



CopERNicus climate change Service Evolution
- CERISE

Comprehensive Assessment of Seasonal Forecasts Across Multiple Models and Model Versions

Núria Pérez-Zanón, Nadia Milders, Carlos Delgado-Torres, Victòria Agudetse, Ángel G. Muñoz, and Francisco Doblas-Reyes
11th of September 2025



Ljubljana & online | 7-12 September 2025



The three most important reasons to verify forecasts are:

- ★ to monitor forecast quality - how accurate are the forecasts and are they improving over time?
- ★ to improve forecast quality - the first step toward getting better is discovering what you're doing wrong.
- ★ to compare the quality of different forecast systems - to what extent does one forecast system give better forecasts than another, and in what ways is that system better?

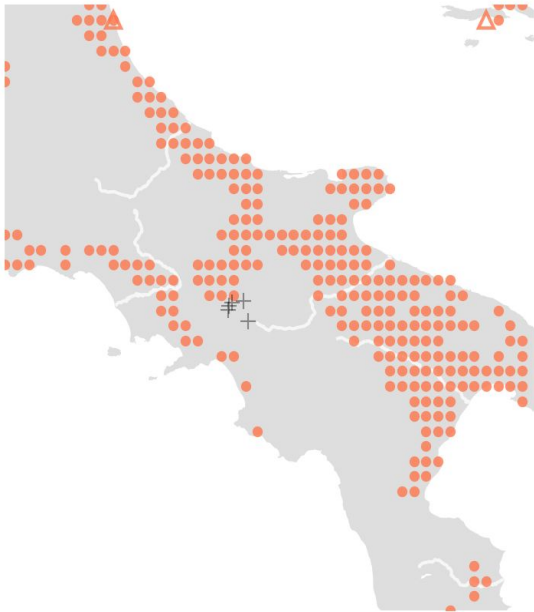
https://www.cawcr.gov.au/projects/verification/#Why_verify

+ To inform users about the quality of the climate products

VitiGEOSS climate forecast service developed for the vineyard management is an example on applying forecast verification to inform users

