Computational Statistics and Data Analysis 95 (2016) 39–56

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A Bayesian hierarchical model for spatial extremes with multiple durations

Yixin Wang, Mike K.P. So *

The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 17 March 2014
Received in revised form 2 August 2015
Accepted 1 September 2015
Available online 30 September 2015

Keywords:
Bayesian analysis
Environmental data
Hierarchical model
Multiple durations
Spatial extreme
Spatial–temporal process

ABSTRACT

Bayesian spatial modeling of extreme values has become increasingly popular due to its ability to obtain relevant uncertainty measures for the estimates. This has implications for the problem of limited data on the study of extreme climatological events. Noticing the abundance of non-daily environmental records, 1-h and 6-h records in particular, we propose a Bayesian hierarchical model that can address multiple durations, in addition to the spatial effects, within a set of extreme records with multiple durations. The generalized Pareto distribution for threshold exceedances and the binomial distribution for exceedance frequencies are adopted for the top-most characterization of extreme data. The duration effect on spatial extremes is characterized by pooling the extremes with different durations and merging the duration into a latent spatial process structure as one of the covariates. Statistical inference is performed using Markov Chain Monte Carlo (MCMC) methods, for which an adaptive tuning algorithm is proposed. The methodology is applied to simulated datasets and real precipitation data for a region around Hong Kong.

1. Introduction

Investigation of the patterns of extreme climatological events plays a crucial role in environmental risk analysis and policy making. It informs the development of precautionary measures and sheds light on long-term regional planning. A careful study of these extreme events is, therefore, crucial. However, the limited availability of records of extreme events has always been a challenging aspect of such studies. Observations are only available from designated stations, where even the commonly studied daily records tend to be limited, let alone the records of extreme events. These limitations have motivated studies on the dependence structure of extreme events, which enable our estimations to borrow strength from data nearby.

Many previous studies have considered the dependence structure of extreme events as characterized by various covariates such as space, time, etc. de Haan (1985) first studied the high-dimensional multivariate dependence structure of extreme observations. Corresponding dependence measures were later developed for the study of multivariate structure of extreme events (Coles et al., 1999; Schlather and Tawn, 2003; Cooley et al., 2006). In addition to many classical works using frequentist approach in spatial modeling, e.g., de Haan and Resnick (1977), de Haan (1984), Resnick (1987), Smith (1989), Tawn (1988, 1990), Coles (2001), and Davison et al. (2012), the hierarchical Bayesian approach also receives attention in spatial and extreme value modeling. Wikle et al. (1998) described the monthly maximum temperature using an underlying state process with site-specific time series models characterizing the state variable; they incorporated both large-scale

* Corresponding author.
E-mail address: immkpso@ust.hk (M.K.P. So).

http://dx.doi.org/10.1016/j.csda.2015.09.001
0167-9473/

variabilities and small-scale space–time dynamics into the model. Cooley and Sain (2010) captured the spatial structure of meteorological extremes via a latent process under the point process representation, whereas Cooley et al. (2007) presented a similar idea with the threshold exceedance characterization. Ghosh and Mallick (2011) considered a spatial–temporal framework by incorporating spatial dependence into the likelihood while keeping the temporal dependence within the latent structure. Sang and Gelfand (2009) generalized the spatial structure by adopting multivariate Markov random field models for the joint spatial dependence of location and scale parameters, while temporal dependence was allowed only for the location parameter. For spatial–temporal modeling, Sang and Gelfand (2010) further proposed continuous spatial process models with mean square continuous realizations to relax the common conditional independence assumption in the top-most hierarchies. More recently, Kunihama et al. (2012) and Nakajima et al. (2012) proposed the state space approach to model temporal dependence in extremes and proposed several efficient algorithms for their corresponding Bayesian inferences.

