Geophysical Research Abstracts
Vol. 17, EGU2015-5231, 2015
EGU General Assembly 2015

On demand processing of climate station sensor data

Stephan Wöllauer, Spaska Forteva, and Thomas Nauss

Environmental Informatics, Faculty of Geography, Philipps-University Marburg, Marburg, Germany
(stephan.woellauer@geo.uni-marburg.de)

Large sets of climate stations with several sensors produce big amounts of finegrained time series data. To gain value of this data, further processing and aggregation is needed. We present a flexible system to process the raw data on demand.

Several aspects need to be considered to process the raw data in a way that scientists can use the processed data conveniently for their specific research interests. First of all, it is not feasible to pre-process the data in advance because of the great variety of ways it can be processed. Therefore, in this approach only the raw measurement data is archived in a database. When a scientist requires some time series, the system processes the required raw data according to the user-defined request.

Based on the type of measurement sensor, some data validation is needed, because the climate station sensors may produce erroneous data. Currently, three validation methods are integrated in the on demand processing system and are optionally selectable. The most basic validation method checks if measurement values are within a predefined range of possible values. For example, it may be assumed that an air temperature sensor measures values within a range of -40 °C to +60 °C. Values outside of this range are considered as a measurement error by this validation method and consequently rejected. An other validation method checks for outliers in the stream of measurement values by defining a maximum change rate between subsequent measurement values. The third validation method compares measurement data to the average values of neighboring stations and rejects measurement values with a high variance. These quality checks are optional, because especially extreme climatic values may be valid but rejected by some quality check method.

An other important task is the preparation of measurement data in terms of time. The observed stations measure values in intervals of minutes to hours. Often scientists need a coarser temporal resolution (days, months, years). Therefore, the interval of time aggregation is selectable for the processing.

For some use cases it is desirable that the resulting time series are as continuous as possible. To meet these requirements, the processing system includes techniques to fill gaps of missing values by interpolating measurement values with data from adjacent stations using available contemporaneous measurements from the respective stations as training datasets.

Alongside processing of sensor values, we created interactive visualization techniques to get a quick overview of a big amount of archived time series data.