Towards large ensembles of local ML-based climate predictions/projections: potential, pitfalls and prognoses

Stefan Sobolowski(NORCE/UiB), Antoine Doury(Meteo-France), Samuel Somot(Meteo-France), Jesus Fernandez(CSIC), José Gutierrez(CSIC), Jorge Baño-Medina(CSIC), Temesgen Asfaw(UiB), Asgeir Sorteberg(UiB), Roberto Bilbao(BSC), Pablo Orteg (BSC), Ole Christiansen(DMI), Rashed Mahmood(DMI) , Shuting Yang(DMI)

UPCLIV Workshop, 19 Nov. Bologna, Italy

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081555

NORCE

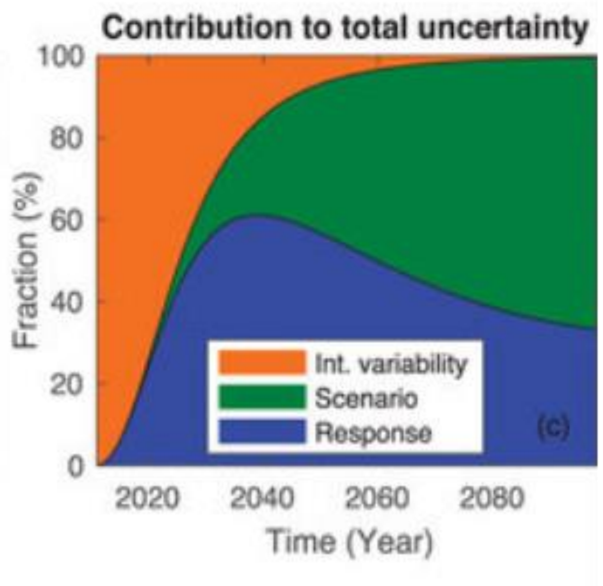
Deser et al, 2023

The twin challenge: climate information at local and decadal scales

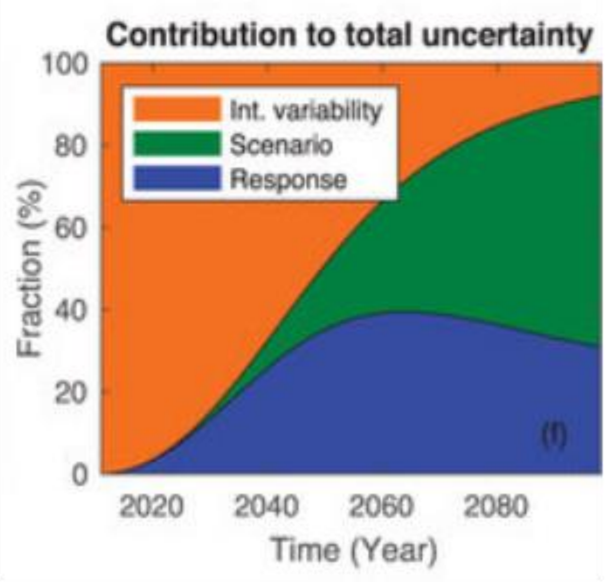
- Internal variability is the dominant uncertainty source at local scales in the coming decades
- Need large ensembles to explore all possible future weather situations and to compute event occurrence probability at given temporal horizons

Temperature Changes

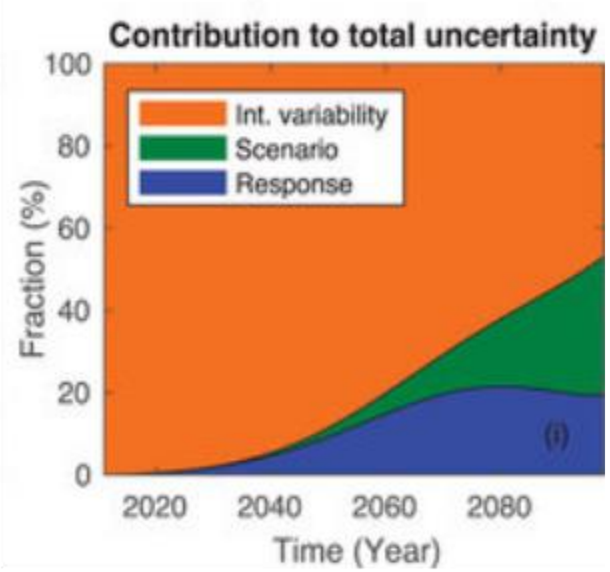
Global



North America



Locally (1 point, Anchorage)



Lehner and Deser, 2023

Outline of talk

1) ML Emulators: using CNRM-Unet as an example;
2) Ensemble approach;
3) Potential: promising early results;
4) Challenges: consistency, transferability, etc.
5) Prognoses: cause for hope?

