E-Infrastructures

H2020-EINFRA-2015-1

EINFRA-5-2015: Centres of Excellence
for computing applications

EoCoE

Energy oriented Center of Excellence

for computing applications

Grant Agreement Number: EINFRA-676629

D2.11 - M36

Calibration of wind-speed ensemble ensemble system prediction

Project and Deliverable Information Sheet

EoCoE	Project Ref:	EINFRA-676629
	Project Title:	Energy oriented Centre of Excellence
	Project Web Site:	http://www.eocoe.eu
	Deliverable ID:	D2.11 - M36
	Lead Beneficiary:	Meteo-France
	Contact:	Filipa Varino
	Contact's e-mail:	filipa.varino@meteo.fr
	Deliverable Nature:	Report
	Dissemination Level:	PU*
	Contractual Date of Delivery:	30/09/2018
	Actual Date of Delivery:	30/09/2018
	EC Project Officer:	Carlos Morais-Pires

* - The dissemination level are indicated as follows: PU – Public, CO – Confidential, only for members of the consortium (including the Commission Services) CL – Classified, as referred to in Commission Decision 2991/844/EC.

Document Control Sheet

Document	Title :	Calibration of wind-speed ensemble ensemble system prediction
	ID :	D2.11 - M36
	Available at:	http://www.eocoe.eu
	Software tool:	LaTeX
Authorship	Written by:	Filipa Varino, Olivier Mestre, Philippe Arbogast, Michael Zamo, Maxime Taillardat
	Contributors:	Meteo-France ,CNRS
	Reviewed by:	Edouard Audit, Nathalie Girard - PEC members

Contents

List of Figures

List of Tables

1. Introduction and objectives

Ensemble forecast models provide some measure of uncertainty when compared with deterministic forecasts. Nevertheless, these models still contain bias and/or errors that can be overcome using statistical post-processing calibration. This is important not only among the meteorological community but also for other meteorological-related activities as for instance, renewable energy production.

Concerning wind power production, accurate wind predictions are obviously determinant on the energy produced. In the case of extreme winds there are safety issues as well. An increase in the number of members (a member is an individual forecast) is then tested to mitigate this problem in the frame of the present project. This approach was adopted by the Forschungszentrum Jülich in Germany, with the development of an ultra-ensemble of one thousand members. In this study calibration methods are applied to this ultra ensemble. The added value of the calibration is evaluated with respect to an operational ensemble. In this report we present a summary of the work achieved during the period from the 1st of November 2017 up to the 30th of September 2018 at Météo-France/ Inria in collaboration with the Forschungszentrum Jülich in the framework of the EoCoE project.

2. PEARP and Ultra-Ensemble Data

The models used in this study are Meteo-France's model ARPEGE global Ensemble Prediction System (PEARP) (Descamps et al. 2015) and the Utra-ensemble Forschungszentrum Jülich WRF-based model.
The first one is an operational probabilistic system with a resolution up to 0.1 degrees in the European region with 35-members. The system accounts for initial condition and model uncertainties thanks to several subgrid-scale parametrization schemes and an ensemble of initial conditions accounting for observations uncertainty (figure 1). This second one is a 1000-member WRF-based model with a spatial resolution of 0.5 degree, running on the Forschungszentrum Jülich supercomputer JUQUEEN in the framework of the project EoCoE.
For this study we recover 512 members of this ultra-ensemble and the 35-members of the PEARP model. The period of study is different for each model, PEARP's data calibration is made form 1st of May 2015 to the 31st of May 2016 and Jülich's data is available between the 1st of May and the 30 of September 2015.