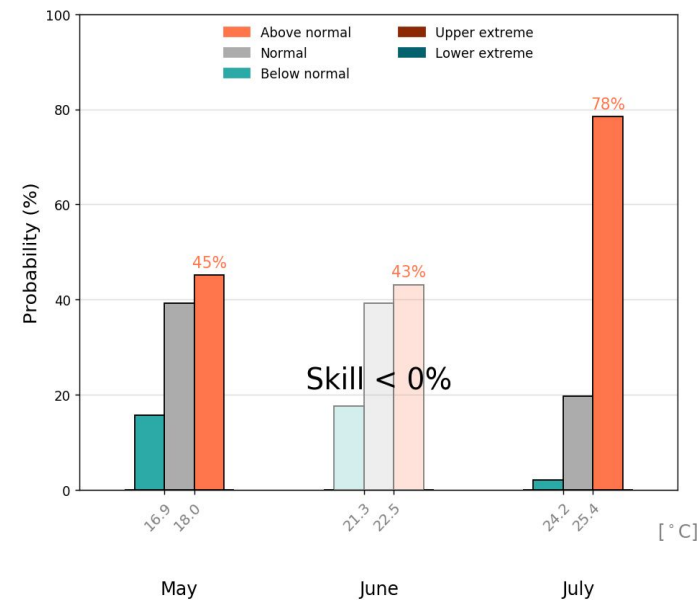
Mask applied during operational service

Valid for Jul 2022, issued on 01 Apr 2022



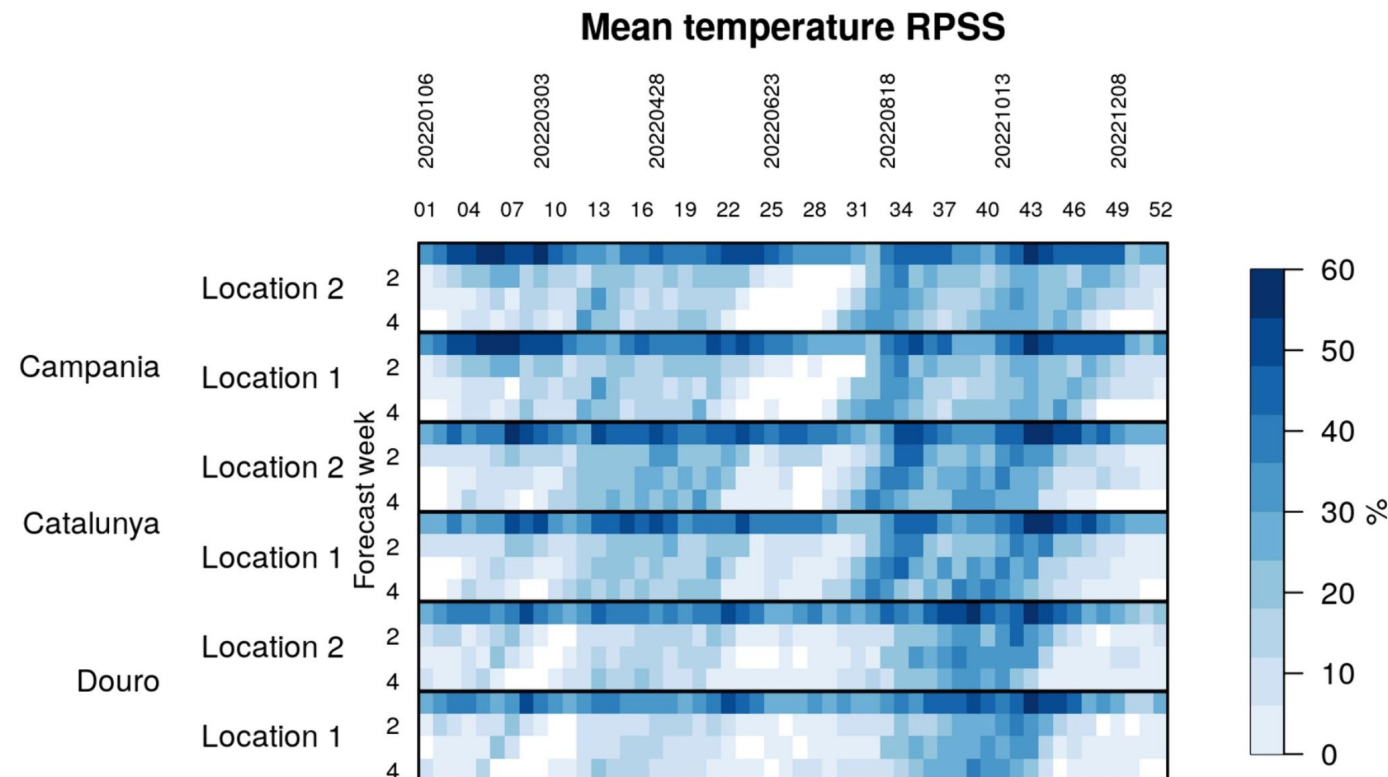
Temperature (Pietradefusi)

Seasonal forecast issued on Apr 2022



VitiGEOSS climate forecast service developed for the vineyard management is an example on applying forecast verification to inform users

