

Daily quality control of climatic variables involved in SPEI drought index calculation

Miquel Tomas-Burguera (1), Sergio M. Vicente-Serrano (2), Santiago Begueria (1), Yolanda Luna (3), Ana Morata (3), Azucena Jiménez (4), and Jose Carlos González-Hidalgo (4)

(1) Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas (EEAD-CSIC), Zaragoza, Spain, (2) Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas (IPE-CSIC), Zaragoza, Spain, (3) Agencia Estatal de Meteorología (AEMET), Madrid, Spain, (4) Departament of Geography, University of Zaragoza, Zaragoza, Spain

In the framework of the DESEMON (Development of drought indices for sectoral applications: improvement of drought monitoring and early warning in Spain) project, a quality control procedure has been implemented in R Software to detect suspicious daily data of 6 climatic variables (maximum temperature, minimum temperature, relative humidity, wind speed, sunshine duration and rainfall) obtained from the historical data base of the Spanish Meteorological Agency (AEMET: Agencia Estatal de METerología). These variables were selected due to their importance in the Standardised Precipitation-Evapotranspiration Index (SPEI) calculation.

The quality control procedure was focused in the detection of two types of suspicious data: i) codification errors and ii) anomalous values. A final step of spatial coherence was tested but not implemented due to its high percentage of false negatives. The coherence between variables is only tested for maximum and minimum temperature. While the general scheme of the implemented controls is the same for all the variables, some specific controls and thresholds exist for each variable. In the control of anomalous values, we classified the errors as: i) values out of the range of the variable, ii) extreme values considered not feasible in Spain and iii) extreme values considered as suspicious values. Moreover, three kinds of codification errors were found in the database: i) exact values appearing at distinct points of the database during a minimum number of days (duplicated data); ii) the same value appearing in "n" consecutive days, depending the threshold "n" on the variable (consecutive values); and iii) values codified using non-standard units for the variable according to AEMET metadata (erroneous units). After the quality control, the results showed that the procedure of detection of codification errors has to be implemented in the whole climate database, due to the detection of duplicated data between distant points, temporally and spatially speaking. While the temporal analysis of detected errors clearly show some specific moments when errors are more present in the database (apparently related with digitalization processes), the spatial analysis does not show a clear pattern.