Journal of Hydrology 498 (2013) 23–35

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

Stochastic generation of multi-site daily precipitation for applications in risk management

Korbinian Breinl [a,*], Thea Turkington [b], Markus Stowasser [c]

[a] Department of Geoinformatics, Paris-Lodron University of Salzburg, Schillerstrasse 30, 5020 Salzburg, Austria
[b] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[c] Allianz SE Reinsurance, Koeniginstrasse 28, 80802 Munich, Germany

ARTICLE INFO

Article history:
Received 22 March 2013
Received in revised form 23 May 2013
Accepted 2 June 2013
Available online 18 June 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Uwe Haberlandt, Associate Editor

Keywords:
Multi-site model
Precipitation
Markov chain
Stochastic simulation
Bootstrap

SUMMARY

Unlike single-site precipitation generators, multi-site precipitation generators make it possible to reproduce the space–time variation of precipitation at several sites. The extension of single-site approaches to multiple sites is a challenging task, and has led to a large variety of different model philosophies for multi-site models. This paper presents an alternative semi-parametric multi-site model for daily precipitation that is straightforward and easy to implement. Multi-site precipitation occurrences are simulated with a univariate Markov process, removing the need for individual Markov models at each site. Precipitation amounts are generated by first resampling observed values, followed by sampling synthetic precipitation amounts from parametric distribution functions. These synthetic precipitation amounts are subsequently reshuffled according to the ranks of the resampled observations in order to maintain important statistical properties of the observation network. The proposed method successfully combines the advantages of non-parametric bootstrapping and parametric modeling techniques. It is applied to two small rain gauge networks in France (Ubaye catchment) and Austria/Germany (Salzach catchment) and is shown to well reproduce the observations. Limitations of the model relate to the bias of the reproduced seasonal standard deviation of precipitation and the underestimation of maximum dry spells. While the lag-1 autocorrelation is well reproduced for precipitation occurrences, it tends to be underestimated for precipitation amounts. The model can generate daily precipitation amounts exceeding the ones in the observations, which can be crucial for risk management related applications. Moreover, the model deals particularly well with the spatial variability of precipitation. Despite its straightforwardness, the new concept makes a good alternative for risk management related studies concerned with producing daily synthetic multi-site precipitation time series.

1. Introduction

Stochastic precipitation generators are tools to produce synthetic time series of precipitation of any length, maintaining the statistics of the observations. They are typically applied when long data series are required but the observations available are too short. Such tools have been applied in engineering design, agricultural, ecosystem as well as hydrological impact studies (Wilks and Wilby, 1999).

The generation of daily precipitation at single sites typically involves two components: the binary precipitation occurrence (precipitation or no precipitation) and the precipitation amount on wet days. Precipitation occurrences are modeled with a Markov chain and precipitation amounts with a parametric distribution (e.g., Katz, 1977; Richardson, 1981), usually ignoring serial correlation. This type of parametric model is often referred to as ‘Richardson type model’. Frequently used parametric distributions include the exponential, mixed exponential, lognormal, gamma, generalized Pareto or Weibull distribution. WGEN (Richardson and Wright, 1984) is a widely applied model applying this two-stage approach. Non-parametric approaches are based on resampling methods such as bootstrapping techniques (e.g., Brandsma and Buishand, 1997a). The idea behind such models is that synthetic time series with the same statistical properties as the observations can be generated by taking samples from the observations (Haberlandt et al., 2011). Semi-parametric approaches combine parametric as well as non-parametric methods. LARS-WG (Semenov et al., 1998) for instance generates dry and wet spells as well as precipitation amounts from semi-empirical distributions. Different models have been proposed for single sites (e.g., Brandsma and Buishand, 1997a; Hayhoe, 2000; Katz, 1977; McKague et al., 2005; Richardson and Wright, 1984; Richardson, 1981; Sharma

* Corresponding author. Tel.: +43 (0)662 8044 7578.
E-mail addresses: korbinian.breinl@sbg.ac.at (K. Breinl), t.a.r.turkington@utwente.nl (T. Turkington), markus.stowasser@allianz.com (M. Stowasser).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.06.015

and Lall, 1999; Stowasser, 2011; Wan et al., 2005; Zheng and Katz, 2008).

Unlike single-site models, multi-site models have to reproduce the space–time variation of precipitation at several sites. The extension of single-site approaches to multiple sites is a challenging task, and has resulted in many different model philosophies for multi-site models. Non-parametric methods include k-nearest neighbor (kNN) algorithms (Apipattanavis et al., 2007; Beersma and Buishand, 2003; Leander and Buishand, 2009), resampling of weather types (Wilby et al., 2003), reshuffling algorithms (Clark et al., 2004) or Neyman–Scott rectangular pulses processes (Burton et al., 2008). Semi-parametric approaches have been proposed by Brandsma and Buishand (1997b), Palutikof et al. (2002), Fowler et al. (2005), Apipattanavis et al. (2007), Kilsby et al. (2007), Cannon (2008) or Leander and Buishand (2009). They all apply various methods including the resimulation of atmospheric circulations (Brandsma and Buishand, 1997b), the combination of Markov processes and the resimulation of weather types (Palutikof et al., 2002), Markov processes and Neyman–Scott rectangular pulses (Fowler et al., 2005; Kilsby et al., 2007), the combination of Markov processes and kNN algorithms (Apipattanavis et al., 2007), Bernoulli–Gamma density networks (Cannon, 2008) or kNN algorithm with perturbation of the highest resampled values (Leander and Buishand, 2009). Parametric approaches have been introduced by Wilks (1998), Hughes et al. (1999), Brissette et al. (2007), Ailliot et al. (2009), Bardossy and Pegram (2009), Khalili et al. (2009), Serinaldi (2009b), Srikanthan and Pegram (2009) or Baigorria and Jones (2010). Amongst these, often cited is the work by Wilks (1998), who uses Markov chains at several sites driven by a correlated random field. While this model is able to reproduce many important statistical properties of the multi-site precipitation records, it does not effectively deal with the spatial variability of the generated precipitation fields. This problem is referred to as the 'spatial intermittence problem', a term first used by Bardossy and Plate (1992). The Wilks (1998) approach was further adapted by Brissette et al. (2007), Khalili et al. (2009) and Srikanthan and Pegram (2009). Copula-based methods are suggested by Bardossy and Pegram (2009) and Serinaldi (2009b) whereas Baigorria and Jones (2010) propose an orthogonal Markov chain to simulate multi-site precipitation occurrences. Hidden Markov models are applied by Hughes et al. (1999) and Ailliot et al. (2009). Comprehensive literature reviews of multi-site models are provided by Srikanthan and McMahon (2001), Serinaldi (2009b) or Baigorria and Jones (2010).

