CSTools R package bringing state-of-the-arts postprocessing methods to seasonal-to-decadal forecast users

Poster EGU2020-7599

Núria Pérez-Zanón[1], Louis-Philippe Caron[1], Carmen Alvarez-Castro[2], Lauriane Batté[3], Susanna Corti[4], Marta Dominguez[5], Federico Fabiano[4], Silvio Gualdi[2], Jost von Hardenberg[4], Llorenç Lledó[1], Nicolau Manubens[1], Paola Marson[3], Stefano Materia[2], Eroteida Sánchez[5], Bert Van Schaeybroeck[6], Verónica Torralba[1], Silvia Terzago[4], Deborah Verfaillie[1], and Danila Volpi[4]

[1]Earth Sciences Department, Barcelona Supercomputing Center (BSC), Spain , [2]Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), [3]Météo-France, [4]Institute of Atmospheric Sciences and Climate, National Research Council of Italy (ISAC-CNR), [5]Agencia Estatal de Meteorología (AEMET), [6]Royal Meteorological Institute of Belgium (RMI)

Introduction

The availability of climate data has never been larger, as evidenced by the development of the Copernicus Climate Change Service. However, availability of climate data does not automatically translate into usability and sophisticated post-processing is often required to turn these climate data into user-relevant climate information allowing them to develop and implement strategies of adaptation to climate variability and to trigger decisions. Developed under the umbrella of the ERA4CS Medscope project by multiple European partners, here we present a R package currently in development, which aims to provide tools to exploit dynamical seasonal forecasts such as to provide information relevant to public and private stakeholders at the seasonal timescale. This toolbox, called CSTools (short for Climate Service Tools), contains process-based methods for forecast calibration, bias correction, statistical and stochastic downscaling, optimal forecast combination and multivariate verification, as well as basic and advanced tools to obtain tailored products.

Development Strategy

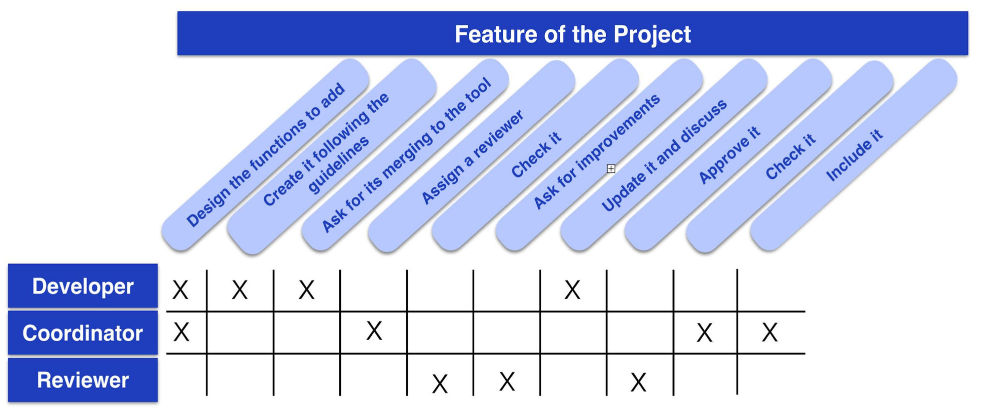


Fig 15: Overview of the workflow to include a functionality considered by the project in the toolbox.

The development strategy defines which are the required roles and the tasks the partners should carry for a safe development. For this project, a cloud service for Git repositories has been used, which allows keeping track of the code and documentation as well as the discussion at every step of the process between all the contributors. Guidelines have been provided to the developers to ensure the quality, the usability and the interoperability of the functions. While the reviewer specially checks the scientific performance of the functionality, the coordinator assesses the global adequacy of the function in the CSTools package.

RainFARM

RainFARM is a stochastic precipitation downscaling method which produces, from large-scale spatio-temporal precipitation fields, ensembles of stochastic realizations at finer spatial resolution (typically 1 km), which preserve the large-scale statistical properties of the original field and with realistic spatial correlation structures. It also corrects precipitation over complex orography using weights based on existing fine-scale precipitation climatology.

