial improvement (2-8%) in surface scores for 2m temperature, 2m dewpoint temperature, 10m wind speed and total cloud cover, when evaluating against observations

- Improvement (1-2%) in z500 (NH), z250 + t500 (NH & SH)

- Deterioration in Z, T and VW at 100hPa level against analysis

- Small deterioration in the tropics (~ 1%) - a general reduction in ensemble spread (~1-2%, expect for 2d, tcc and 10ff)

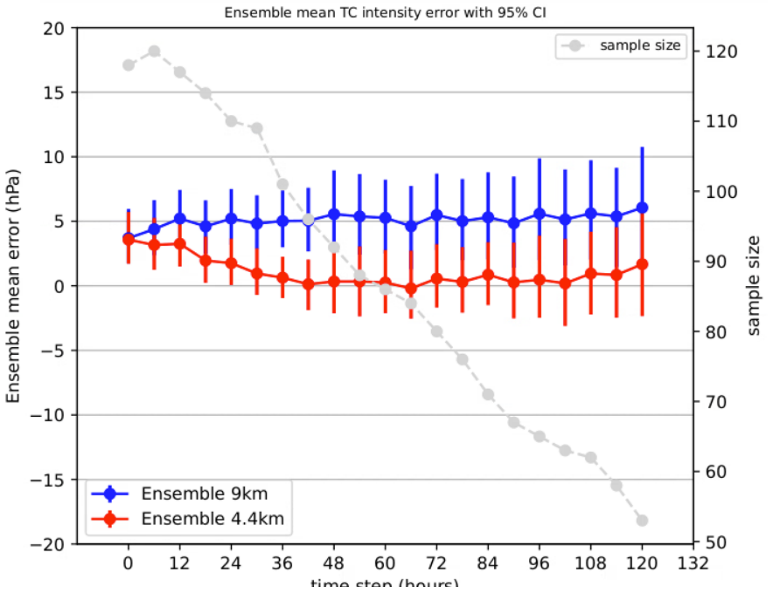
Scores produced for periods July – Sept. 2023 and Dec. 2023 – Feb. 2024, i.e. 92 initializations days

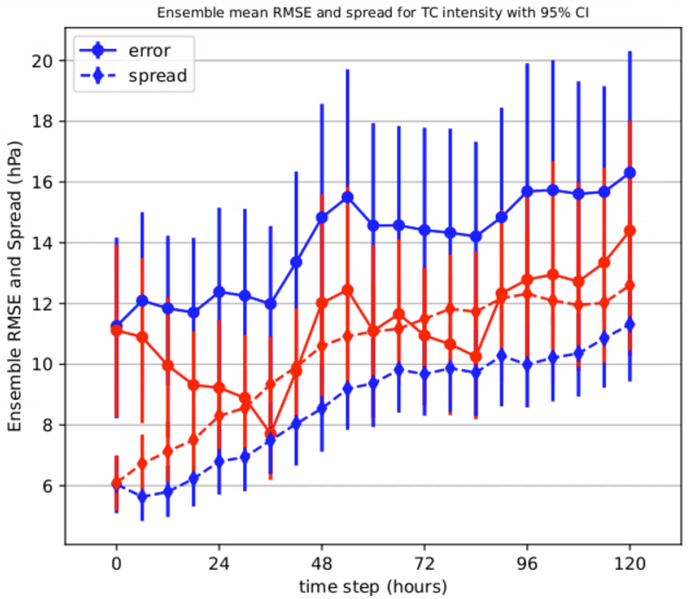


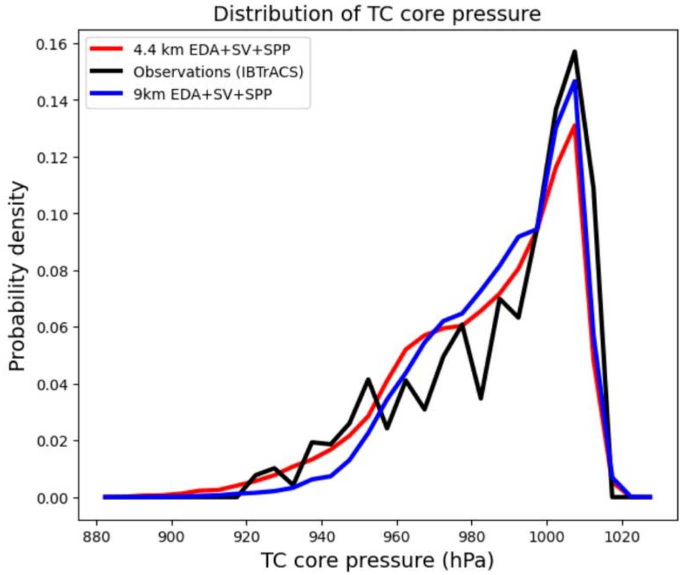
Improvements in the Representation of Tropical Cyclones

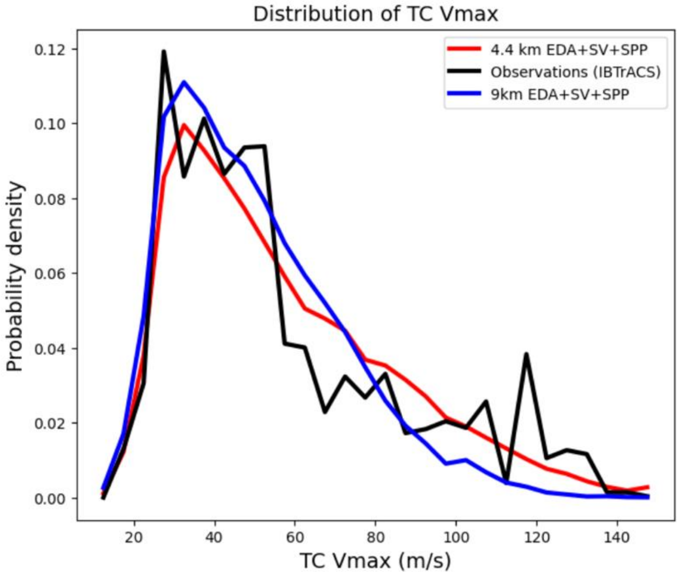
Verification against IBTrACS (July – September 2023)

- Reduction of ensemble mean bias to nearly zero
- Reduction of RMSE by up to 30% compared to 9km ensemble
- Increase in ensemble spread









Better match for TC core pressure and Vmax with the 4.4 km ensemble, especially for the tails of the distribution (i.e., more extreme events)

Impact of ensemble size for extreme events at km-scale

Valencia flood case (29th October 2024)

