Example of operational post-processing using machine learning

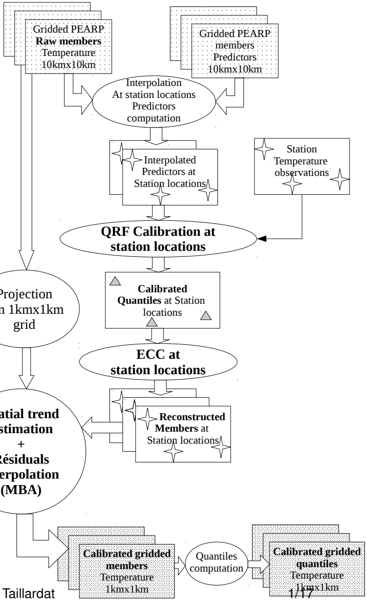
Maxime Taillardat
Olivier Mestre

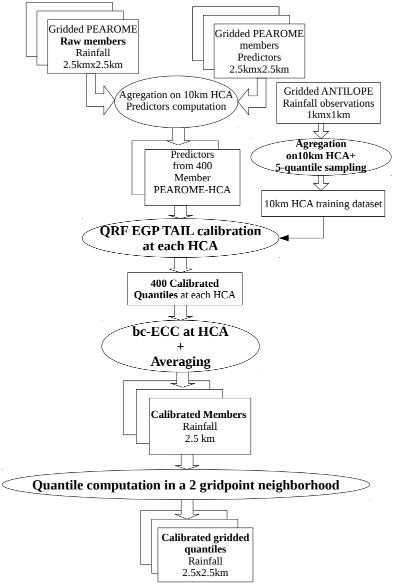
Météo-France / CNRM UMR 3589

maxime.taillardat@meteo.fr

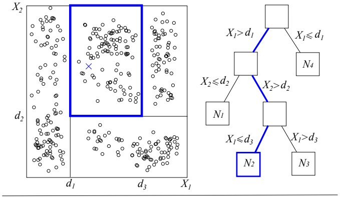
EGU Meeting 2020 - May, the 8th

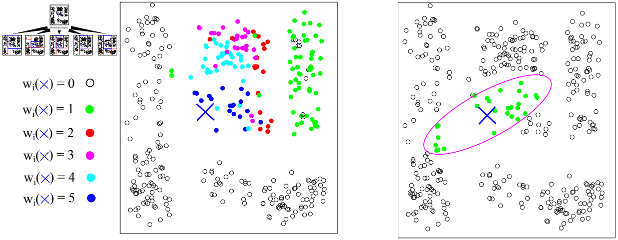
Two examples of PP at an industrial scale





Method of PP employed : QRF, another way to find analogues





$$\widehat{\mathbb{P}}(Y \leq y | X = x) = \frac{1}{N} \sum_{i=1}^{N} w_i(x) \mathbf{1}\{Y_i \leq y\}$$

Method of PP employed : QRF, another way to find analogues

Pros

- No assumptions on the target variable
- Self-selction of the most useful predictors, interpretable
- hyperparameters tuning quite easy and stable over locations vs. other ML techniques

Cons

- Potentially big models (need massive HPC optimization, and storage capacities)
- QRF cannot go "beyond the range of the data"
- available archives: 2 years

Taillardat, Maxime, Olivier Mestre, Michaël Zamo, and Philippe Naveau. "Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics." *Monthly Weather Review* 144, no. 6 (2016): 2375-2393.

