Available online at www.sciencedirect.com

ScienceDirect

Advances in Climate Change Research 7 (2016) 70–81

www.keaipublishing.com/en/journals/accr/

# A multi-state weather generator for daily precipitation for the Torne River basin, northern Sweden/western Finland

David RAYNER*, Christine ACHBERGER, Deliang CHEN

*Regional Climate Group, Department of Earth Sciences, University of Gothenburg, 40530 Gothenburg, Sweden*

Received 10 December 2015; revised 22 June 2016; accepted 23 June 2016
Available online 29 June 2016

## Abstract

This paper describes a new weather generator – the 10-state empirical model – that combines a 10-state, first-order Markov chain with a non-parametric precipitation amounts model. Using a doubly-stochastic transition-matrix results in a weather generator for which the overall precipitation distribution (including both wet and dry days) and the temporal-correlation can be modified independently for climate change studies. This paper assesses the ability of the 10-state empirical model to simulate daily area-average precipitation in the Torne River catchment in northern Sweden/western Finland in the context of 3 other models: a 10-state model with a parametric (Gamma) amounts model; a wet/dry chain with the empirical amounts model; and a wet/dry chain with the parametric amounts model. The ability to accurately simulate the distribution of multi-day precipitation in the catchment is the primary consideration.

Results showed that the 10-state empirical model represented accumulated 2- to 14-day precipitation most realistically. Further, the distribution of precipitation on wet days in the catchment is related to the placement of a wet day within a wet-spell, and the 10-state models represented this realistically, while the wet/dry models did not. Although all four models accurately reproduced the annual and monthly averages in the training data, all models underestimated inter-annual and inter-seasonal variance. Even so, the 10-state empirical model performed best. We conclude that the multi-state model is a promising candidate for hydrological applications, as it simulates multi-day precipitation well, but that further development is required to improve the simulation of interannual variation.

*Keywords:* Weather generator; Multi-state; Torne River; Precipitation

## 1. Introduction

A weather generator (WG) is a stochastic model that is designed to generate synthetic weather time-series with the same statistical properties as observed data. WGs can provide additional data when the observed climate record is insufficient with respect to completeness, spatial coverage or length to reliably estimate of the probability of extreme events (Jones et al., 2011; Kilsby et al., 2007; e.g. Wilks and Wilby, 1999). WGs can be used to simulate short-term weather (e.g. at daily or sub-daily scales) for the past or future, and have become a common tool for studying impacts of climate change on ecosystems and human settlements (Jones et al., 2011; e.g. Wilks, 2010). One particular advantage of using WGs is that they can generate time-series for an extended period without significant computational investment.

Within the broad family of WGs that have been developed for simulating daily precipitation, two categories of models dominate the literature, which we call the “Richardson”

---

* Corresponding author.

*E-mail address:* David.Rayner@gvc.gu.se (RAYNER D.).

Peer review under responsibility of National Climate Center (China Meteorological Administration).

Production and Hosting by Elsevier on behalf of KeAi

http://dx.doi.org/10.1016/j.accre.2016.06.006

(Richardson and Wright, 1984; Richardson, 1981) and the "serial" (Racsko et al., 1991; Semenov et al., 1998) types. The Richardson-type WG − applied in this study − simulates daily precipitation in two separate steps, the first to simulate rainfall occurrence and the second to estimate the rainfall amount on wet days. With the classical Richardson model, the first step is accomplished using a first-order, two-state Markov chain, which describes the probability of a wet day following a dry day, a wet day following a wet day, etc. The transition probabilities can be estimated from the observed data.

Once a certain day is modeled as wet, an "amounts model" simulates the precipitation amount for that day. Parametric approaches (e.g. Chen et al., 2015; Chen and Brissette, 2014) use pre-specified functions to approximate the observed precipitation distribution. With non-parametric approaches, the observed precipitation distribution itself is used as the basis for the amounts model. The simplest non-parametric approach is to resample directly from the observed sequence, or from a sub-set of the sequence which represents the "nearest neighbors" with-respect-to weather conditions (e.g. Sharma and Lall, 1999). Kernel-density smoothing the observed distribution allows non-parametric methods to generate a continuous distribution of precipitation, and also to generate values higher than observed historically (e.g. Harrold et al., 2003; Mehrotra and Sharma, 2007a). Non-parametric methods allow a WG to generate precipitation sequences that match the observed distribution with arbitrary-high precision, at the cost of introducing arbitrarily-many parameters. Non-parametric methods allow more flexibility for including new forms of conditional dependence; they have the ability to reproduce features such as non-linearity, asymmetry, or multi-modality in observed records (Mehrotra et al., 2006); and they do not make any strong assumptions about the precipitation distribution (Mehrotra and Sharma, 2007a).

