On the visualization of water-related big data: extracting insights from drought proxies' datasets

Vitali Diaz (1,2), Gerald Corzo (1), Henny A.J. van Lanen (3), Dimitri Solomatine (1,2)

(1) UNESCO-IHE Institute for Water Education, Hydroinformatics chair group, Delft, the Netherlands (v.diazmercado@unesco-ihe.org), (2) Delft University of Technology, Water Resources section, Delft, the Netherlands, (3) Hydrology and Quantitative Water Management Group, Centre for Water and Climate, Wageningen University, Wageningen, Droevendaalsesteeg 3a, 6708 PB, the Netherlands

Big data is a growing area of science where hydroinformatics can benefit largely. There have been a number of important developments in the area of data science aimed at analysis of large datasets. Such datasets related to water include measurements, simulations, reanalysis, scenario analyses and proxies. By convention, information contained in these databases is referred to a specific time and a space (i.e. longitude/latitude).

This work is motivated by the need to extract insights from large water-related datasets, i.e. transforming large amounts of data into useful information that helps to better understand of water-related phenomena, particularly about drought.

In this context, data visualization, part of data science, involves techniques to create and to communicate data by encoding it as visual graphical objects. They may help to better understand data and detect trends. Base on existing methods of data analysis and visualization, this work aims to develop tools for visualizing water-related large datasets.

These tools were developed taking advantage of existing libraries for data visualization into a group of graphs which include both polar area diagrams (PADs) and radar charts (RDs). In both graphs, time steps are represented by the polar angles and the percentages of area in drought by the radios.

For illustration, three large datasets of drought proxies are chosen to identify trends, prone areas and spatio-temporal variability of drought in a set of case studies. The datasets are (1) SPI-TS2p1 (1901-2002, 11.7 GB), (2) SPI-PRECL0p5 (1948-2016, 7.91 GB) and (3) SPEI-baseV2.3 (1901-2013, 15.3 GB). All of them are on a monthly basis and with a spatial resolution of 0.5 degrees. First two were retrieved from the repository of the International Research Institute for Climate and Society (IRI). They are included into the Analyses Standardized Precipitation Index (SPI) project (iridl.ldeo.columbia.edu/SOURCES/.IRI/.Analyses/.SPI/). The third dataset was recovered from the Standardized Precipitation Evaporation Index (SPEI) Monitor (digital.csic.es/handle/10261/128892).

PADs were found suitable to identify the spatio-temporal variability and prone areas of drought. Drought trends were visually detected by using both PADs and RDs. A similar approach can be followed to include other types of graphs to deal with the analysis of water-related big data.

Key words: Big data, data visualization, drought, SPI, SPEI