Multi-model ensemble clustering at Météo-France for operational seasonal prediction

Damien Specq[1], Shan Li[1], Lauriane Batté[2], Frédéric Gayrard[2], Pierre Bonnin[2]

[1]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[2]Direction de la climatologie et des Services Climatiques, Météo-France, Toulouse, France
Contact: damien.specq@meteo.fr

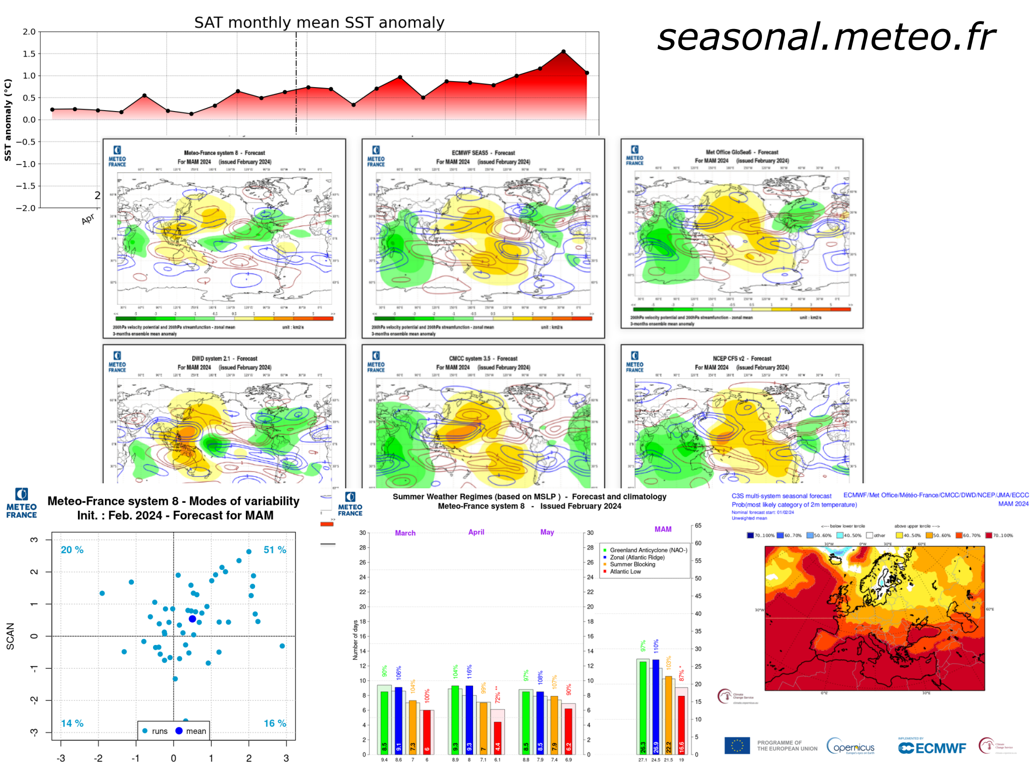
MedCOF workshop
2 July 2024

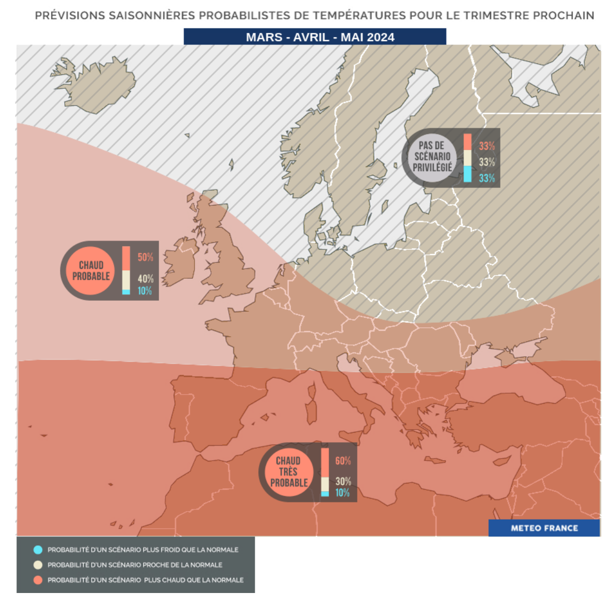
This work is funded by the Copernicus Climate Change Services (C3S) contract with Météo-France on Operational Seasonal Prediction

Context

- Météo-France **seasonal forecast bulletins** produced every month
- Based on **multimodel C3S seasonal forecast** outputs (init. 1st of month)
- Main target: **next season** lead time 1, e.g MAM for February init.
- **Synthesis maps** of temperature (T2m) and precipitation (RR) over Europe

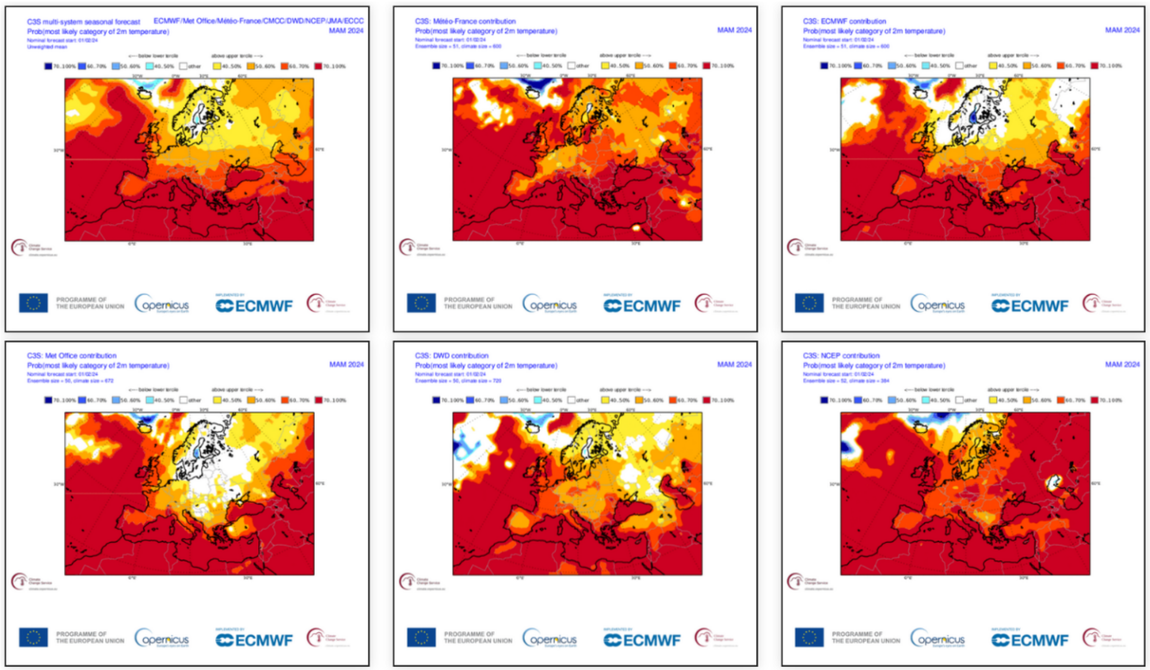
seasonal.meteo.fr





▲ *Expertised operational T2M synthesis map for MAM 2024*

Motivation for seasonal forecast scenarios



C3S multi-models probability map (top left) and MF-S8, ECMWF-SEAS5, UKMO , DWD, CMCC models.

◄ *Forecast T2m anomalies for MAM 2024 in the C3S models*

- What are the main “points of disagreement” between the ensemble members?

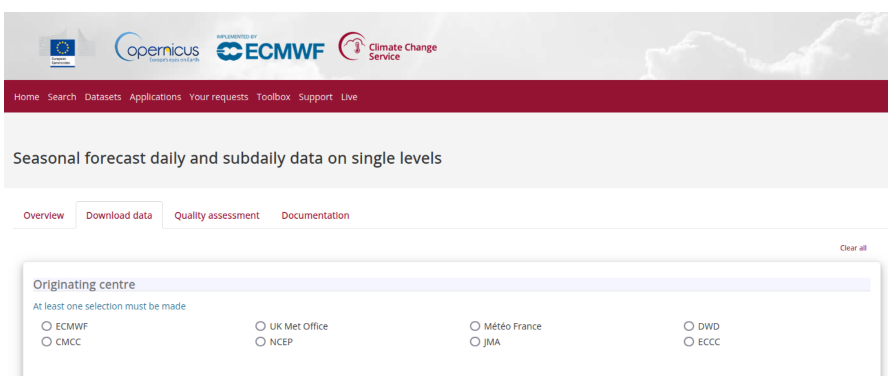
- Providing **a representation of seasonal forecasts** that:
  - Is more refined than the multimodel ensemble mean
  - Is **more relevant to summarize uncertainty** than the collection of individual model forecasts