The choice of a model often relies on practical aspects including the quantity and quality of available data and the straightforwardness of a model (Serinaldi, 2009b). The latter was of particular importance in this research. We therefore propose an alternative multi-site generator for daily precipitation and risk management related applications. Thanks to its semi-parametric nature, the proposed model allows generating daily precipitation amounts larger than the ones in the observations. In risk management, non-observed extremes are of particular interest, as extreme precipitation events can trigger floods or mass-movements. To model multi-site precipitation occurrence, a univariate discrete Markov process is fitted to the time series of catchment-wide occurrence vectors. Precipitation amounts are simulated by first resampling observations, followed by sampling from parametric distribution functions. Following an idea by Clark et al. (2004), the precipitation amounts from the parametric distributions are subsequently reshuffled according to the ranks of resampled observations in order to maintain important statistical properties of the observation network. The design of the model allows sampling with correlated random numbers having the same correlation structure as the observations, thereby considerably simplifying the implementation of the algorithm. To model precipitation occurrence, the algorithms by Wilks (1998) and Brissette et al. (2007) for example require random numbers with stronger correlation compared to the observed ones.

The paper is structured as follows: first, the proposed model is explained in detail. Second, all steps of the model setup and evaluation are described. Third, we apply the model to two small rain gauge networks in France and Austria/Germany and discuss the results. The paper ends with a conclusion and some remarks.

2. Proposed multi-site precipitation generator

The proposed model arises by merging a Markov process (e.g., Ailliot et al., 2009; Apipattanavis et al., 2007; Brissette et al., 2007; Hughes et al., 1999; Khalili et al., 2009; McKague et al., 2005; Richardson and Wright, 1984; Richardson, 1981; Stowasser, 2011; Wilks, 1998) and a reshuffling algorithm (Clark et al., 2004; Mehrotra and Sharma, 2009) into a new promising framework.

2.1. Univariate Markov process for multi-site precipitation occurrences

The focus in the generation of precipitation occurrence is shifted from a single-site perspective towards a catchment-wide perspective. The concept of separate Markov chains at single sites as for example suggested by Wilks (1998), Brissette et al. (2007) or Khalili et al. (2009) is abandoned in favor of a univariate discrete Markov process. Similar attempts have been made by Hughes et al. (1999) who related a set of atmospheric variables to precipitation occurrence at multiple locations via a finite number of hidden or unobserved weather states (Mehrotra et al., 2006). More recently, a similar concept was suggested by Ailliot et al. (2009) who used a regional weather type model to simulate the temporal dependence combined with censored, power transformed Gaussian distributions, to model the spatial dependence of precipitation occurrence and amounts.

In this research, a comparatively straightforward approach is pursued without taking into account atmospheric processes. We treat catchment-wide occurrence vectors on any day as single catchment-wide events. The occurrence vector is the equivalent of the precipitation field defined by the observation network. For example, in a hypothetical three-station observation network, the occurrence vector on any given day can vary between (1,1,1) if there is precipitation at all stations at once and (0,0,0) if all stations are dry. It is straightforward to fit a single discrete Markov chain to this time series of occurrence vectors.

A Markov chain represents a stochastic process that describes time series of discrete random variables and can be characterized by two main properties, the 'state' and the 'order'. The 'state' defines the number of values the variable can take on. The order of a Markov chain defines the number of previous values to determine the probabilities of transitioning from one state to another. These probabilities are often referred to as 'transition probabilities'. The transition probabilities are defined in a transition matrix, where each row lists the probabilities of transitioning from a given state represented by that row, to each of the various possible states. An mth order Markov chain where the transition probabilities depend on the previous m days is defined by

$$PR\{X_{t+1}|X_t, X_{t-1}, X_{t-2}, \ldots, X_1\} = PR\{X_{t+1}|X_t, X_{t-1}, \ldots, X_{t+m}\} \qquad (1)$$

Likelihood measures such as the Akaike information criterion AIC (Katz, 1981) or the Bayesian information criteria BIC (Schwarz, 1978) can help to determine the most appropriate order for a Markov chain. In the simplest case at single sites, modeling the occurrence of precipitation can be achieved with a two-state first-order Markov chain for binary states of 0's and 1's indicating dry (0) and wet (1) days (e.g., Katz, 1977; Richardson, 1981; Stern and Coe,