Journal of Hydrology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Hydrology

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

A conditional stochastic weather generator for seasonal to multi-decadal simulations

Andrew Verdin [a,*], Balaji Rajagopalan [a,b], William Kleiber [c], Guillermo Podestá [d], Federico Bert [e]

[a] Dept of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, CO, United States
[b] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, United States
[c] Dept of Applied Mathematics, University of Colorado, Boulder, CO, United States
[d] School of Marine & Atmospheric Sciences, University of Miami, Miami, FL, United States
[e] Facultdad de Agronomía, Universidad de Buenos Aires – CONICET, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Generalized linear models
Stochastic weather generator
Conditional simulation
Downscaling seasonal forecasts
Daily precipitation
Daily temperature

SUMMARY

We present the application of a parametric stochastic weather generator within a nonstationary context, enabling simulations of weather sequences conditioned on interannual and multi-decadal trends. The generalized linear model framework of the weather generator allows any number of covariates to be included, such as large-scale climate indices, local climate information, seasonal precipitation and temperature, among others. Here we focus on the Salado *A* basin of the Argentine Pampas as a case study, but the methodology is portable to any region. We include domain-averaged (e.g., areal) seasonal total precipitation and mean maximum and minimum temperatures as covariates for conditional simulation. Areal covariates are motivated by a principal component analysis that indicates the seasonal spatial average is the dominant mode of variability across the domain. We find this modification to be effective in capturing the nonstationarity prevalent in interseasonal precipitation and temperature data. We further illustrate the ability of this weather generator to act as a spatiotemporal downscaler of seasonal forecasts and multidecadal projections, both of which are generally of coarse resolution.

1. Introduction

Scientific and technological advances, together with awareness of the importance of climate on human endeavors, are creating increased worldwide demand for climate information. Fortunately, our ability to monitor and predict variations in climate has increased substantially (Barnston et al., 2010; Stockdale et al., 2010). A number of groups now forecast climate conditions a few seasons ahead (Goddard et al., 2003; Saha et al., 2006). Emerging developments may enable climate projections 10–20 years into the future, a scale intermediate between seasonal forecasts and manmade climate change projections (Haines et al., 2009; Hurrell et al., 2009; Meehl et al., 2009). These advances, however, must be matched by a better understanding of how science can inform climate-resilient planning and development (Stainforth et al., 2007).

To support public and private adaptation and mitigation responses, climate information must be credible, legitimate and, especially, salient – e.g., relevant to the needs of decision makers (Cash et al., 2003). Needs include not only predictions or projections[1] (Bray and von Storch, 2009) of regional climate: *potential outcomes of adaptation actions are probably more relevant to stakeholders than raw climate information*. Thus, an enhanced capacity is needed to "translate" climate information into distributions of outcomes for risk assessment and management (Hansen et al., 2006).

Process models (e.g., crop biophysical models, hydrological models) can be useful tools to assess likely impacts on climate-sensitive sectors of society, and to evaluate the outcomes of alternative adaptive actions (Ferreyra et al., 2001; Berger, 2001; Berger et al., 2006; Happe et al., 2008; Freeman et al., 2009; Schreinemachers and Berger, 2011; Bert et al., 2006, 2007, 2014). These models, however, typically require daily weather data. Although historical daily weather can be used, getting long-term daily weather is laborious and costly at best and, in some cases, impossible. Typically, historical observations have missing data

[1] Following Bray and von Storch (2009), ***prediction*** conveys a sense of certainty whereas ***projection*** is associated more with the possibility of something happening given a certain set of plausible, but not necessarily probable, circumstances. A prediction can be used to design specific response strategies, while a projection, or more precisely a series of projections, provides a range on which to consider a range of response strategies.

* Corresponding author.
E-mail address: andrew.verdin@colorado.edu (A. Verdin).

http://dx.doi.org/10.1016/j.jhydrol.2015.12.036
0022-1694/

that are not accepted by impact models. Similarly, point measurements may not represent the true spatial variability of a nonstationary natural process (e.g., daily precipitation). Most importantly, observed sequences provide a solution based on only one realization of the weather process (Richardson, 1981).

The use of seasonal forecasts of regional climate and its impacts can help decision-makers to lessen the adverse effects of unfavorable conditions or, alternatively, to capitalize on favorable conditions. Nevertheless, a major obstacle to broader use of seasonal climate forecasts is their coarse spatial and temporal resolution. Similarly, 10–20 year projections of regional climate conditions have been identified as important to infrastructure planners, water resource managers, and many others (Hurrell et al., 2009). Unfortunately, projections of regional monthly precipitation and temperature from climate models not only are coarse in space and time – as seasonal forecasts – but also involve considerable uncertainty, which requires exploration of the impacts of alternative, plausible trajectories. Stochastic weather generators have long been used for risk assessment and adaptation, as they can provide a rich variety of plausible climatic scenarios. Moreover, weather generators can produce spatially consistent series that can be used to downscale larger-scale scenarios.