Photo: Bård Bøe

CNRM-UNET RCM Emulator: concept / training

Concept

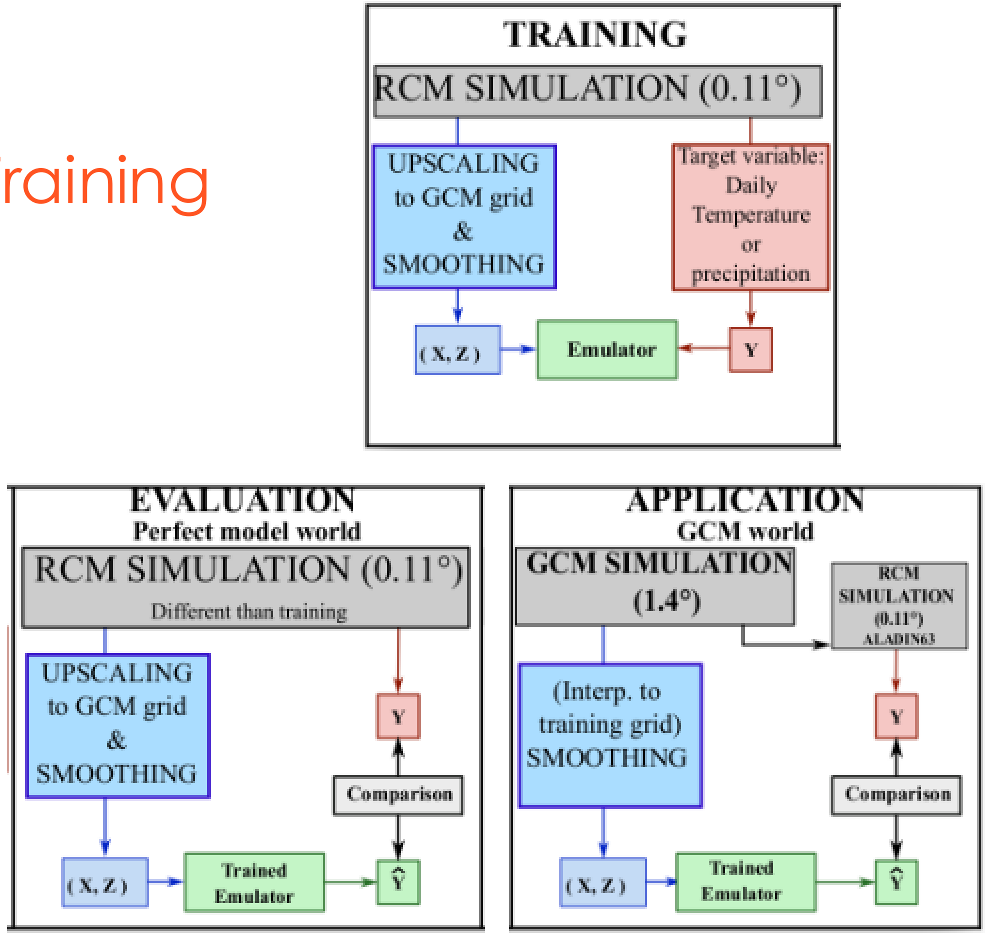
RCM: ALADIN63 (12km, driven by CMIP5 runs)

Target variables : Daily Temperature & Precipitations

Conceptual/Technical aspects :
Training in Perfect model framework
To ensure consistency between inputs and outputs as RCM tends to modify GCM large scale information

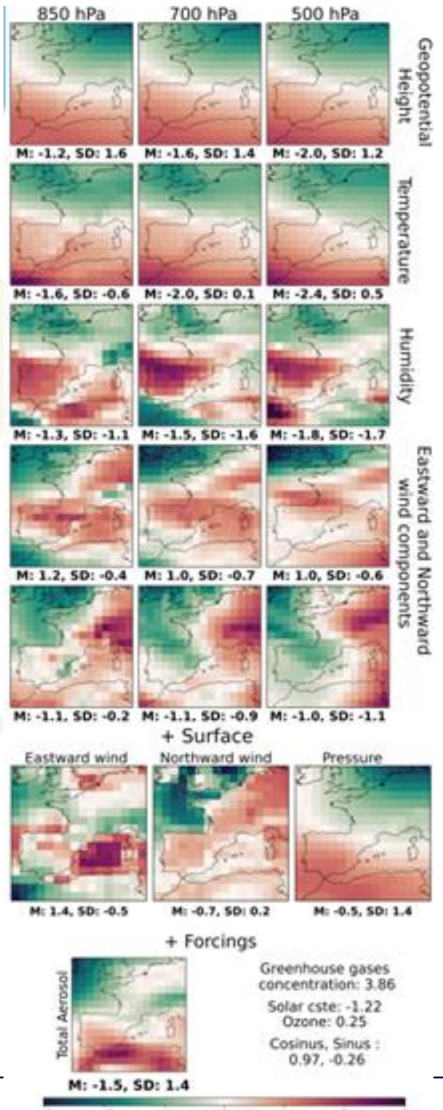
Doury et al. 2023, 2024

Training

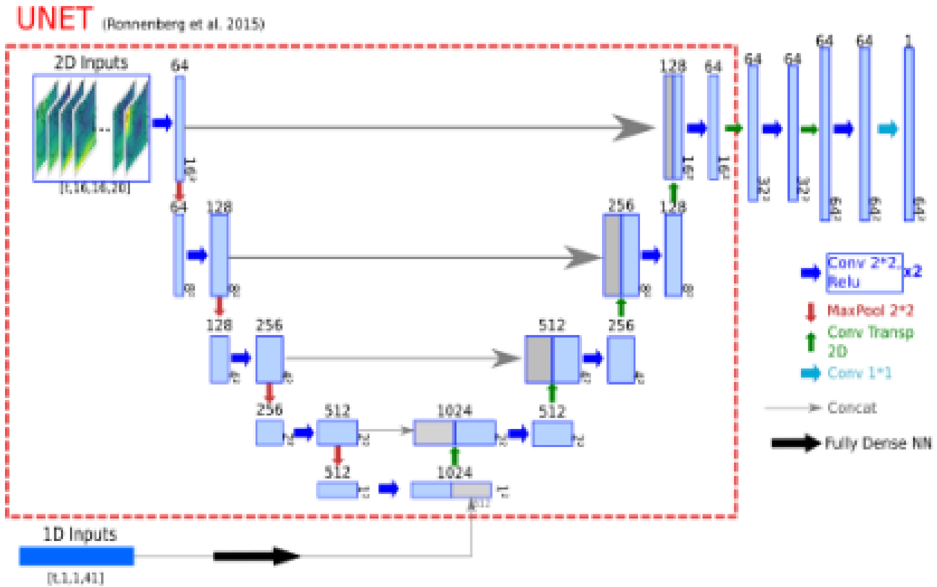


CNRM-UNET RCM Emulator : in practice

Input data



U-Net architecture



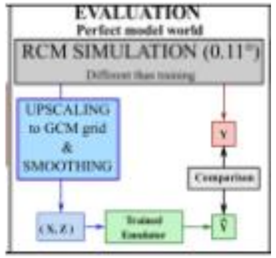
- ~ 25 million parameter
- ~ 1h to train on GPU (depends on the target domain size)
- ~ 1 min to predict

