Gaussian mixture models for clustering and calibration of ensemble weather forecasts

Jouan Gabriel [1] Cuzol Anne [2] Monbet Valérie [3] Monnier Goulven [4]

[1]University of Rennes / Scalian DS

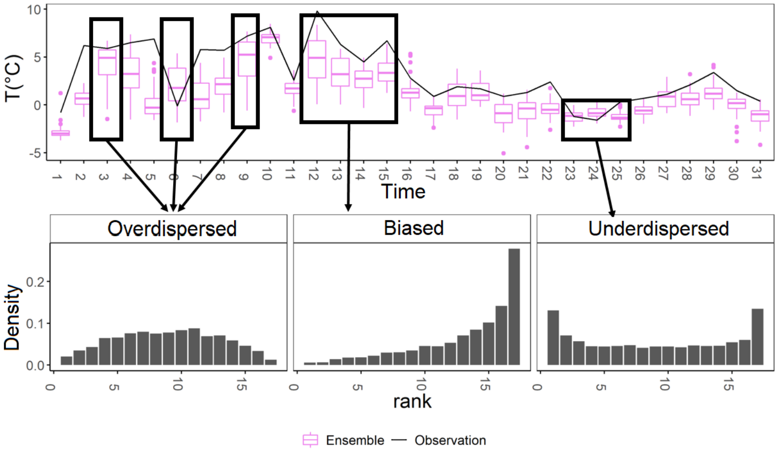
[2]University of Bretagne Sud

[3]University of Rennes

[4]Scalian DS

24 mai 2022

Introduction



- Different error types identified ;
- Extension of the Nonhomogeneous Gaussian Regression (NGR) with a revisited Gaussian Mixture Model (GMM) applied to ensemble data.

NGR-Z : a two steps calibration model

First step : Cluster ensemble data into K weather regimes.

Three strategies proposed :

1. "Empirical statistics" ;
   - GMM is fitted on a vector $S_i = (\overline{x}_i, s_i^2)$ of empirical statistics derived from the ensemble X_i^* at the time i.
2. "Super sample" ;
   - Classical GMM applied on each ensemble members defined as i.i.d element from the dataset.
3. "Exchangeable variables" ;
   - GMM defined on $(X_1, \cdots, X_M)$, a vector of exchangeable gaussian random variables.

NGR-Z : a two steps calibration model

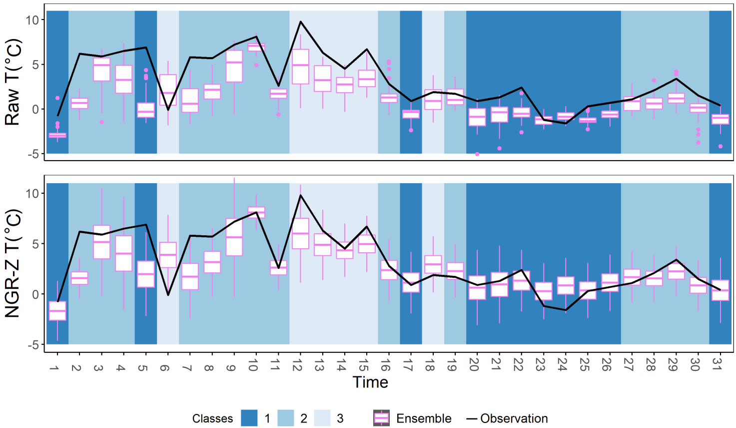
First step : Cluster ensemble data into K weather regimes.

Three strategies proposed :

1. "Empirical statistics" ;
   - GMM is fitted on a vector $S_i = (\overline{x}_i, s_i^2)$ of empirical statistics derived from the ensemble X_i^* at the time i.
2. "Super sample" ;
   - Classical GMM applied on each ensemble members defined as i.i.d element from the dataset.
3. "Exchangeable variables" ;
   - GMM defined on $(X_1, \cdots, X_M)$, a vector of exchangeable gaussian random variables.

Second step : Fit an *NGR* on each $k \in \{1, \cdots, K\}$ regime.

Conclusion



- Biase reduced for $t \in \{12, 13, 14, 15\}$;
- Dispersion improved for $t \in \{23, 24, 25\}$.

Work published into "Discrete and Continuous Dynamical Systems - Series S" [1].

1. https://www.aimsciences.org/article/doi/10.3934/dcdss.2022037?viewType=html#b31

Conclusion

Thank you for your attention.

contact : gabriel.jouan@scalian.com