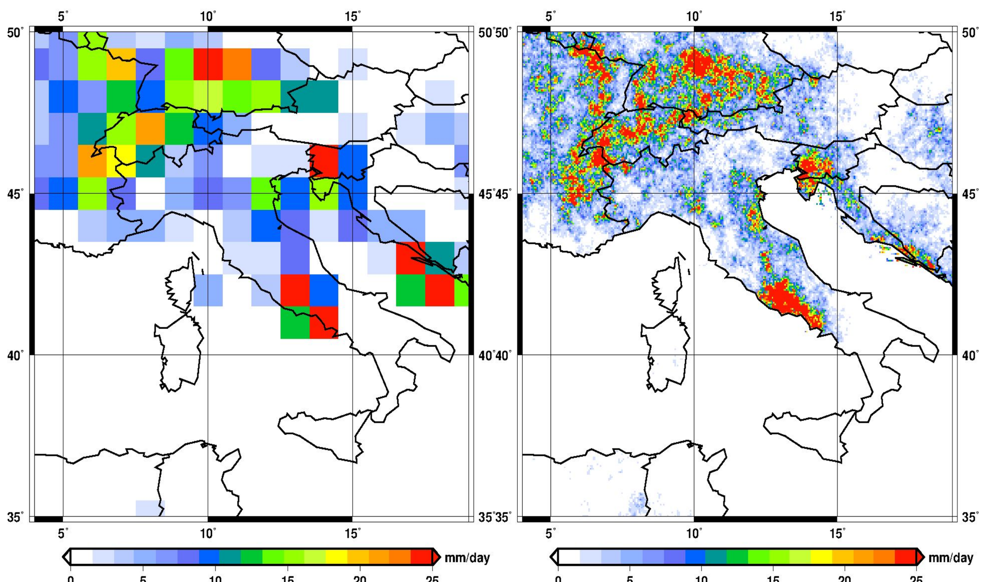


Fig 1: Application of the RainFARM method to an event which occurred on 2010-08-17. On the left is the original ECMWF System 5 precipitation forecast at 1° resolution and on the right is the precipitation downscaled to a target resolution of 0.05°, using a fine-scale precipitation climatology for orographic correction.

Terzago, S., Palazzi, E., and von Hardenberg, J. (2018). Stochastic downscaling of precipitation in complex orography: a simple method to reproduce a realistic fine-scale climatology, Nat. Hazards Earth Syst. Sci., 18, 2825-2840, doi: https://doi.org/10.5194/nhess-18-2825-2018

PlotMostLikelyQuantile

This function produces a map with the probability of the most likely category (e.g. terciles) for a particular forecast. It also allows complementing this information with a skill metric to mask those regions where the forecasts are not skilful.

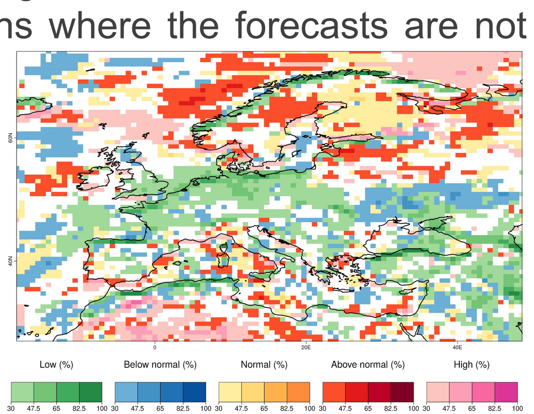


Fig 2: Probabilistic most likely quintile map of 10-m wind speed for ECMWF System 4 seasonal forecast for DJF 2016-2017. The predictions were issued the 1st of November 2016. The most likely category and its percentage of probability to occur is shown. White colour indicates that the forecasts probabilities are below the 30% for all five categories. The reference dataset is ERA-Interim and the climatological period 1981-2015.

