Fiorella Acquaotta fiorella.acquaotta@unito.it
Diego Guenzi diego.guenzi@unito.it
Diego Garzena diego.garzena@unito.it
Simona Fratianni simona.fratianni@unito.it

University of Turin, department of Earth Sciences

Co.Rain: a new semi-automatic procedure for a better assessment of parallel precipitation measurements networks

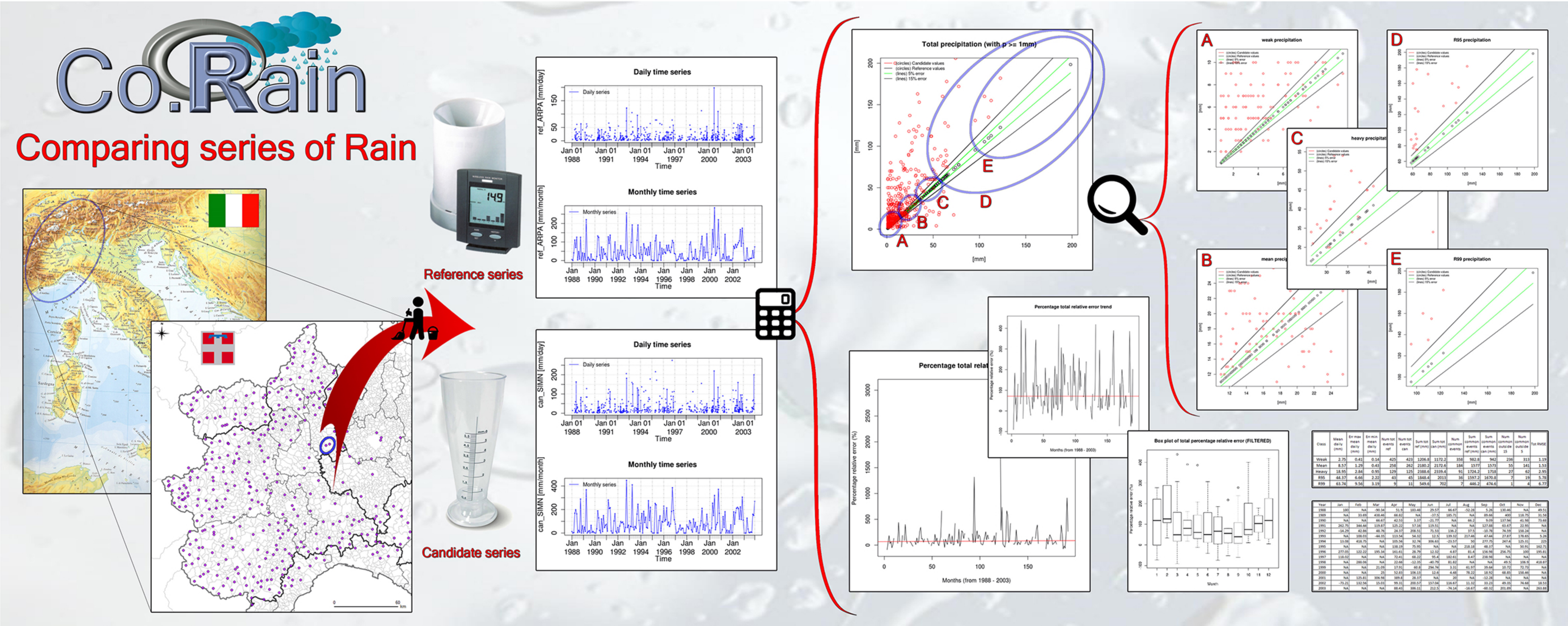


FIGURE 1: overview of the main steps and features of Co.Rain program - cleaning of data, comparison of rain series and organization in classes

Introduction

The study of long-term climate datasets is essential for different hydrological applications related to water resource planning, power production, irrigation and flood control, in addition to the climate changes. However, all these applications require series of high quality of data. Long historical climate records, usually, contain non-climatic changes that can influence the observed behavior of meteorological variables. The availability of parallel measurements offers an ideal occasion to study these discontinuities. The World Meteorological Organization (WMO), and in particular the Commission for Instruments and Methods of Observation (CIMO), have recognized the need to conduct a series of intercomparison of instruments in order to highlight and classify these discontinuities in precipitation recordings. The instrumental transition from manual to automatic measurements, typical of some regions of the world (among which we have Italy), produce differences and a spurious change in the precipitation, thus showing the importance of having a homogeneous dataset to identify real climate variations.