Data & methodology

Data

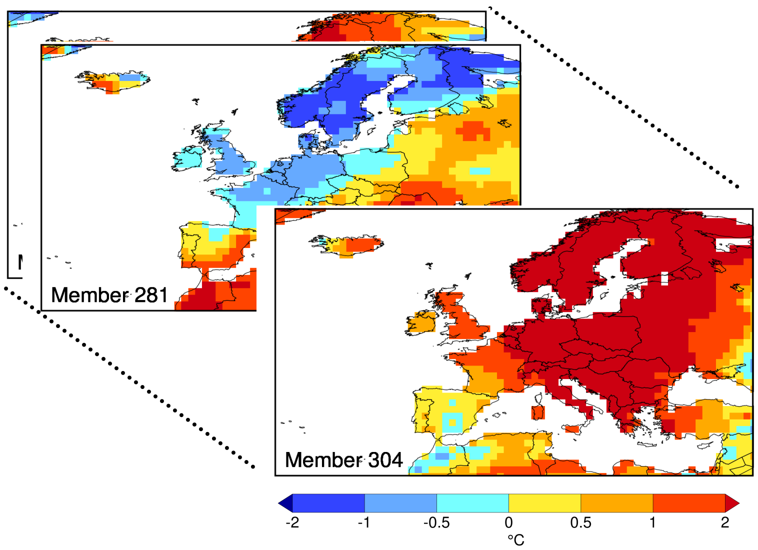
- C3S seasonal forecast outputs from 6 models
- Surface parameters (T2m, RR) and atmospheric circulation parameters (Z500)
- 1° archiving resolution
- European focus: 29.5°W-40.5°E; 30.5°N-70.5°N
- Reforecast period: 1993-2016 (24 years)
- Real-time forecast period: 07/2021 – present

Model	CMCC	DWD	ECMWF	MF	NCEP	UKMO	**Multi-model**
Real-time ensemble size	50	50	51	51	52	50	**304**
Reforecast ensemble size	40	30	25	25	24	28	**172**



◀ *cds.climate.copernicus.eu*

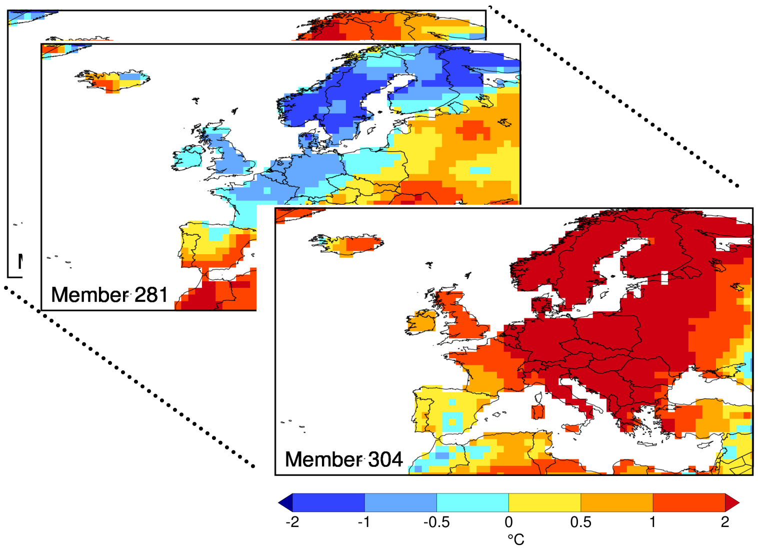
Defining scenarios: a cluster analysis



◀ *T2m MAM 2024 anomaly maps in the 304 members of the C3S multimodel real-time forecasts issued February 2024*

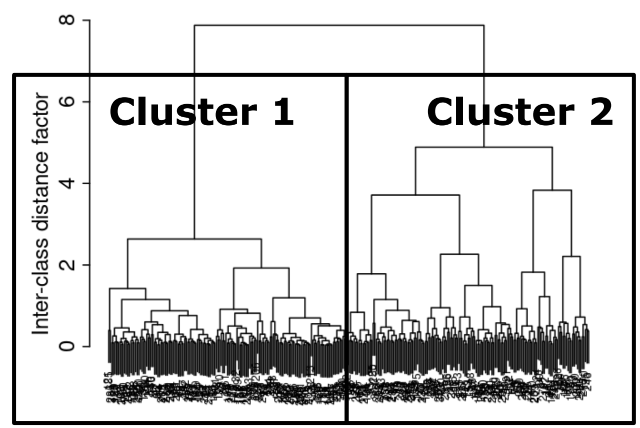
- 304 members to be grouped according to how close they are
- Cluster analysis : 1 cluster = 1 scenario

Defining scenarios: one retained solution



Dissimilarity between maps T_i and T_j
d(i,j) = 1 – ACC(i,j)
(e.g Nakaegawa and Kanamitsu, 2006)

Hierarchical clustering
Ward's method



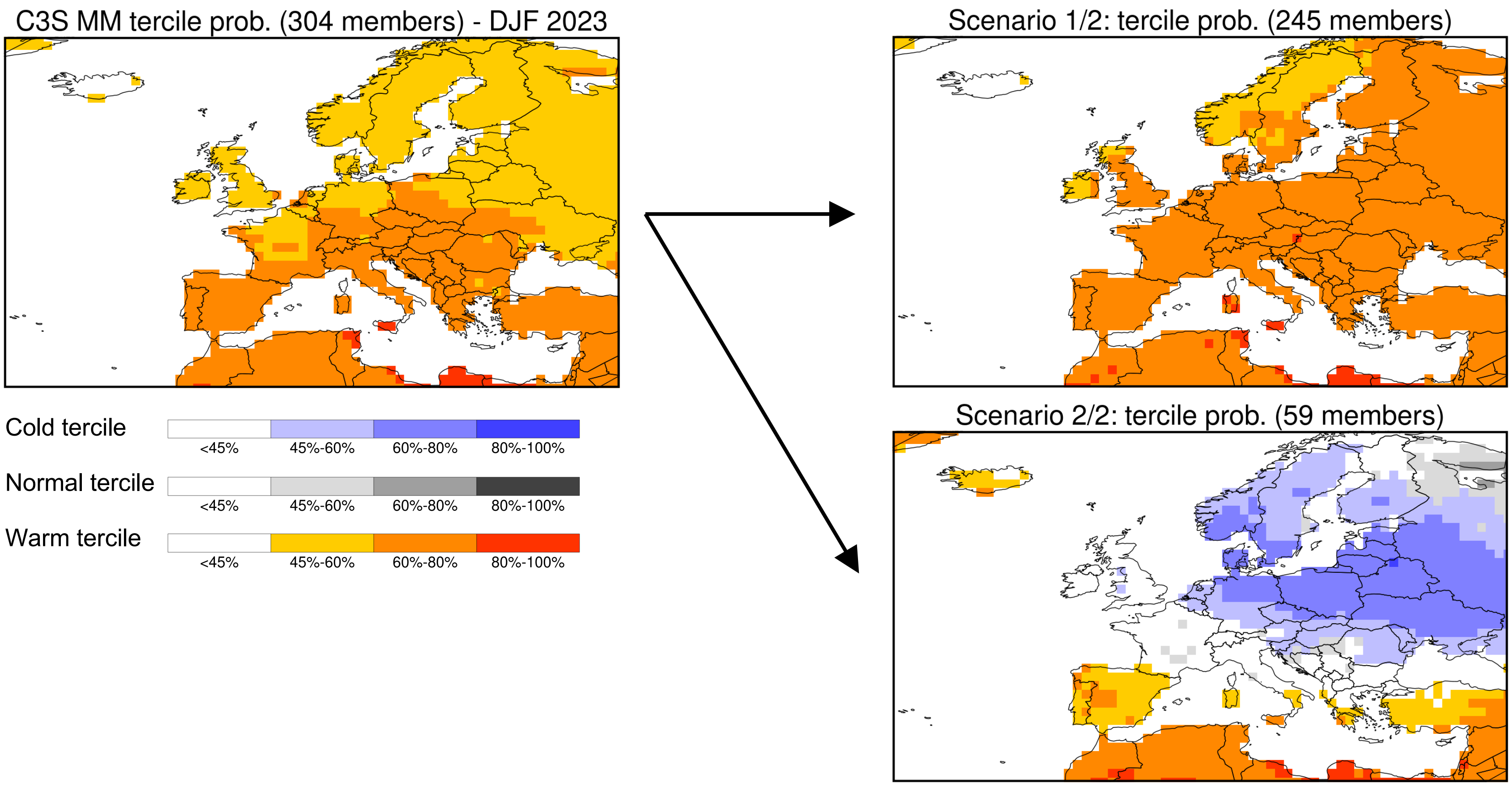
▲ *Dendrogram of the 304-member clustering*

- **Clustering on T2m** anomaly maps
- Dissimilarity **based on ACC** → members with similar patterns are grouped together, irrespective of strength of anomalies

Case study 1: DJF 2023-2024 forecasts (issued November 2023)

DJF 2023/24 issued November 2023

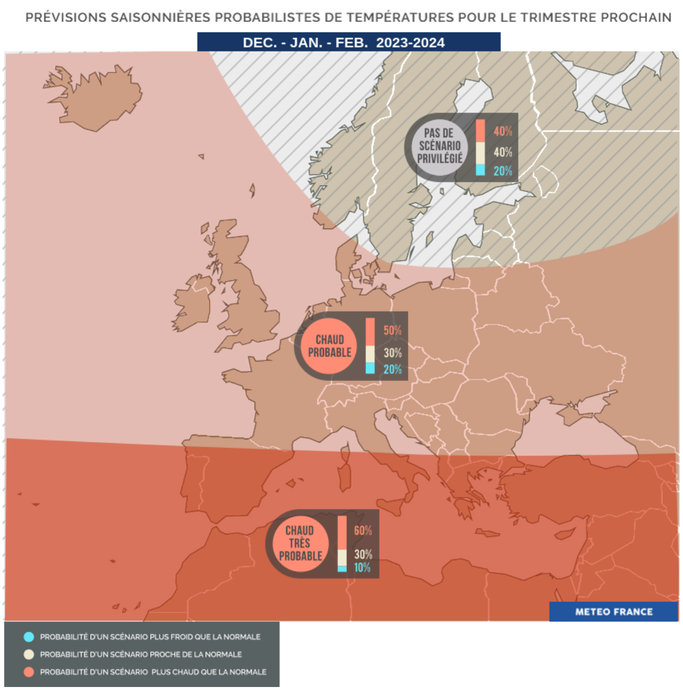
- Scenario composites for T2m (i.e the clustering variable)