Recall that setting up a dependence structure via a latent spatial process in hierarchical Bayesian spatial modeling helps in the study of pooled records from different locations. In particular, it helps to strengthen the parameter estimation at a particular location by borrowing strength from records at neighboring stations. Continuing with this thinking, we borrow additional strength from records with different durations. The use of multiple duration records is common in the study of intensity–duration–frequency relationships (Koutsoyiannis et al., 1998) and in constructing intensity–duration–frequency curves. Baldassarre et al. (2006) used extreme rainfall with durations ranging from 15 min to 24 h to index storm estimation. Feng et al. (2007) studied extreme precipitation in China using 2-day, 5-day, and 10-day moving sums of precipitation in addition to daily values. Back et al. (2011) modeled precipitations of short durations. We pool the records from different locations and durations, and integrate duration as a covariate into a latent dependence structure.

The proposed model adopts a typical three-layer Bayesian hierarchy that is similar in spirit to that of Cooley et al. (2007). The characterization of extremes is divided into two parts—the generalized Pareto distribution (GPD) for threshold exceedances and the binomial distribution for exceedance frequencies. A latent Gaussian process, involving duration and spatial covariates, is assumed for each distribution parameter. Essentially, if the spatial characterization includes longitude and latitude coordinates, for instance, then duration is weighted to serve as a third coordinate, where this weight is left as a parameter to estimate. This integration of duration into the model helps to avoid potentially nonsensical return-level estimates, such as a 6-h return-level estimate that is higher than the 24-h one, which can be the result when modeling data of different durations separately. In terms of statistical inference, we obtain approximate draws from the posterior distribution by adopting the adaptive random walk Metropolis–Hastings (MH) algorithm within a Gibbs sampler. An adaptive tuning algorithm is proposed to facilitate the mixing of Markov Chain Monte Carlo (MCMC) chains and hence the implementation of our approach.

There are two reasons for taking duration into consideration. First, the characterization of extreme records, especially those of precipitation, should always specify how long the time coverage is. In addition, most previous studies of meteorological extremes focus only on those of a particular duration, such as daily precipitation. However, in addition to daily records, there are usually many more records with other durations, namely 1-h or 6-h ones. Studying all of these records together can ease the data scarcity problem to a large extent and thus achieve a potential gain in model efficiency. In the Atmospheric and Environmental Real-time Database of the Institute for the Environment (IENV) at the Hong Kong University of Science and Technology (HKUST), for example, some stations only have 1-h observations. Hence, over 96% of the records fall into the 1-h category, whereas only around 2% are 6-h, and less than 1% are 24-h. Apart from this abundance of data, records with shorter durations are especially valuable for studying extremes as extreme events, precipitation in particular, seldom sustain their peak level over 12 h. Moreover, the devastating damage caused by severe events is mostly due to extreme events over a relatively short period rather than cumulative events over a longer duration. Therefore, we propose a model to study the entire set of records with multiple durations and we believe the inclusion of duration in the dependence structure is worth investigating.

The proposed approach is illustrated by an application to a dataset of 1-h, 6-h, and 24-h precipitation records in a small region around Hong Kong with 89 stations. In this application, we keep the shape parameter of the GPD constant throughout the study region for model simplicity. One striking outcome is that in the latent Gaussian process, the duration has a negative coefficient in the mean structure but a relatively large one in the covariance structure. This suggests that most extreme precipitation events occur with strong intensity but short durations. In addition, the subtlety of choosing multiple thresholds evolves naturally from the pooling of records with different durations. We compare a one-threshold model (100 mm for 1/6/24-h), a two-threshold model (50 mm for 1/6-h and 100 mm for 24-h), and a three-threshold model (25 mm for 1-h, 50 mm for 6-h and 100 mm for 24-h). The three models produce GPD scale and shape parameter estimates that fit similarly well to the data, which suggests a certain robustness of the model at least in this particular application. With the use of the proposed adaptive tuning algorithm, the addition of duration to the model does not impede a reasonably fast mixing of Markov Chain Monte Carlo (MCMC) chains when applied to both the real data and the simulated data.

The paper is organized as follows. Section 2 reviews extreme value statistics and discusses the spatial extremes. Section 3 describes the construction of the Bayesian hierarchical spatial model with multiple durations. Section 4 discusses the sampling scheme for Bayesian inference and sketches the proposed adaptive tuning algorithm. Section 5 demonstrates, via simulation studies, how the Bayesian estimator and the proposed adaptive tuning algorithm perform in general. Section 6 presents an application of the proposed model to real precipitation extremes. We conclude with a discussion in Section 7.