Many WGs using Markov-approaches have been found to have 3 partially-related deficiencies: they underestimate the frequency of extended drought periods (Mehrotra and Sharma, 2007a), they often ignore temporal correlations within wet-spells (Harrold et al., 2003), and they underestimate low-frequency (usually described by inter-annual) variability (e.g. Gregory et al., 1993; Katz and Zheng, 1999; Srikanthan et al., 2005). These issues appear to be partially related, as a proportion of low-frequency variability in precipitation can be accounted for by the short-lag correlation (Gregory et al., 1993) or by the rainfall occurrence process (Mehrotra and Sharma, 2007a).

The tendency of purely Markov-based WGs to underestimate inter-annual variability has been attributed to climatic non-stationarities, for example the influence of the El Niño Southern Oscillation (Harrold et al., 2003). One way to increase the simulated inter-annual variability is to condition the WG parameters on a physical, slowly-varying index representing atmospheric circulation or SST (Katz and Parlange, 1993; Wilby et al., 2002). Another method is to use longer-period, aggregated precipitation as the conditioning index, either explicitly (Harrold et al., 2003; Mehrotra and Sharma, 2007a, 2007b) or via wavelet decomposition (Steinschneider and Brown, 2013). Finally, the low-frequency signal can be increased by postulating dependence on a "hidden" index, whose variation must be estimated using iterative methods (Katz and Zheng, 1999).

Consecutive days with similar rainfall amounts are often clustered in time, a property which is not represented by two-state Markov chains that distinguish only between dry and wet. This property can be represented using a multi-state Markov chain model that models transitions between different precipitation bands (e.g. Boughton, 1999; Gregory et al., 1993; Haan et al., 1976; Srikanthan and McMahon, 2001). The state boundaries can be defined using geometric progression, resulting in increasing class widths (e.g. Haan et al., 1976; Srikanthan et al., 2005) and a relatively even number observations in each state.

Simulation of catchment runoff often requires multiple precipitation time-series, each representing a different sub-catchment or gauge, with realistic spatial correlations. There are many approaches to generating such series. One is to drive a collection of individual models (representing different locations) with a common "random" number series, which is in-turn derived from an index of larger-scale atmospheric circulation. Another is to feed independent models with serially-independent but spatially correlated random series (Wilks, 1998). Finally, it is possible to simulate the catchment-average precipitation as a single time-series (Chen et al., 2012), which is the approach used in this study.

This paper introduces a new WG that combines a 10-state, first-order Markov chain and a non-parametric precipitation amounts model. Our innovation is to adopt a doubly-stochastic transition-matrix, rather than manually defining transition thresholds, which gives the model the property that the overall precipitation distribution (defined as including both wet and dry days) is independent of the transition-matrix. The paper first describes the new WG and its implementation for the Torne River catchment in northern Sweden/western Finland. The paper then quantitatively-assesses whether the model's performance is an improvement over simpler two-state approaches. The properties assessed are those that are important for hydrological modeling: inter-annual, inter-seasonal and multi-day precipitation distributions; lengths of wet and dry spells; and the variations of precipitation within multi-day events.

## 2. Site description and data

### 2.1. The Torne River catchment

The Torne River catchment (Fig. 1) straddles the border of Sweden and Finland and covers 40,157 km$^2$. The catchment extends from the northern mountains of Sweden and north-western Finnish Lapland, south-east down through marshes and lowlands to the Gulf of Bothnia in the Baltic Sea. The lower reaches of the Torne River comprises the border between Sweden and Finland, with the closely-connected towns of Haparanda (Sweden) and Tornio (Finland) near the river mouth together having around 23,000 inhabitants. The Torne River is essentially unregulated and the catchment is sparely