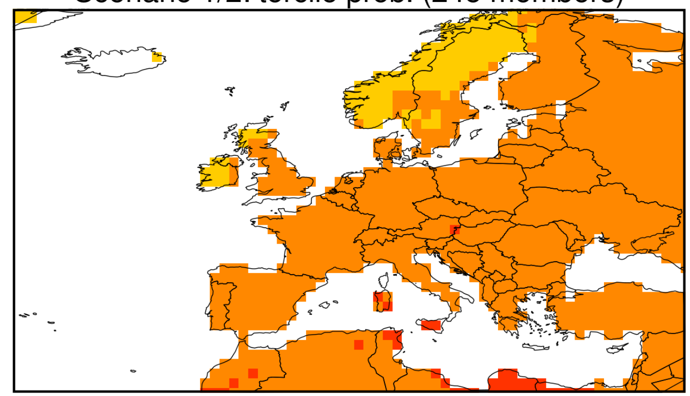
- Strong imbalance between scenarios

DJF 2023/24 issued November 2023

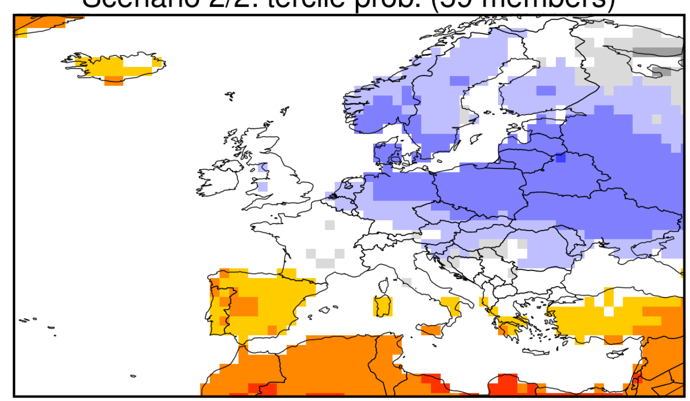
- Scenarios helped devise the synthesis map



Scenario 1/2: tercile prob. (245 members)

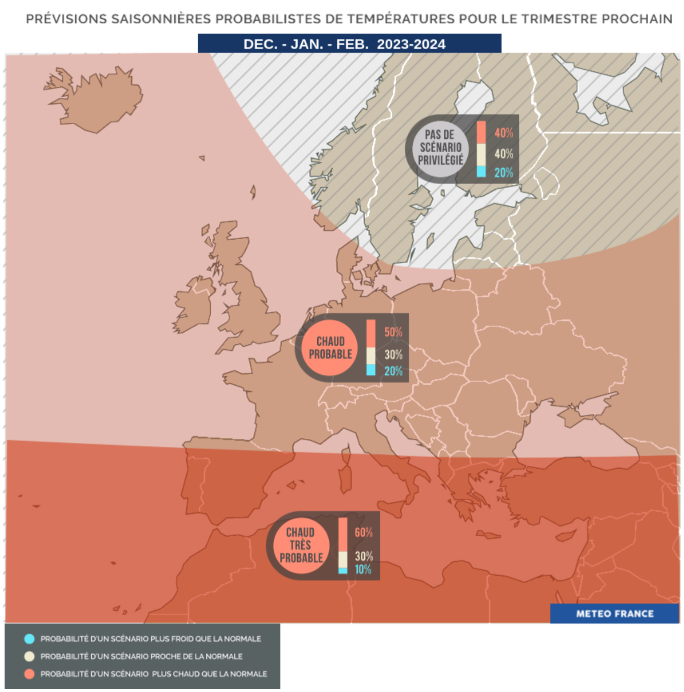


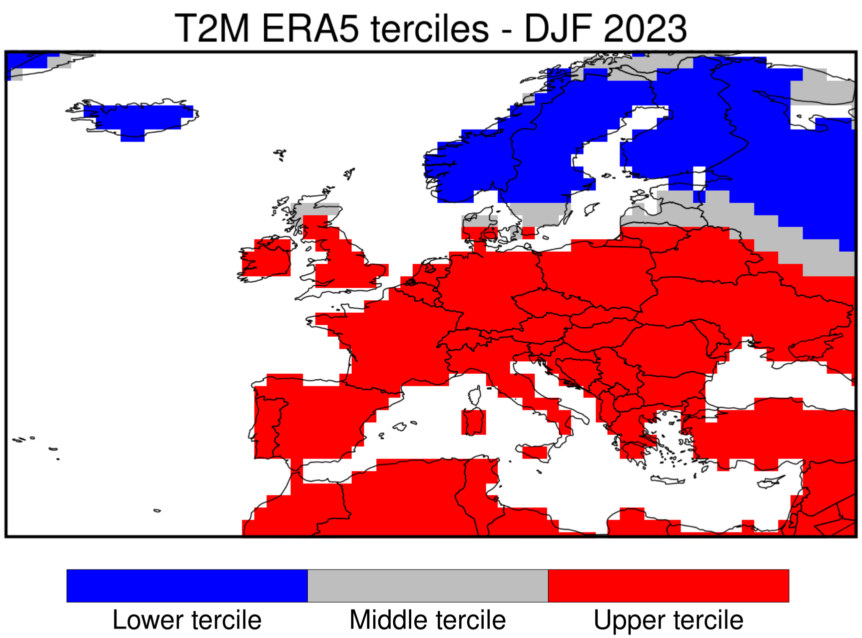
Scenario 2/2: tercile prob. (59 members)



DJF 2023/24 issued November 2023

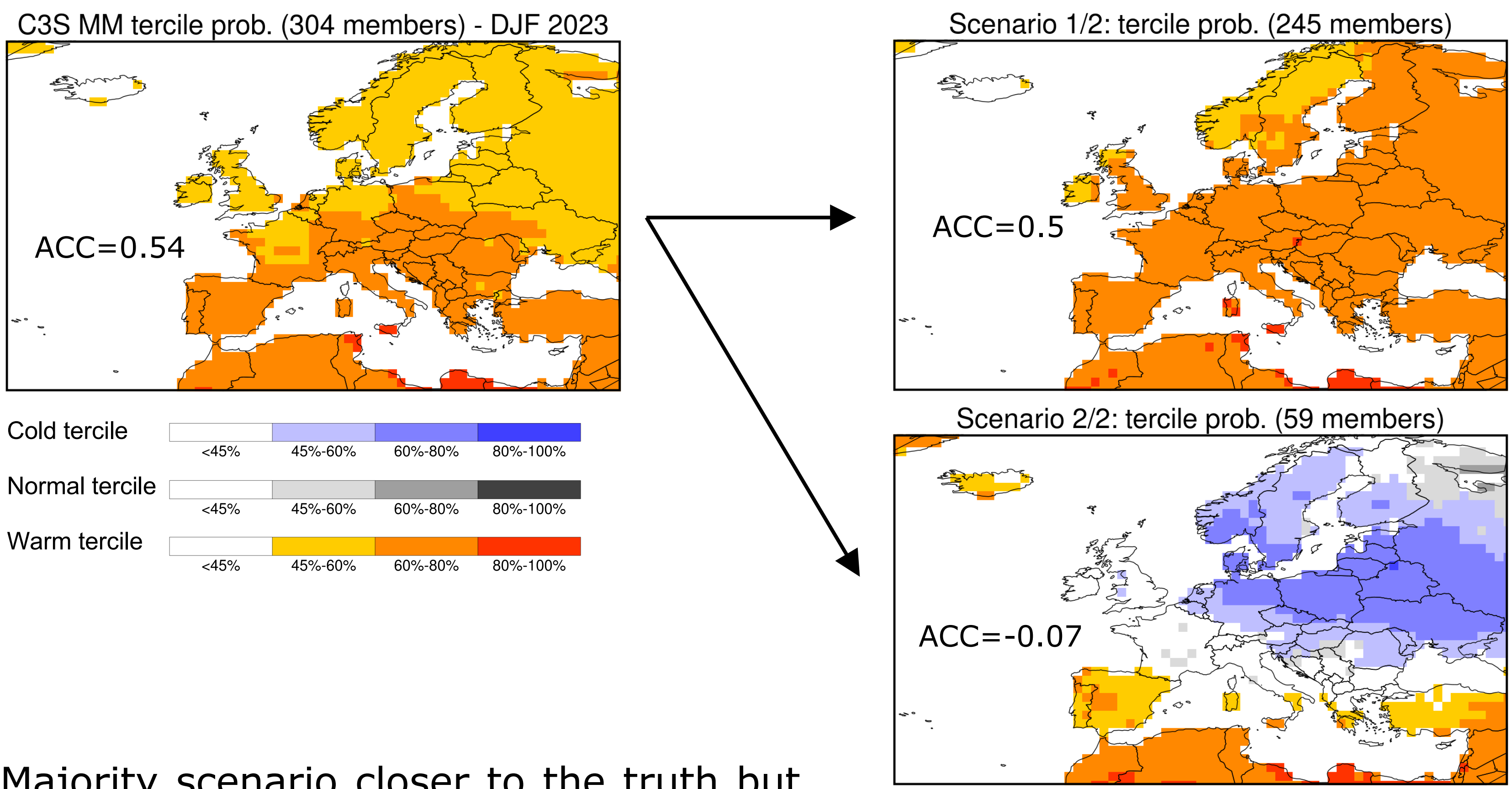
- Comparison with the actual outcome: it was wise to be careful about Scandinavia!