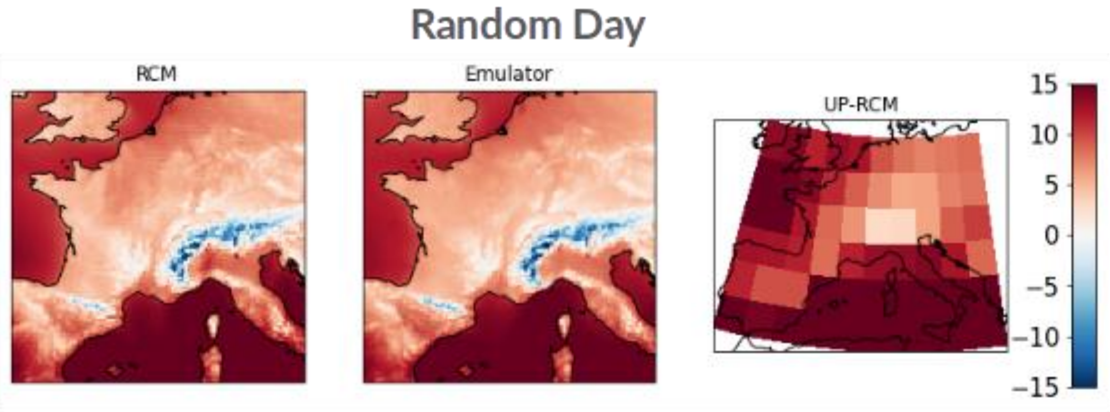
Doury et al. 2023

CNRM-UNET RCM Emulator: first evaluation

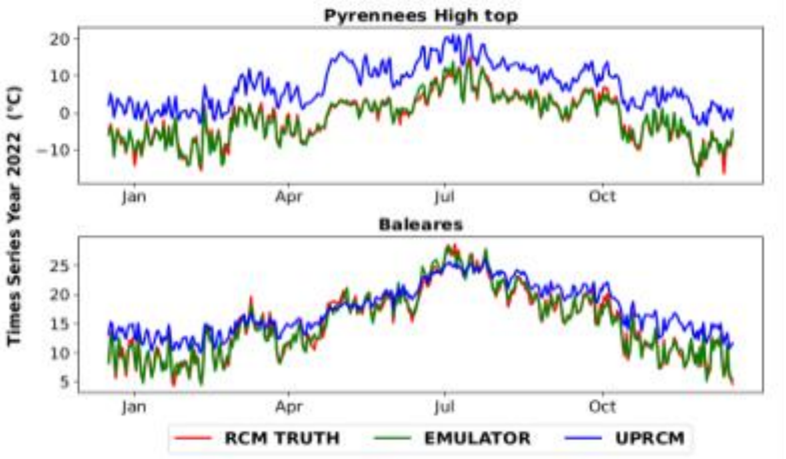
Perfect model evaluation : Temperature

Doury et al, 2023





- Good reproduction of the high resolution structure
- Very high spatial correlation (0.99)
- Correct range of temperature

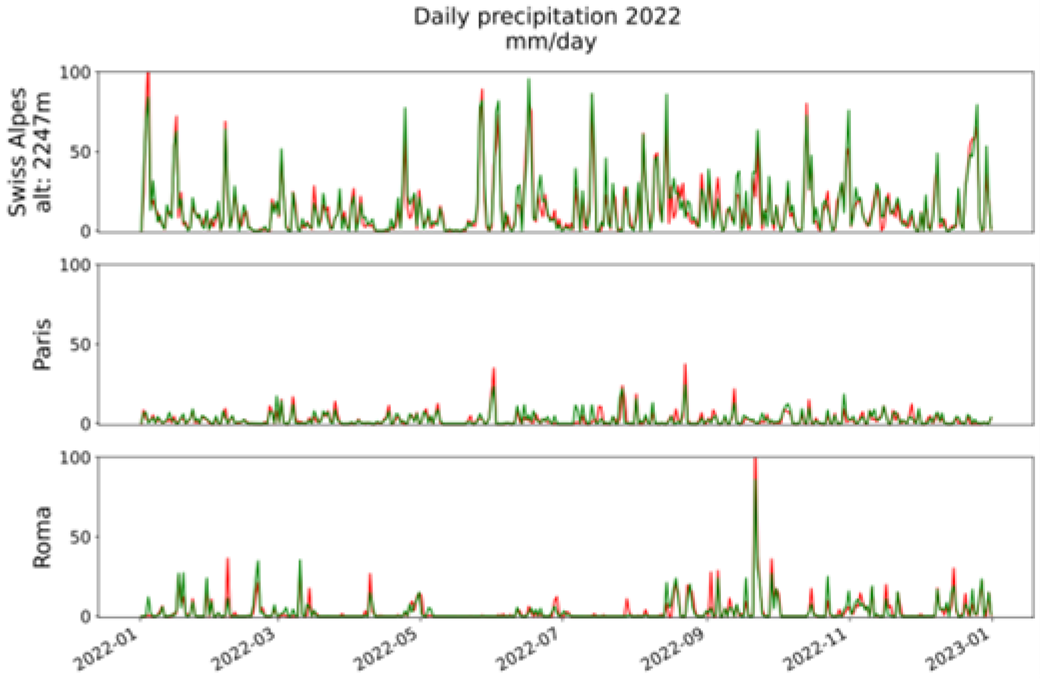


- Good temporal correlation at the grid point scale
- Good reproduction of the temporal variance