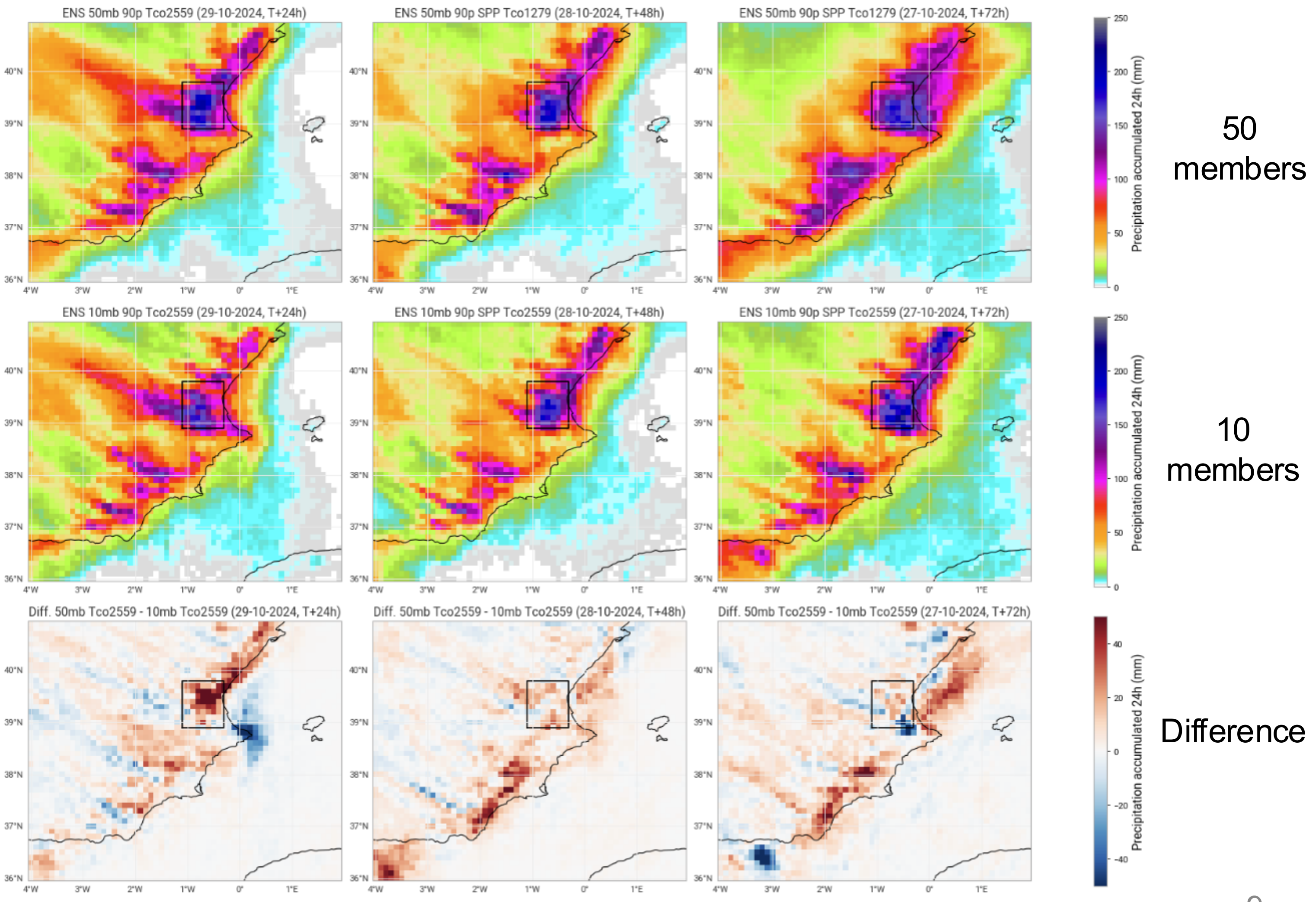
- Testing impact of ensemble size on 90th precipitation percentile at km-scale ensemble

Initialization dates	Ens 90% 10 mb 4km	Ens 90% 50 mb 4km	Ens 90% 10 mb 9km	Ens 90% 50 mb 9km
27-10-2024	232	**216**	243	**192**
28-10-2024	192	**196**	128	**176**
29-10-2024	176	**209**	158	**162**

Max value of 90th percentile in the area within the black box

Less consistency for extreme precipitation across initialization dates with reduced ensemble size

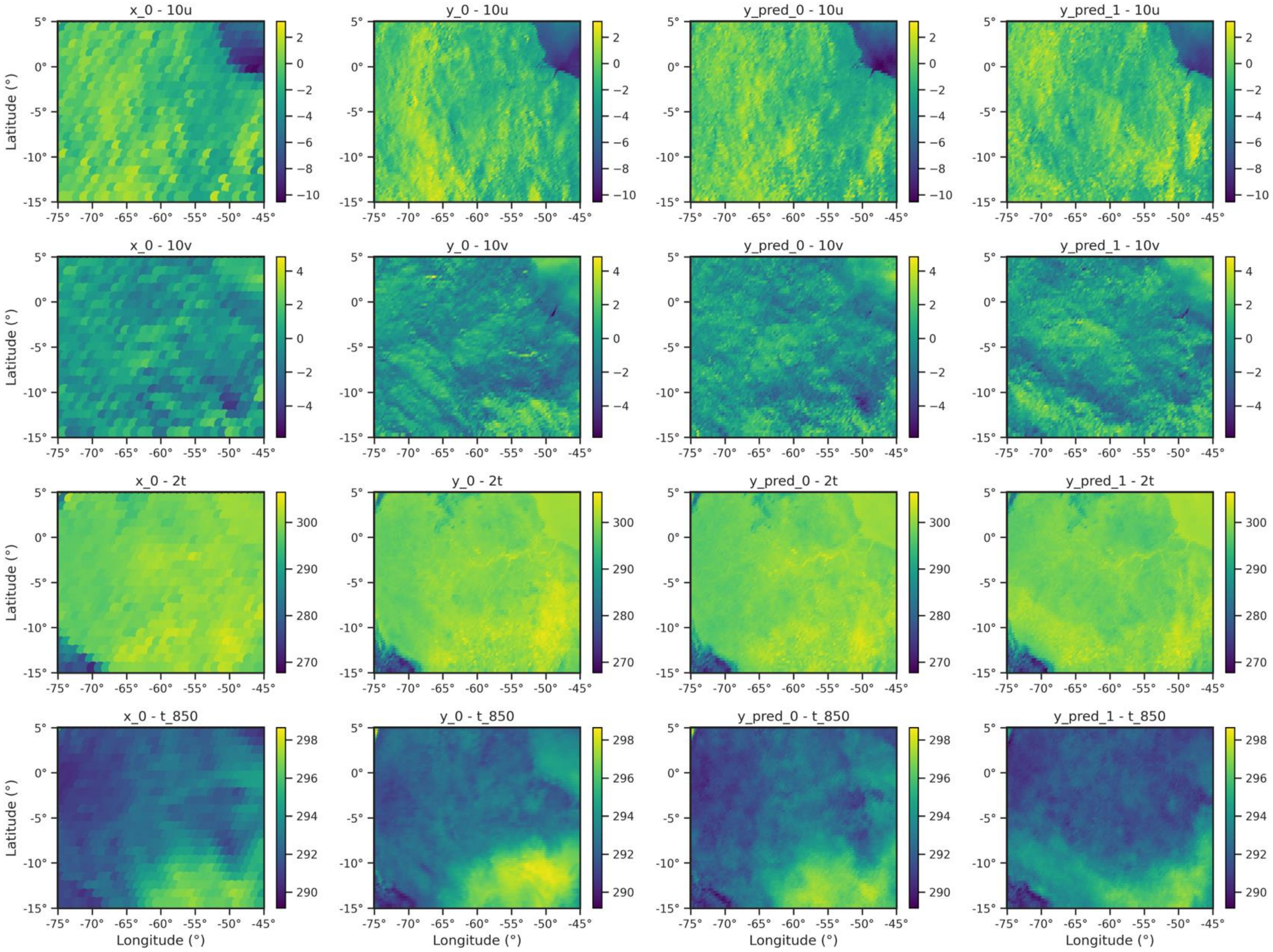
More ensemble members would be better for more reliable forecasts of extreme weather events