DJF 2023/24 issued November 2023

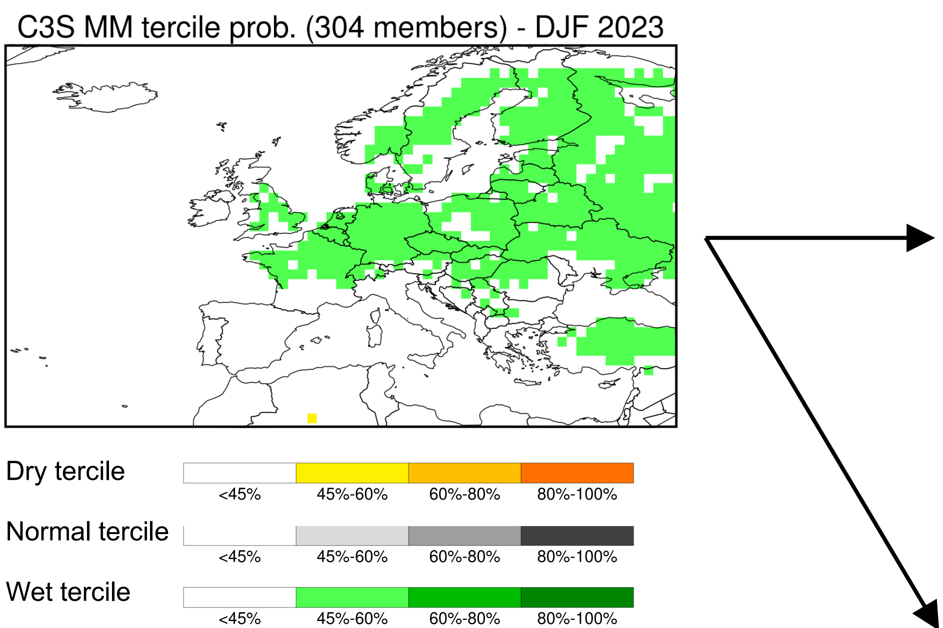
- Comparison with the actual outcome: spatial correlation ACC

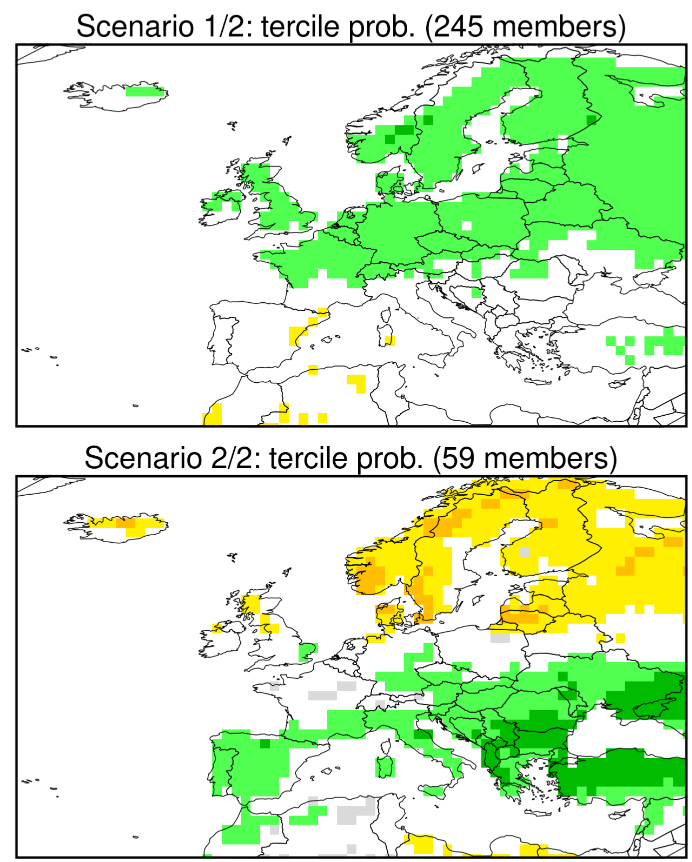


- Majority scenario closer to the truth but no better than the ensemble mean

DJF 2023/24 issued November 2023

- How does T2m clustering reflect on **rainfall composites**?

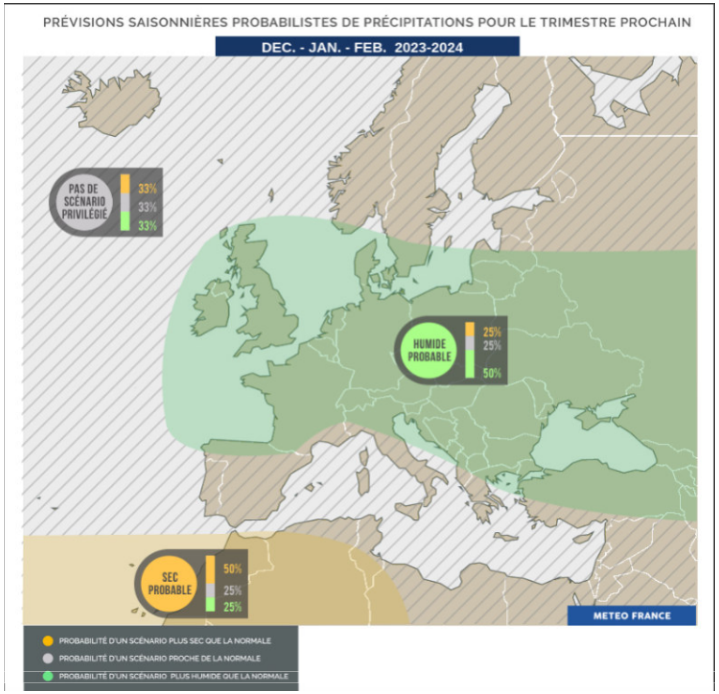




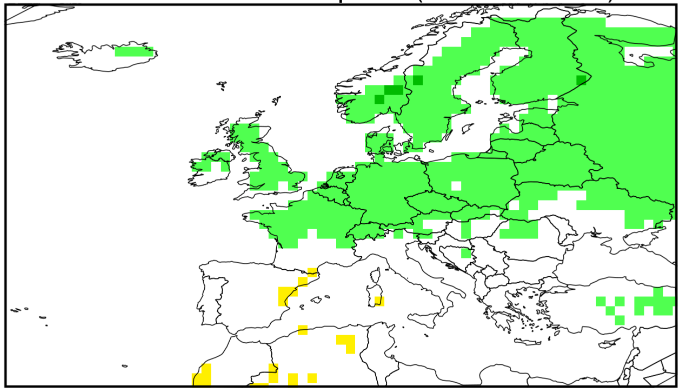
- Strong and consistent differences in precipitation as well
- Good agreement with T2m-RR relationships in winter (e.g "colder in drier")

DJF 2023/24 issued November 2023

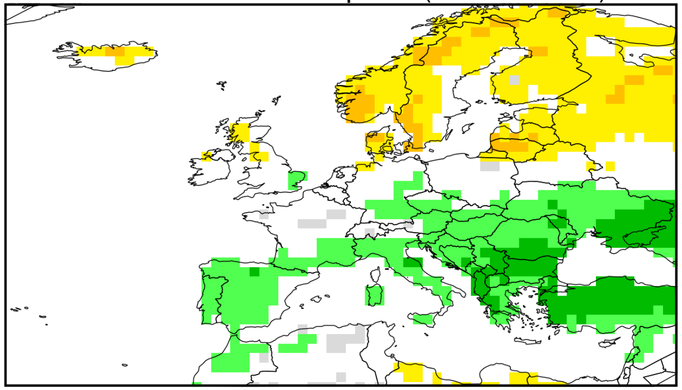
➢ Rainfall synthesis map vs Scenarios: wet winter is forecast where both scenarios agree



Scenario 1/2: tercile prob. (245 members)