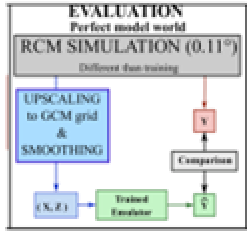
Doury et al. 2023

CNRM-UNET RCM Emulator: precipitation

Precipitation : illustration



RCM EMULATOR

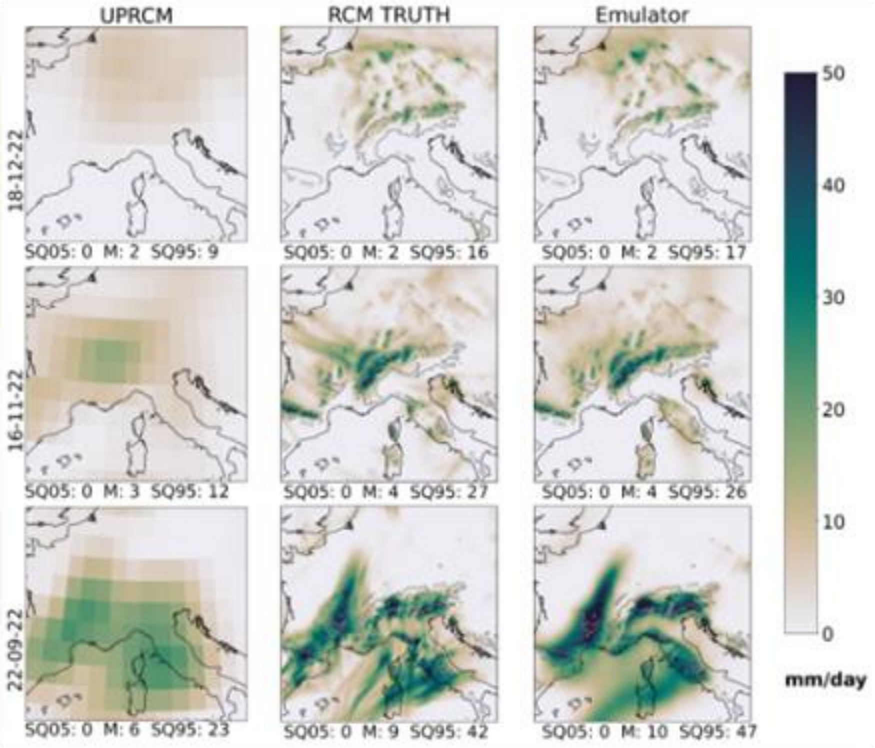


- Good representation of the local series
- Good temporal correlation
- Reproduction of the different variances
- Reproduction of extremes (ex: Roma)

Doury et al. 2024

CNRM-UNET RCM Emulator: precipitation

Precipitation : illustration



- 'Precipitation lookalike map'
- Good spatial correlation
- The precipitation object are correctly reproduced :
    - Well located
    - Correct intensity
    - Too "smooth" or bleeding

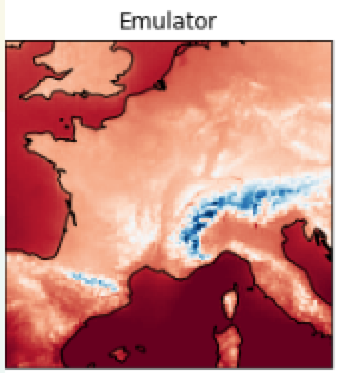
Due to UNeT.

Doury et al. 2024

CNRM-UNET RCM Emulator: large-ensemble production

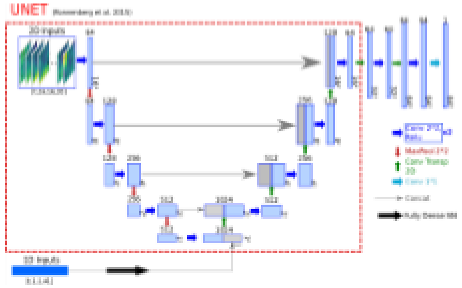
- After training on existing runs, RCM Emulators are cheap tools allowing to produce large ensembles at local scale
- WP3 is currently developing at least 5 different emulators of RCMs, able to produce large ensembles (12km to 2.5km, daily to hourly, precipitation and near-surface temperature)
- First illustration with the first large ensemble produced with the CNRM-UNET RCM emulator

Training dataset



incl. Paris, Barcelona, Prague

EURO-CORDEX CNRM-ALADIN63@12km
Two GCM-RCM runs for a total of 300 years
- RCP8.5 / CNRM-CM5 / 1950-2100
- RCP8.5 / MPI-ESM-LR / 1950-2100



Doury, Dubuisson, Somot, pers. comm.

Large ensemble production

CMIP6 GCM # of member	Historical	SSP126, SSP245, SSP370, SSP585
	1950-2014	*2015-2039*
CNRM-CM6-1	22	96
MPI-ESM1-2-LR	10	40
UKESM1-0-LL	5	20
IPSL-CM6A-LR	15	33
CanESM5	20	80
Total	72	269