In a forecast automation context

Goal: Be skillful for extremes events subject to a good overall performance

Temperature

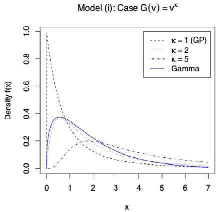
Work on forecast anomalies (w.r.t the ensemble mean for example)

Hourly rainfall

Use QRF outputs to fit a distribution which would:

- Model jointly low, moderate and heavy rainfall
- Be flexible

- Use of an Extended GP distribution (EGP3) (Papastathopoulos and Tawn, 2013 ; Naveau et al., 2016 ; Tencaliec et al. 2019)



A semi-parametric approach for hourly rainfall

Our final distribution is:

$$G(x) = f_0 + (1 - f_0)\left[1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}\right]^{\kappa}$$

Strategy

1. Run QRF to get $\widehat{F}(y|X = x) = \widehat{\mathbb{P}}(Y \leq y|X = x)$
2. Keep the probability of no rain $\widehat{f_0} = \widehat{\mathbb{P}}(Y = 0|X = x)$ from QRF outputs
3. Estimate $(\widehat{\kappa}, \widehat{\sigma}, \widehat{\xi})$ from non-zero QRF quantiles

Taillardat, Maxime, Anne-Laure Fougères, Philippe Naveau, and Olivier Mestre. "Forest-based and semi-parametric methods for the postprocessing of rainfall ensemble forecasting" *Weather and Forecasting* (2019).

Post-processing of Temperature post-processing

- Observations available on 2000 stations locations across Western Europe
- Raw model resolution: 10km

Goal

- Station-wise post-processing with ECC
- Target resolution: 1km (Downscaling step), 4000000 points

Procedure time has to be inferior to 15min for operational constraints.

Towards high resolution temperature fields

Similar to regression-kriging

Regression phase member by member

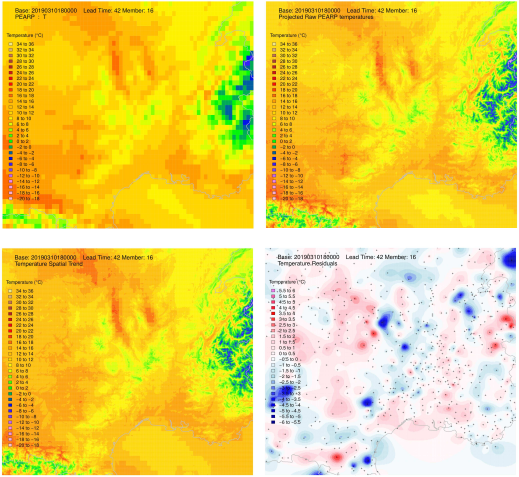
- between PP values and (downscaled) raw NWP values
- On homogeneous climate zones
- With geomorphological predictors (altitude, distance to coast, PCA on topography...)

Regression equation applied to the whole grid: spatial trend estimation

Spatialization of residuals

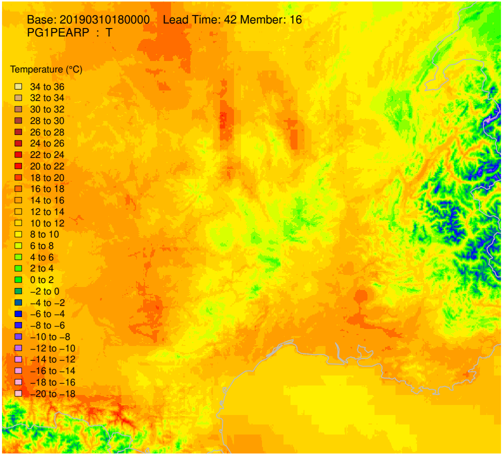
- using multi-resolution B-splines (MBA ; Lee et al., 1997)

Towards high resolution temperature fields: Illustration



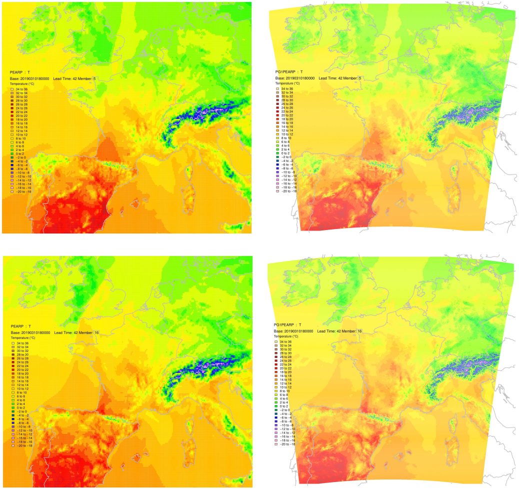
Step-by-step procedure illustrated over the southeast of France: raw member temperatures on 10km grid (upper left panel), raw projected temperatures on a 1km grid (upper right panel), spatial trend estimation using regression model on subdomains (lower left panel), field of residuals interpolated using a MBA procedure (lower right panel).