Downscaling to Destination Earth

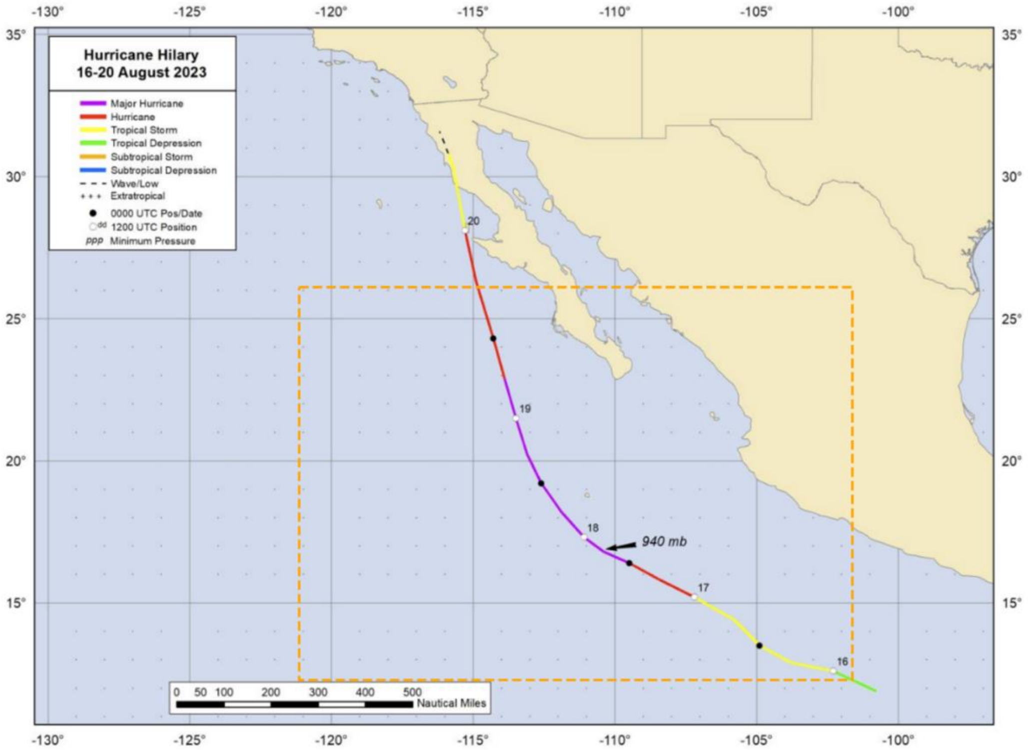
Problem: Producing kilometre-scale global ensemble forecasts is essential for assessing the probabilities and intensities of extreme events, but physical models render this computationally prohibitive.

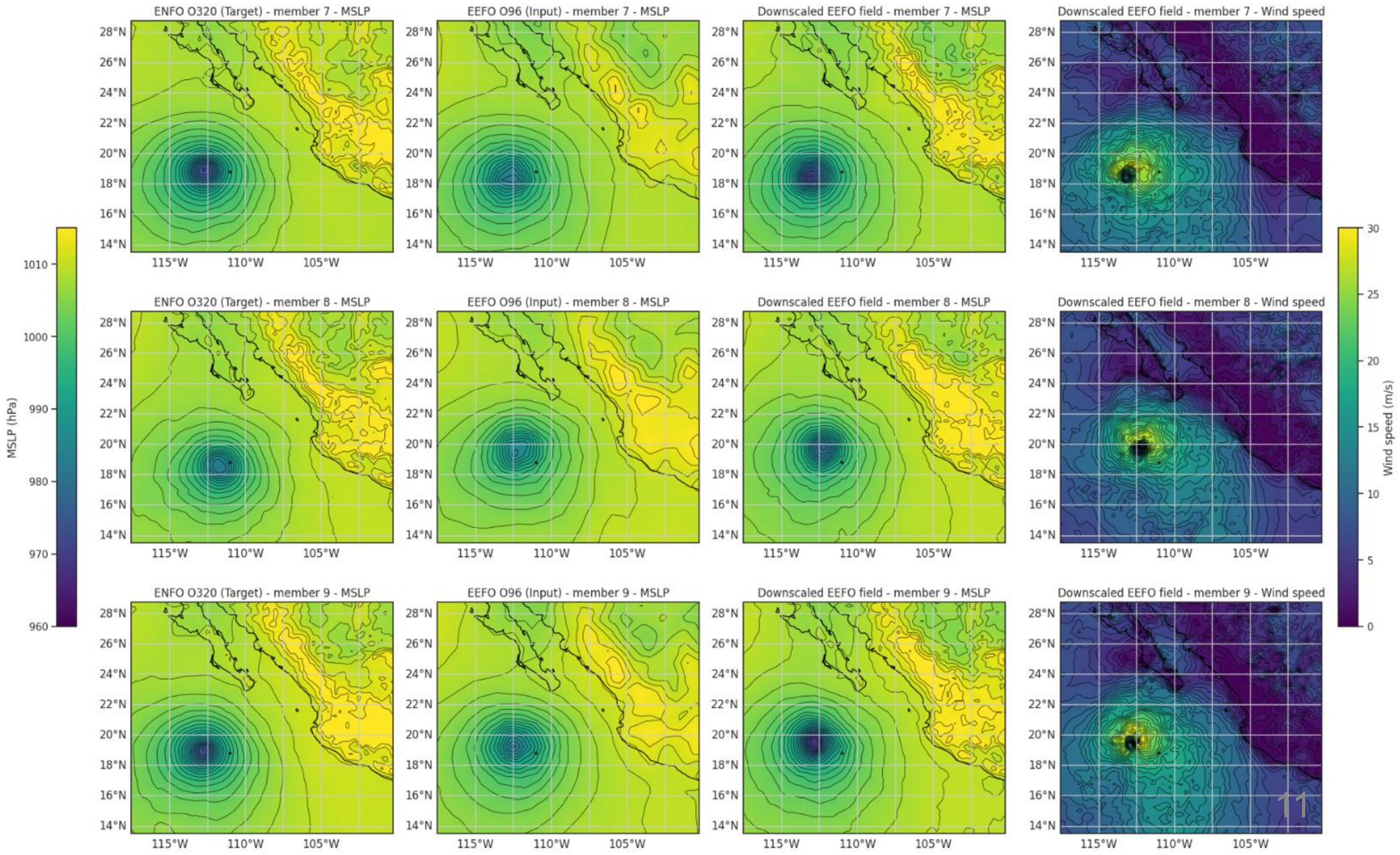
Objective: Develop a high-resolution ensemble forecasting system based on ECMWF's operational forecasts, leveraging advanced AI generative models (e.g. diffusion models) to capture uncertainty, improve predictive skill, and better characterise extreme event risks.