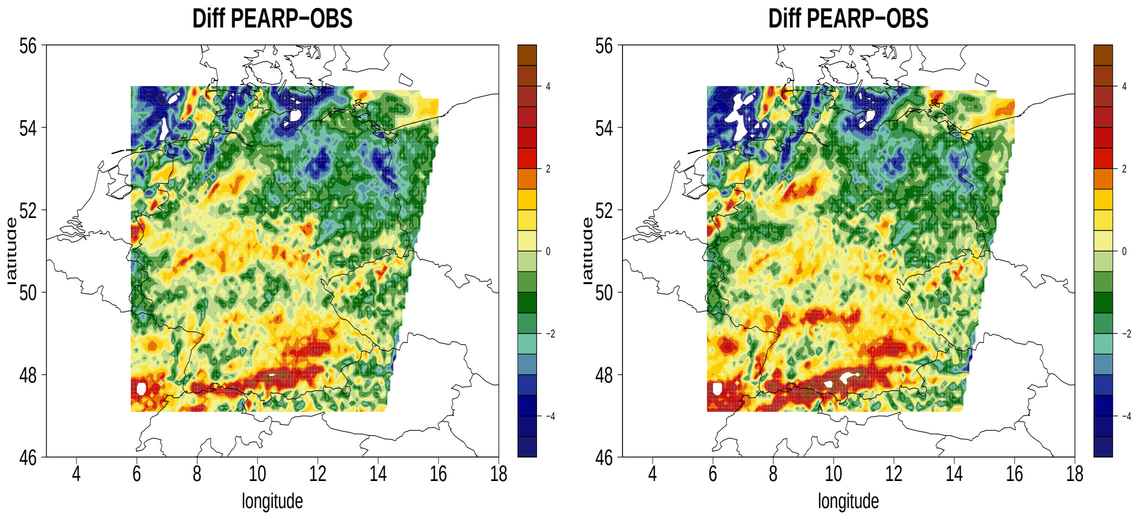


Figure 1: Example of grid point differences between two different PEARP members and the interpolated observation. Initial conditions create slightly different members outputs.

Finally, since the main objective is to improve wind turbine energy production in Germany, we focused only on Germany wind speed calibration.

2.1 Methodology and calibration

All calibration tools have been implemented on the Meteo-France's Bull supercomputer. First, data recovered from the Forschungszentrum Jülich is adapted to a standard format in order to be read by a pre-created calibration algorithm. Concerning the ultra-large ensemble, only wind speed have been retrieved due to time constraints. It should be stressed that the use of more dynamical predictors could have improved the skill of the calibration. Secondly, observation data an all PEARP data is retrieved for the period and area aforementioned. The method of calibration used for the study is the Quantile Regression Forest, a non-parametric post-processing statistical method suitable for ensemble systems. Preliminary results applied on PEARP forecasts show promising results

(Taillardat et al. 2016).

3. Results - Quantile regression forest calibration

3.1 PEARP

Figure 2a) presents the error between the ensemble mean of PEARP and the interpolated observations for the same period and same lead-times as the Ultra-Ensemble runs, i.e ,between 1st of May and 30th of September 2015. The figure 2b) shows the same information but for the calibrated version of PEARP. The bias is clearly reduced by the calibration.
The calibrated model of PEARP takes as predictors temperature, sea level pressure, the standard deviation and the 20, 50 and 80% percentiles of the wind speed ensemble for each day and each forecast range.