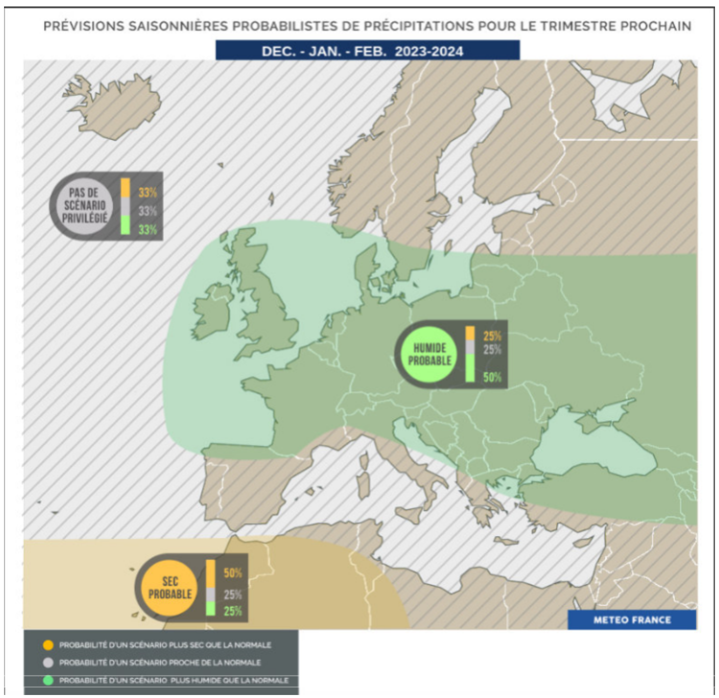


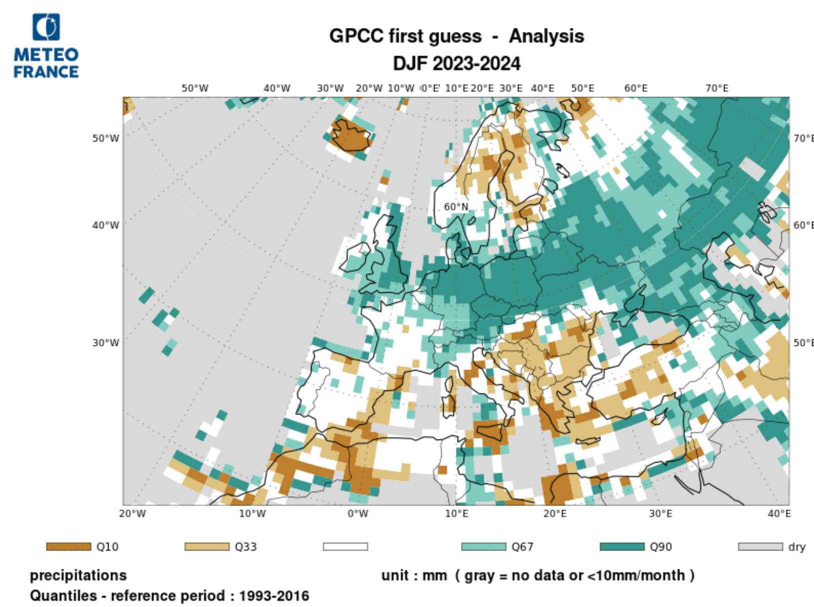
Scenario 2/2: tercile prob. (59 members)



DJF 2023/24 issued November 2023

➢ Comparison with the actual outcome: it was wise to be careful about Scandinavia again!

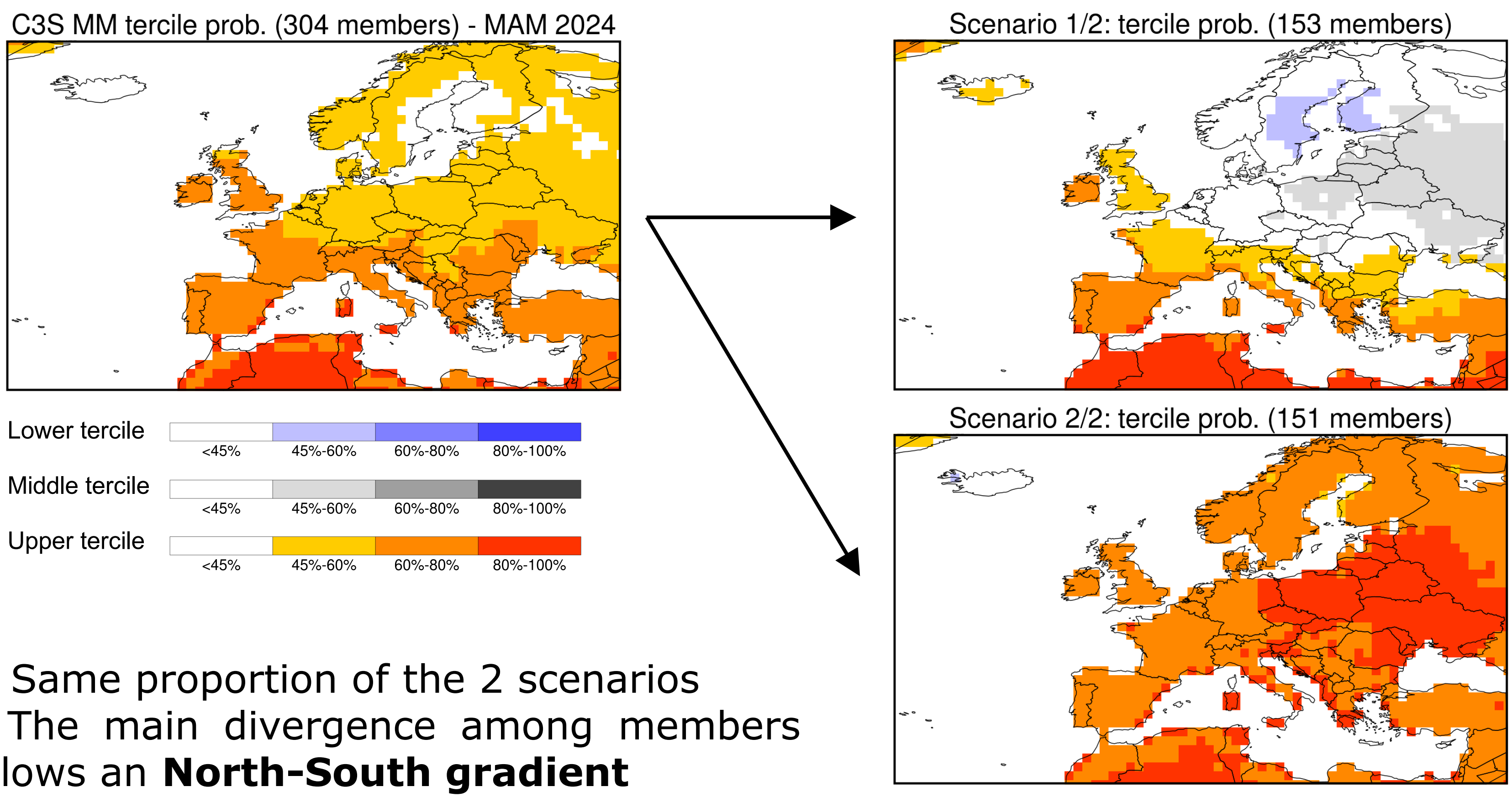




Case study 2: MAM 2024 forecasts (issued February 2024)

MAM 2024 forecasts issued February 2024

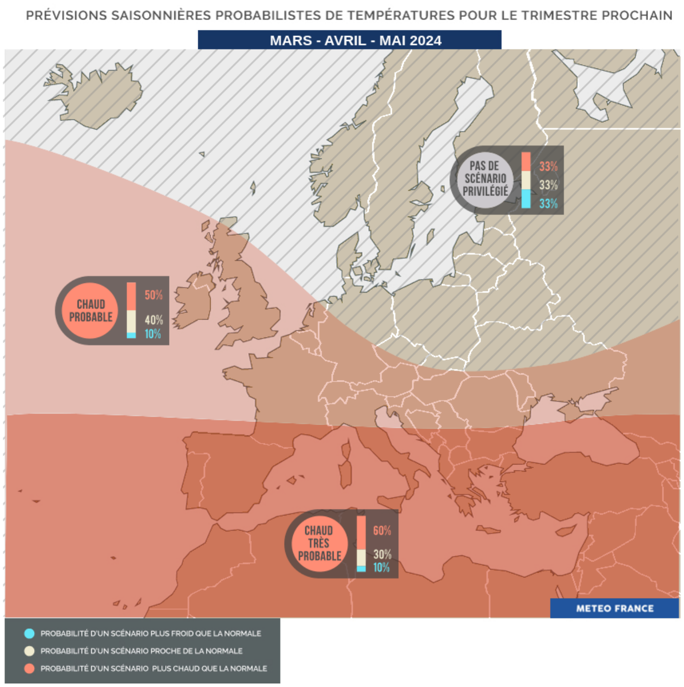
- Scenario composites for T2M (i.e the clustering variable)



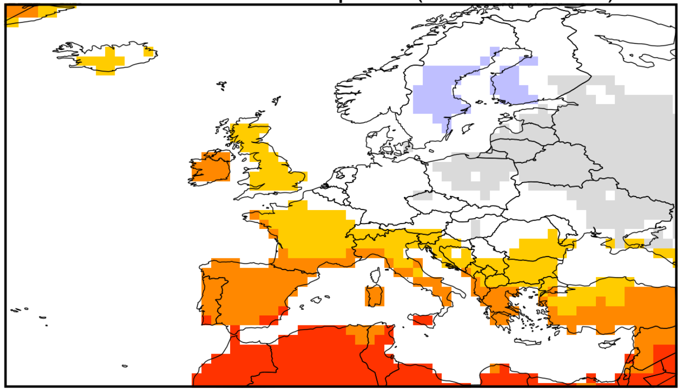
- Same proportion of the 2 scenarios
- The main divergence among members follows an **North-South gradient**
- Western Europe and the Mediterranean **warmer than normal in all cases**

MAM 2024 forecasts issued February 2024

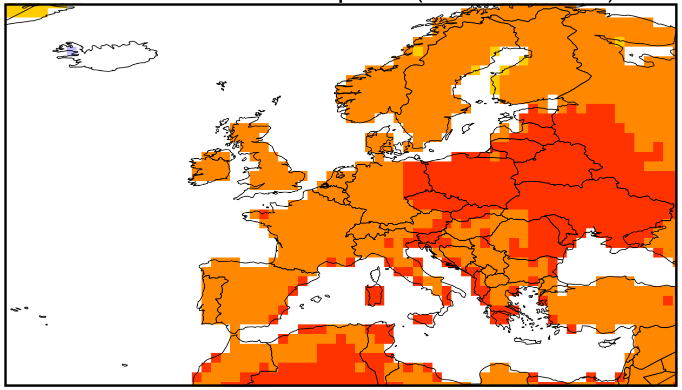
- Scenarios helped devise the synthesis map



Scenario 1/2: tercile prob. (153 members)