Amazonian rainforest, 2023.08.01, eefo o96 -> enfo o320

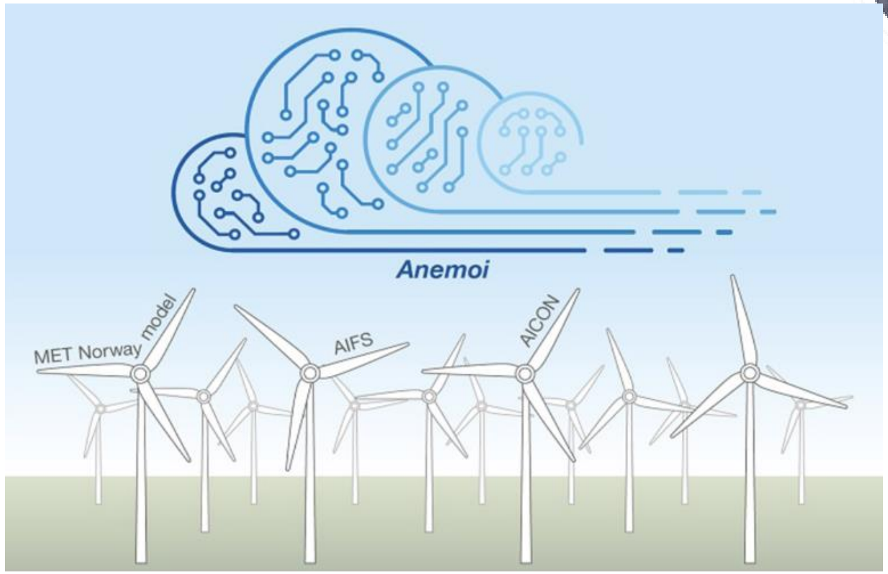
Recovery of extremes (Tropical Cyclone Hilary)

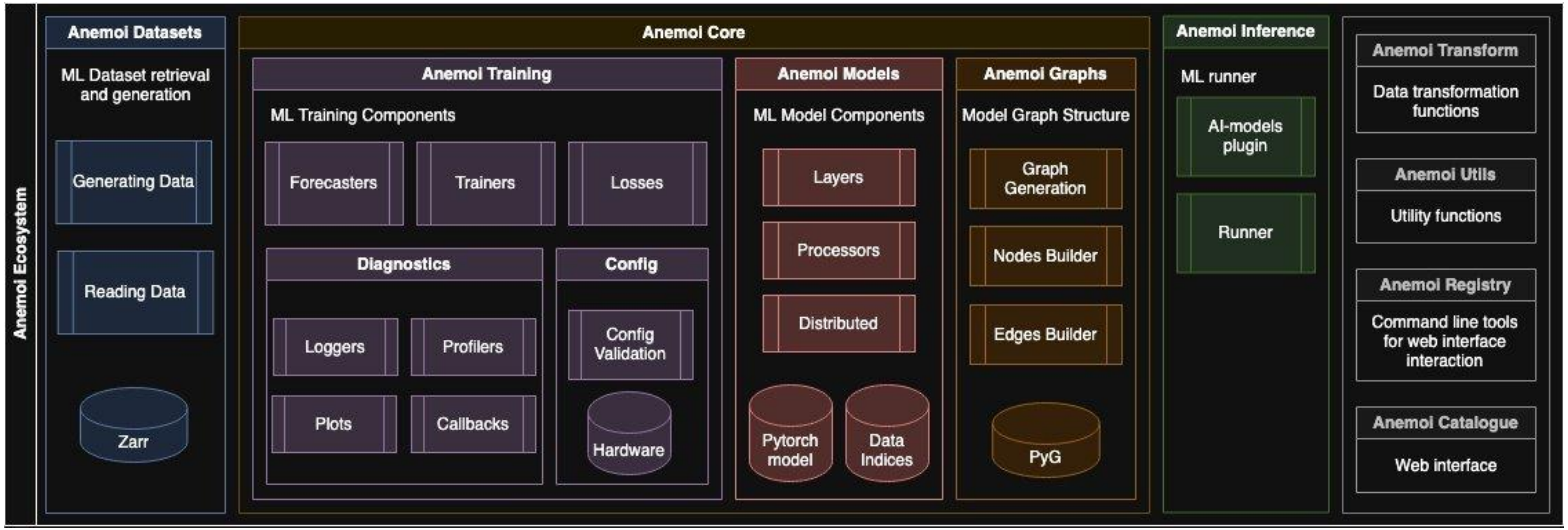




Future implementation on anemoi

- Downscaling will be implemented in Anemoi, an open-source framework providing a full toolkit for developing data-driven weather models.