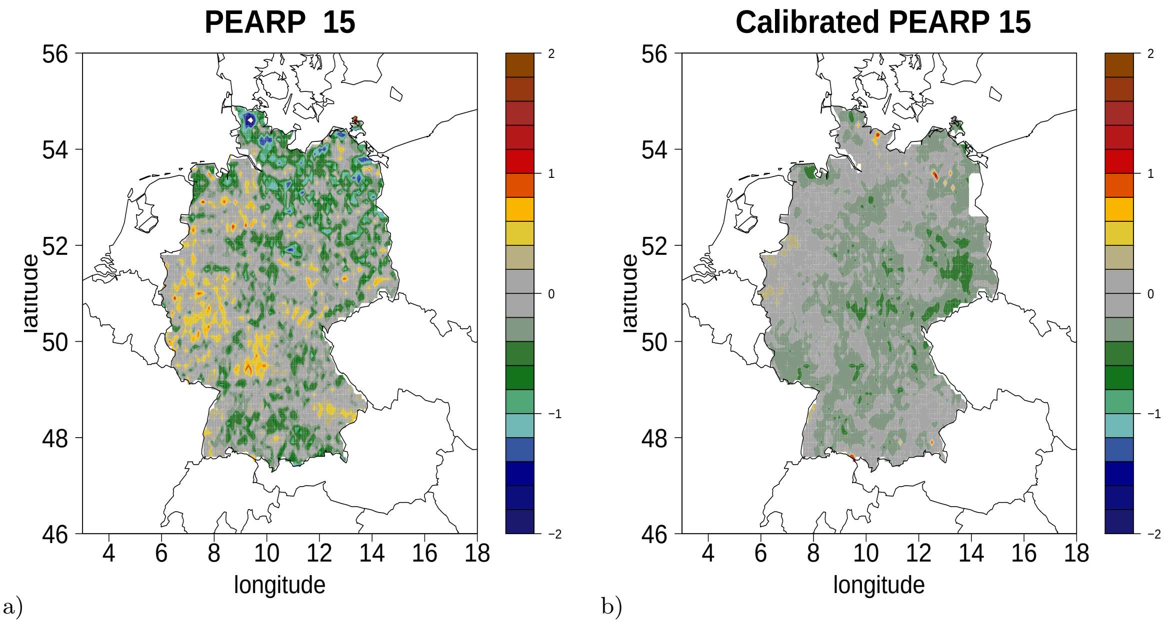
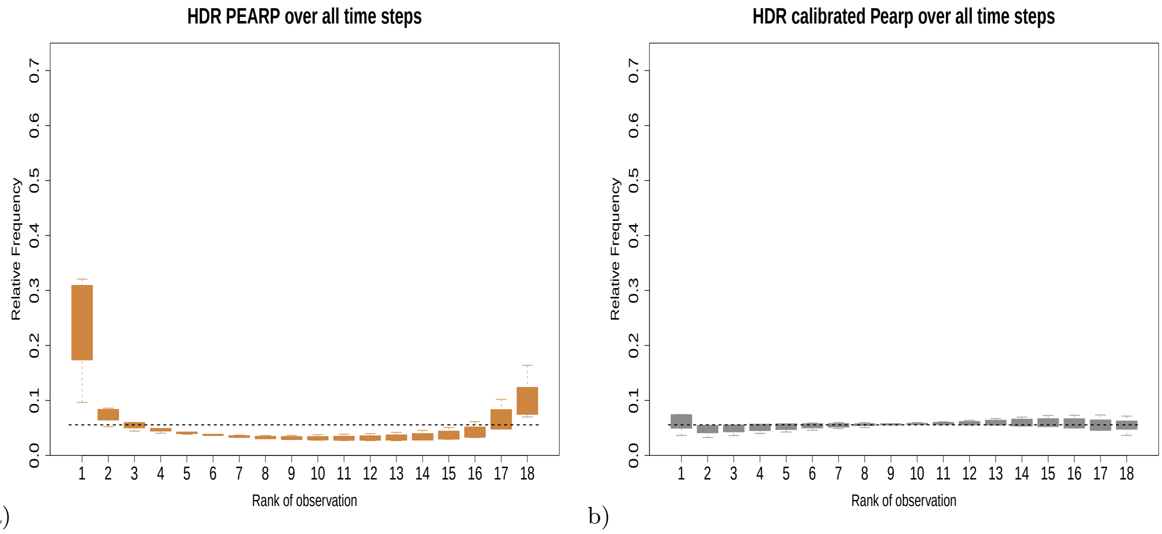


Figure 2: Grid point difference between Ensemble mean of the model and observations for the lead time of 15 hours for PEARP raw model (a) and calibrated PEARP model (b). Mean computed for all time steps of the study period (1st May -30 September 2015)

At this point we just have an idea of the significant bias reduction. The purpose of this WP is also to adjust the probability distribution function (pdf hereafter) sampled by the ensembles so that the reliability (perfect reliability is obtained when the observation and the ensemble come from the same pdf) and sharpness (ability to predict an event) are improved. We then introduce scores dedicated to this purpose.

Rank Histograms

The increase of reliability can be assessed using a standard tool called rank histogram. *Rank histograms (RH), also called Talagrand diagrams were developed independently by Anderson (1996), Talagrand et al. (1997), and Hamill and Colucci (1997). We employ RH to check the reliability of an ensemble forecast or a set of quantiles. An RH is built by ranking observations according to associated forecasts. Reliability implies that each rank should be filled with the same probability. Calibrated ensemble prediction systems should result in a flat RH* (Taillardat et al., 2016). Figure 3a) and b) represent rank histograms of the raw and calibrated versions of PEARP respectively. The PEARP



a) b)

Figure 3: Rank histogram for the raw (left in orange) and calibrated (right in grey) version of PEARP model

histogram (orange, on figure 3a)) is typical of a biased (the histogram is not symmetric) and underdispersive (U-shaped) ensemble.

The calibrated version of the model on figure 3b) shows a different pattern. The underdispersion is pretty well corrected as well as the bias.