Expanding this protocol to emulator multi-verse

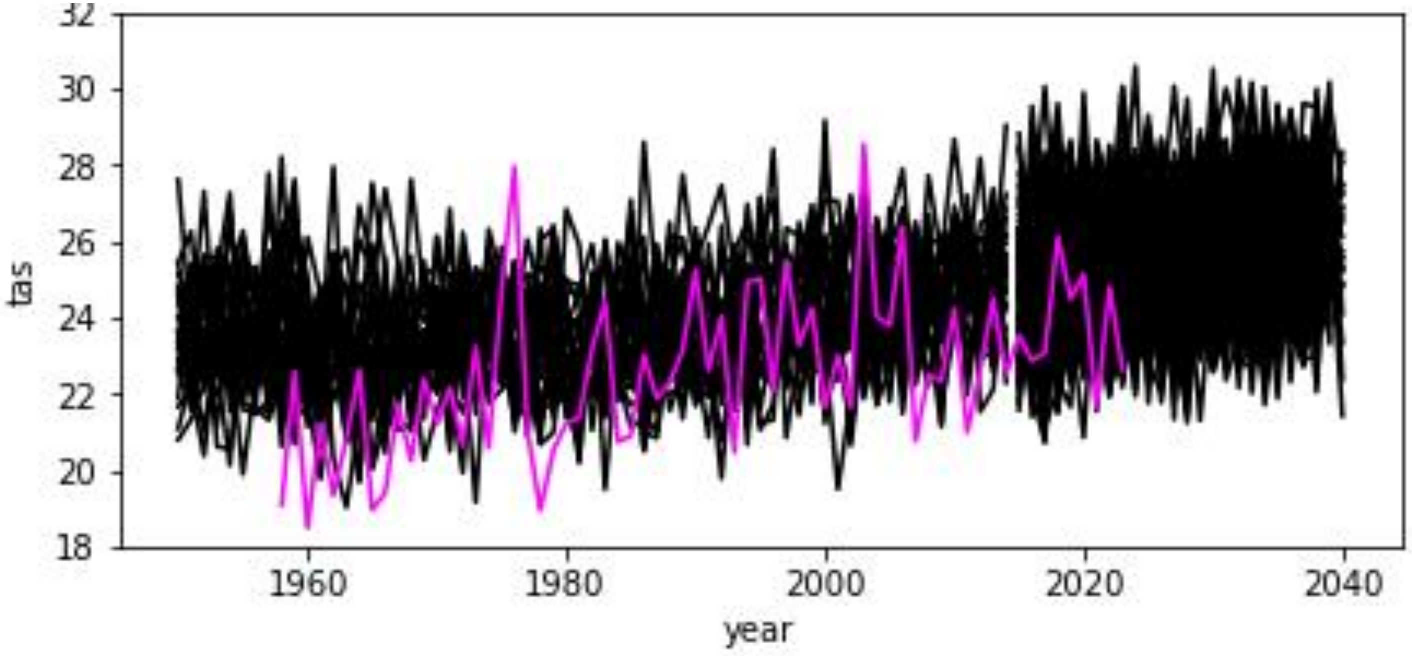
Tableau2

PLEASE REFER TO THE D3.4 FOR DEFINING YOUR SOURCE_ID: https://impetus4change.eu/wp-content/uploads/2024/05/I4C-D3.4-Description-of-Data-

Institute	Emulator name (source id)	Output variables	Temporal resolution	Temporal coverage	Spatial resolution	Domain	Demonstrator covered	Reference
CNRS-MF	CNRM-ALADIN63-emul-CNRM-UNET11-tP22	pr, tas	daily	min: 1980-2040, max: 1850-2100	12 km	ALPX-12	Prague Barcelona Paris	Doury et al. 2023, 2024
CNRS-MF	CNRM-AROME46t1-emul-CNRM-UNET11-tP12	pr, tas	daily	min: 1980-2040, max: 1850-2100	2.5 km	ALPX-3	Prague Barcelona Paris	Doury et al. 2023, 2024
CNRS-CERFACS	CONSTANA	tas, pr	daily	min: 1980-2040, max:	2.5 km	ALPX-3	Prague Barcelona Paris	Boé et al. 2023
CSIC	WRF451R-CI4C-emul-DeepESD1-t1	tas, tasmax, tasmin, pr	daily	min: 1980-2040, max:	3 km	ALPX-3	Prague Barcelona Paris	Bano-Medina et al. 2024
ICTP	GNN4CD	pr	hourly	min: 1980-2040, max:	3 km	ALPX-3 + NSEA-3	Prague Barcelona Paris Bergen	Blasone et al. (2024) Blasone et al. (2025)
UiB	HCLIM-ALADIN-emul-CNRM-UNET11-tP22	tas, pr	daily	min: 1980-2040, max:	12 km	NSEA-12	Bergen	Doury et al. 2023, 2024
UiB	WRF451R-CI4C-emul-CNRM-UNET11-tP22	tas, pr	daily	min: 1980-2040, max:	2.5 km	NSEA-3	Bergen	Doury et al. 2023, 2024
DMI	ANOVA	tas, pr, sfcwind	monthly	min: 1980-2040, max:	3 km	NSEA-3 + ALPX-3	Bergen Prague Barcelona Paris	Christensen and Kjellström, 2022

An encouraging example: Emulator ensemble reliably captures 2003 magnitude events

Temperature of the warmest fortnight of the year
(°C, Paris)