Traditional weather generators (stemming from Richardson, 1981) model precipitation occurrence as a chain-dependent process (Katz, 1977) and thus are capable of generating physically realistic prolonged wet and dry spells. The remaining weather variables (e.g., precipitation intensity and temperature) are parameterized using probability distributions (for precipitation intensity) and linear time series models (for temperature), which capture historical climatological variability and linear relationships between variables but fail to capture extremes (e.g., extreme drought or flooding). In order to capture the variability of weather attributes in any specific season, the simulations need to be conditioned on appropriate covariates. One approach is to estimate the parameters of the generator conditionally by considering ENSO (El Niño Southern Oscillation; Trenberth and Stepaniak, 2001) phase, or any other teleconnection to a region's climate, which enables simulation of skillful sequences (Grondona et al., 2000; Ferreyra et al., 2001; Wilby et al., 2002; Meza, 2005; Katz et al., 2002). Wilks (2008) illustrated the capability of interpolating weather generator parameters to arbitrary locations (e.g., on a grid) using local weighted regressions; Wilks (2009) subsequently offered a method to synchronize gridded synthetic weather series on observed weather data. Approaches to producing weather sequences that deviate from climatology have included the implementation of seasonal correction factors, perturbation of parameters or input data, and spectral approaches (Caron et al., 2008; Kilsby et al., 2007; Hansen and Mavromatis, 2001; Schoof et al., 2005; Qian et al., 2010).

Nonparametric weather generators have an improved ability to capture nonlinearities between variables and sites. Included in this subclass are the k-nearest neighbor (k-NN) bootstrap resampling method (Brandsma and Buishand, 1998; Rajagopalan and Lall, 1999; Buishand and Brandsma, 2001; Beersma and Buishand, 2003; Yates et al., 2003; Sharif and Burn, 2007) and kernel density based estimators (Rajagopalan et al., 1997; Harrold et al., 2003; Mehrotra and Sharma, 2007). Caraway et al. (2014) first applied a clustering algorithm to identify regions of similar climatology before applying the k-NN approach, which has shown good performance in regions of complex terrain. Apipattanavis et al. (2010) modified the k-NN approach to create a semi-parametric weather generator that better captures the duration of wet and dry spells via Markov chain modeling. Modifications of the k-NN based weather generator to incorporate seasonal precipitation forecasts (Apipattanavis et al., 2010) and multi-decadal projections (Podestá et al., 2009) have also been proposed. In these situations, the resampling is weighted to reflect the projected distribution of regional climate conditions. These methods are simple and powerful, however their main drawback is that they cannot generate values outside the range of historical data. More importantly, it is not easy to generate weather sequences at locations other than those with historical observations.

Pioneered by Stern and Coe (1984), generalized linear models (GLMs) are able to straightforwardly model non-normal data through a suite of link functions. Relevant to this research, GLMs can be used to model and simulate daily weather sequences, and have paved the way for generating space–time weather sequences at any desired location (Kleiber et al., 2012, 2013; Furrer and Katz, 2007; Kim et al., 2012; Yan et al., 2002; Yang et al., 2005; Chandler, 2005; Verdin et al., 2015). Recently Verdin et al. (2015) incorporated these developments into a robust space–time weather generator and demonstrated its capability to generate realistic weather sequences at arbitrary locations in the Pampas of Argentina – also the region targeted by this paper. The GLM framework offers several advantages – mainly they reduce the effort in modeling non-normal variables and are parsimonious (McCullagh and Nelder, 1989), especially for discrete and skewed variables (e.g., precipitation occurrence and intensity, respectively). Coupled with spatial processes, GLMs can generate sequences at any spatial resolution – which is important for resource management. Furthermore, covariates such as ENSO information, seasonal climate forecasts, and annual cycles can easily be incorporated in the GLMs to refine or narrow the distribution of expected values (e.g., Chandler and Wheater, 2002; Wheater et al., 2005; Furrer and Katz, 2007; Kim et al., 2012).

As motivated earlier in this section, skillful and realistic sequences of daily weather in any given season are essential for efficient planning and management of agricultural resources. One method of obtaining such sequences requires generating space–time weather sequences that are consistent with, and conditioned on, coarse climate information from seasonal to decadal time scales. To this end, here we propose a modification to the stochastic weather generator presented in Verdin et al. (2015) to include the coarse scale information as covariates. We refer to the weather generator of Verdin et al. (2015) as "original"; that of this research will be called the "modified" weather generator. The paper is organized as follows: the study region and data are described in Section 2; Section 3 contains a brief summary of the modified methodology. In Section 4 we discuss the results, and in Section 5 we conclude with a summary of the research and future work.

2. Study region and data

Application of this methodology is focused on a network of seventeen weather stations located in and around the Salado *A* basin of the Pampas of Argentina (see Fig. 1). The Salado is part of the large Río de la Plata basin (Herzer, 2003). Note the study region differs from that of Verdin et al. (2015).

The *A* basin is an agriculturally productive sub-basin within the Salado River basin where maize, soybean, and wheat are grown. The Salado Basin has very flat topography and a poorly developed and disintegrated drainage system. The western basin (Salado *A*) includes mega-parabolic dunes separated by depressions that constrain evacuation of surface water (Aragón et al., 2010; Viglizzo et al., 2009, 1997). Since colonial times, the Salado has shown alternating floods and droughts that displace populations and disrupt productive activities and livelihoods for extended periods. Floods were frequent during the late 19th and early 20th centuries, a relatively wet epoch. In contrast, extensive droughts were more frequent during the drier 1930s–1950s (Herzer, 2003; Seager et al., 2010). Partly in response to rain increases since the 1970s, severe