ENSClus

This function, based on a clustering algorithm, takes N ensemble members from one (or more) forecasting system(s) and groups together those that show similar seasonal anomalies of a given variable (for example 2m air temperature) over the Mediterranean region. The number of clusters and the variable can be selected by the user. However, since the clustering is intended as a summary of the ensemble information, the maximum number of clusters is supposed to be at least an order of magnitude smaller than the ensemble size. Each cluster is represented by one of its members: the forecast closest to the centroid of the cluster. The representative members of the clusters are referred to as "Seasonal Forecast Scenarios".

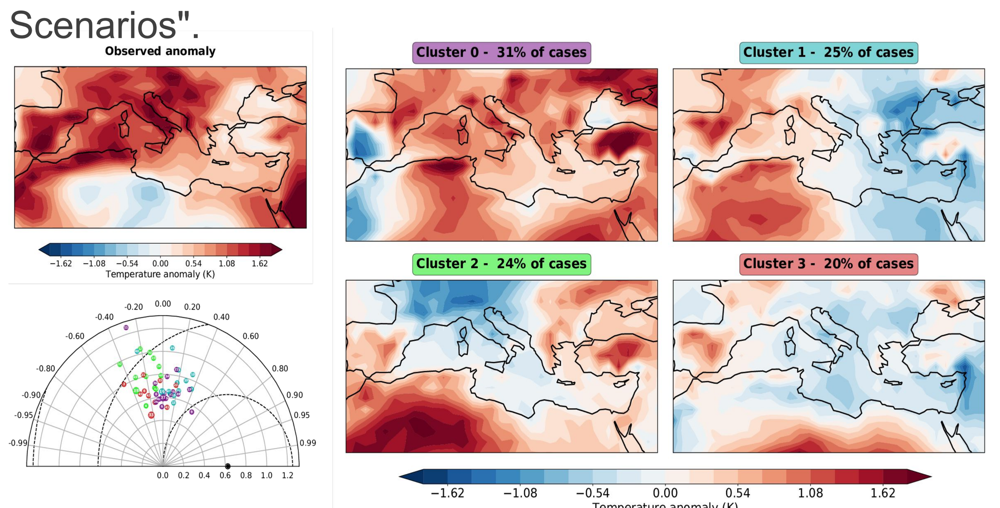


Fig 3: EnsClus has been applied here to the ECMWF seasonal forecasts (System 5) of 2m temperature in the Mediterranean area relative to Summer 2017 (JJA). The 51 ensemble forecasts anomalies are grouped in 4 scenarios and compared with the observed anomaly from ERA-Interim reanalysis (top left panel). The best representative members of the 4 scenarios are shown on the right panels. The Taylor diagram in the bottom left panel shows the relative agreement between the forecasts and the observation: observation is shown in black and the ensemble members are plotted according to their standard deviation (radial axis) and correlation coefficient (polar axis). The member performing better in foreseeing the observed anomaly pattern belongs to cluster 0.

Ensemble calibration

This function calibrates the ensemble forecast by adjusting the ensemble mean, the total forecast variance and the ensemble variance to obtain an accurate but also reliable ensemble. There is no Gaussian assumption underlying this calibration and it preserves the spatio-temporal correlation structure of the original ensemble forecast.