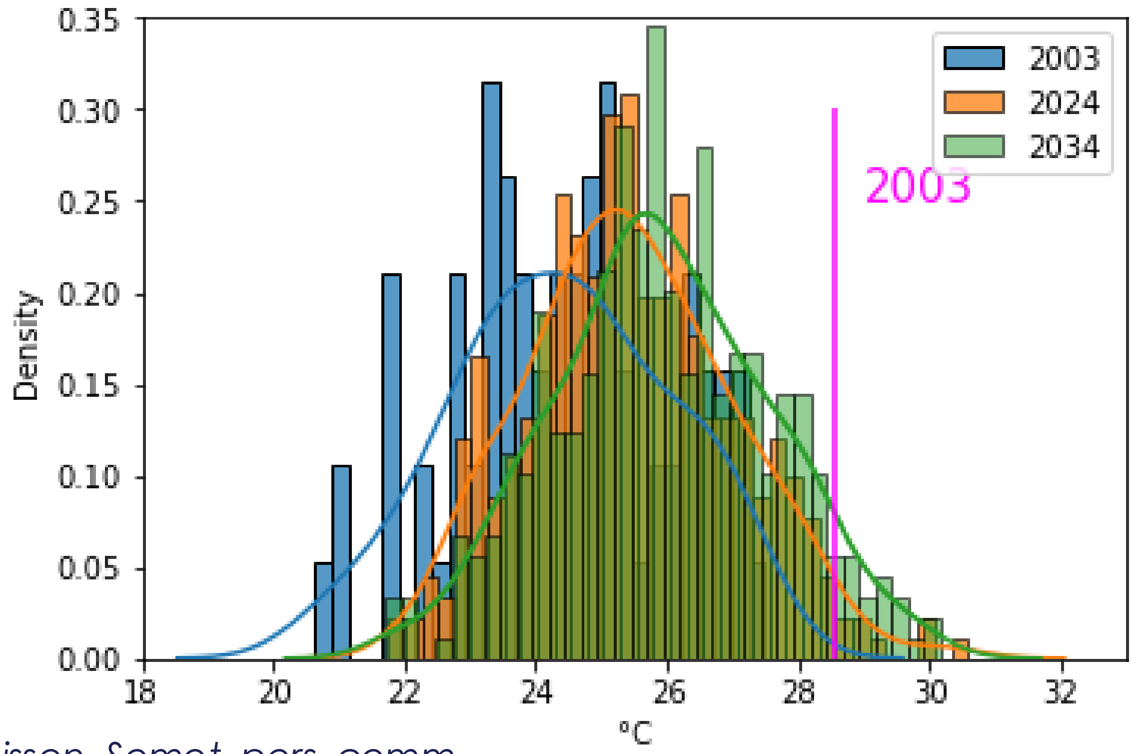


- Warmest observed record: 2003 with 28,5°C
- Values obtained only 4 times during the historical period (over 1232 years) → 0,3 %
- Values never obtained before 1985
- Ensemble mean and spread fits well with the observed time series

Doury, Dubuisson, Somot, pers. comm.

Potential applications: Near term probabilistic assessments of climate hazards

PDF of the warmest fortnight temperature
(°C, Paris)

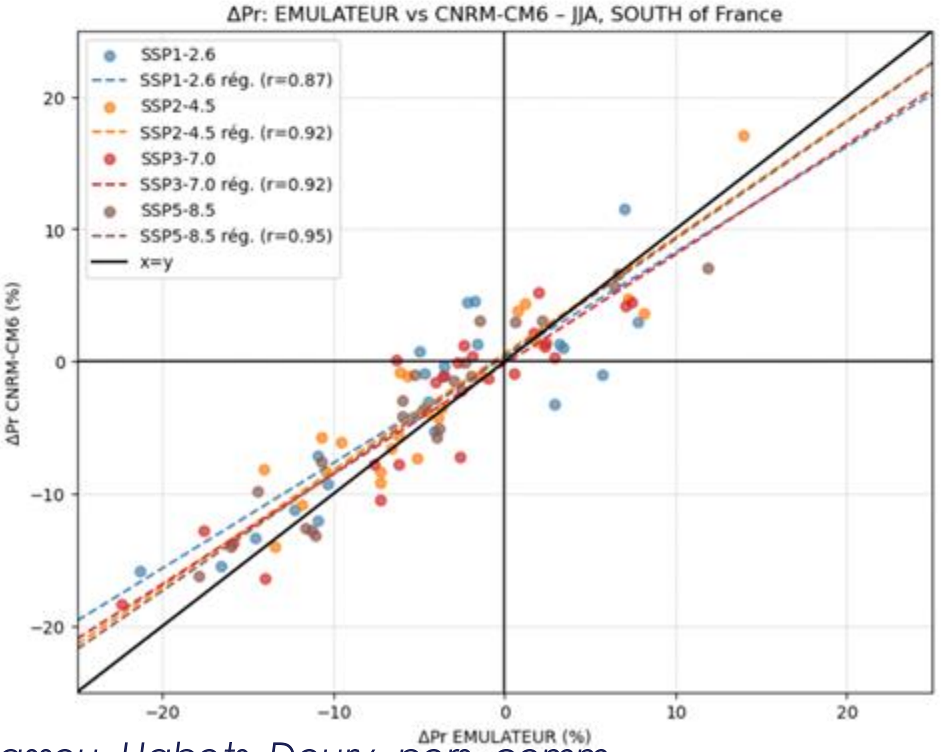


Doury, Dubuisson, Somot, pers. comm.

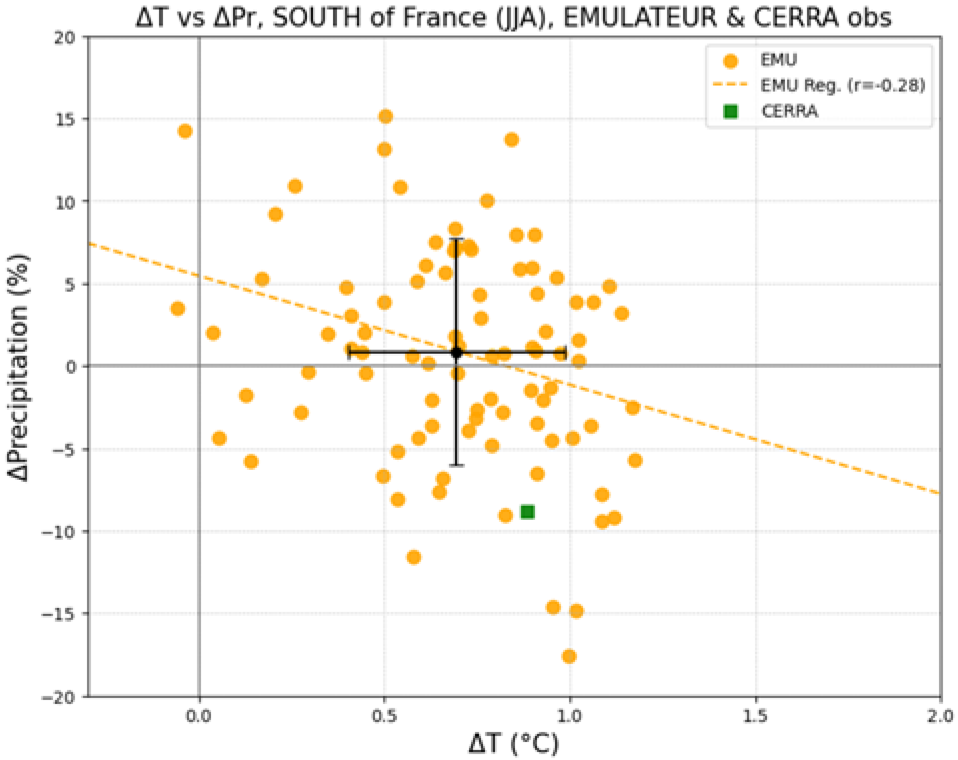
- Warmest observed record: 2003 with 28,5°C
- Probability to occur in :
  2003 : 1 %
  2024 : 4 %
  2034 : 7 %

Challenge: Consistency between emulator and driving GCM/Obs

Climate Change Consistency between the emulator and its driving GCM for precipitation



Climate Change Consistency between the emulator and the observations



Tribou, Cassou, Habets, Doury, pers. comm.