3.2 Ultra-Ensemble

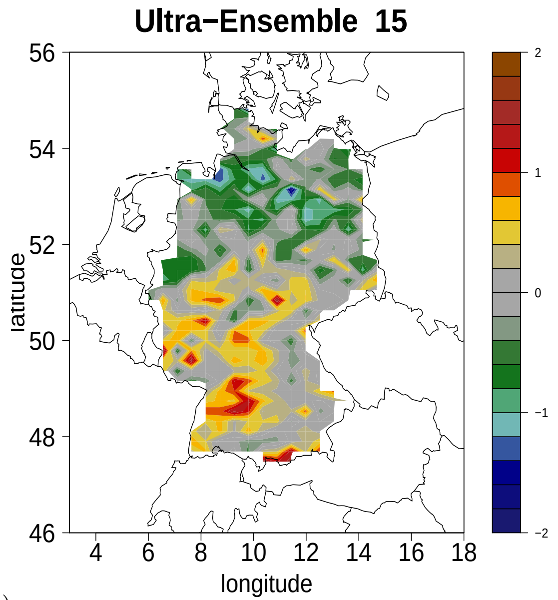
The error between the ensemble mean of the ultra-ensemble 512 members and the interpolated observations for the 5 months of study are shown in figure 4a). The is higher than the PEARP (figure 2), which is probability related to the coarser spatial resolution of the ultra-ensemble.

The calibration error is presented on figure 4b). As in PEARP's calibration it shows a decrease in the bias. Only statistics derived from wind data are used as predictors: the 20, 50 and 80% quantiles of the wind ensemble as well as its standard deviation.

Rank Histograms

The rank histograms for the raw and calibrate version of the Ultra-Ensemble are presented in figures 5a) and 5b).
First, since the resolution of the raw model is different from the interpolated observations, the rank histogram on figure 5a) is computed in a different way. For each grid point of the Ultra-ensemble the five nearest points of the observations are selected to compute a rank histogram.
The histogram shows a clear bias and a strong underdispersion that are must higher than for the PEARP model (figure 3a)).

The rank histogram of the calibrated version of the Ultra Ensemble (figure 5b)) shows a substantial improvement of the forecast but there are still some bias and under-



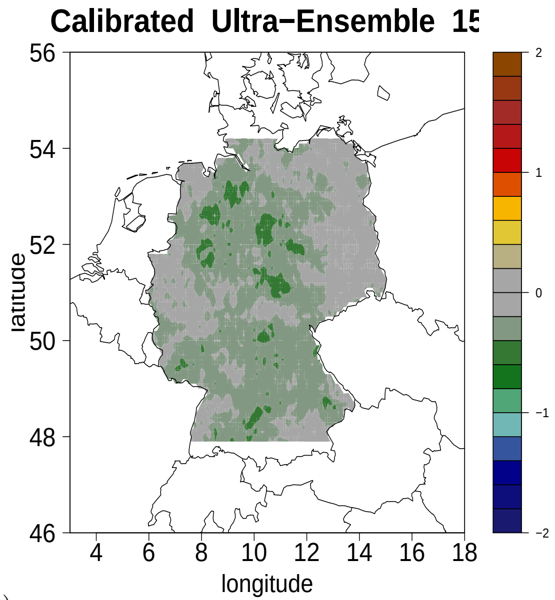


Figure 4: Same as figure 2 but for the 512-member Ultra- ensemble (left) and the calibrated version (right).

dispersion.

Ultra-Ensemble + PEARP

Finally, a last calibration is made using both Ultra-ensemble and PEARP's predictors. The output of this calibration help us understanding the role of both models in calibration. The rank histogram is presented in figure 6 in blue. The additional PEARP predictors have a positive effect on calibration which leads to the idea that there is still room for calibration improvement of the ultra-ensemble. We will go deeper into this intercomparison in next section.