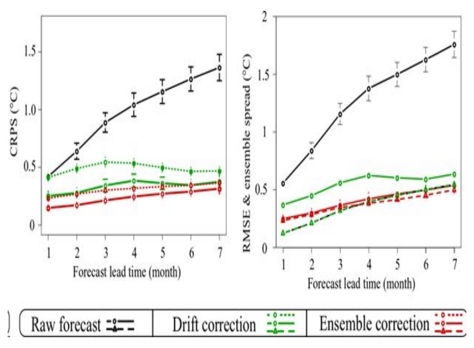


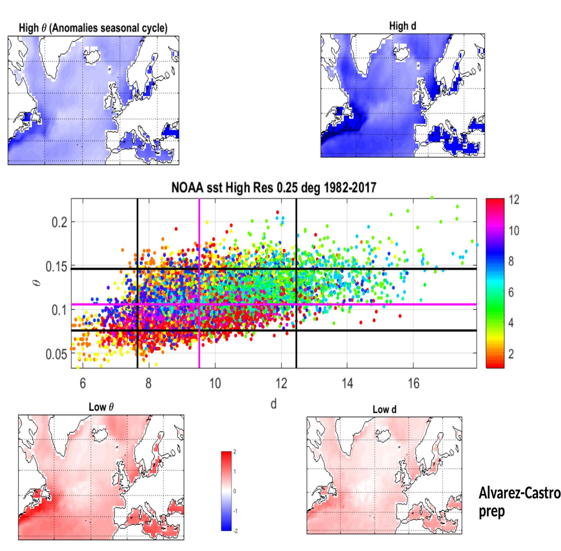
Fig 4: Forecast verification of the ECMWF IFS raw forecasts (black lines) of the Nino 3.4 index, the drift-corrected forecasts (green lines) and the ensemble-corrected forecasts (red lines). Left: Continuous ranked probability score (CRPS) against lead time. The full green and red line are obtained by using four calibrations, one for each season, while the dotted lines are obtained using one calibration, using all available data. Right: RMSE (circles) and ensemble standard deviation (triangles) against lead time. Uncertainty intervals delineate the 95% confidence intervals assuming normal error statistics.

DynBiasCorrection

This function performs a quantile mapping based on a dynamical classification. Following a non-linear approach (Faranda et al, 2019), the function computes two dynamical properties (distance and persistence) of the underlying attractor (SLP/SST). Those proxies are then used to classify the data in terciles. Once the data is classified, a simple quantile mapping approach is applied.

Fig 5: This figure shows an example of the phase-space diagram of the dynamical proxies (local dimension and persistence) computed for the observations of SST. Maps show the SST anomalies in each of the four clusters: (1) high persistence (theta) associated to colder SST anomalies, (2) high local dimension (d) associated to colder SST anomalies, (3) low persistence (theta) associated to warmer anomalies, and (4) low local dimension (d) associated to warmer anomalies. These four parameters in SST will be used to classified the data and apply an empirical quantile map to the precipitation/temperature field associated. Thus, as has been demonstrated in Faranda et al 2019 the predictability (low local dimension) of the system increase with a warmer ocean, a key information for seasonal forecast.

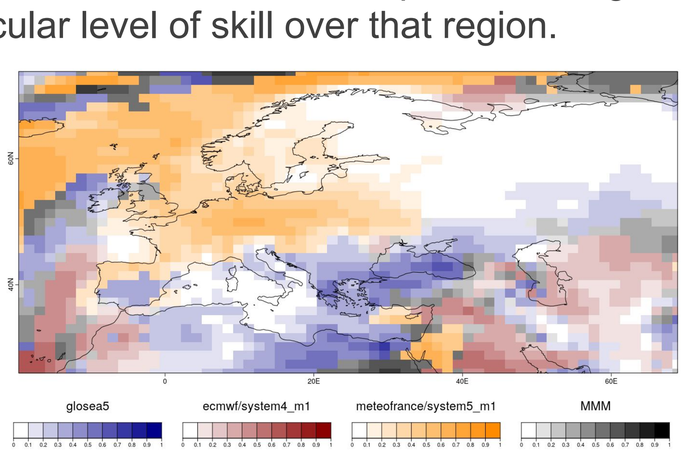
Faranda, D., M. Carmen Alvarez-Castro, G. Messori, D. Rodrigues and P. Yiou. The hammam effect or how a warm ocean enhances large scale atmospheric predictability. Nature Communications, March 2019. https://doi.org/10.1038/s41467-019-09305-8.