Overall expectation along the year

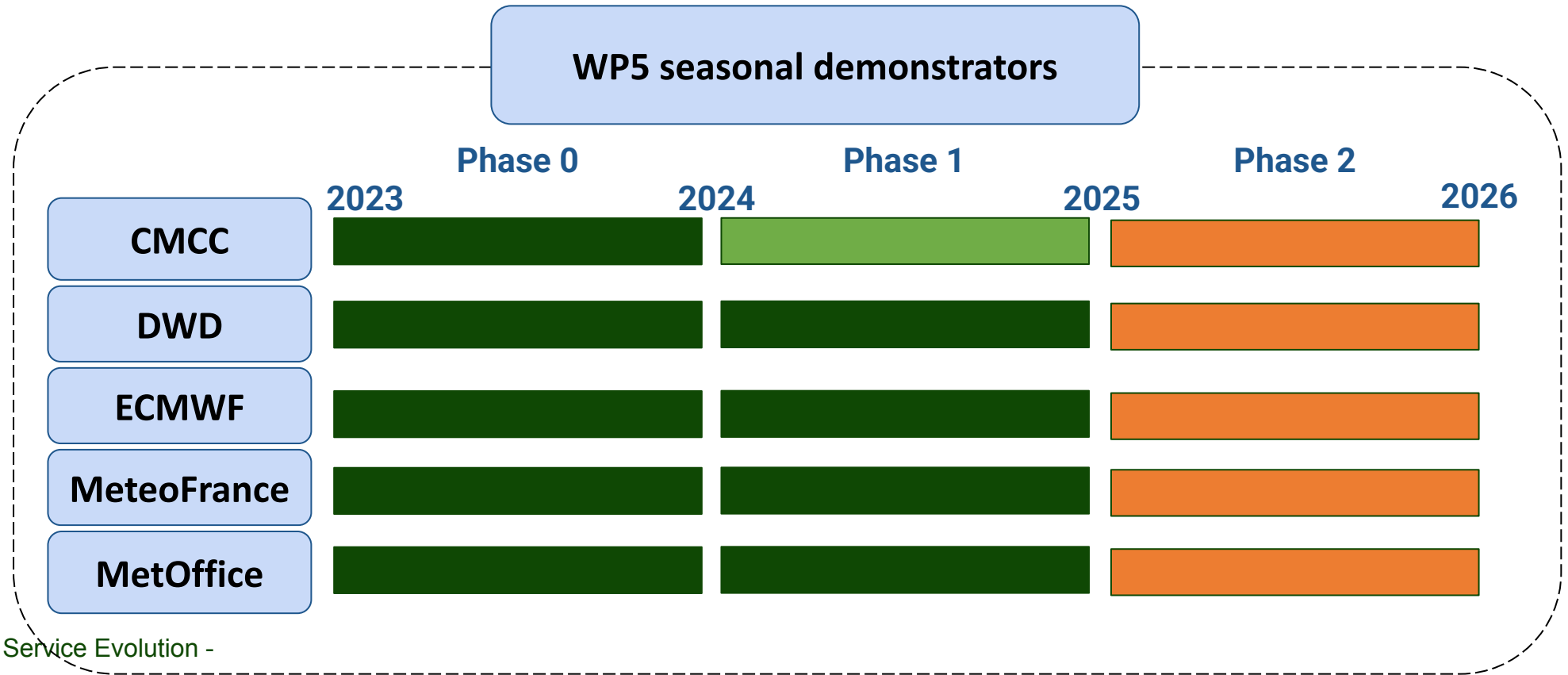


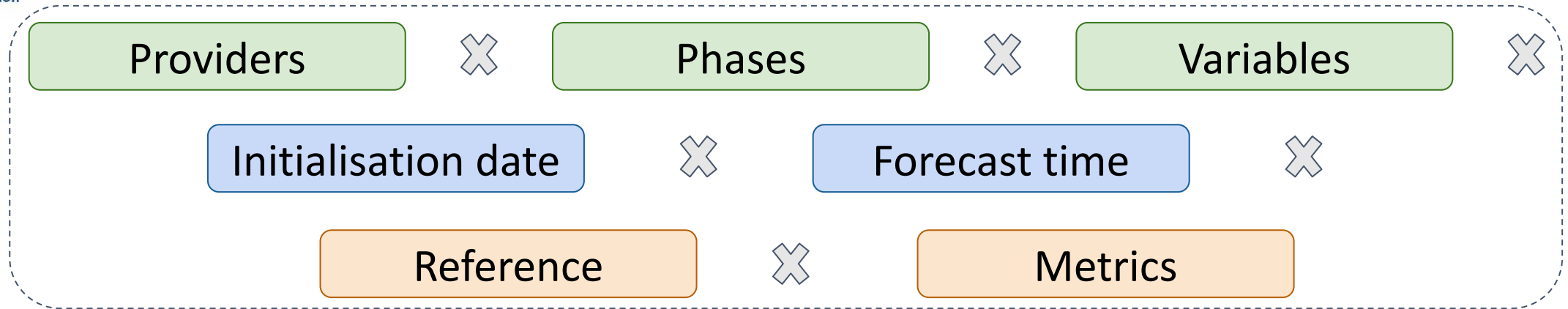
CERISE aims to **support the long-term evolution** of the C3S for :

→ regional and global climate **reanalysis** and

→ multi-system **seasonal prediction**,

towards an Earth system approach, with a **focus on land-atmosphere coupling**.





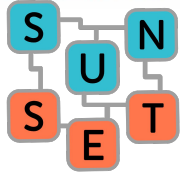
Phase 0 = $5 \times 1 \times 9 \times 12 \times 6 \times 1 \times 7 = 68040$ = 78120

Phase 1 and 2 = $5 \times 2 \times 9 \times 4 \times 4 \times 1 \times 7 = 10080$

Only in Task 5.3, we are expecting ~ 80000 maps



Funded by the
European Union



SUBseasonal to decadal climate forecaSt post-processing and assEssment suiTe

Pérez-Zanón, N., Agudetse, V., Palma, L., Ho, A.-C., Delgado-Torres, C., Milders, N., Duzenli, E., Llabrés-Brustenga, A., de Paula Kinoshita, B., Bretonnière, P.-A., and G. Muñoz, Á.: SUNSET: Subseasonal to decadal climate forecast post-processing and asSEssmenT suite, EMS Annual Meeting 2024, Barcelona, Spain, 1–6 Sep 2024, EMS2024-361, <https://doi.org/10.5194/ems2024-361>, 2024.

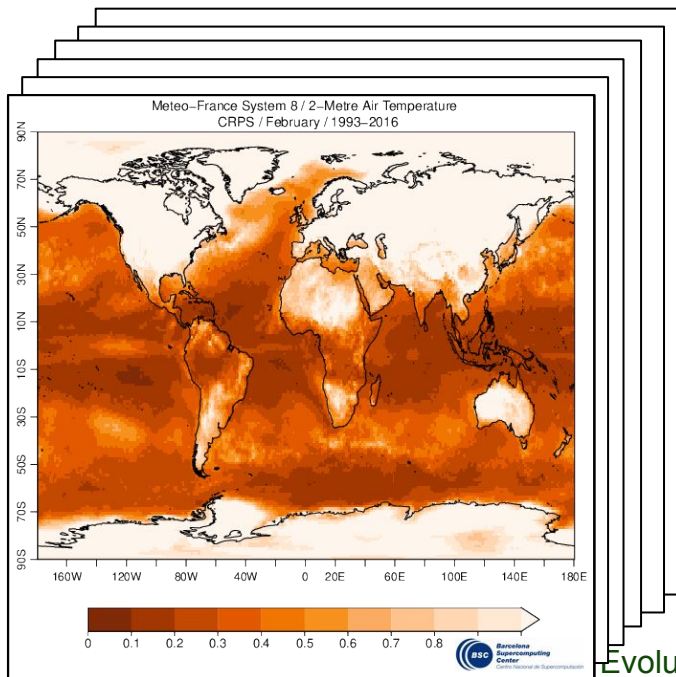
- > Post-processing climate forecast (e.g. climate indices)
- > Assessment of climate forecast
- > Visualization