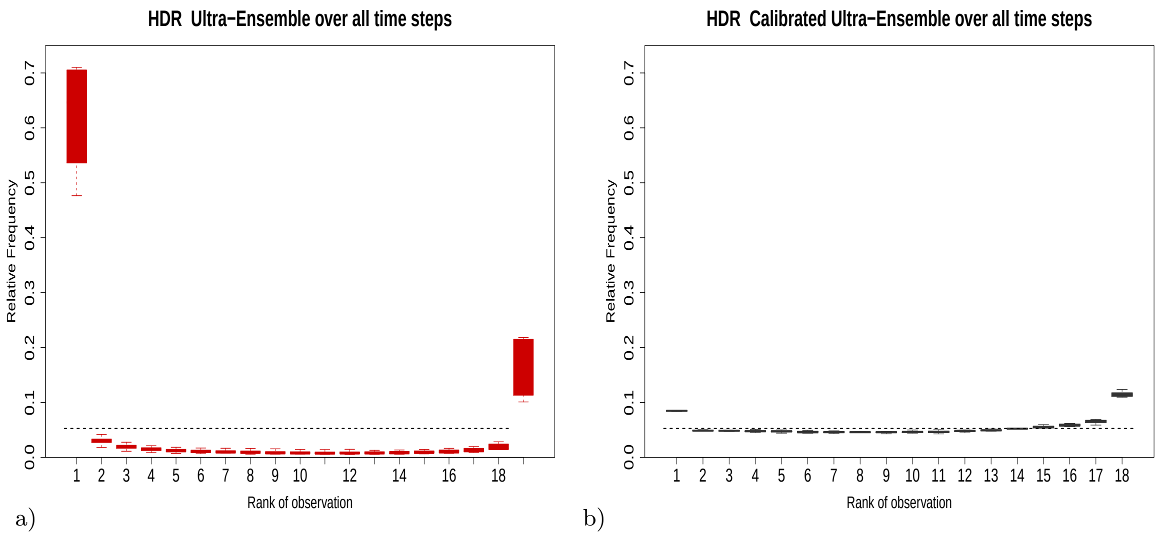


Figure 5: Same as figure 3 but for the Ultra ensemble (left in red) and its calibration (right in dark grey)

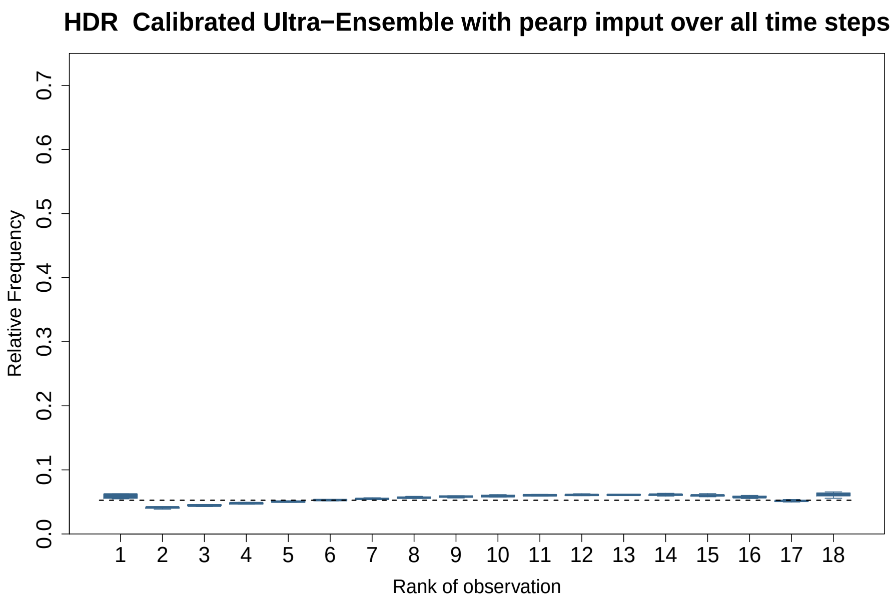


Figure 6: Same as figure 3 but for a calibration using both Ultra ensemble and PEARP imput.

3.3 Cumulative Ranked probabilistic Score

The cumulative Ranked probabilistic Score is one of the most used scores in ensemble systems. It measures the distance between the observed and forecast cumulative density distributions as in:

$$CRPS = \int_{-\infty}^{+\infty} [F(y) - F_o(y)]^2 dy$$

and the lower the better.
In figure 7 we show the CRPS of all models and calibrations computed at each grid point over all time steps and presented per lead time.