MultiMetric

This function calculates the anomaly correlation coefficient (ACC), the root mean square error (RMS) and the root mean square error skill score (RMSSS) of individual models and multi-model ensemble forecasts. It can also be used to identified the best model/forecast over a particular region, as well as the particular level of skill over that region.

Fig 6: Spatial representation of the highest correlation values for each grid point obtained for three different models - GloSea5 (blue), ECMWF System 5 (red) and Météo-France System 5 (yellow) seasonal forecasting systems as well as the ensemble mean (MMM - grey) versus a reference dataset.



The **CSTools** R package is on the CRAN repository:
https://cran.r-project.org/package=CSTools

This icon indicates that the functionality is currently available in the package.

SMOP

Statistical method for the spatialization and downscaling of precipitation in mountainous areas. This function performs an interpolation and a statistical downscaling. A sub-grid refinement is produced by combining local scale processes causing orographic rainfall (analytical) and large-scale precipitation component (from climate models), in a spatial autoregressive framework. The relative contribution of local and large-scale sources is adjusted with observations (Marson et al, in review)

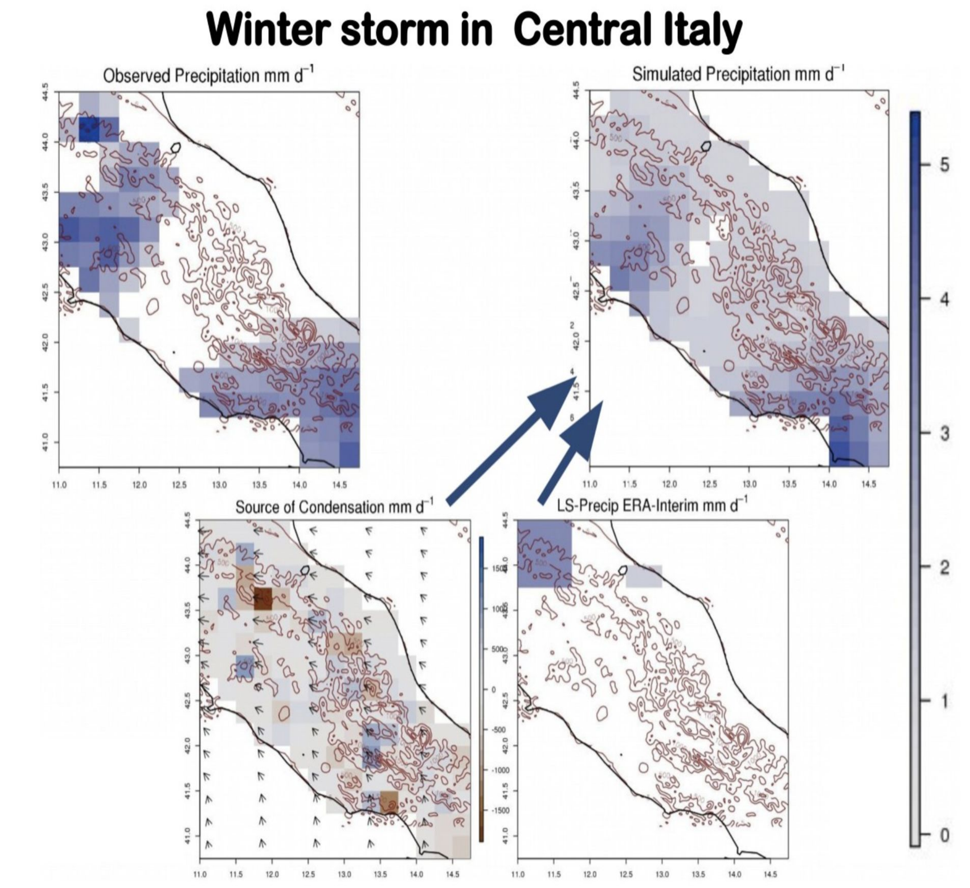


Fig 7: Sub-grid refinement by combining local scale processes causing orographic rainfall (analytical) and large-scale precipitation component (from climate models) in a spatial autoregressive framework. The relative contribution of local and large-scale sources is adjusted with observations. The approach may be used as kernel for predictive downscaling techniques.