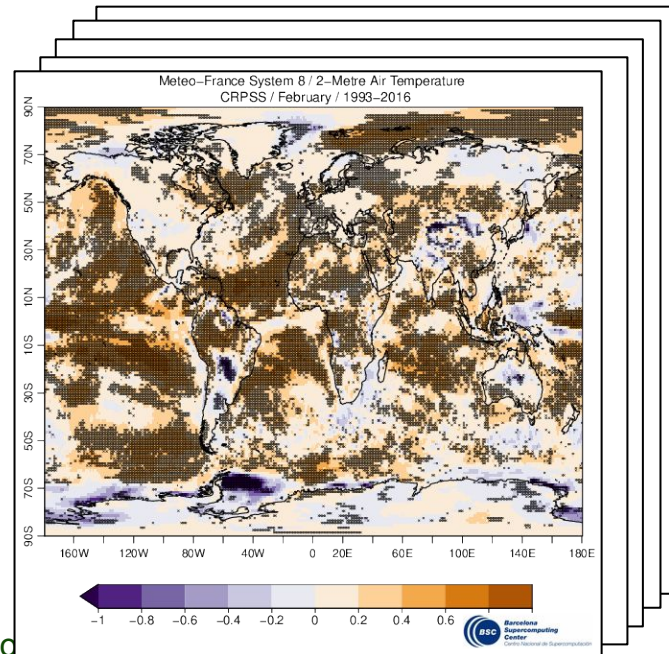
Consider the evaluation of 1 variable, for 1 phase 0 demonstrator,
we will obtain 504 maps for all 5 metrics that characterise different forecast
features.

Are we able to summarise them?

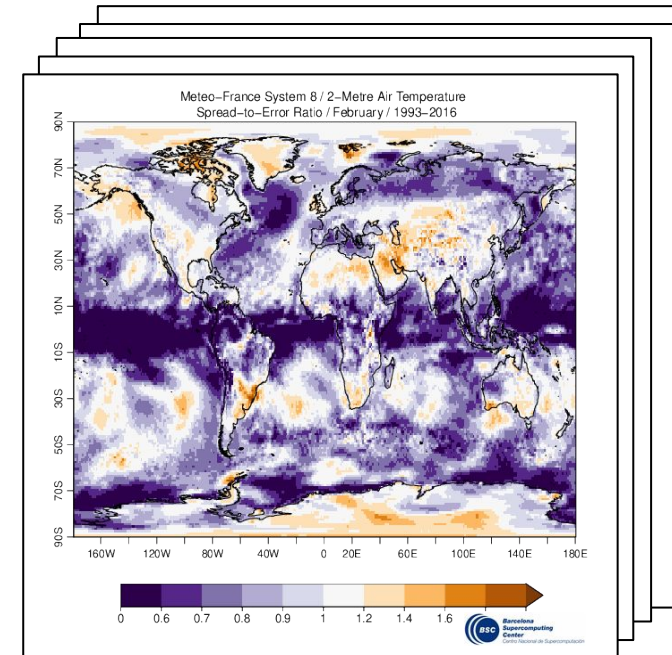
Mean bias



CRPSS



Spread-to-error ratio



...

Get a summary of the metrics for each forecast system and variable: **Scorecards**

Mean bias → systematic error (over- or under-estimation)

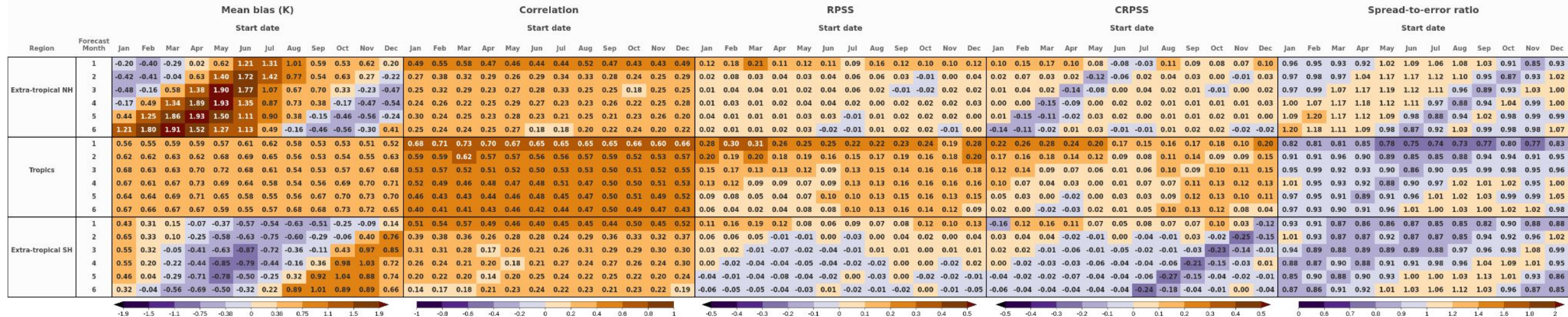
Correlation → association (linear relationship with observations)

