EGU22-2012
https://doi.org/10.5194/egusphere-egu22-2012
EGU General Assembly 2022

Gap filling in air temperature series by matrix completion methods

Benoît Loucheur[1], Pierre-Antoine Absil[1], and Michel Journée[2]
[1]Université Catholique de Louvain, ICTEAM, INMA, Belgium (benoit.loucheur@uclouvain.be)
[2]IRM-KMI, Belgium

Quality control of meteorological data is an important part of atmospheric analysis and prediction, as missing or erroneous data can have a negative impact on the accuracy of these environmental products.

In Belgium, the Royal Meteorological Institute (RMI) is the national meteorological service that provide weather and climate services based on observations and scientific research. RMI collects and archives meteorological observations in Belgium since the 19th century. Currently, air temperature is monitored in Belgium in about 30 synoptic automatic weather stations (AWS) as well as in 110 manual climatological stations. In the latter stations, a volunteer observer records every morning at 8 o'clock the daily extreme air temperatures. All observations are routinely checked for errors, inconsistencies and missing values by the RMI staff. Misleading data are corrected and gaps are filled by estimations. This quality control tasks require a lot of human intervention. With the forthcoming deployment of low-cost weather stations and the subsequent increase in the volume of data to verify, the process of data quality control and completion should become as automated as much as possible.

In this work, the quality control process is fully automated by using mathematical tools. We present low-rank matrix completion methods (LRMC) that we used to solve the problem of completing missing data in daily minimum and maximum temperature series. We used a machine learning technique called Monte Carlo cross-validation to train our algorithms and then test them in a real case.

Among the matrix completion methods, some are regularised by graphs. In our case, it is then possible to represent the spatial and temporal component via graphs. By manipulating the construction of these graphs, we hope to improve the completion results. We were then able to compare our methods with what is done in the state of the art, such as the inverse distance weighting (IDW) method.

All our experiments were performed with a dataset provided by the RMI, including daily minimum and maximum temperature measurements from 100 stations over the period 2005-2019.