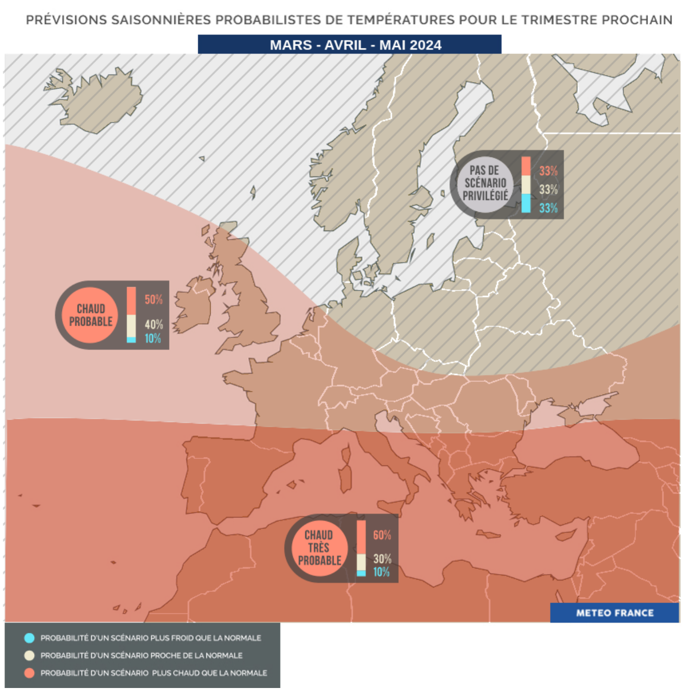


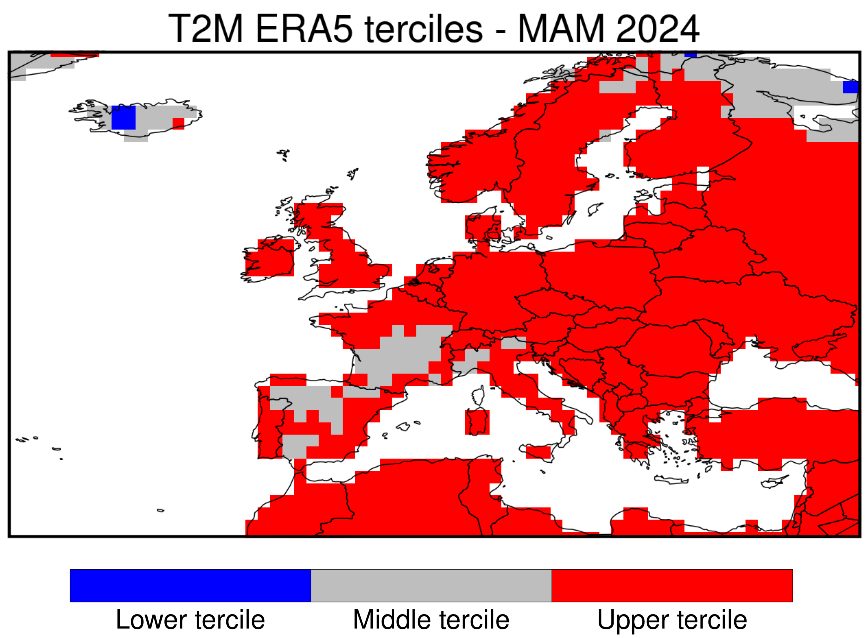
Scenario 2/2: tercile prob. (151 members)



MAM 2024 forecasts issued February 2024

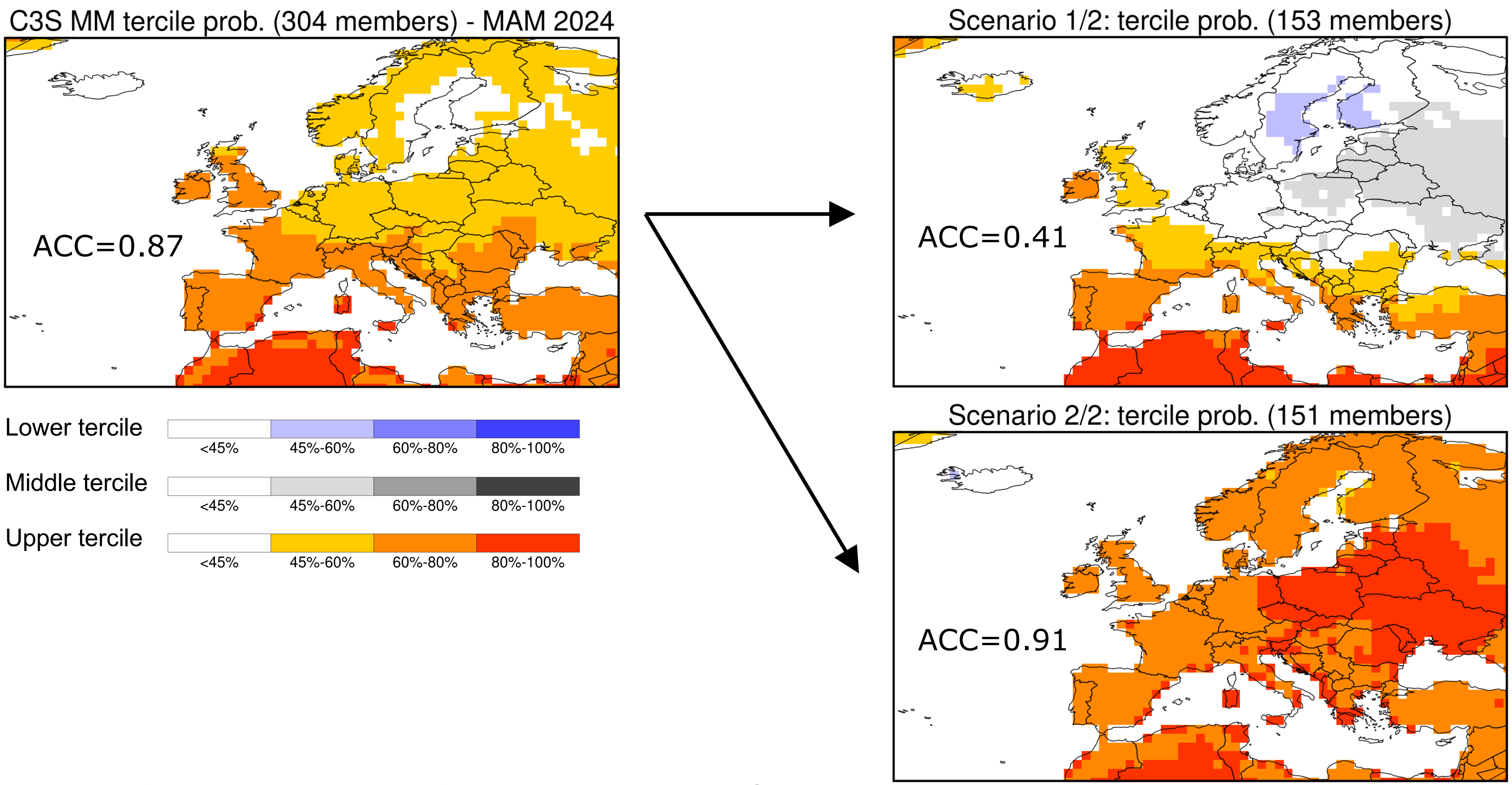
➢ Comparison with the actual outcome: this time, we were too conservative over northeastern Europe





MAM 2024 forecasts issued February 2024

- Comparison with the actual outcome: spatial correlation ACC



- Ensemble mean and Scenario 2 far better than Scenario 1

Take-home messages: interest for real-time seasonal forecasting

- Scenarios **are a tool providing additional guidance** to seasonal forecasters:

    1. Scenarios provide a condensed yet precise representation of the multi-model ensemble spread

    2. Similarities and differences between scenarios reveal **where the seasonal forecast is more confident**, and where it is more uncertain

    3. Scenarios help forecasters interpret at one glance the signals in the multi-model ensemble mean

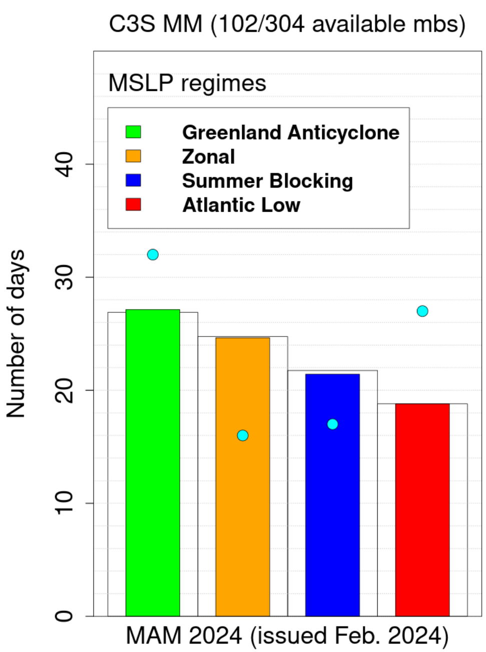
- Scenarios **are not** an approach to systematically improve seasonal forecast skill

Thank you for your attention

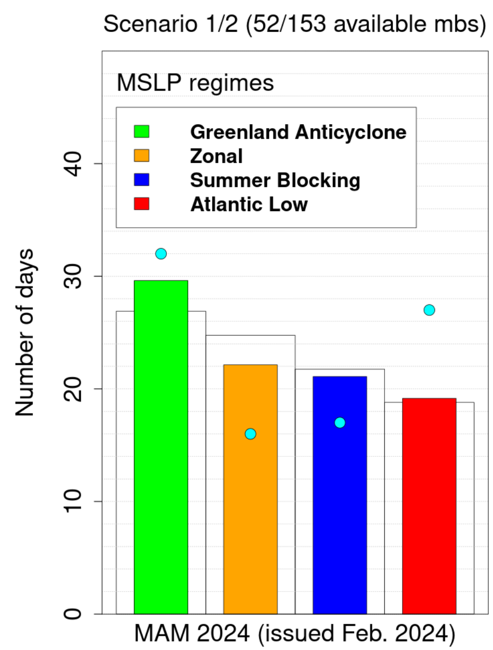
Questions?

