Ensemble dispersion improvement auto-tune (EDIT), a generic post processing step for Machine Learning regression results with quantified uncertainty

Jouke de Baar, Cees de Valk, Gerard van der Schrier (jouke.de.baar@knmi.nl)

Challenge

An objective of machine learning is to provide information to support data-driven decision making. It has been shown that people can make better decisions if we provide them with a quantification of the uncertainty in our results.[2]

An ensemble approach is a powerful way of propagating and communicating uncertainty of gridded data sets, obtained from spatial regression. However, the uncertainty our ensemble provides us is not always realistic. Therefore, we aim to improve the ensemble dispersion.

Approach

We introduce ensemble dispersion improvement auto-tune (EDIT). Using cross-validation, we compare the station prediction error with the local ensemble. We then correct the ensemble dispersion.

The local correction factor is a function of proxies (in our application, the proxies are 'distance to nearest station' and 'local terrain complexity'). We determine the coefficients of the correction function using multi-objective optimisation (in our application, the objectives are a flat rank histogram and an accurate proxy response).

"Perhaps nothing is completely true, and not even thát."[1]

Application

We apply this approach to the gridded E-OBS data set for daily mean wind speed, which provides daily wind speed over Europe for the period 1980-now and is updated monthly. [3,4]

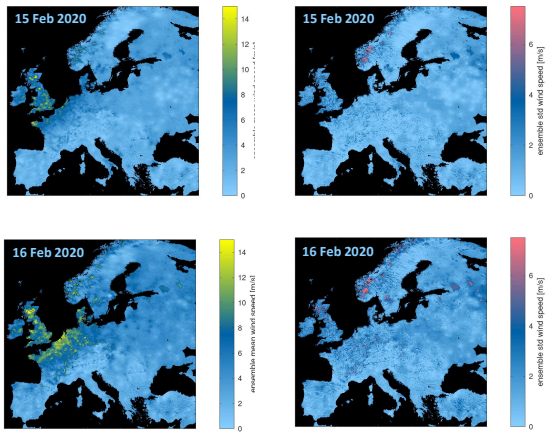

Figure: E-OBS daily mean wind speed for storm Dennis (2020), showing (left) ensemble mean and (right) EDITed ensemble standard deviation. [4]

Results

When we apply EDIT to the ensemble of the E-OBS gridded data set for daily mean wind speed, we see a clear reduction of the relative standard deviation of the station cross-validation-based rank histogram. This reduction seems consistent over the entire period. Also, we see a reduction of the relative rms error of the response of the uncertainty to changes in the proxies ('distance to nearest nearest station' and 'local complexity of terrain'). This indicates that the dependency of the ensemble spread on these proxies is realistic. [4]

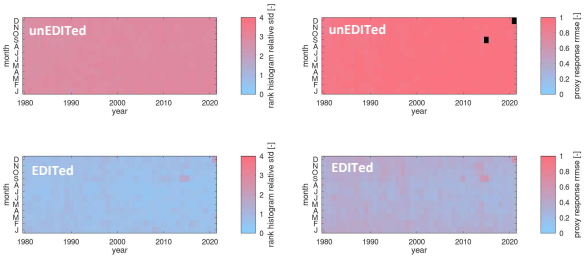

Figure: EDIT results, showing (left) rank histogram relative standard deviation and (right) proxy-response relative rms error. Both objectives are reduced by EDIT. [4]

References

1. Multatuli. Ideën van Multatuli, Zesde herziene druk. G.L. Funke, Amsterdam; 1879.
2. Ripberger J, Bell A, Fox A, Forney A, Livingston W, Gaddie C, Silva C, Jenkins-Smith H, Communicating Probability Information in Weather Forecasts: Findings and Recommendations from a Living Systematic Review of the Research Literature. Weather Climate and Society (2022)
3. Cornes RC, van der Schrier G, van den Besselaar EJM, Jones PD. An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. Journal of Geophysical Research: Atmospheres (2018)
4. de Baar JHS, van der Schrier G, van den Besselaar EJM, Garcia-Marti I, de Valk C, A new E-OBS gridded data set for daily mean wind speed over Europe (under review)

Outlook

EDIT is a generic post-processing step for ensemble regression and prediction. It relies on cross validation, and is such is only mildly intrusive. Besides applying EDIT to other members of the E-OBS family, it might be worth investigating the applicability of EDIT in other applications.

Copernicus Europe's eyes on Earth | European Commission | ECMWF