Geophysical Research Abstracts
Vol. 19, EGU2017-18343, 2017
EGU General Assembly 2017

Using R for analysing spatio-temporal datasets: a satellite-based precipitation case study

Mauricio Zambrano-Bigiarini (1,2)

(1) Universidad de La Frontera, Faculty of Engineering and Sciences, Civil Engineering, Temuco, Chile
(mauricio.zambrano@ufrontera.cl), (2) Center for Climate and Resilience Research, Universidad de Chile, Santiago, Chile

Increasing computer power and the availability of remote-sensing data measuring different environmental variables has led to unprecedented opportunities for Earth sciences in recent decades. However, dealing with hundred or thousands of files, usually in different vectorial and raster formats and measured with different temporal frequencies, impose high computation challenges to take full advantage of all the available data. `R` is *a language and environment for statistical computing and graphics* which includes several functions for data manipulation, calculation and graphical display, which are particularly well suited for Earth sciences.

In this work I describe how `R` was used to exhaustively evaluate seven state-of-the-art satellite-based rainfall estimates (SRE) products (TMPA 3B42v7, CHIRPSv2, CMORPH, PERSIANN-CDR, PERSIAN-CCS-adj, MSWEPv1.1 and PGFv3) over the complex topography and diverse climatic gradients of Chile. First, built-in functions were used to automatically download the satellite-images in different raster formats and spatial resolutions and to clip them into the Chilean spatial extent if necessary. Second, the `raster` package was used to read, plot, and conduct an exploratory data analysis in selected files of each SRE product, in order to detect unexpected problems (rotated spatial domains, order or variables in NetCDF files, etc). Third, `raster` was used along with the `hydroTSM` package to aggregate SRE files into different temporal scales (daily, monthly, seasonal, annual). Finally, the `hydroTSM` and `hydroGOF` packages were used to carry out a point-to-pixel comparison between precipitation time series measured at 366 stations and the corresponding grid cell of each SRE. The modified Kling-Gupta index of model performance was used to identify possible sources of systematic errors in each SRE, while five categorical indices (PC, POD, FAR, ETS, fBIAS) were used to assess the ability of each SRE to correctly identify different precipitation intensities.

In the end, `R` proved to be and efficient environment to deal with thousands of raster, vectorial and time series files, with different spatial and temporal resolutions and spatial reference systems. In addition, the use of well-documented `R` scripts made code readable and re-usable, facilitating reproducible research which is essential to build trust in stakeholders and scientific community.