Software description

For this purpose, we adopted a new methodology (Acquaotta et al. 2016), implemented in the free and open source software called Co.Rain, which is written in the multi-platform R language and is available online under GPL license. The program takes as input a plain text file with two series of daily information on precipitation, normally obtained from different networks and starts its elaboration, which is divided in three steps. All outputs produced by each of the software's steps are either CSV or PNG files.

Step 1

In the first step, the program cleans the input data, removing all values smaller than 1 mm/day and setting the same missing values on both series; after that, it computes a statistical analysis on the two series, asking the user's agreement to proceed (since the cleaning process could be too aggressive and remove significant portions of data).

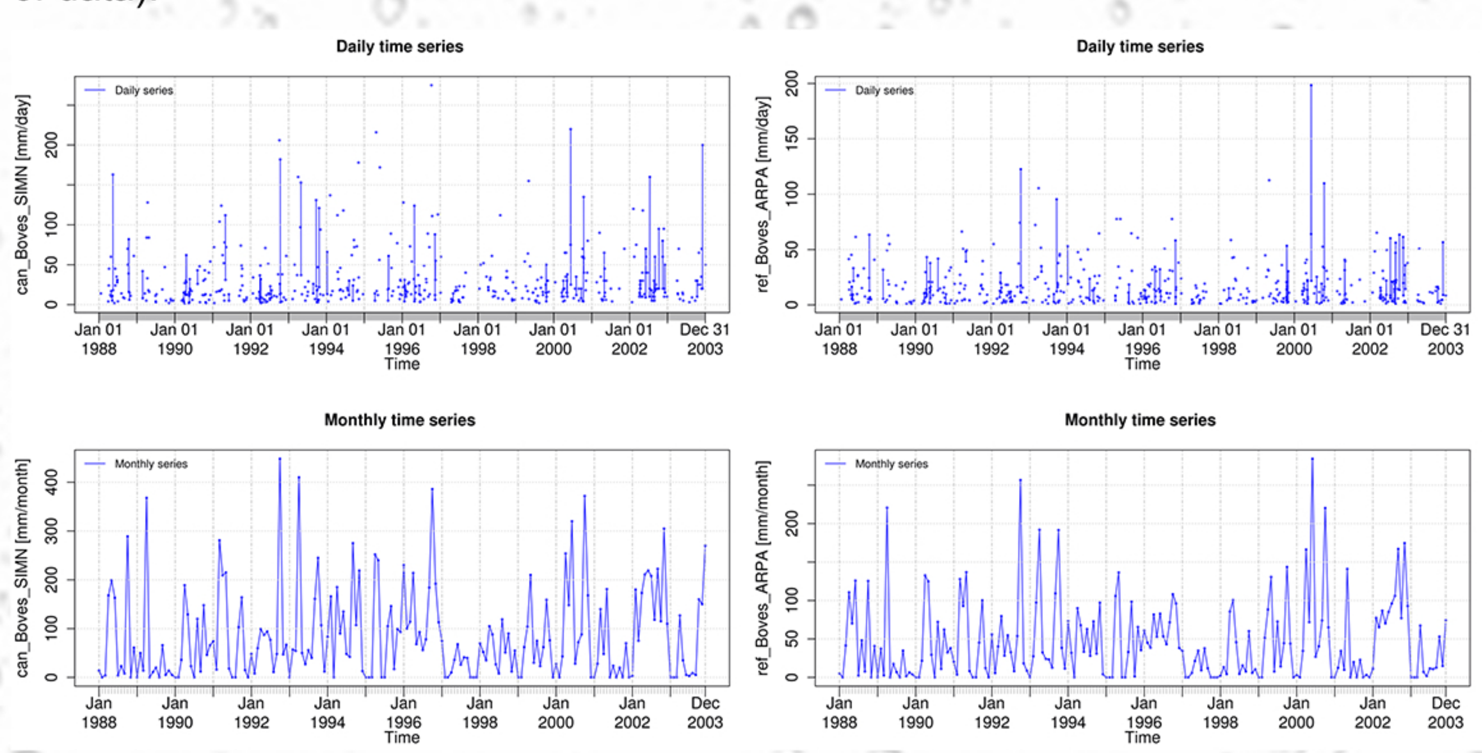


FIGURE 2: the daily and monthly behavior of two parallel series of Boves, in the south of Piedmont. One station (called SIMN and being the candidate series obtained from the old manual station) is located at 590 m a.s.l and measures an average of 1322.9 mm of rain per year; the other one (called ARPA and being the reference series obtained from the new tipping-bucket automated station) is located near the previous one but at 575 m a.s.l. and measures an average of 1106.9 mm/year. The two stations are distant 1240 m and have an overlapping period ranging from 1988 to 2003