How atmospheric circulation shapes the scenarios?

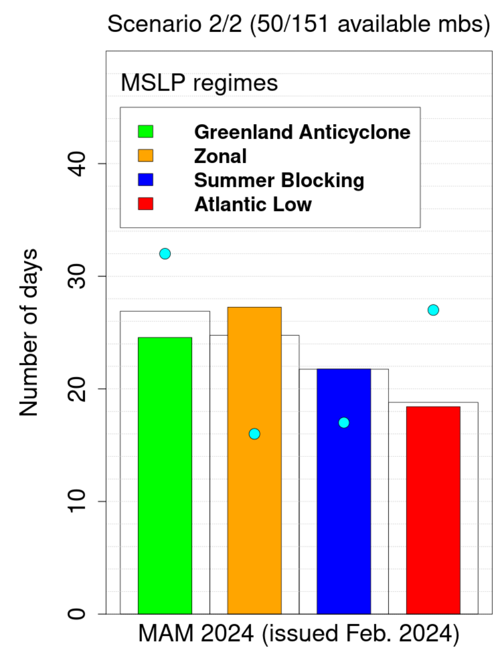
➢ What atmospheric circulation anomalies lead scenarios to diverge?
Use of in-house products: weather regimes



Full ensemble



Scenario 1

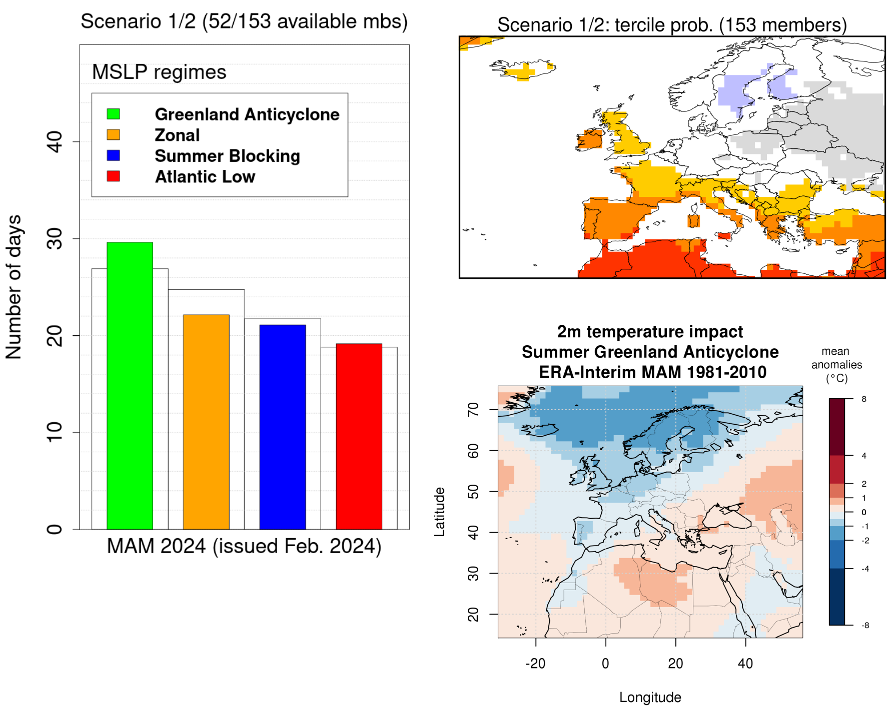


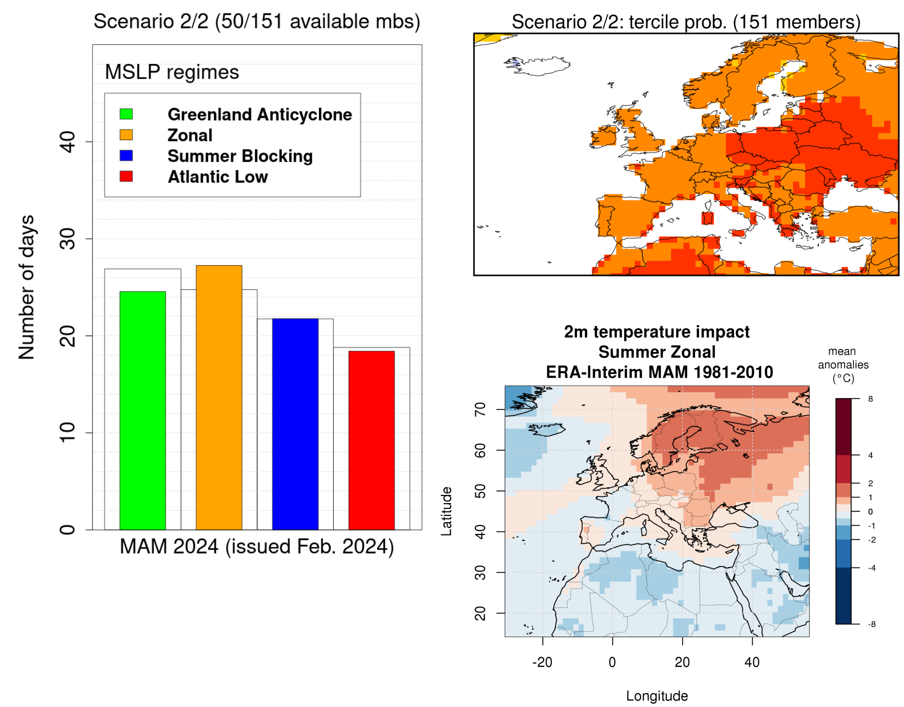
Scenario 2

➢ Ensemble mean close to climatological occurrences of weather regimes
➢ Scenarios are discriminated by a seesaw between Greenland Anticyclone and Zonal regimes

How atmospheric circulation shapes the scenarios?

- Do scenarios match with the expected impacts of weather regimes?





- The temperature anomalies of scenarios are broadly consistent with the anomalies associated with each discriminant regime

How do scenarios verify against reference data?

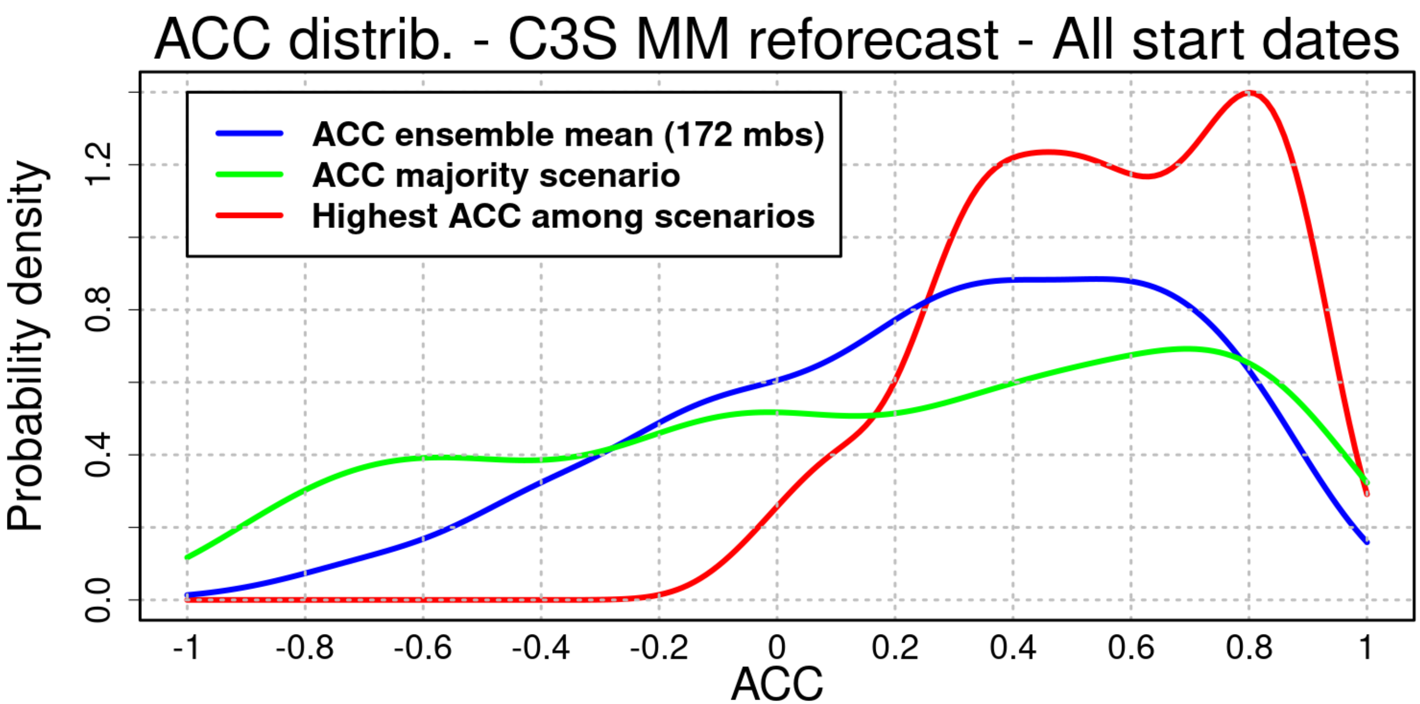
Do scenarios verify better than the ensemble mean?