Towards high resolution temperature fields: Illustration



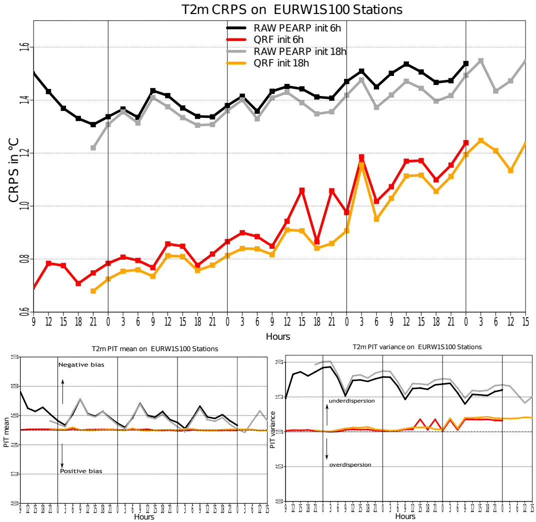
Resulting member.

Towards high resolution temperature fields: Illustration



Raw member 6 temperature field (upper left panel), the same after calibration, ECC and interpolation phase (upper right panel) together with raw (lower left panel) and post-processed temperature field (lower right panel) for member 16.

Performance on stations



Results of post-processing of temperature in the 2056 stations with averages CRPS (top), and mean and variance of PIT statistic, related to rank histograms. The validation is made by a 2-fold cross-validation on the two years of data (one sample per year).

Operational framework for hourly rainfall

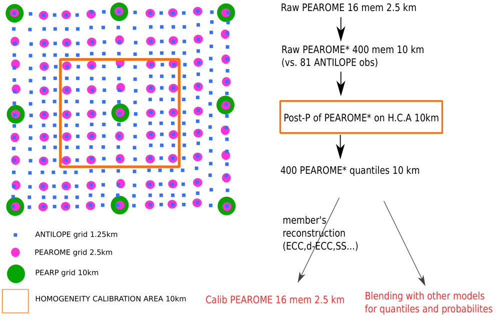
- French grid : 300000 gridpoints
- PEAROME (16 members, 2.5 km), lead times from 1 to 45 hours
- Observations: Radar+rain gauges ANTILOPEJP1H (1 km)
- Semi-parametric QRF
- Restore scenarios (post-processed members).

Predictors

- Max, q50, q10, q90, sd, mean, Proba rain, Proba >5mm/h RR1
- Max, Proba rain RR1 lead time before
- q10, q90 de reflectivity max.
- Mean CAPE_INS
- Mean ICA
- q10, q90 of HU 500m, 700hPa, TCC
- mean FX 10m
- mean U,V, FF 700hPa

Architecture

- Data pooling: We consider high res. errors homogeneous on 10km boxes (spatial penalty). PP is made on these HCA: number of statistical models reduced by a factor 25. (14000 HCA)
- Data boosting: As observation is at 1km, observation is a distribution. Instead of taking one upscaled observation, the empirical quantiles of order 0, 0.25, 0.5, 0.75, 1 of ANTILOPE distribution in the HCA are taken. The length of the training sample is inflated by a factor 5.



What sort of members do we want ?