Step 2

The second step is the comparison between the two cleaned series, where different statistical tests are applied (Student's t-test, Kolmogorov-Smirmov test, Wilcoxon rank sum test, Kruskal-Wallis test, correlation coefficient computed using Spearman's method, root mean square error...), generating a scatter plot of daily values, calculating the minimum and maximum errors and other plots that help in the data visualization.

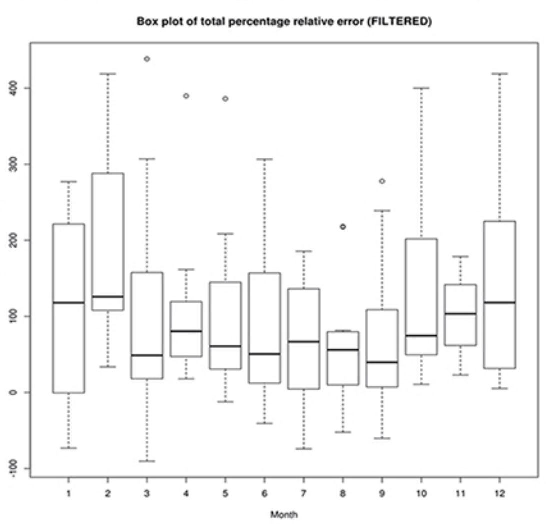


FIGURE 3: box plot of total percentage relative error grouped by month, using the Boves series previously described

Step 3

Class name	Range (percentile)
Weak rain (A)	R < 50th
Medium rain (B)	50th ≤ R < 80th
Heavy rain (C)	80th ≤ R ≤ 95th
Very heavy rain (D)	R > 95th
Extreme rain (E)	R > 99th

TABLE 1: names and ranges of the five precipitation classes. The thresholds were calculated from the percentiles estimated using the reference series

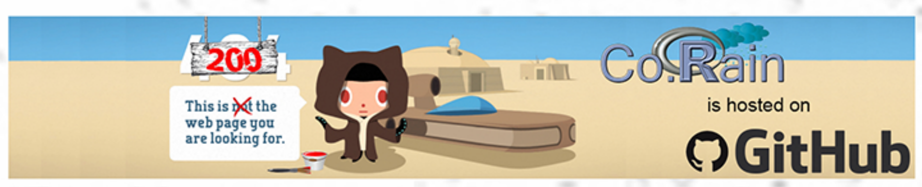
In the final step, the software calculates the percentiles of the series to identify the thresholds for the classification of the events. Every rainy day is classified as weak (A), medium (B), heavy (C), very heavy (D) or extreme (E). For each class, it creates a scatter plot and a summary of useful information like thresholds, mean value of maximum and minimum errors, number of events analyzed for each series, precipitation amounts, number of events recorded in the same day, number of events outside the range for maximum and minimum error, root mean square error and others.

Results

The code has been tested on different rain series in the Piedmont region (in the northwestern part of Italy). That region is perfect for running our software because it counts more than 350 active tipping-bucket rain gauges in a 25.000 Km^2 territory, all managed by the local agency for environmental protection (ARPA). Beside this, Piedmont has many long daily precipitation series available and its database contains parallel measurement from 1987 to 2003; meteorological stations are very close and the mean overlapping period is 12 years. Our study showed its importance in identifying climate variations and instrumentation errors. In addition, this procedure allowed us to discern the locations with or without discrepancies and evaluated the correctness of joining consecutive time data series obtained from different neighboring instruments.

References

For details on methodology, see *Acquaotta F., Fratianni S., Venema V. (2016) Assessment of parallel precipitation measurements networks in Piedmont, Italy. International Journal of Climatology* (doi: 10.1002/joc.4606)



To get the free source code of the software, scan the QR code above or browse https://github.com/UniToDSTGruppoClima/CoRain (doi: 10.5281/zenodo58478)