Best NAO weighting

This function applies the statistical estimation theory to obtain the best linear unbiased estimation of NAO. Two Gaussian distributions, modelling the predicted winter NAO pdf, are used as prior estimates. They represent the ECMWF System 5 after bias correction and a skilful empirical relationship, respectively. Forecasted ECMWF System 5 members are then weighted according to the a posteriori NAO Gaussian pdf.

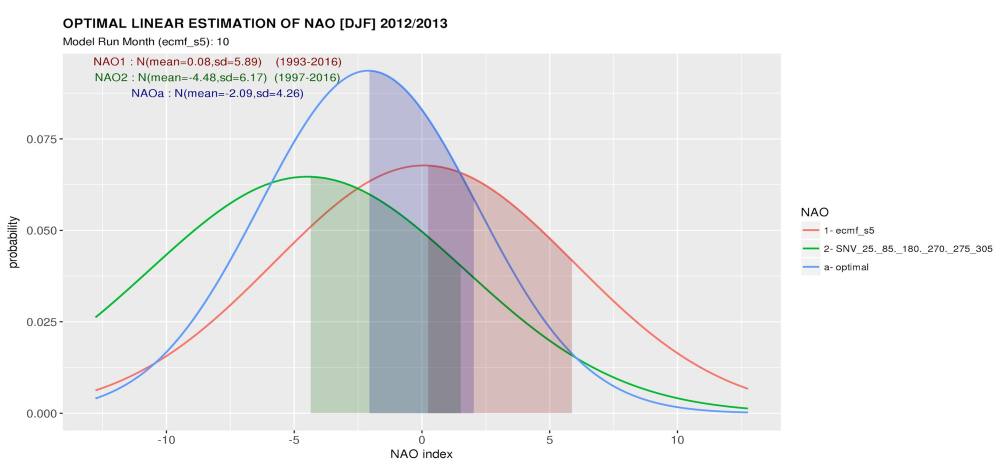


Fig 8: Gaussian pdfs representing the forecasted NAO distribution for winter 2012-2013 by the bias corrected ECMWF System 5 (red), S-ClimWare empirical system (green) and the a posteriori Best NAO estimate (blue).

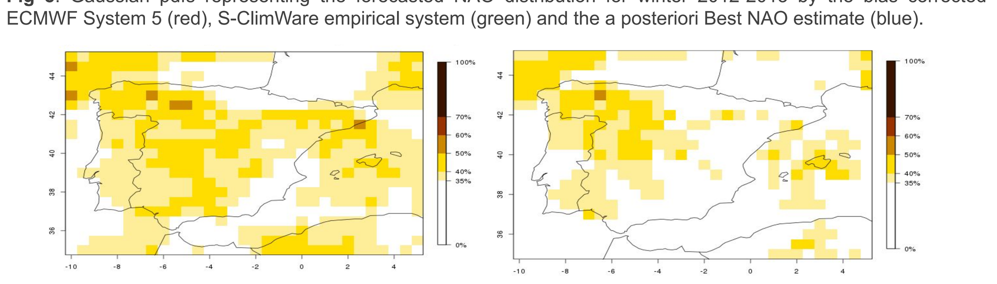


Fig 9: Left: Original probability from ECMWF Seasonal Forecast System 5 that the total precipitation from November 2012 to March 2013 will be in the lower tercile. Right: Probability that the total precipitation from November 2012 to March 2013 will be in the lower tercile based on the best NAO pdf estimate.

Downscale Analog

This function makes successive use of analogs based on Euclidean distance and regression to downscale maximum and minimum temperature and precipitation. It requires historical observations based on a new 5 km resolution gridded dataset covering the whole Iberian Peninsula (or other regions).