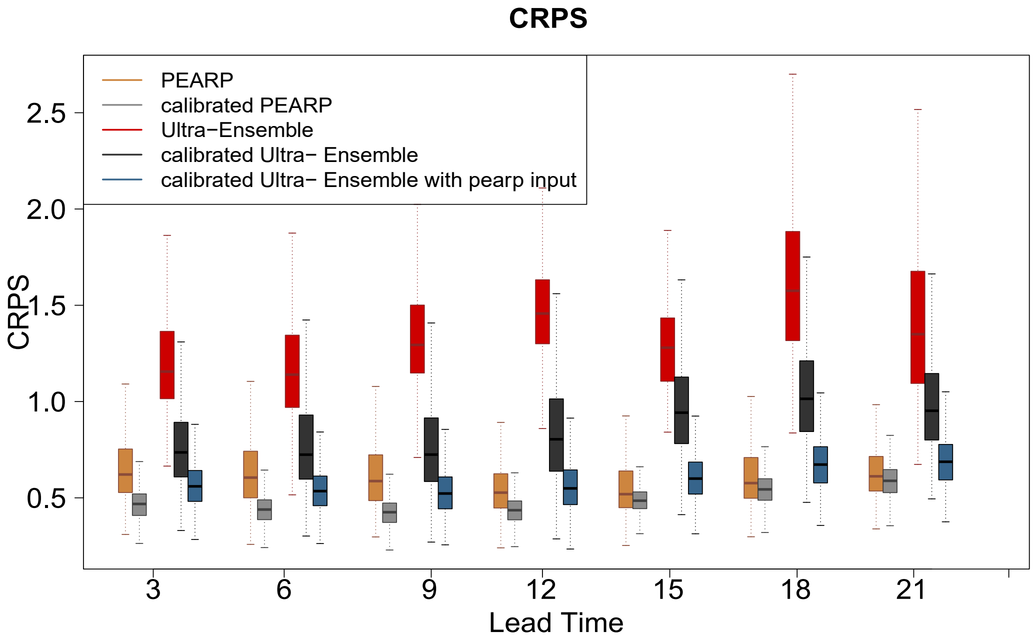


Figure 7: Computed crps for all time steps and each lead time for all models.

As expected, the calibrated PEARP and calibrated Ultra Ensemble (light and dark grey) have lower CRPS than their non-calibrated versions (PEARP in orange and Ultra Ensemble in red). The Ultra-Ensemble model is the one with the highest CRPS while the calibrated PEARP is the one with the lowest.
The blue boxplot represents the calibrated forecast using both models predictors as mentioned before. It is systematically better then the Ultra-ensemble calibration alone (dark grey boxplots) but not as good as the PEARP calibration (light grey). This is not surprising since raw PEARP's CRPS forecast is than better than the Ultra Ensemble and since spatial resolution is also higher. Clearly PEARP benefits from decades of tuning and feedback from operational validations.

4. Conclusions and Perspectives

This originality of this study lies in the fact that it is the very first attempt of calibration and comparison of a ultra-large ensemble and operational ensemble together. In this study we have demonstrated that wind forecasts of PEARP (Meteo-France) and Ultra-Ensemble model (Forschungszentrum Jülich) are significantly improved through the use of statistical calibration. This was obtained for the Ultra-Ensemble even when using only predictors derived from wind ensembles. Better calibrations can be obtained in various aspects:

- -in increasing the variety of predictors by the introduction of variables such as temperature and sea level pressure. This is postponed due to time constrains.
- -in using a larger study period. Optimal performance using the quantile regression forest method is obtained when using at least one year of data.
- - in focusing on extreme events.
- - optimising the ratio between size and spatial resolution of the ensemble.
- - instead of calibrating wind speed, a direct calibration of energy production using production data as variable of interest could be performed.

To that aim an increase in the computation power is clearly required . Finally, we have benefited from a better understanding of methods and techniques studied in the framework of the EoCoE project (Taillardat et al., 2016 and 2017, and Zamo, 2016).

5. References

Descamps, L. , Labadie, C. , Joly, A. , Bazile, E. , Arbogast, P. and Cébron, P. (2015): PEARP, the MétéoFrance shortrange ensemble prediction system.; Q.J.R. Meteorol. Soc, 141: 1671-1685. doi:10.1002/qj.2469

Taillardat, M., O. Mestre, M. Zamo, and P. Naveau, (2016): Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics.; Mon. Wea. Rev., 144, 2375–2393, https://doi.org/10.1175/MWR-D-15-0260.1

6. Publications made in the framework of EoCoE

Maxime Taillardat, Anne-Laure Fougères; Philippe Naveau; Olivier mestre (2017) : Forest-based methods and ensemble model output statistics for rainfall ensemble forecasting" (WAF-D-18-0149) ; released in arxiv (https://arxiv.org/pdf/1711.10937.pdf)

Zamo, M., L. Bel, O. Mestre, and J. Stein, (2016): Improved Gridded Wind Speed Forecasts by Statistical Postprocessing of Numerical Models with Block Regression.; Wea. Forecasting, 31, 1929–1945 https://doi.org/10.1175/WAF-D-16-0052.1