RPSS / CRPSS → accuracy + resolution + reliability

Spread-to-error ratio → reliability

2-metre air temperature of Meteo-France System 8

(Ref: ERA5 1993-2016)





Funded by the
European Union

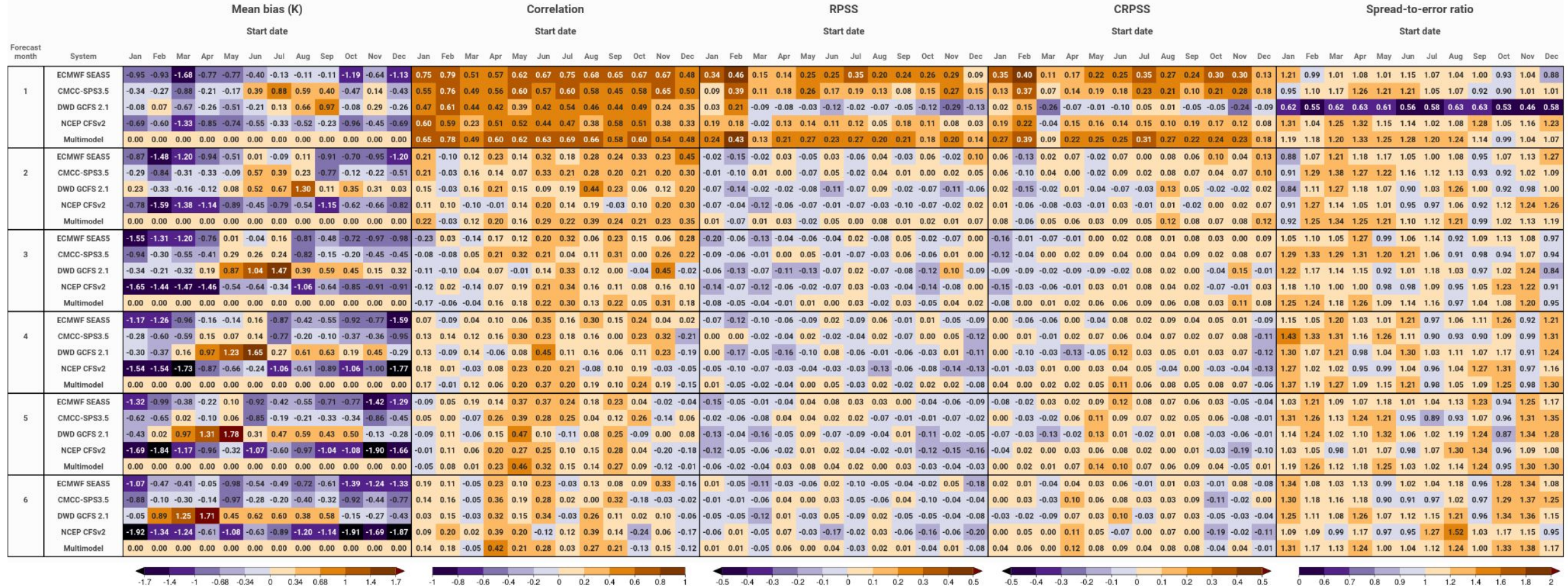


AI4S
BSC AI4Science Fellowships



Near-Surface Air Temperature. Region: Spain (Interpolation = to system, Aggregation level = skill, Cross-validation = anomalies & terciles & crps_clim)

(Ref: ERA5 1993-2016)



How to fill an scorecard?

It depends on the product that we provide to the user:

- a) If we provide the average over a region, a time series needs to be verified
- b) If we provide a gridded forecast over a region, all grid points need to be verified