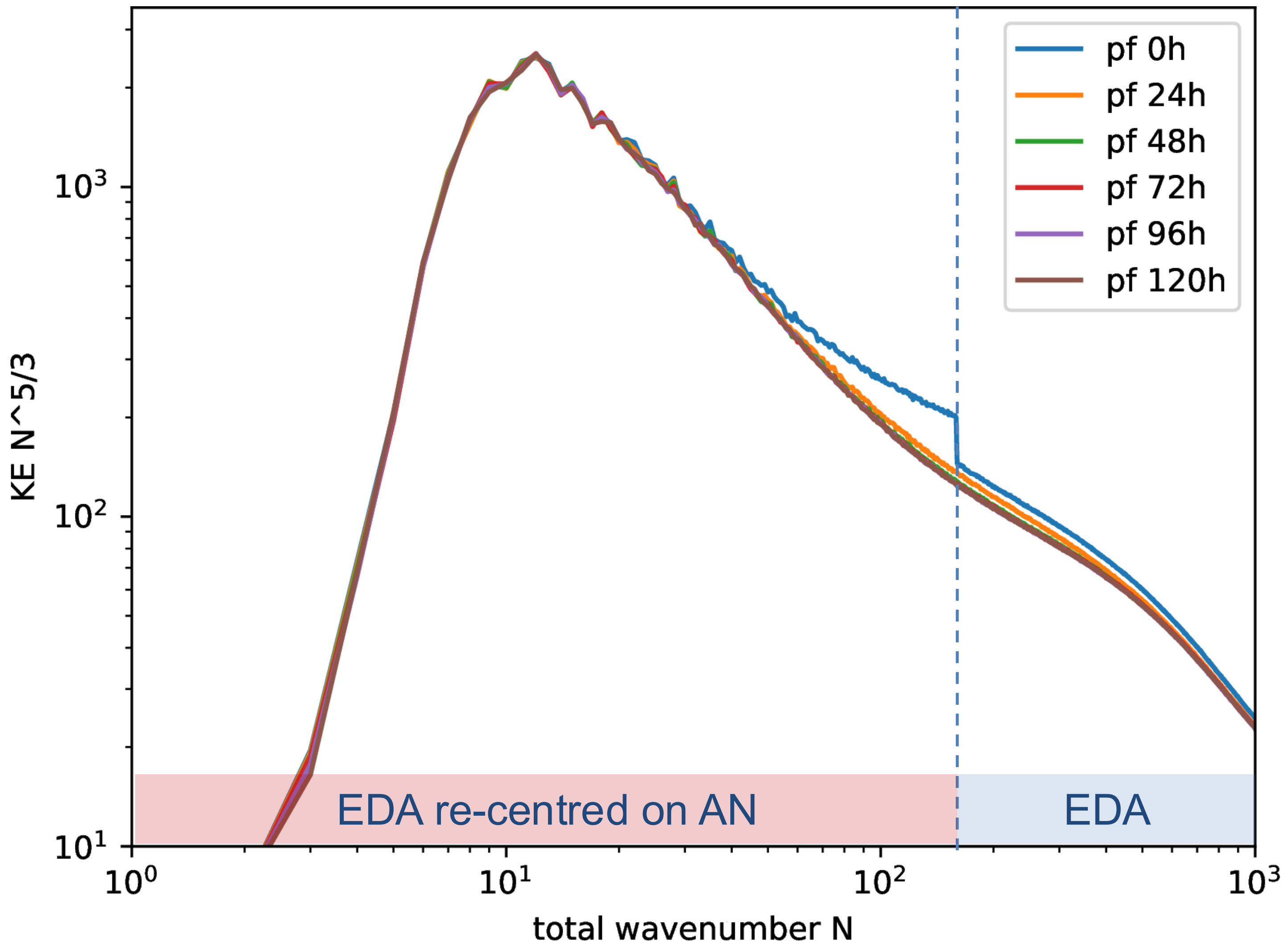
Ongoing work and steps ahead

- Test the impact of providing km-scale later boundary conditions from Global DT ensemble to the sub-km scale On-Demand ensemble (ongoing)
- Investigate a few extreme weather cases with a 10+1 member, 2.8 km ensemble (ongoing)
- Identify optimal model configuration for generating nudged datasets to be used for downscaling training (early development)

Part 2

General Developments in Uncertainty Representation at ECMWF

Scale-dependent EDA re-centring: KE spectra at 250 hPa



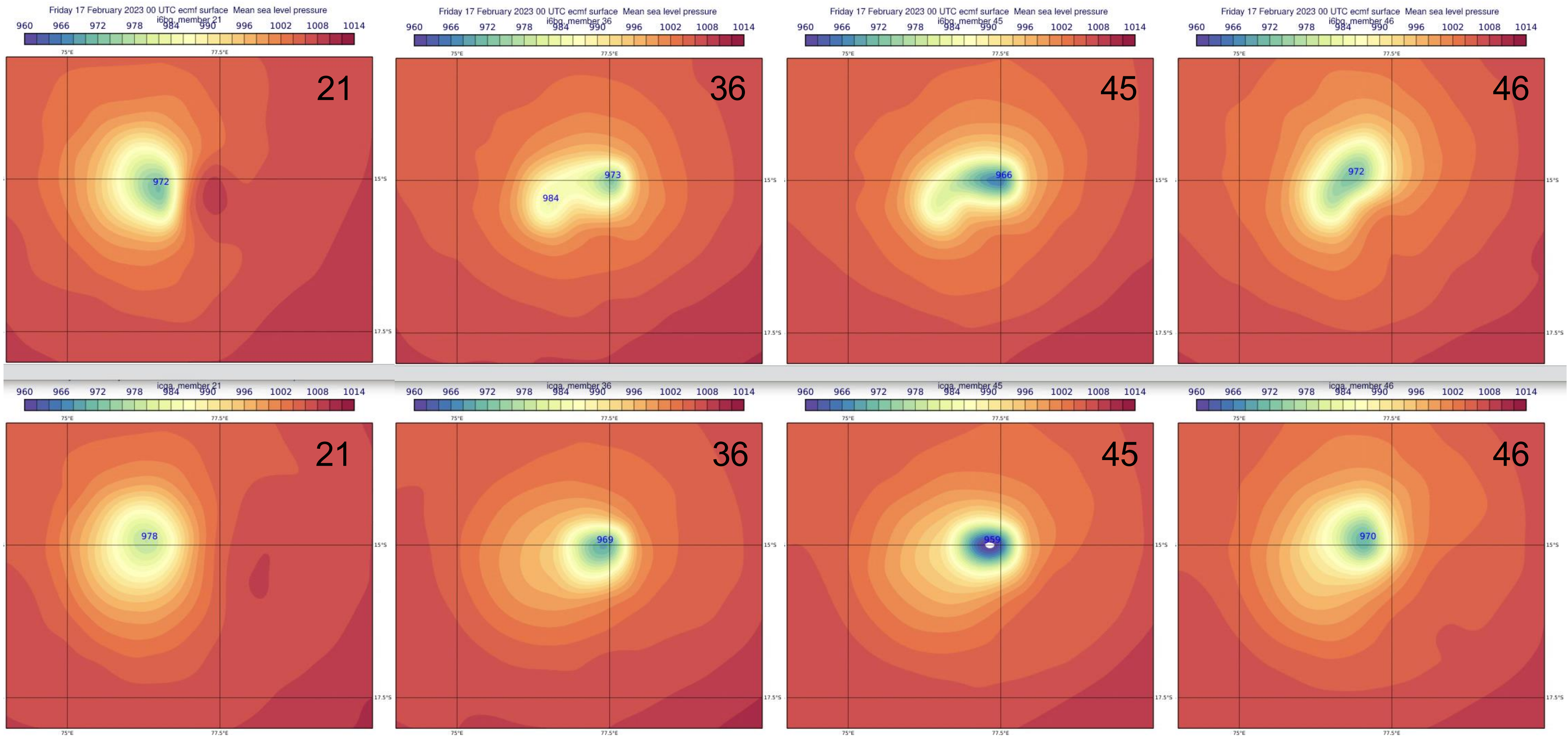
TC Freddy, 17 Feb 2023, 00UTC, mean sea level pressure t=0h

