Journal of Hydrology 512 (2014) 285–302

Contents lists available at ScienceDirect

Journal of Hydrology

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

Simulating daily rainfall fields over large areas for collective risk estimation

Francesco Serinaldi *, Chris G. Kilsby

School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK
Willis Research Network, 51 Lime St., London EC3M 7DQ, UK

ARTICLE INFO

Article history:
Received 3 June 2013
Received in revised form 16 October 2013
Accepted 17 February 2014
Available online 7 March 2014
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Emmanouil N. Anagnostou, Associate Editor

Keywords:
Rainfall modeling
Generalized linear models
Generalized additive models
Gaussian random fields
Large scale risk assessment

SUMMARY

Large scale rainfall models are needed for collective risk estimation in flood insurance, infrastructure networks and water resource management applications. There is a lack of models which can provide simulations over large river basins (potentially multi-national) at appropriate spatial resolution (e.g., 5–25 km) that preserve both the local properties of rainfall (i.e., marginal distributions and temporal correlation) and the spatial structure of the field (i.e., the spatial dependence structure). In this study we describe a methodology which merges meta-Gaussian random fields and generalized additive models to simulate realistic rainfall fields at daily time scale over large areas. Unlike other techniques previously proposed in the literature, the suggested approach does not split the rainfall occurrence and intensity processes and resorts to a unique discrete–continuous distribution to reproduce the local properties of rainfall. This choice allows the use of a unique meta-Gaussian spatio-temporal random field substrate that is devised to reproduce the spatial properties and the short term temporal characteristics of the observed precipitation. The model is calibrated and tested on a 25 km gridded daily rainfall data set covering the 817 000 km^2 of the Danube basin. Standard and ad hoc diagnostics highlight the overall good performance over the whole range of rainfall values at multiple scales of spatio-temporal aggregation with particular attention to extreme values. Moreover, the modular structure of the model allows for refinements, adaptation to different areas and the introduction of exogenous forcing variables, thus making it a valuable tool for classical hydrologic analyses as well as for new challenges of network and reinsurance risk assessment over extensive areas.

1. Introduction

Dealing with large geographic areas, the modeling of the spatio-temporal evolution of rainfall is a challenging task that however must be tackled to provide realistic scenarios to be used as an essential input of water resource or flood risk assessment analyses. Over large areas, extreme events cannot be simply defined as the occurrence of an observation or a cluster of high values over a spatially coherent geographic zone, but more realistically as a set of rainfall fields evolving in time (generally driven by large scale climate patterns). The combination of local rainfall intensities, spatial extension and temporal persistence creates rainfall scenarios that can overload the basin system resulting in critical conditions of saturation and surface runoff. For instance, persistent and spatially extended rainfall events with medium intensity can be more dangerous than short and highly intensive events, when the phenomenon affects wet soils already saturated by previous events. Modeling these conditions therefore plays a key role in a well-devised risk assessment procedure and requires moving from a static point of view (rainfall frequency analysis) to a dynamic perspective (stochastic modeling).

A large number of spatio-temporal rainfall models has been suggested in the literature. Overviews were provided by Wilks and Wilby (1999), Srikanthan and McMahon (2001), Mehrotra et al. (2006) and more recently, by Baigorria and Jones (2010), Maraun et al. (2010) and Haberlandt et al. (2011). Even though several hybrid versions are available, it is possible to attempt a rather general classification based on the underling backbone technique of each method. Following Haberlandt et al. (2011), we can distinguish between:

1. Alternating renewal processes.
2. Time series models.

* Corresponding author at: School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK. Tel.: +44 01912085956; fax: +44 01912086502.

E-mail address: francesco.serinaldi@ncl.ac.uk (F. Serinaldi).

http://dx.doi.org/10.1016/j.jhydrol.2014.02.043
0022-1694/© 2014 The Authors. Published by Elsevier B.V.

3. Markov chain models.
4. Linear parametric models.
5. Point processes.
6. Disaggregation models.
7. Resampling techniques.

Each approach exhibits both suitable properties and shortcomings. We give a brief description of each method highlighting the aspects that are potentially attractive to model rainfall over large areas and thus the reasons that have led to the approach adopted in this study.

The alternating renewal processes separate the precipitation process into alternating wet and dry spells plus an internal structure of the rainfall pattern. The process is described by four random variables, namely dry spell duration, wet spell duration, wet spell amount and wet spell intensity. The approach is devised for time series modeling and can be extended to a multisite/gridded set-up via optimization techniques. Dependence of the model parameters on exogenous variables can be introduced. The method is suitable for fine scale temporal resolution by introducing "internal" rainfall patterns within each storm event, whereas the spatial domain might depend on the computational time required by the optimization/resampling procedures (see e.g., Haberlandt et al., 2008, 2011, and references therein).