Funded by the
European Union



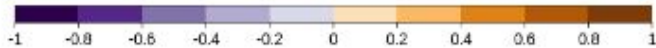
AI4S
BSC AI4Science Fellowships



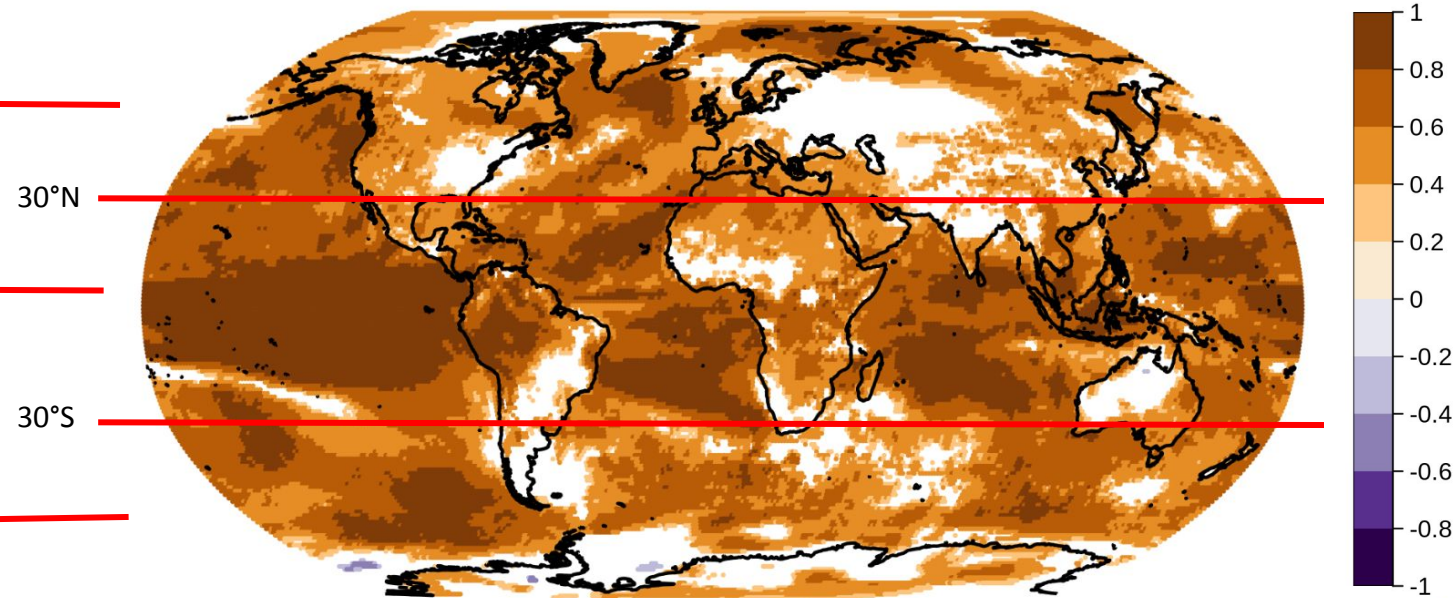
2-metre air temperature of Meteo-France System 8

(Ref: ERA5 1993-2016)

		Correlation											
		Start date											
Region	Forecast Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Extra-tropical NH	1	0.49	0.55	0.58	0.47	0.46	0.44	0.44	0.52	0.47	0.43	0.43	0.49
	2	0.27	0.38	0.32	0.29	0.26	0.29	0.34	0.33	0.28	0.24	0.25	0.29
	3	0.25	0.32	0.29	0.23	0.27	0.28	0.33	0.25	0.25	0.18	0.25	0.25
	4	0.24	0.26	0.22	0.25	0.29	0.27	0.23	0.23	0.26	0.22	0.25	0.28
	5	0.30	0.24	0.25	0.23	0.28	0.23	0.21	0.25	0.21	0.23	0.26	0.20
	6	0.25	0.24	0.24	0.25	0.27	0.18	0.18	0.20	0.22	0.24	0.20	0.22
Tropics	1	0.68	0.71	0.73	0.70	0.67	0.65	0.65	0.65	0.65	0.66	0.60	0.66
	2	0.59	0.59	0.62	0.57	0.57	0.56	0.56	0.57	0.59	0.52	0.53	0.57
	3	0.53	0.57	0.52	0.51	0.52	0.50	0.53	0.53	0.50	0.51	0.52	0.55
	4	0.52	0.49	0.46	0.48	0.47	0.48	0.51	0.47	0.50	0.50	0.51	0.53
	5	0.46	0.43	0.43	0.44	0.46	0.48	0.45	0.47	0.50	0.51	0.49	0.52
	6	0.40	0.41	0.41	0.43	0.46	0.42	0.44	0.47	0.50	0.49	0.47	0.43
Extra-tropical SH	1	0.51	0.54	0.57	0.49	0.46	0.40	0.45	0.45	0.44	0.50	0.45	0.52
	2	0.39	0.38	0.36	0.26	0.28	0.28	0.24	0.29	0.36	0.33	0.32	0.37
	3	0.31	0.31	0.28	0.17	0.26	0.21	0.26	0.31	0.29	0.29	0.30	0.30
	4	0.26	0.24	0.21	0.20	0.18	0.21	0.27	0.24	0.27	0.26	0.24	0.30
	5	0.20	0.22	0.20	0.14	0.20	0.25	0.24	0.22	0.25	0.22	0.20	0.24
	6	0.14	0.17	0.18	0.21	0.23	0.24	0.22	0.23	0.21	0.23	0.22	0.19



Meteo-France System 8 / 2-Metre Air Temperature Ensemble Mean Correlation / January / 1993-2016



Nominal start date: 1st of January
Forecast month: 01
Reference: ERA5
 $\alpha = 0.05$

How to fill each scorecard cell?