re-centred
EDA

49r1

scale-
dependent
re-centring
of EDA

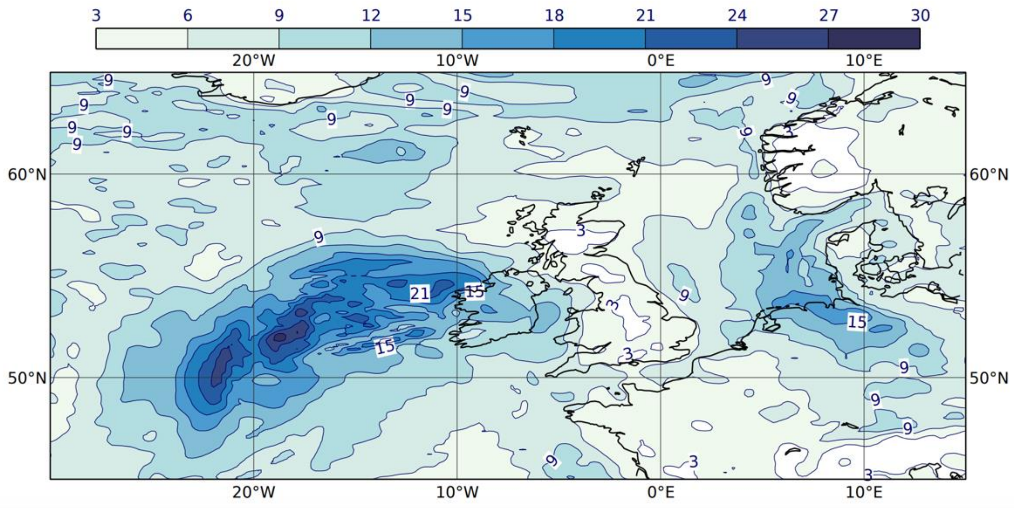
(50r1)



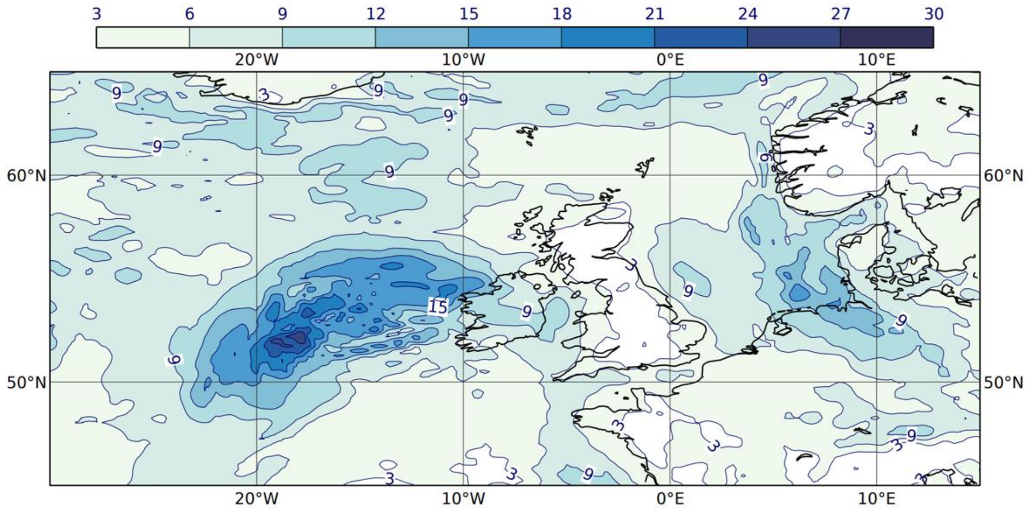
Assessing the impact of SPP perturbations - an example:

Ensemble range* for 10-metre windspeeds (m/s), t=36 h, for Storm Darragh

CY49R1 settings



Revised SPP settings



Valid time: Friday 6 December, 12 UTC (Storm Darragh)

- TCo399 (~29km grid-spacing)
- 50 perturbed ensemble members
- initialised with initial conditions from the operational forecast

*At each location, ensemble range equals max value across ensemble minus min value

Train AIFS-ENS with multi-scale loss formulation:

https://arxiv.org/abs/2506.10868

A multi-scale loss formulation for learning a probabilistic model with proper score optimisation

Simon Lang, Martin Leutbecher, and Pedro Maciel

ECMWF, Reading, UK

12 June 2025

Abstract

We assess the impact of a multi-scale loss formulation for training probabilistic machine-learned weather forecasting models. The multi-scale loss is tested in AIFS-CRPS, a machine-learned weather forecasting model developed at the European Centre for Medium-Range Weather Forecasts (ECMWF). AIFS-CRPS is trained by directly optimising the almost fair continuous ranked probability score (afCRPS). The multi-scale loss better constrains small scale variability without negatively impacting forecast skill. This opens up promising directions for future work in scale-aware model training.