Challenge: Regional transferability, GCM transferability

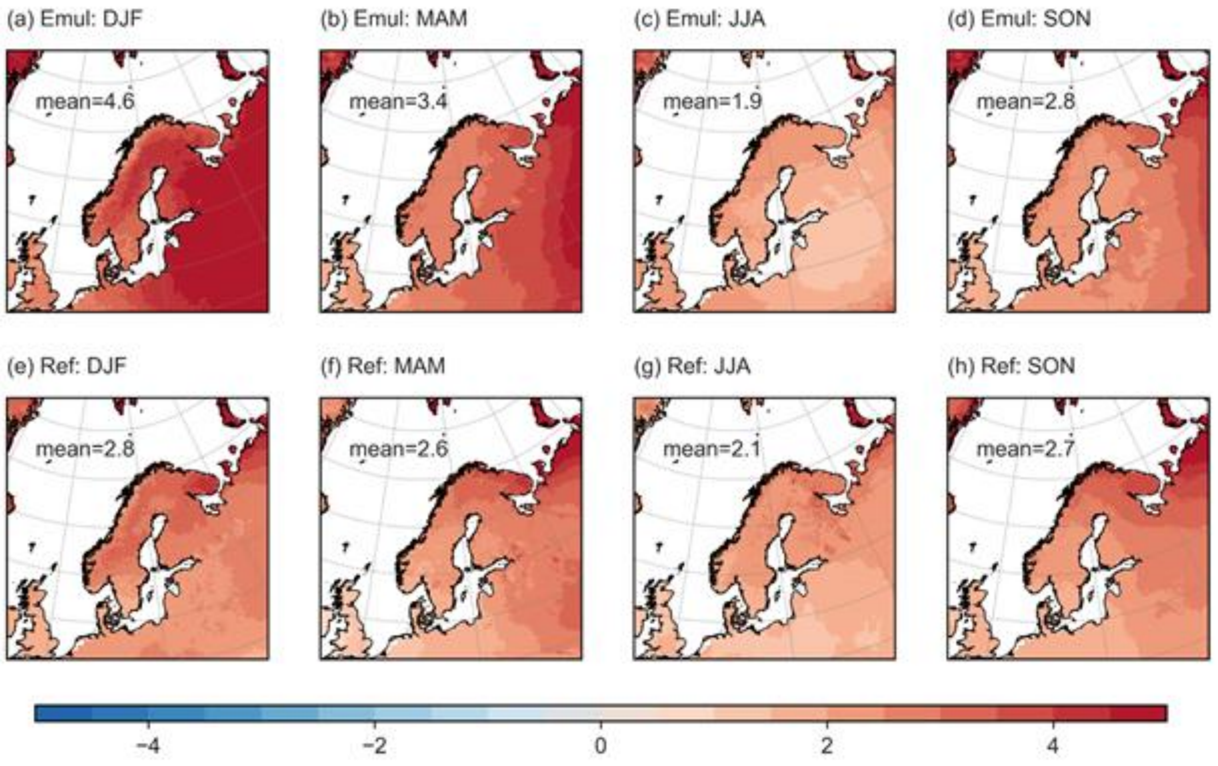
Methodology: Perfect model strategy

Training Phase: The U-net learned the relationship between low-resolution (Upscaled to GCM resolution) and high-resolution target variables (temperature and precipitation) from RCM simulation driven by GFDL

Prediction Phase: low-resolution (Upscaled to GCM resolution) Atmospheric variables from RCM simulation driven driven by EC-EARTH were used as inputs to obtain high resolution target variables

Evaluation: U-Net outputs were evaluated against high-resolution target variables from RCM simulation driven by EC-EARTH as a reference

Climate Change Signal (2041-2060)



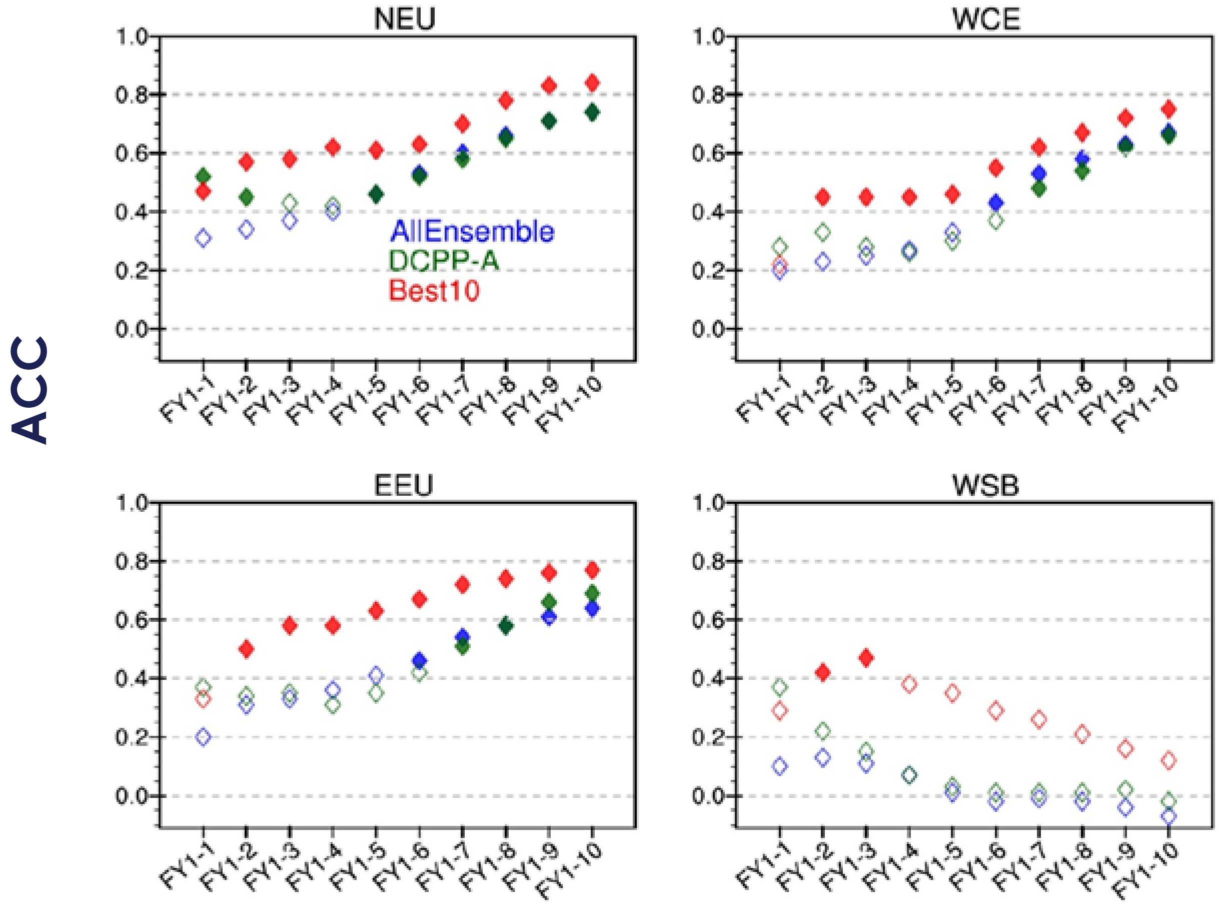
Challenge: lack of driving data; Solution: Constrain the NAO-temperature teleconnection in CMIP6 simulations

