Journal of Hydrology: Regional Studies xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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

Joint modelling of drought characteristics derived from historical and synthetic rainfalls: Application of Generalized Linear Models and Copulas

Fatih Tosunoğlu[a,*], Christian Onof[b]

[a] Department of Civil Engineering, Erzurum Technical University, Erzurum, Turkey
[b] Department of Civil and Environmental Engineering, Imperial College London, United Kingdom

ARTICLE INFO

Keywords:
Drought
Standard precipitation index (SPI)
General linear models (GLMs)
Copulas
Çoruh basin
Turkey

ABSTRACT

Study region: Çoruh Basin in Northeastern Turkey.
Study focus: In recent years, copulas have been widely used to model the joint distribution function of duration and severity series which are the major characteristics of a drought event to be considered in the planning and management of water resources systems. However, as the copula functions are typically fitted to the drought series that are derived from a limited amount of observed data, it may be insufficient to characterize the full range of the analyzed drought characteristics. Therefore, General Linear Models (GLMs) were used to model and simulate rainfall data in this study. The Standard Precipitation Index (SPI) method was used to obtain the drought characteristics from simulated and historical rainfall series. Four Archimedean copulas, namely Ali-Mikhail-Haq, Clayton, Frank and Gumbel-Hougaard, were evaluated to model the joint distribution functions of these characteristics.
New hydrological insights for the region: The Gumbel-Hougaard copula was found to be the most suitable copula in modelling the joint dependence structure of the drought characteristics at five stations in the basin. The derived Gumbel-Hougaard copulas for each station were employed to obtain joint and conditional return periods of the historical and generated drought characteristics. The drought risks that are estimated based on bivariate return periods for different circumstances can provide useful information in planning, management and in assessing adequacy of the water structures in the basin.

1. Introduction

A drought is generally considered as a natural phenomenon that occurs when rainfall is significantly lower than the normal value. To identify when a drought starts and terminates is always an important task. Therefore, the drought status is usually defined using various indicators according to the purpose of the analysis (Mishra and Singh, 2011). Among these drought indicators, the Standardized Precipitation Index (SPI) is a widely used indicator to characterize meteorological droughts in any region (McKee et al., 1993). The SPI method has many advantages that have been comprehensively described by researchers over the world (Bonaccorso et al., 2003; Sirdas and Sen, 2003; Vicente-Serrano, 2006; Bacanli et al., 2009; Keskin et al., 2009; Kumar et al., 2009; Zhang et al., 2012b; Liu et al., 2013). Utilizing the SPI series, a drought event is defined by means of the theory of runs that was introduced by Yevjevich (1967). According to this theory, a threshold level is selected to determine drought events which are mainly characterized

* Corresponding author.
E-mail address: ftosunoglu@erzurum.edu.tr (F. Tosunoğlu).

https://doi.org/10.1016/j.ejrh.2017.11.001
Received 4 July 2017; Received in revised form 5 October 2017; Accepted 1 November 2017

by duration and severity. As the drought characteristics are random variables, they are evaluated and modelled with probability theories (Sarlak et al., 2009). The drought and severity series are also mutually correlated variables and the proper way to analyze droughts is by means of multivariate approaches. Hence, researchers have used traditional multivariate distributions to derive the joint distribution of drought characteristics (e.g. Nadarajah, 2009). However, the traditional multivariate distributions generally have various constraints. For instance, the marginal distribution of the drought characteristics must come from the same distribution type. These difficulties have been recently overcome by using copulas. A copula is a function that compounds various univariate cumulative distribution functions (CDFs) into their joint CDF (Graler et al., 2013; Grimaldi et al., 2016). The main advantages of copulas are that they are capable of developing the multivariate distribution in simple terms regardless of the univariate distribution type and preserving the dependence structure among variables. Hence, copulas have gained wide popularity in multivariate drought modelling and successful results have been obtained in real case studies (Shiau and Modarres, 2009; Liu et al., 2011; Ganguli and Reddy, 2012; Chen et al., 2013; Zhang et al., 2013; Rauf and Zeephongsekul, 2014; Liu et al., 2016; Tosunoglu and Can, 2016 and many others). In most previous studies, copula functions were fitted to drought characteristics derived from historical records which are usually inadequate in length. To obtain more reliable drought risk assessment, a sufficient sample of data must be provided (Yoo et al., 2013). To overcome this issue, stochastic based models are frequently applied to generate synthetic data series that preserve the statistical properties of historical data at one or more locations. Thanks to the generated series, a set of alternative drought events that may occur in the future can be projected, statistical behavior of the drought characteristics can be comprehensively analyzed and their usefulness for risk and reliability assessment in the management and design of water resources systems can be evaluated.

