TECHNISCHE UNIVERSITÄT DRESDEN

Disaggregation of meteorological input data from daily to hourly resolution by means of gradient boosting

Philipp Körner

Objectives

Meteorological values in hourly resolution are a prerequisite for the operation of numerous hydrological models as input data. In comparison to meteorological data in daily resolution, hourly data are available for fewer stations. In addition, the time series are significantly shorter. For longer historical calculations or projections, the required hourly values are disaggregated from daily values. The disaggregation is to be carried out multivariate, i.e. for several meteorological quantities, on the one hand, and for several stations, on the other hand, so that the relationships between the elements and stations are preserved.

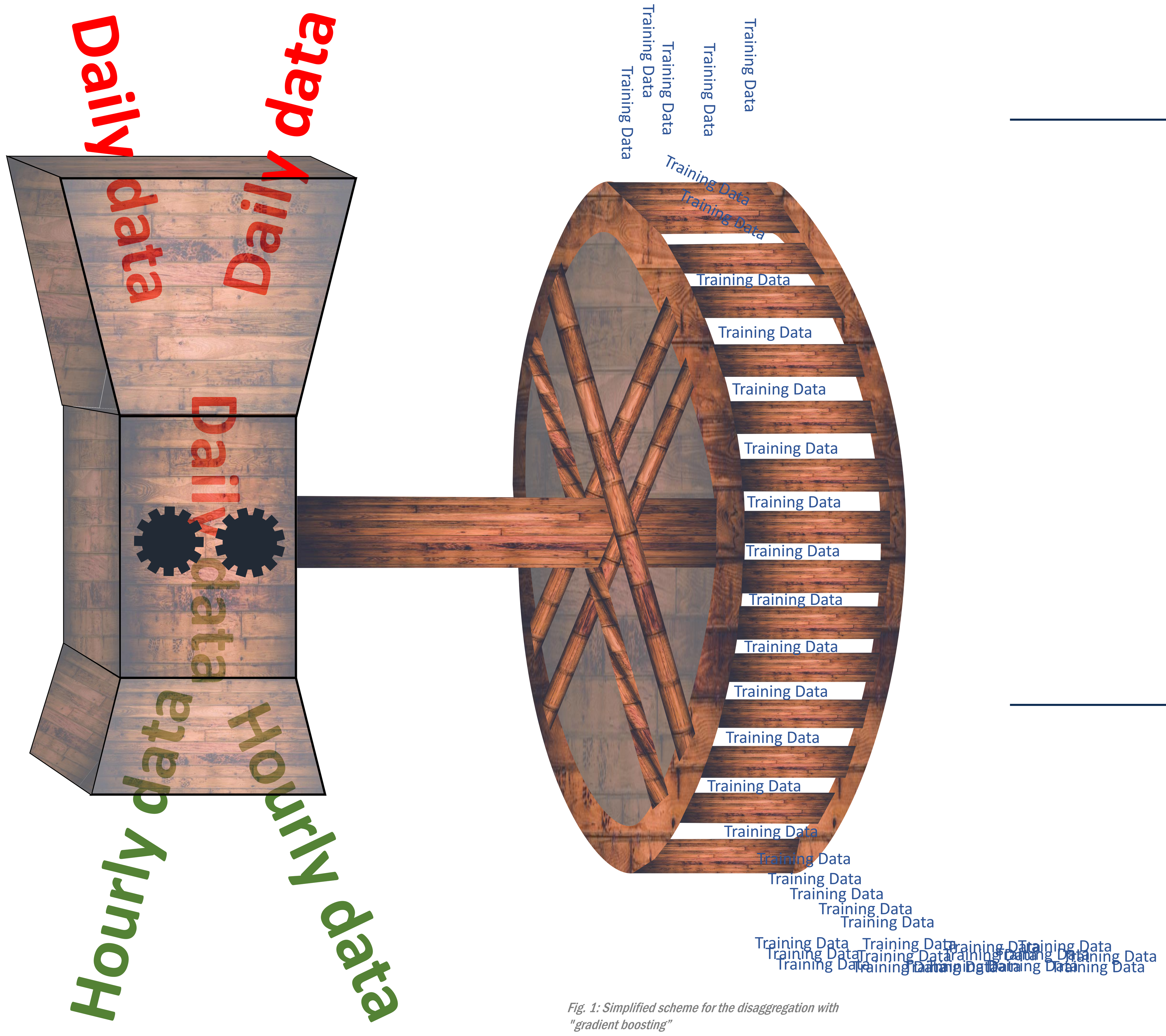

Fig. 1: Simplified scheme for the disaggregation with "gradient boosting"

Methods

Gradient boosting:

- Model is trained for the variables temperature, precipitation, relative humidity, sunshine duration and wind speed from historical hourly data.