ECMWF

CRPS is computed point-wise on the full output field. However, atmospheric processes are inherently multi-scale, and different scales contribute to a different degree to the loss function -> introduce scale-aware loss

Scale-unaware:
$$\mathcal{L} = c \int_{\mathcal{M}} \mathcal{S}([x_j \mid j = 1, \ldots M], y)\, \mathrm{d}\mu$$

scoring rule $\mathcal{S}$ for scalars

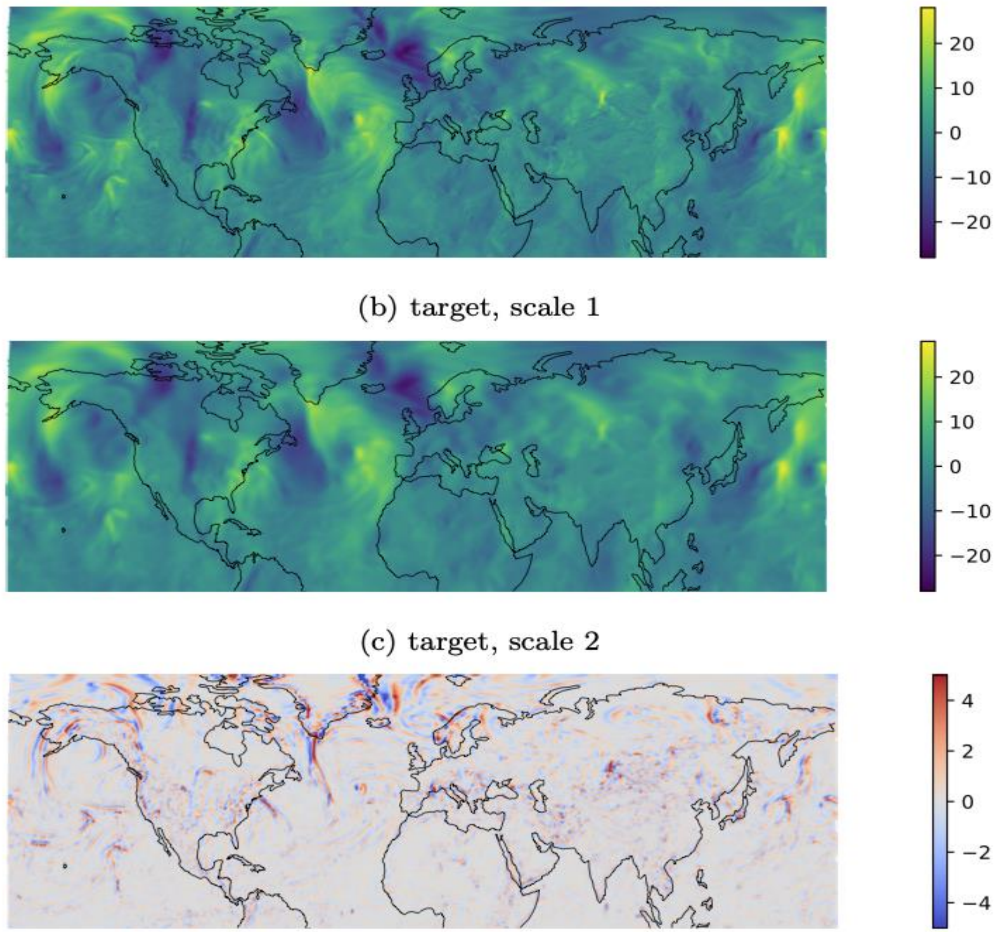
Scale-aware:
$$\mathcal{L}_{n\text{-scale}} = \sum_{i=1}^{n} \zeta_i\, c \int_{\mathcal{M}} \mathcal{S}([x_{j,\mathrm{scale}\, i} \mid j = 1, \ldots M], y_{\mathrm{scale}\, i})\, \mathrm{d}\mu$$

with weight $\zeta_i > 0$ for scale i.

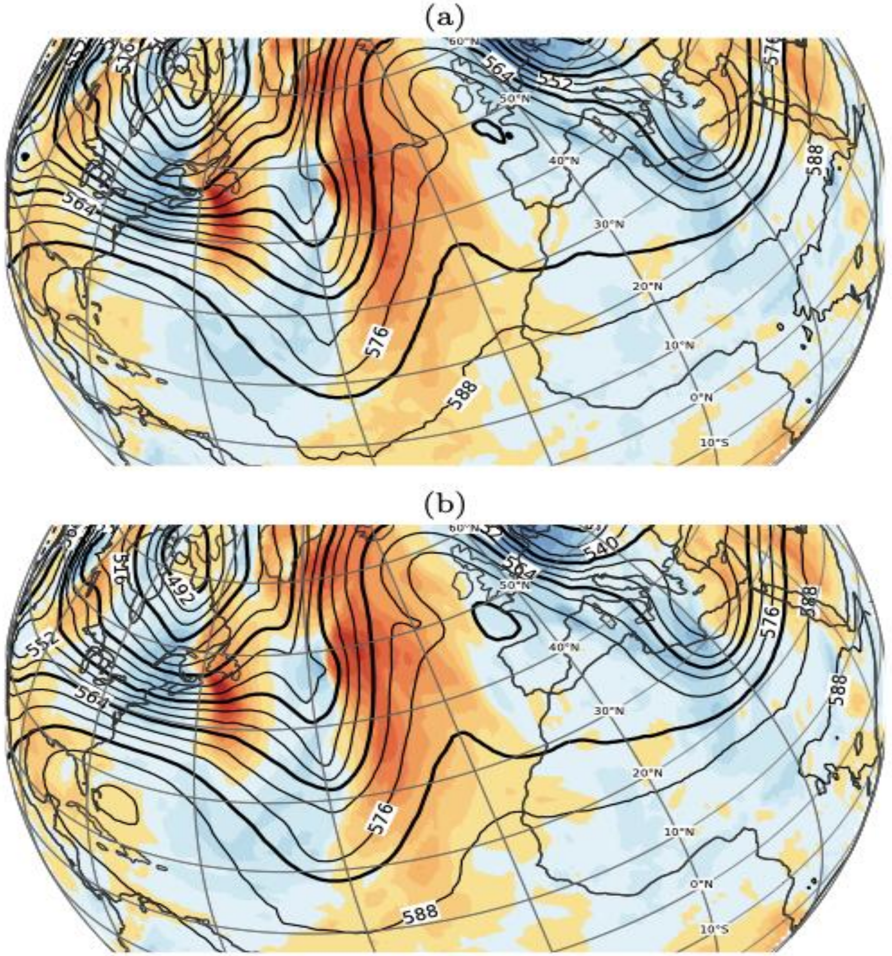
With ordered smoothing operators D_i, that partition of a function ϕ on the manifold M into n scales:

$$\begin{aligned}
\phi_{\mathrm{scale}\,1} &= D_1(\phi) \\
\phi_{\mathrm{scale}\,2} &= D_2(\phi) - D_1(\phi) \\
&\vdots \\
\phi_{\mathrm{scale}\,n} &= \phi - D_{n-1}(\phi)
\end{aligned}$$

Multi-Scale Loss removes noise in predictions (here two scales):



ERA5 v-component of wind (in m s^{-1}) at 850 hPa, full field (a), field after filtering (b) and difference between filtered and full field (c).