Two alternative forecasting “strategies”:

1. Forecasting the **majority scenario**
2. A thought experiment: forecasting the **“best” scenario**, i.e the one that is the more spatially correlated with verifying reanalysis → What would be the gain if we could guess it in advance?

Do scenarios verify better than the ensemble mean?

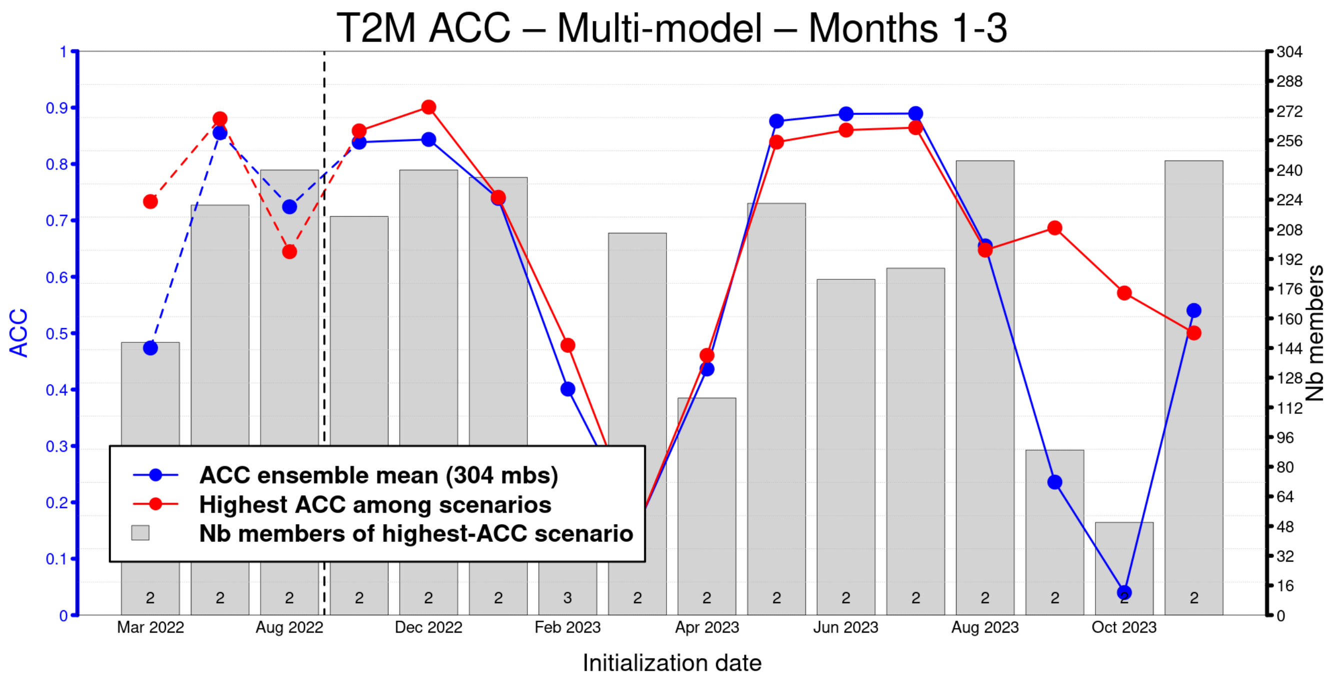
Two alternative forecasting “strategies”:
→ Judging from the full reforecast sample, forecasting the majority scenario **is not a good strategy** to improve forecast skill (unsuprisingly!)



▲ *Distribution of the Anomaly Correlation Coefficient (ACC) of the multi-model seasonal reforecasts over Europe against ERA5 for all initializations in 1993-2016 (sample size 12 x 24 = 288)*

Do scenarios verify better than the ensemble mean?

Two alternative forecasting "strategies"
→ Judging from the recent real-time forecasts:
- The "best" scenario is not necessarily better than the ensemble mean
- Some rare cases where the "best" scenario is a strongly minority scenario



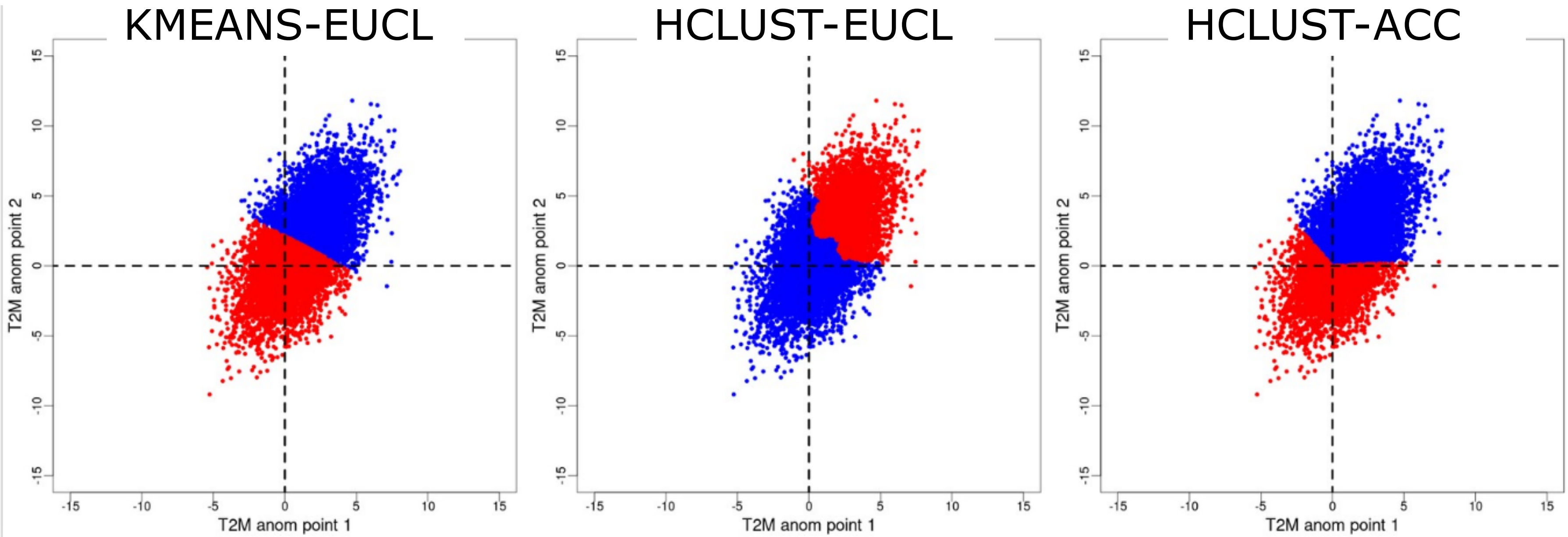
▲ *Anomaly Correlation Coefficient (ACC) of recent multi-model real-time forecasts. Blue curve: Full ensemble mean (304 members). Red curve: Scenario with highest ACC among scenarios.*

Determination of the clustering methodology

Which clustering algorithm?

➢ Three options considered:

1. KMEANS-EUCL: k-means algorithm, Euclidean distance
2. HCLUST-EUCL: Hierarchical clustering, Euclidean distance
3. **HCLUST-ACC**: Hierarchical clustering, dissimilarity between members i and j is d(i,j) = 1 – ACC(i,j), where ACC measures spatial correlation



▲ *Illustration of the behavior of the three algorithms on synthetic two-dimensional data*

Which clustering algorithm?

- Three options considered
- Let’s compare them on the full reforecast sample: 12 x 24 = 288 forecasts

	KMEANS-EUCL	HCLUST-EUCL	HCLUST-ACC
KMEANS-EUCL		287/288 (99%)	286/288 (99%)
HCLUST-EUCL			270/288 (94%)
HCLUST-ACC			

Number of similar clusterings of the T2M anomaly maps of the multi-model reforecasts (1993-2016, 12 initialization months) obtained with different methods.

→ The final clustering result is **not very sensitive** to the choice of the algorithm

Which climate parameters to choose?

- T2M, RR, Z500 and any possible combination (e.g T2M+RR, etc.)
- Comparison of the options on the full reforecast sample (12 x 24 = 288 forecasts): do they give the same clustering results?

	T2M	RR	T2M + RR
T2M		134/288 (46%)	**280/288 (97%)**
RR			213/288 (74%)
T2M + RR			

Number of similar clusterings of the multi-model reforecasts (1993-2016, 12 initialization months) obtained with different sets of variables: T2M anomalies alone, RR anomalies alone and T2M+RR anomalies combined.

→ When T2M and RR are combined, T2M drives the clustering

Which climate parameters to choose?

- T2M, RR, Z500 and any possible combination (e.g T2M+RR, etc.)
- Comparison of the options on the full reforecast sample (12 x 24 = 288 forecasts): do they give the same clustering results?

	Z500	T2M+Z500	RR+Z500	T2M+RR+Z500
T2M	166/288 (58%)	**280/288 (97%)**	180/288 (62%)	**259/288 (90%)**
RR	231/288 (80%)	211/288 (73%)	**263/288 (91%)**	237/288 (82%)
Z500		228/288 (79%)	**272/288 (94%)**	242/288 (84%)

Number of similar clusterings of the multi-model reforecasts (1993-2016, 12 initialization months) obtained with different sets of variables: T2M anomalies alone, RR anomalies alone, Z500 anomalies alone and combinations involving Z500.

→ Whenever T2M is present, it drives the clustering
→ RR and Z500 make another group of variables
→ **Two main options retained: T2M alone, RR+Z500 together**