- In addition to the meteorological data in daily resolution, predictors are the day of the year, the respective hour and the geographical coordinates including the terrain height.
- Cross validation was done for a total of about 200 stations in Germany with 5 folds each.

Results

The quality of the results can be measured in different ways. First, the disaggregated hourly values can be directly compared with the measured ones. This comparison is shown in Table 1.

In a second step, the multivariate correlation can be compared. For this purpose, two comparisons of the correlations (R) of two parameters each are presented here as examples (Fig. 2).

In the third step, the correlation between the stations is investigated. This comparison is illustrated by two variables in Figure 3.

Tab. 1.: Statistical parameters of disaggregation

	RMSE	R^2
Temperature (°C)	1.18	0.98
Relative Humidity (%)	6.15	0.86
Wind Speed (m/s)	1.05	0.74
Sunshine Duration (min)	10.97	0.78
Precipitation (mm/h)	0.54	0.12

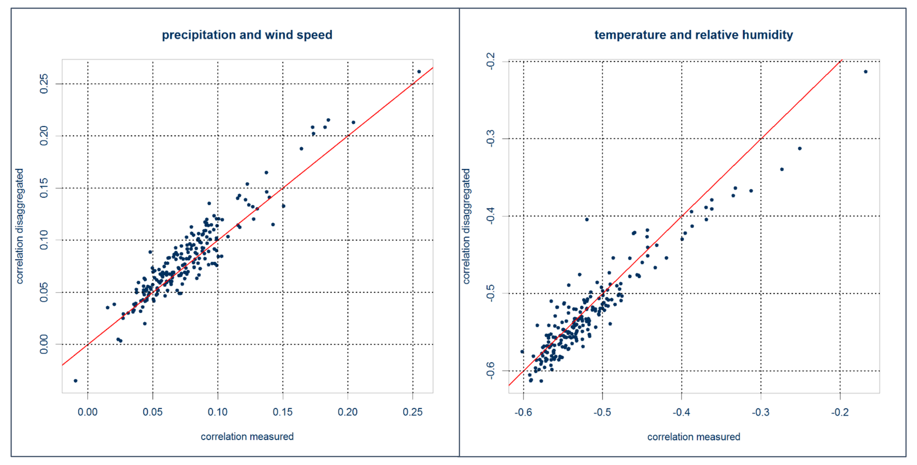

Fig. 2: Multivariate consistency using two examples. Each point represents one station.

