Geophysical Research Abstracts
Vol. 13, EGU2011-1525-1, 2011
EGU General Assembly 2011

A Dynamic Method for Gap Filling in Long Term Temperature Series

Gianmarco Tardivo and Antonio Berti
Padova, Environmental Agronomy and Crop Science, Legnaro (Pd), Italy (gian0812@gmail.com, +39 0498272839)

The reconstruction of missing weather data can be done using both data collected before the gap and those collected after the gap. Still, the dependencies between the missing and available weather variables are expected to be complex, and advanced data analysis tools are needed to find and to express these dependencies with sufficient accuracy.
This paper proposes a variant a multiple regression model which matches closely with the potential of numerical processing available today, for the reconstruction of daily temperature data.
The data used in this study are acquired by a relatively dense network of meteorological stations (114 stations), distributed across the Veneto Region (North-eastern Italy). Data from 01 January 1993 to December 31, 2008 were considered, for a total of 5'844 days. Three variables were considered (maximum, minimum and mean temperature), for a total of 1'998'648 data. There are 4'440 intervals of missing data, corresponding to a total of 18'063 missing days.
98% of missing intervals does not exceed 20 days duration (with a single maximum of 213 days); and 86% does not exceed 5 days. The relative shortness of the missing intervals is a characteristic that allows a data reconstruction by whole intervals and not day by day.
In most cases, before or after gaps, there are long periods of existing data that are used in the method for cross-validation testing. The method identify the stations used for data reconstruction considering a maximum of 10 stations with available data, among them a set of max 4 stations is selected. The final set of stations is selected on the basis of their correlation with the target station, performing the research both before and after the gap.
The length of the period used for the final regressions varies between a minimum and a maximum that can be set by the user, but the final length is automatically selected for each reconstruction in order to minimise the absolute mean error during the reconstruction. A specific reconstruction error is then estimated simulating reconstructions in the neighbourhood of each gap.
The main trait of the method is then to select the station used for gap filling in a dynamic way, identifying for each gap a specific set of stations, an optimised length of the dataset used and a coupling period in the past or in the future of the gap.
The method proved to be highly accurate and reliable, despite the morphological and altimetric differences in the considered network of meteorological stations.
The median error was of 0.006 °C and the 95th percentile of the absolute mean error did not exceeded 0.111 °C. Furthermore, the number of temperature inversions (i.e. Tmin> Tmed; Tmax<Tmed) was almost negligible, with 1 inversion over more of 6'000 data reconstructions.