Schaake Shuffle (SS) and MD-SS(see e.g. Clarke, 2005 ; Scheuerer, 2018)

- We need an observations archive, we lose the model "signature"

Ensemble Copula Coupling-like methods (ECC) (see e.g. Schefzik et al., 2013 ; Ben Bouallègue et al., 2017)

- Using (potentially wrong) physical structures of the raw ensemble

ECC and rainfall: it is not so simple...

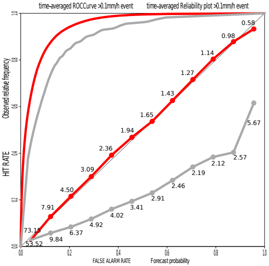
Bootstrapped-Constrained Ensemble Copula Coupling (bc-ECC)

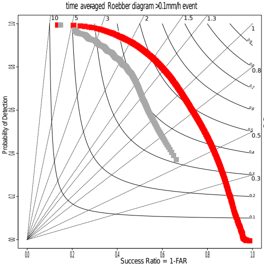
We do ECC many times (here 250 times by HCA) and average values :

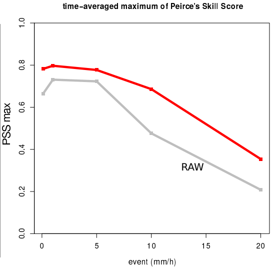
- If raw zeros > calib. zeros : smallest non-zero calib. rainfall are assigned and averaged on raw zeros
- a raw zero becomes a non-zero member IF there is a raw non-zero member in a 2 grid point neighborhood

Calibration : 1 distribution on 1 HCA $\xrightarrow{bc-ECC}$ 16 members / grid point