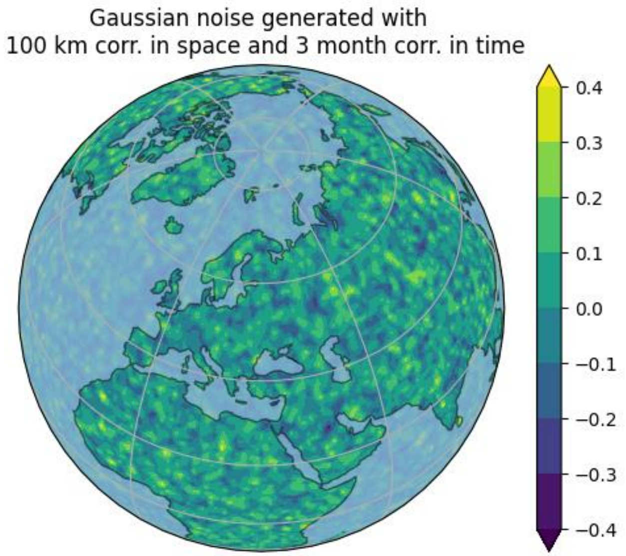


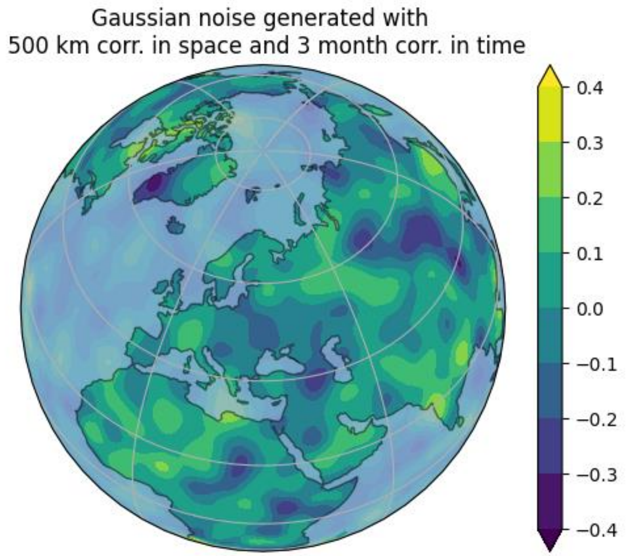
Geopotential at 500 hPa (dam, contours) and v-component of wind at 850 hPa (shaded) of (a) the scale-unaware loss experiment, (b) the multi-scale loss experiment Shown are 24 h forecasts of member 1, initialised on 2019-01-01 00 UTC (a, b) and verifying analysis on 2019-01-02 00 UTC (c).

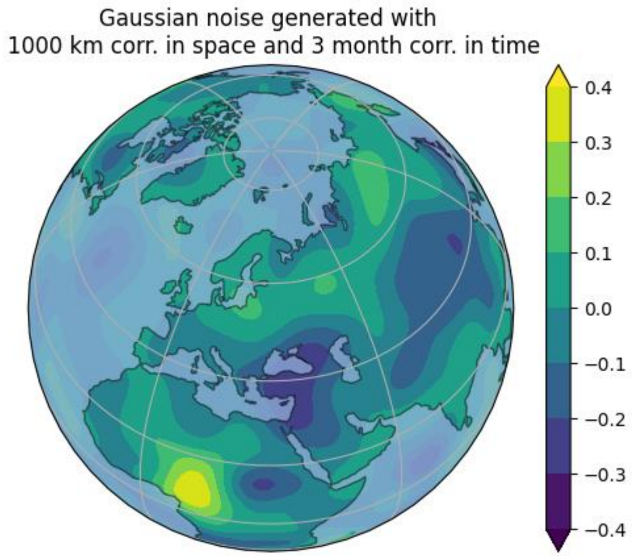
ECMWF

Other model uncertainty representations
(on-going work)

- Land surface uncertainties with **SPP** (with coupled DA & land surface teams)
  - Currently: developing perturbations to LAI, vegetation cover fraction (CV)
  - Testing:
    - Sensitivity to random number horizontal correlation scale: 100 / 500 / 1000 km (see examples below)
    - Exploring alternative distributions for the random numbers (to preserve physical constraints on LAI/CV)







ECMWF

Other model uncertainty representations
(on going work)

Representation of uncertainties in TKE with **SPP**
(Special thanks for Ivan Bastak Duran)

TKE_ LSC	multiplicator of the BL89 length scale above the surface layer (a candidate for replacing RKAP scaling): L=TKE_LSC*LBL89
ALMA VE	the asymptotic value of the length scale in free atmosphere (not in the surface layer): L=MAX(L,ALMAVE)
TKE_ ST_P	determines the shape of the stability function for heat/moisture: phi_3=(1-Rif)/(1-Rif/P)
ALMA VX	the asymptotic value of the length scale in deep convection cloud (interaction of turbulence with deep convection, higher ALMAVX->stronger incloud turbulence transport): L=MAX(L,ALMAVX)
TKE_ EQR	weigth for equilibrium estimate of the TKE source terms from 1st order closure (first guess/equilibrium estimate of exch. coefficients: KM0, KH0): Shear term= KM*S^2, Buoyancy term=KH*N^2, KM=KM0*TKE_EQR+KM*(1-TKE_EQR),KH=KH0*TKE_EQR+KH*(1-TKE_EQR)
TKE_ LAS_ ST	the asymptotic value of the length scale in the PBL outside of the StCu or Cu regime
TKE_ LSL_ RPBL	the relative heigth (in relation to PBL heigth) of the surface layer. The length scale (L) in the surface layer is equivalent to l=kappa*z . This has impact on near surface quantities.
TKE_ ST_D RP	determines the shape of the stability function for momentum: phi_3=(1-Rif)/(1-Rif/R), R=TKE_ST_P+TKE_ST_DRP*(1.0-TKE_ST_P)

Very promising initial results when compared against EDA+SV only ensemble (ran with TKE activated)!

ECMWF

Thank you!