Time series models such as Markov chain models, generalized linear models (GLM; McCullagh and Nelder, 1989), generalized additive models (GAM; Hastie and Tibshirani, 1990), linear parametric models (ARMA and their extensions; Hipel and McLeod, 1994) are widely used to simulate hydrological variables. A well established theoretical framework allows accounting for exogenous variables, temporal dependence and wet/dry conditions via a suitable parameterization. First or higher order Markov chains are used to model the alternation of rainfall states (wet/dry states in a two-states set-up). Multisite/gridded extensions can be built via latent spatial processes with uniform marginals. This type of models is commonly applied to daily data (see e.g., Wilks, 1998; Yang et al., 2005; Serinaldi, 2009; Kleiber et al., 2012). Examples of rainfall simulation have been provided for grids spacing ≈ 20 km over areas up to $\approx 700\,000$ km^2 (Kleiber et al., 2012).

The models based on point processes describe rainfall events as an arrival process of rectangular pulses. The superposition of the cells generates clusters that define the rainfall events. Multisite extensions are available via two-dimensional rain cells and spatial point processes. Dependence between climate and catchment properties and model parameters can be introduced. This class of models is devised for sub-daily temporal resolution (1 h) and has been applied up to nation-wide scales (e.g., UK) via domain partitioning (required because these models are usually stationary in occurrence in space). Data at different time scales are required to calibrate these models (e.g., Cowpertwait, 2006; Burton et al., 2010).

Disaggregation models are devised for space–time disaggregation of rainfall series according to the scaling properties of the rainfall. Fine scale high quality data are required to fit the models. Dependence between climate and catchment properties and model parameters can be introduced; however, in these cases, the models are no longer properly scaling/fractal/multifractal. The spatial domain depends on the range of scales in which the scaling properties reasonably hold (see e.g., Schertzer and Lovejoy, 1987; Gupta and Waymire, 1993; Over and Gupta, 1996; Deidda, 2000).

Resampling models do not model rainfall but sample the observed values according to suitable rules that preserve the spatio-temporal statistical properties of the rainfall measurements. Dependence on climate and catchment properties can be introduced by modifying the resampling rules according to a suitable data stratification. The approach is data-driven and non-parametric, thus avoiding any model misspecification. The method does not allow generation of values more extreme than those observed and can be time expensive for large areas and fine (sub-daily) time scales (see e.g., Brandsma and Buishand, 1998; Buishand and Brandsma, 2001; Apipattanavis et al., 2007; Mehrotra and Sharma, 2009; Mezghani and Hingray, 2009).

Based on this overview, some models are less suitable than others to describe and simulate rainfall fields over large areas. For example, disaggregation models need extensive good quality rainfall information at fine time scale, which is rarely available. Resampling procedures do not always perform satisfactorily if the focus is on extreme events. Models based on point processes have a well defined mathematical framework; however their extension and incorporation of exogenous variables is not always easy; moreover, even in this case, good quality data at multiple time scales required for the model calibration are rarely available for large areas. In addition, a rainfall model for large areas must fulfill some requirements, such as ease of implementation, interpretation and extension (by incorporating exogenous variables), a reasonable simulation speed, and adaptability. The latter property refers to the possibility of tuning specific components to tailor the model according to specific areas, climate regions, and also to improve the performance in terms of specific aspects that might be of interest for design purposes.

Based on the above remarks, we opted for a parametric approach that falls into the class of the time series models for daily data. In particular, we combine Markov chain models with GAM components for marginals and spatio-temporal meta-Gaussian random fields. The modeling framework is described in Section 2. Section 3 introduces the data sets used in the case study (rainfall data and covariates), whereas the model set-up is presented in Section 4. The model performance is discussed in Section 5 and concluding remarks are reported in Section 6.

2. Model structure

As is mentioned in the introduction, the problem of simulating rainfall over large areas is tackled here by adopting a parametric method for daily rainfall data. The choice of the time scale is dictated by the availability of rainfall data and climate covariates for large areas. Indeed, good quality gauge and gridded daily rainfall data sets are provided by several institutions worldwide an can be used for the model calibration. Moreover, the aim is to develop a tool useful for risk analysis at country scale (at least), meaning that fine time scale details may be less important than an overall picture of the rainfall phenomenon. Daily rainfall data are also a suitable input for rainfall–runoff models for large basins. On the other hand, using methods relying on fine scale or multiple scale data sets can be impractical because of data requirement and the general low flexibility and adaptability of the corresponding modeling frameworks.

Therefore, the modeling approach proposed in this study belongs to the class of models proposed by Wilks (1998) and Chandler and Wheater (2002) (see also Yang et al., 2005; Segond et al., 2006, 2007) and further developed by Baigorria and Jones (2010) and Kleiber et al. (2012). The basic idea behind these models is to split at-site occurrence process (the transition between wet and dry days) and rainfall amount process (positive rainfall values in wet days). Both processes are therefore modeled by suitable GLM/GAM that describe the at-site marginal distribution of the rainfall process. The spatial correlation is introduced in the simulation stage by hidden meta-Gaussian processes which enable the simulation of spatially correlated random numbers with uniform marginal distributions. These correlated random numbers are then plugged in the GLM/GAM expressions and transformed into values that preserve the spatial correlation and follow the at-site rainfall discrete–continuous marginal distributions