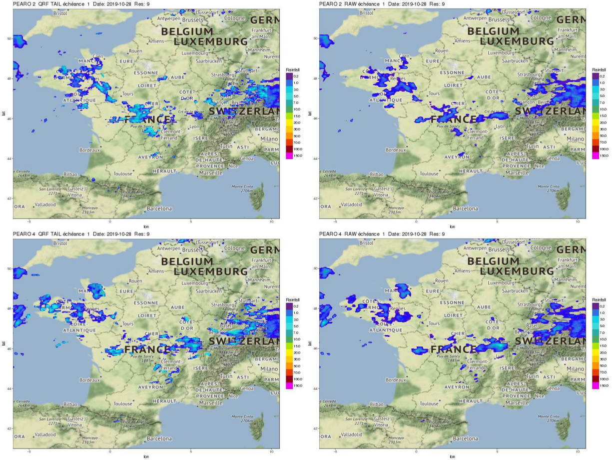
Rain discrimination results







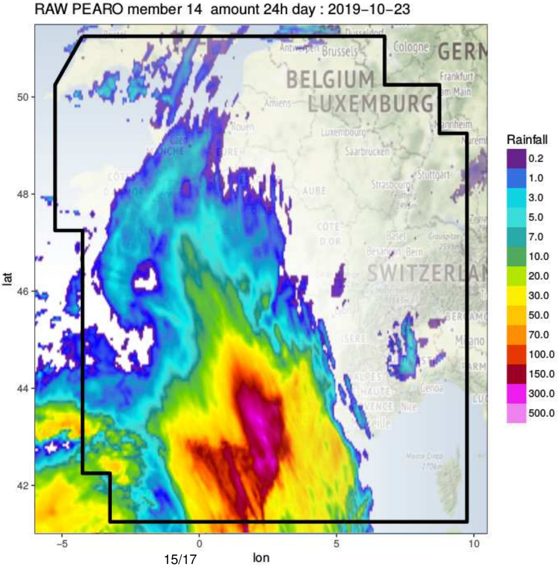
ECC + post-processing visualization



2 PP members (left) with their associated raw members (right)

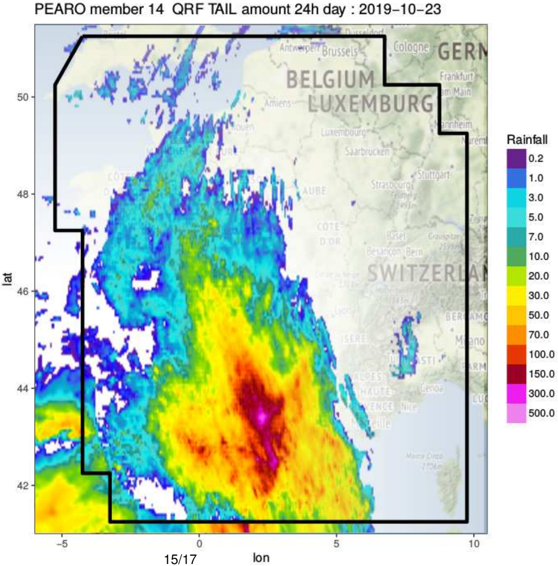
Visual inspection on a heavy Mediterranean event

The best member of the raw ensemble for this event vs. the PP one vs. the radar obs.



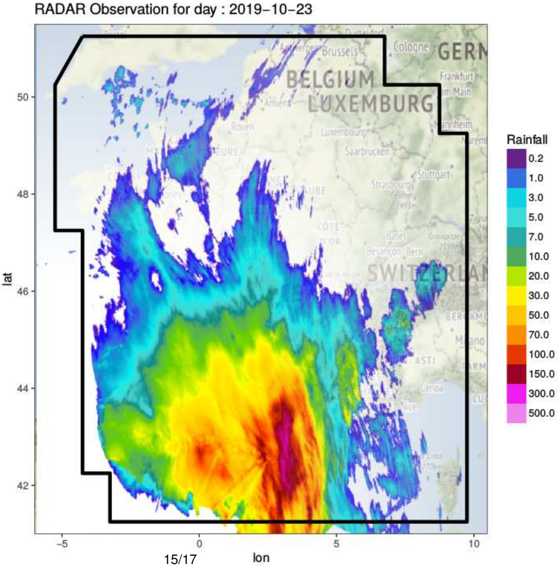
Visual inspection on a heavy Mediterranean event

The best member of the raw ensemble for this event vs. the PP one vs. the radar obs.

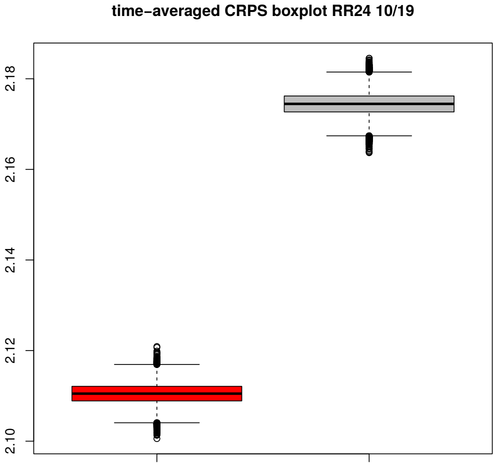


Visual inspection on a heavy Mediterranean event

The best member of the raw ensemble for this event vs. the PP one vs. the radar obs.

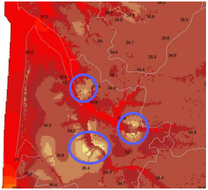


Are RR24 generated by PP RR1 + bc-ECC good ?



Conclusion

- No "absolute" method
- Tuning takes time
- Different goals, computing capabilites, skills = different algorithms to consider
- Methods should be interpretable, robust, with easy/universal set-up.
- Forecast automation : avoiding big/unphysical mistakes. (not seen by classical scoring rules). Must do visual inspections.



Reference

- *Taillardat, Maxime, and Olivier Mestre. "From research to applications–Examples of operational ensemble post-processing in France using machine learning." Nonlinear Processes in Geophysics Discussions (2020): 1-27.*

- QRF: good, easy to tune, but big models (here several hundreds of Gb). Deep Learning/ (C)NN is coming... Is the triptych "performance/tuning/model size" better with U-net/CNN ?
- PP strategies highly depend on NWP archive data policy/capacities...