Environmental Modelling & Software 97 (2017) 243–258

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A rainfall analysis and forecasting tool

Javier Diez-Sierra, Manuel del Jesus*

Environmental Hydraulics Institute, Universidad de Cantabria - Avda. Isabel Torres, 15, Parque Científico y Tecnológico de Cantabria, 39011, Santander, Spain

ARTICLE INFO

Article history:
Received 24 February 2017
Received in revised form
19 July 2017
Accepted 23 August 2017

Keywords:
Rainfall modeling
Weather types
Spatial analysis
Kriging
Rainfall extremes

ABSTRACT

MENSEI-L is a stand-alone software tool for the automatic analysis of pluviometric networks, that also provides three-day rainfall forecasts based on weather types. The software tool, implemented in Python and R, is able to fill missing values in original daily data series and to generate synthetic pluviometers in ungauged locations, by means of kriging techniques. MENSEI-L also characterizes punctual and spatial, average and extreme distributions of precipitation for the complete pluviometric network. Tenerife (Canary Islands, Spain) is used as study site to evaluate the capabilities of MENSEI-L and the implicit rainfall analysis methodology that it implements. MENSEI-L proves to be a useful tool to extract information from dense observation networks where manual analysis is not practical.

1. Introduction

Rainfall is the main driver of the continental portion of the water cycle below parallel 50°. Rainfall characterization is important for an accurate modeling of other hydrological fluxes such as runoff and evapotranspiration, as rainfall controls their statistical behavior, both average and extreme distributions. Even more, in small arid and semiarid watersheds, where runoff and discharge measurements may not be practical, an accurate spatio-temporal rainfall characterization becomes an essential part of the assessment of hydrological fluxes. These fluxes may be required for infrastructure design or risk assessments, effectively turning rainfall into an important part of the design and the decision-making processes (Xu et al., 2010; Paixao et al., 2011; Arnbjerg-Nielsen et al., 2013; Renschler et al., 1999; Sugiyama et al., 1995). Furthermore, the incorporation of quantitative precipitation forecasting (QPF) may also be required for many hydrological applications especially for real-time flood forecasting, as it is a critical component of public safety and quality of life. The potential risk associated to flash floods is related to two main characteristics: their rapid occurrence and to the spatial dispersion of the areas affected by these floods. Both characteristics limit the ability to issue timely flood warnings (Borga et al., 2011) turning flash floods into one of the most dangerous natural hazards (Blöschl et al., 2008).

Rainfall is measured by pluviometers, which only capture rainfall data for a very small portion of the territory, and thus, tend to be deployed in networks to cover larger areas. On top of collecting rainfall data for a large area, pluviometric networks provide useful information for the spatio-temporal characterization of rainfall and are thus important to inform engineering design and decision-making processes. In addition, the availability of dense pluviometric networks offers the opportunity to build short-term stochastic rainfall forecasting systems making predictions that cover the entire area of interest. Rainfall data collection and treatment, however, presents a series of complications. Measurements for specific events may be missing, due to equipment malfunctions or other reasons, creating discontinuities in the time series. Equipment upgrades may lead to differences in time resolution for a given gauge for different time periods. Newer gauges, with better quality data, provide shorter time series, limiting their usability. Data collection from pluviometric networks leads to generating large databases from which manually extracting and interpreting information is complicated and time consuming. Moreover, the implementation of statistical short-term forecasting systems often leads to deal with sophisticated techniques and large atmospheric databases. For these reasons, statistical tools, machine learning techniques and software applications become indispensable to convert rainfall and atmospheric data into information, ready to support professional practitioners and managers.

* Corresponding author.
E-mail address: manuel.deljesus@unican.es (M. del Jesus).

http://dx.doi.org/10.1016/j.envsoft.2017.08.011
1364-8152/

Nowadays, different software applications exist to assist technicians in the process of rainfall analysis. STORMS 2010 (Sabourin and Associates Inc., 2010) is a commercial application able to analyze local rainfall information to fit average and extreme distributions. It computes Intensity-Duration-Frequency (IDF) curves, performs intensity trend analysis and is able to create design storms. CLIMATOL (Guijarro, 2016) is a time series homogenization software distributed as an R (R Core Team, 2015) package under an open source license. It was designed to deal with general climatological time series, not only with rainfall. Similar software applications are MASH (Szentimrey, 2006), SPLIDHOM (Mestre et al., 2011) and EXTRAQC (Aguilar and Prohom, 2011), tools designed for homogenizing and checking the internal coherence of daily time series, and AnCli (Stepanek, 2005), which also provides statistical characterization tools.

Software libraries and packages have also been developed for the treatment of rainfall data. One interesting example is *reddPrec* (Serrano-Notivoli et al., 2017), an R package that contains a set of functions for controlling the quality of rainfall time series, filling missing values and creating virtual series (and gridded datasets) in ungauged locations. Other softwares such as RainSim (Burton et al., 2008) or MRS (Mehrotra et al., 2015a) are able to simulate stochastic rainfall at multiple locations preserving the most important statistics (Burton et al., 2013; Kretzschmar et al., 2014; Mehrotra et al., 2015b).