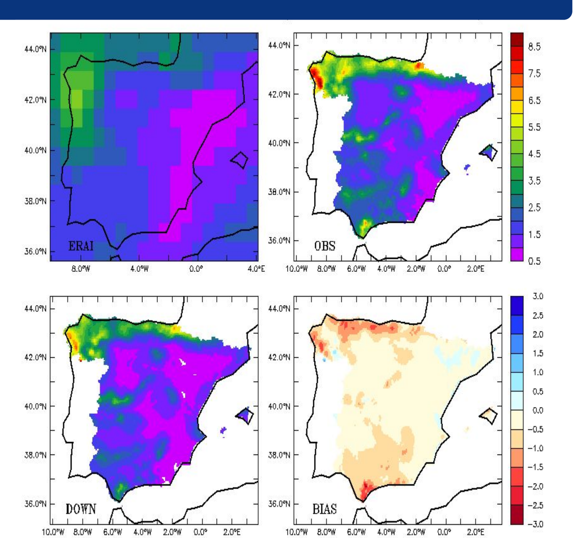


Fig 10: Example of the spatial distribution of the downscaling daily winter precipitation (mm/day) of ERA Interim (ERAI, upper-left panel), AEMET observational database (OBS, upper-right panel), downscaling ERA Interim-based (lower-left panel) and of the bias (lower-right panel), averaged over the period 1997-2016.

Acknowledgments

MEDSCOPE is co-funded by the H2020 ERA-Net ERA4CS European Research Area for Climate Services (Grant 690462).

Horizon 2020 European Union Funding for Research & Innovation

This poster can be downloaded at https://earth.bsc.es/wiki/doku.php?id=library:external:posters

Corresponding authors:
louis-philippe.caron@bsc.es, nuria.perez@bsc.es

PlotForecastPDF

This function plots one or several probabilistic ensemble forecasts side by side. For each forecast, it displays the ensemble members and a probability distribution function obtained from dressing the ensemble members. The probabilities for the three tercile categories and the above P10 and below P90 categories are also displayed. The observation can be added as well.