The overall aim of this study is to obtain joint and conditional return periods of drought duration and severity. To achieve this aim, the following objectives are defined: (1) to generate synthetic rainfall series that will be served as a basis for drought analyzing by means of the Standardized Precipitation Index (SPI), which is able to produce a set of realizations of possible drought characteristics (duration and severity). To achieve this aim, General Linear Models (GLMs) are used to model daily rainfall series of five rainfall stations in Çoruh basin, Turkey and they are then applied to generate 100synthetic rainfall series each being the same length as historic time series; (2) to identify marginal distribution for drought duration and severity series; (3) to identify the most suitable copula function for joint modelling of drought duration and severity series. Archimedean copulas, including Ali-Mikhail-Haq, Clayton, Frank and Gumbel-Hougaard, are evaluated for this purpose.

2. Methods

2.1. General linear models (GLMs) for modelling daily rainfall at a single site

Generalized Linear Models (GLMs) are one of the models that have been widely used in rainfall modelling and generation (e.g.Chandler and Wheater, 2002; Yang et al., 2006; Kenabatho et al., 2012; Chun et al., 2013). The main advantages of GLMs are the ability to incorporate autocorrelation, seasonality, location, as well as the interaction of such properties, and the inclusionof external drivers such as atmospheric circulation indices. During the application of GLMs to simulate daily rainfall, a probability distribution is predicted for rainfall of each day at every site of interest, by relating the distribution to the values of various predictor variables (Mirshahi, 2010). A GLM for daily rainfall consists of modelling rainfall occurrence is first identified, and then one for the rainfall amounts is defined.

For the occurrence model, logistic regression is used to estimate the rainfall occurrence probability for each day of at a location and defined as;

$$ln\left(\frac{p_i}{1 - p_i}\right) = x_i \cdot \beta \tag{1}$$

where p_i indicates the probability of rainfall occurrence on the ith day, x_i is the corresponding (1 × n) predictor vector, β is the related (n × 1) vector of model coefficients.

For the amounts model, the rainfall amount on each wet day is modelled by a gamma distribution using a logarithmic link function. This model is defined as:

$$ln(\mu_i) = W_i \cdot \gamma \tag{2}$$

where μ_i is the mean of the gamma distribution on the i-th wet day, W_i is predictor vector (1xn) on a wet day and γ is a coefficient vector (nx1). Here, the shape factor (or dispersion parameter) of the gamma distribution is assumed to be constant for each wet day. The GLIMCLIM software, which was introduced by Chandler (2002), is used for modelling and simulating daily rainfall series in this study (please see http://www.ucl.ac.uk/~ucakarc/work/glimclim.html for R package (rglimclim)). One of the main advantages of the GLIMCLIM software is to allow covariates to provide daily, monthly and yearly effects (Chun, 2010). For model fitting purposes, temporal, spatial and external predictors are included in the model, starting with a constant or a single predictor. These predictors for both models (occurrence and amounts) are mainly periodic functions along with certain month effects to imply seasonality and transformations of rainfall values on previous days to account for persistence (Leith, 2005). There are several criteria to compare models with different number of predictors. The Akaike Information (AIC) and Bayesian Information Criterions (BIC) are the widely used criterions in hydrological applications. The value of AIC is calculated as;

$$AIC = -2 \cdot logL + 2 \cdot k \tag{3}$$