To the best of the author's knowledge, there are not too many tools specifically developed to perform short-term rainfall forecasting (Neiman et al., 2009; Wright et al., 2017). In most cases rainfall forecasting is dealt with by means of numerical weather prediction models or the extrapolation of remote sensing observations (radar data and satellite images) to the current weather conditions. Unfortunately, while the first do not seem able to provide accurate rainfall forecasts at the temporal and spatial resolution required by many hydrologic applications, neither the outputs from satellite or radar images allow a satisfactory assessment of rainfall intensity yet (Toth et al., 2000; Krzysztofowicz, 1995; Liu et al., 2013; Moser et al., 2015).

All the reviewed applications and techniques serve to characterize, homogenize and interpolate punctual rainfall time series, to generate stochastic rainfall series or to forecast rainfall. However, no software package provides an integrated methodology to automatically analyze the spatio-temporal characteristics of rainfall for a pluviometric network and generate short-term rainfall forecasting. This paper presents such methodology, implemented into an automated rainfall analysis software tool called MENSEI-L, that allows to generate point and spatial characterizations of average and extreme distributions of rainfall, as well as to make short-term stochastic rainfall forecasts using an analogs method based on weather typing. An application to Tenerife (Canary islands, Spain) is used as a calibration and validation step for the methodology developed in the present work.

In order to make the exposition clearer and easier to follow, the theoretical basis of each section will be presented first and the specific implementation for the Tenerife study case will follow. For this reason, the paper begins with a description of the study area and the databases used in the study. Then, a description of the base methods used to carry out the rainfall analysis is presented. These methods are grouped in three categories: the generation of weather types, the characterization of point statistics of rainfall, and the implementation of a spatio-temporal interpolation scheme based on geostatistics. The paper continues presenting the work flow followed to carry out the rainfall analysis, explaining how the basic methods are used to construct robust rainfall analysis tools. Afterwards, the scheme of the stochastic rainfall forecasting system is described. The main body of the paper ends with a description of the implementation details of MENSEI-L, to finally close with future developments, limitations and the conclusions of the work.

2. Description of the study area and the information sources

2.1. Description of the study area

Tenerife is one of the seven islands that form the Spanish archipelago of the Canary Islands. Tenerife (see Fig. 1) is located in the Atlantic Ocean, 300 Km west of the African coast, between parallels $28^\circ N$ and $29^\circ N$, and between meridians $16^\circ W$ and $17^\circ W$. Tenerife covers a surface area of 2034.38 Km^2, with a population of $891,111$ inhabitants. It is the largest, and the most populated of the Canary Islands and of the whole Macaronesia region.

Tenerife displays a strong topographic gradient from the coast to the central part of the island. The central region of the island is a mountain range, where *El Teide* (3.718 m) is the highest peak, that separates the island in two well differentiated climatic areas: a northern region, relatively wet due to the effect of the Trade winds, and a southern region, drier due to the blocking induced by the orography. Spatial heterogeneities of climate are also present within each of the regions, configuring a widely varied range of climates within a relatively small area.

The precipitation in Tenerife is very scarce, with an annual average of 233 mm for the whole island. However, its temporal and spatial distribution is very heterogeneous since a very high percentage of precipitation falls in short periods of time and with a great spatial inhomogeneity. The singular orographic and climatic characteristics of the island combined with the urbanization processes carried out in the last years have transformed flash floods into a serious problem. If fact, in the twentieth century, almost one hundred floods were recorded in Tenerife causing important losses both materials and humans.

2.2. Description of data sources

Information regarding synoptic patterns of atmospheric variables is obtained from two sources. On the one hand, historical information is obtained from the global reanalysis NCEP Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010), together with its continuation, the Climate Forecast System Version 2 (CFSV2), developed by the National Oceanic and Atmospheric Administration (NOAA). In this paper, the term CFSR will be used to refer both databases, CFSR and CFSV2. CFSR is a third-generation reanalysis database, generated with a global high-resolution land-ocean-atmosphere model coupled with an ice sheet model. CFSR provides 6-hourly data. The spatial resolution of the atmospheric model is 38 Km horizontally, with 64 vertical levels. The spatial resolution of the oceanic model is approximately 0.25° near the Equator, and 0.5° beyond the Tropics, with 40 vertical levels. The land model counts with 4 soil levels, and the ice model with 3 levels. CFSR and CFSv2 together cover the period 1979–2015.

On the other hand, forecasting information is obtained from the Global Forecast System (GFS), also developed by NOAA. GFS provides 16 days-into-the-future forecasts with a model similar to the one used to generate CFSR.

Rainfall data is provided by *Consejo Insular de Agua de Tenerife (CIATF)*, the water planning and managing agency for Tenerife Island. CIATF maintains a database of rainfall observations (Melián et al., 2011) that includes the observation network of *Agencia Estatal de Meteorología (AEMET)*, the Spanish national meteorological agency, and *AgroCabildo*, a local agency for agriculture development. The CIATF database contains information for 125 gauges, combining daily and sub-daily values, with an average coverage of 15 years, although the longest series start in 1890.