Methodology:

1. Compute NAO-temperature teleconnection based on their first order linear regression in observational (i.e. ERA5) data and separately for each individual model member for a 20 year winter window.

2. Compute pattern correlation between the observed regression (i.e. teleconnection) and the model member regression, resulting in one pattern correlation value for each member. The pattern correlation is calculated over the whole masked region as mentioned above. Based on these pattern correlation values the members were ranked from highest to lowest ranking members. The top ranking members are used to predict for the next 10 years.

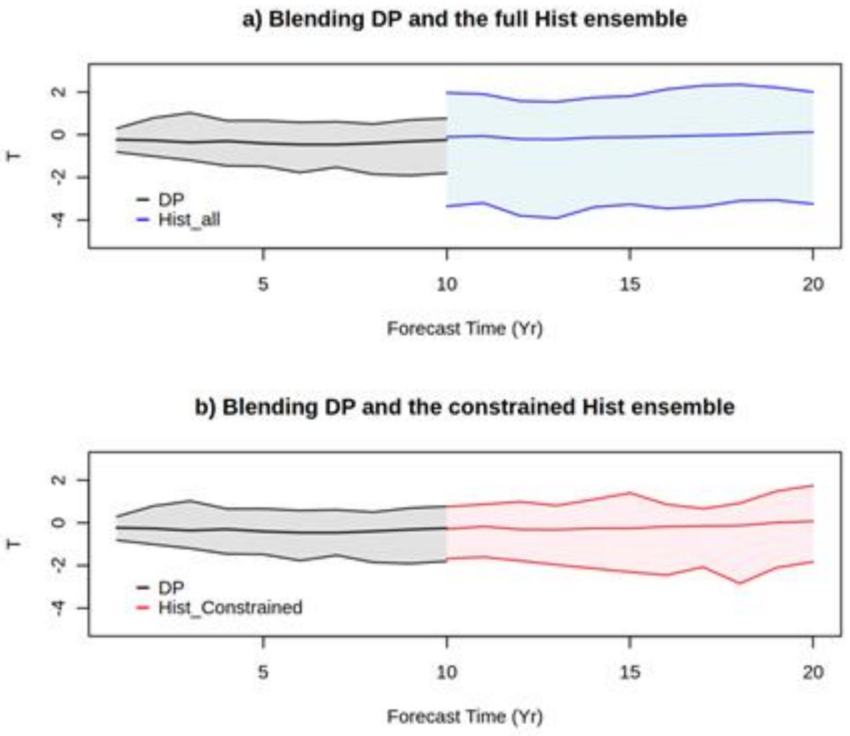
Skill of the constrained ensemble in predicting winter temperature compared to full ensemble and also the DCPP-A ensembles



Mahmood et al. (2025)

Challenge: Inconsistency between prediction and projection ensembles; Solution: Analog-based constraints

- Decadal predictions are generally more skillful than climate projections, but they usually extend only up to about 10 years into the future. Beyond that range, one might consider appending projections to predictions; however, this can result in inconsistencies in the transition period.
- We have developed a novel blending approach based on an **analog-based constraining** method. For each decadal forecast ensemble member the projection member in closest agreement is selected, thereby generating an ensemble of climate projections that matches the size of the decadal predictions ensemble.
- This approach aims to maintain the **physical and statistical coherence** of both ensembles, minimizing inconsistencies across the transition period.



Example of blended decadal predictions and historical simulations SST in the Sub-polar North Atlantic for a) the full historical ensemble and b) constrained ensemble. Spread represents the 5th-95th percentiles of the member ensembles.

Prognosis and next steps

- ML-based emulators show promising results, at least comparable to their dynamical counterparts and at a fraction of the cost.
    - Several challenges remain however (extrapolation, extremes, transferability), and these techniques are not a panacea for poor skill.
- Large ensembles of these can be produced with CMIP, DCCP, or any other multi-model product that outputs the requisite variables.
    - Constraining CMIP could be a promising avenue for near term predictions
- A multi-emulator approach is likely needed
    - ML emulators like their dynamical counterparts exhibit a wide range of capabilities; there is no one "best" model.
- Establishing benchmarks and best practices for ML techniques is crucial (cf. CORDEXBench)
- Need to move to multivariate, physically consistent emulators at sub-daily scale