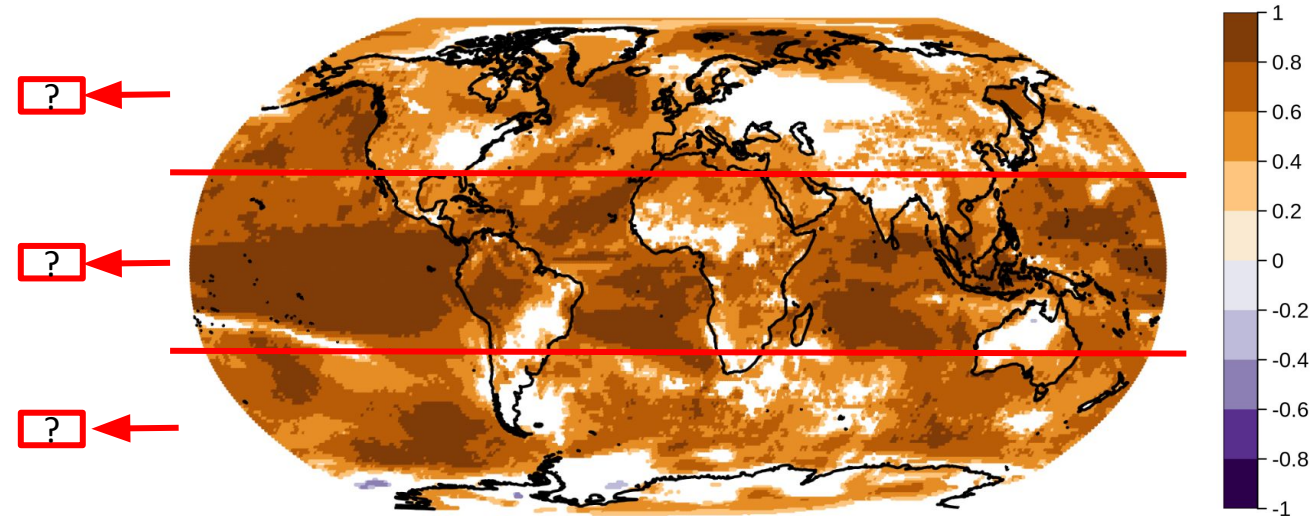
Averaging correlations for each region is not mathematically correct.



Alternative:

1. compute the standard deviation and covariance at each grid point,
2. compute the weighted mean of the variance and covariance for the region
3. compute the correlation

Meteo-France System 8 / 2-Metre Air Temperature
Ensemble Mean Correlation / January / 1993-2016



Nominal start date: 1st of January
Forecast month: 01
Reference: ERA5
alpha = 0.05

**On-going
development**

$$\text{Correlation} = \frac{\text{Cov}(x, y)}{\sigma_x * \sigma_y}$$

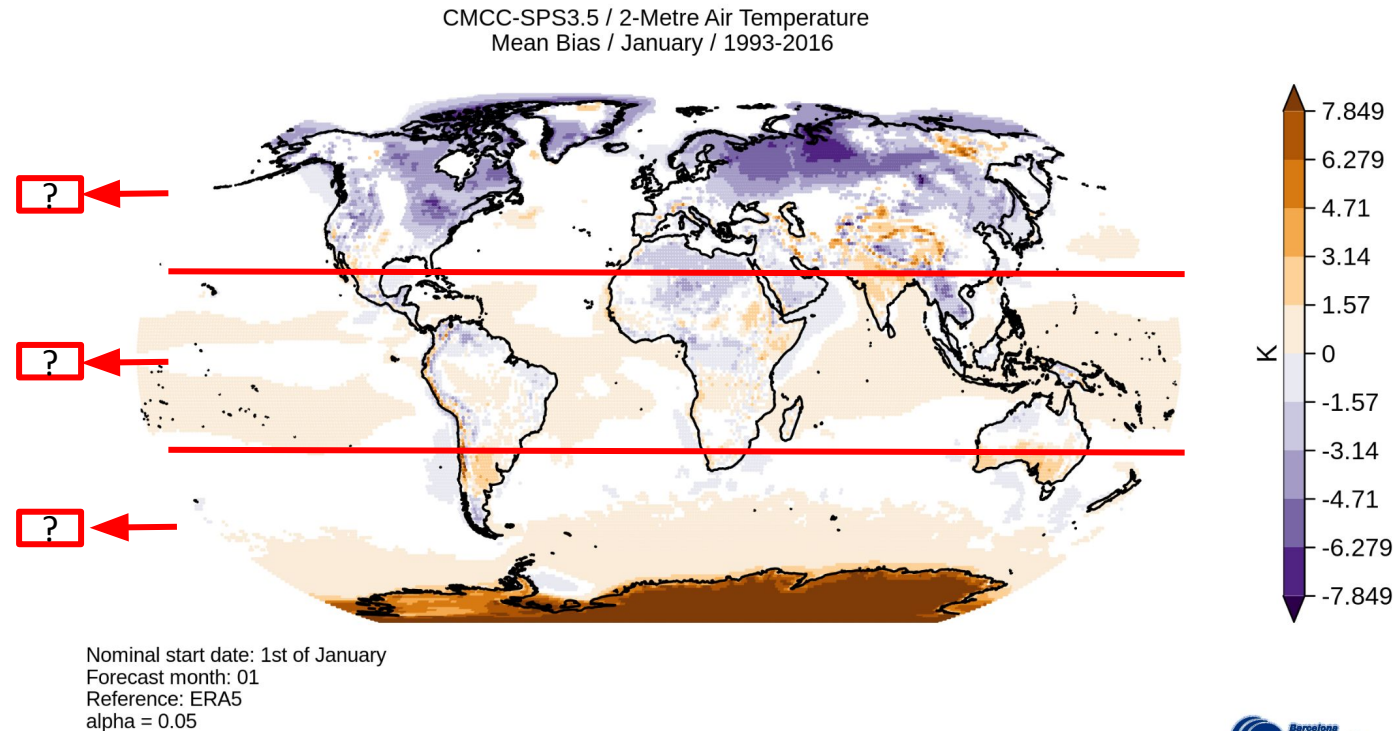
How to highlight a result that is statistically significant?