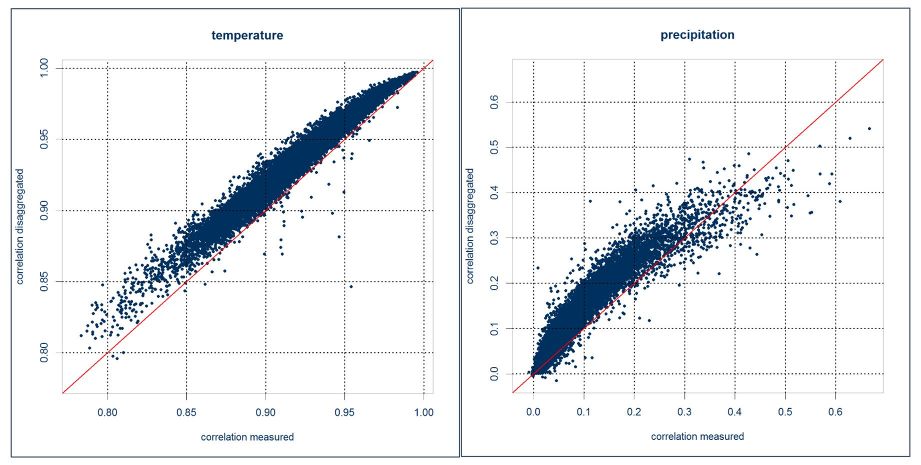

Fig. 3: Spatial consistency using two examples. Each point represents one station.

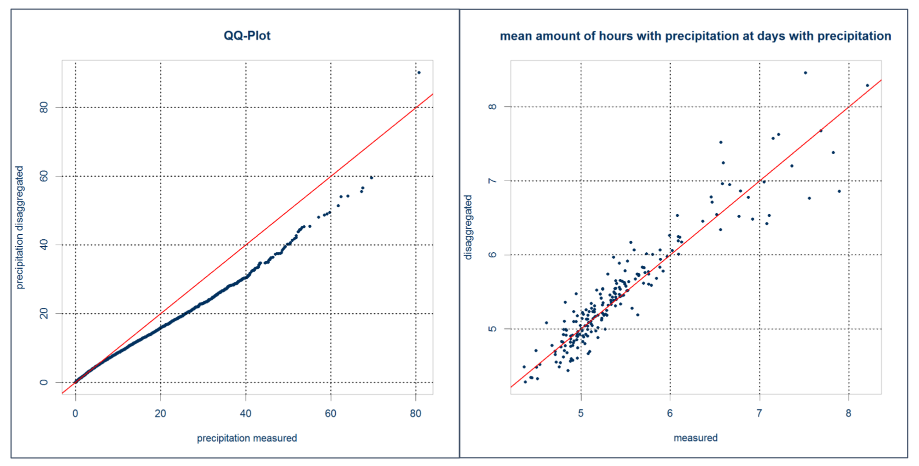

Fig. 4: Representation of the quality of precipitation as a QQ plot and for the mean number of precipitation hours per precipitation day for each station.

Discussion

The method requires measured hourly values of the involved parameters in the past. However, these do not have to be measured at the station itself. The cross validation shows the quality of the disaggregation without using historical data of the stations to be disaggregated in the training phase. Compared to other methods (e.g. Förster et al., 2016), cross-validation shows very good results. The multivariate consistency is given, as shown as an example in Fig. 2. However, there is a lack of comparison possibilities to other methods, as well as for spatial consistency. The correlation (R) of the individual stations is somewhat higher for the disaggregated data than for the measured data. There are deviations for precipitation time series which are spatially very close to each other. Here the correlation of the disaggregated data is slightly lower. Due to its heterogeneity, precipitation should be considered separately: The QQ plot (Fig. 4) shows that high precipitation on average is somewhat underestimated. The precipitation hours per day are hit with a Heidke Skill Score (HSS) of 0.35, the mean number of precipitation hours per precipitation day is also disaggregated appropriately.

Sources:

1. Chen, T. et al. xgboost: Extreme Gradient Boosting. (2018).

2. Förster, K., Hanzer, F., Winter, B., Marke, T. & Strasser, U. An open-source MEteoroLOgical observation time series DISaggregation Tool (MELODIST v0.1.1). Geoscientific Model Development 9, 2315–2333 (2016).

3. Körner, P., Kronenberg, R., Genzel, S. & Bernhofer, C. Introducing Gradient Boosting as a universal gap filling tool for meteorological time series. Meteorologische Zeitschrift 369–376 (2018). doi:10.1127/metz/2018/0908

IHM METEOROLOGIE
DRESDEN concept
Philipp Körner
Institute of Hydrology and Meteorology, Chair of Meteorology
philipp.koerner@tu-dresden.de