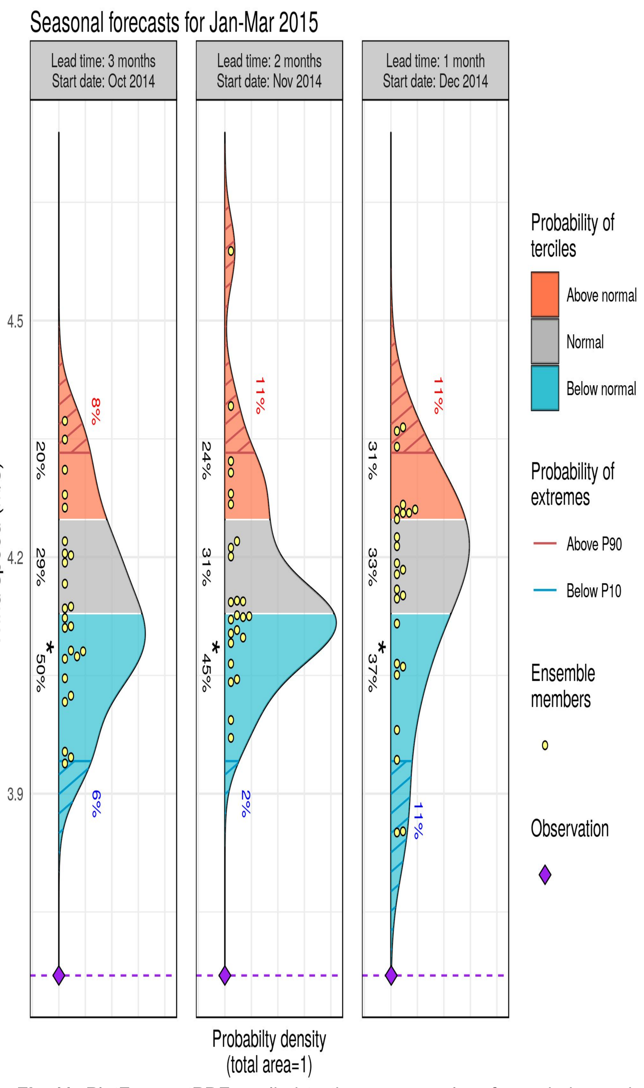


Fig 11: PlotForecastPDF applied to three seasonal surface wind speed forecasts. Each panel corresponds to a different start date. Each ensemble member (yellow circle) and the observation for that month (purple diamond) are drawn for each forecast. The probability of each tercile is shown in different colored shadows: above normal (brown), normal (grey), below normal (blue) and their value is specified on the left axis. The probabilities above 90th (below 10th) percentile are displayed with a red (blue) striped background. An asterisk marks the tercile with the highest probability.

ADAMONT

The ADAMONT method is a quantile mapping method for statistical adjustment of climate simulations uses analogs based on weather regimes to provide sub-daily disaggregation of data. It can be used for providing input to certain type of models, such as hydrological models.

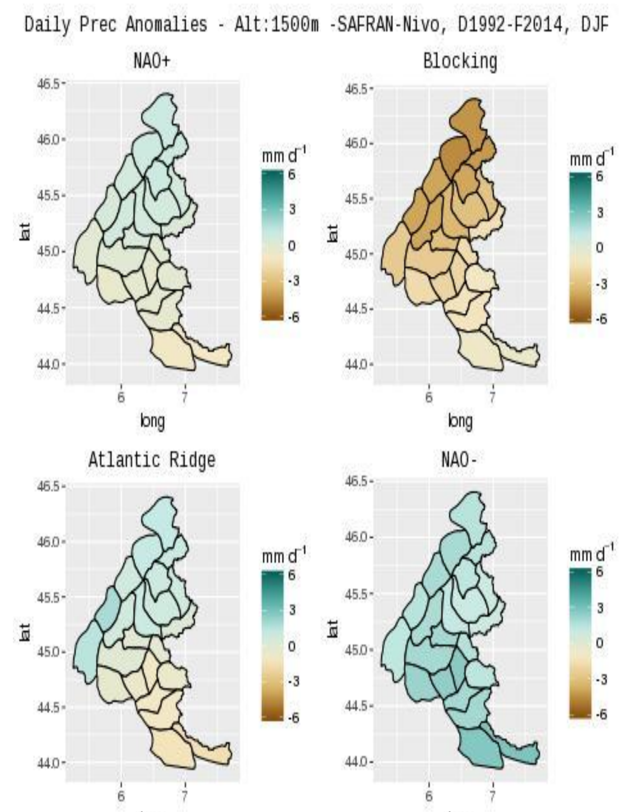


Fig 12: Daily precipitation anomalies at 1500m over the Alps massifs in the SAFRAN-Nivo reanalysis according to the North Atlantic weather regimes for DJF 1992-2014. This highlights the interest of bias-correcting seasonal predictions over the Euro-Mediterranean region with a dependency on weather regimes as will be possible with ADAMONT.

MultiVarRMSE

This function calculates the RMSE from multiple variables at once. The multivariate RMSE is computed as the mean of each variable's RMSE scaled by its observed standard deviation. The variables can also be weighted based on their relative importance, as defined by the user.

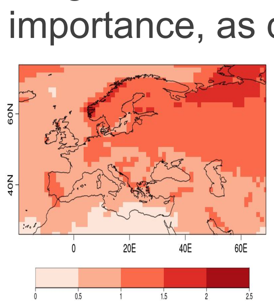


Fig 13: Example of the multivariate RMSE for surface air temperature and precipitation combined, from 1992 to 2012, using the Glosea5 seasonal forecasting system over Europe. In this example, temperature is assigned a weight of 2 and precipitation a weight of 1.