💡 For mean bias, we could use the **Welch's t-test** that considers the variance of observations and ensemble mean in the region

$$t = \frac{\Delta \bar{X}}{s_{\Delta \bar{X}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_{\bar{X}_1}^2 + s_{\bar{X}_2}^2}},$$

$$s_{\bar{X}_i} = \frac{s_i}{\sqrt{N_i}},$$

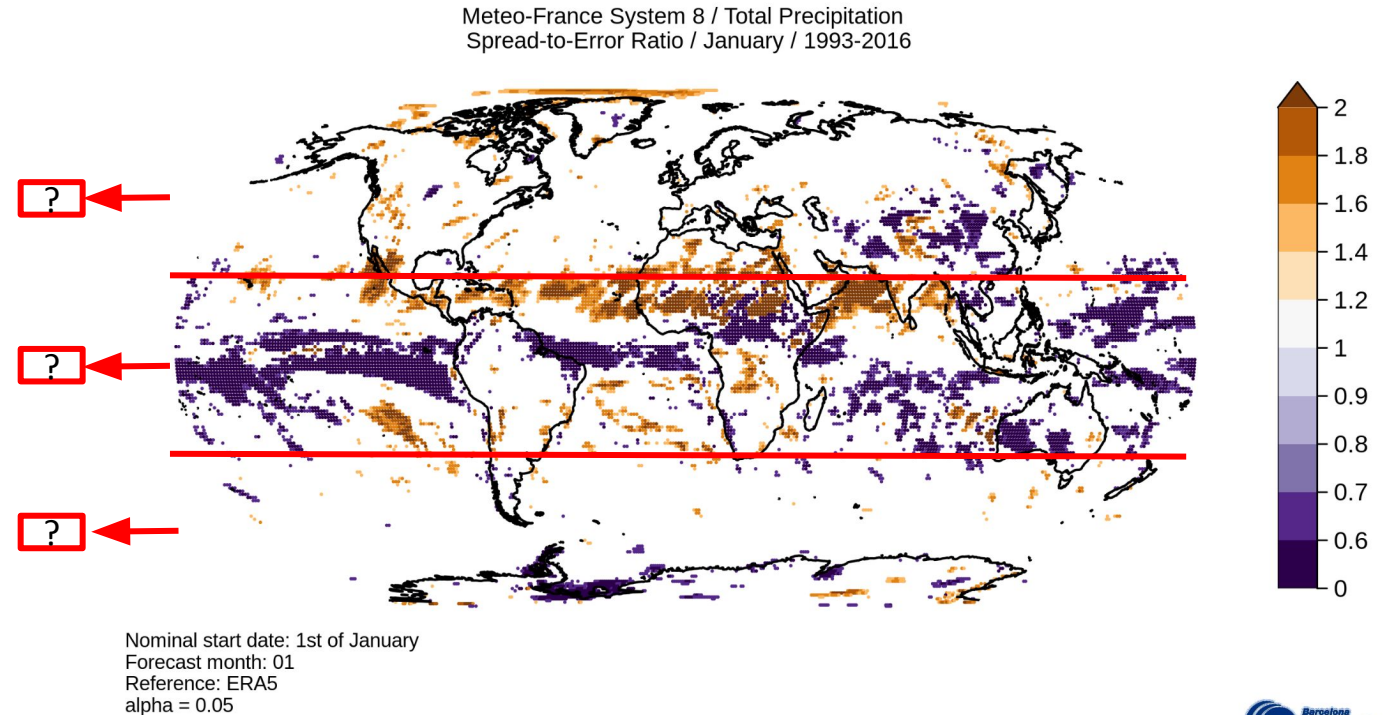
**On-going
development**



<https://earth.bsc.es/gitlab/es/s2dv/-/blob/master/R/Bias.R>

How to highlight a result that is statistically significant?

How to know if the spread-to-error ratio over a region is significantly different from 1?



**On-going
development**



Funded by
the European Union



CopERNicus climate change Service Evolution
- CERISE

Thank you!

nuria.perez@bsc.es



The CERISE project (grant agreement No 101082139) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

Núria Pérez-Zanón acknowledges her AI4S fellowship within the “Generación D” initiative by Red.es, Ministerio para la Transformación Digital y de la Función Pública, for talent attraction (C005/24-ED CV1), funded by NextGenerationEU through PRTR.