# Certainties and Uncertainties in Observational and Reanalysis Data in the Studies of Extreme Precipitation Events Over La Plata Basin

José Leandro Campos
Tércio Ambrizzi

Institute of Astronomy, Geophysics and Atmospheric Sciences – University of São Paulo

Currently many scientific papers have been using reanalysis data produced by weather and climate prediction centers and regular gridded interpolated data derived from surface stations. Due to these wide use and lack of reliable and easy access to time series, is advisable to check the capacity of these data in represent climate extremes of precipitation. The aim of this study is to analyze the existing deviations in these data sets in relation to "real" station data, when considering extreme precipitation events.

Six data sets of daily precipitation were used in this work: the NCEP/NCAR and ECMWF/ERA40 reanalysis; three interpolated gridded data, the CU/CIRES with resolution of 1.0° and 2.5°, the CPC with resolution of 0.5°; and data from 108 meteorological stations along the La Plata basin from CLARIS-LPB project (A Europe South America Network For Climate Change Assessment and Impact Studies in La Plata Basin), in the period of 1979-2002, where all the data sets were compared with the CLARIS-LPB data (hereafter referred to as observational data). Some specific indices, such as, the prcptot (total precipitation in wet days in one year), Rx1day (monthly maximum daily precipitation), SDII (simple monthly intensity precipitation index) and CWD (yearly number of consecutive wet days) were calculated and a statistical analysis (correlations, scatter plots and cumulative density functions) were also applied in the data sets.

The correlations results indicated low correlations (around 0.4) between CLARIS-LPB and NCEP/NCAR and ECMWF/ERA40, high correlation (around 0.8) for regular interpolated gridded data between CLARIS-LPB, CPC and CU/CIRES 1.0° and a correlation of 0.6 between CLARIS-LPB and CU/CIRES 2.5°. The correlation differences are probably related to the interpolation process between the interpolated gridded data and the reanalysis which do not well represent the daily precipitation.

To verify how well the reanalysis and interpolated gridded data represent the precipitation extreme, a cumulative probability density function (CDF) was calculated for the six data sets where it is verified the distance among the curves and their "rate of growth". It was found that the greatest rate of growth occurred for the ECMWF/ERA40 and CU/CIRES 2.5° and consequently they had the largest distance from the CLARIS-LPB data curve. On the other hand, CPC, CU/CIRES 1° and NCEP/NCAR data were closer to the observational data. This result suggest that the first two data sets, (ECMWF/ERA40 and CU/CIRES) have more days with light rain than the others, thus the extreme daily events in these analysis are smoothed and underestimated when compared with the CLARIS-LPB for some selected stations.

The analysis from the indices Rx1day and SDII showed that the reanalysis and regular interpolated gridded data are smaller than those obtained using station data. The better data set considering these two indices was the CPC data. When the total precipitation in wet days in one year is considered (prcptot index) it was found that the reanalysis and interpolated gridded data tend to overestimate the precipitation amount compared to the observation. The result suggests that though these data sets underestimate extreme rainfall events, the precipitation accumulated along the year, even if it is light rain, is overestimated and it does not well represent the observational data. For the CWD index the result similar to that found for the prcptot i.e., the number of consecutive wet days in reanalysis and interpolated gridded data is greater than in the CLARIS-LPB data, meaning that there are more days with precipitation, which agree with the bigger volume accumulated in one year. However, it is clear that this dataset also misrepresents the total monthly or daily rainfall amount or extremes.

The present results suggest that the effects of interpolation and resolution in interpolated gridded data as well as the poor representation of daily precipitation in reanalysis data may have a significant impact on the analysis of extreme precipitation based on climate indices. It was found that the monthly indices (Rx1day and SDII) are generally underestimated and the yearly indices (prcptot and CWD) are overestimated when compared to station data.

**Corresponding Author:**

**Name:** José Leandro Campos and/or Tércio Ambrizzi
**Organization:** Department of Atmospheric Sciences - University of São Paulo
**Address:** Rua do Matão, 1226
São Paulo, 05508-090
Brazil