Spatial Statistics 26 (2018) 1–20

Contents lists available at ScienceDirect

Spatial Statistics

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

Bayesian spatial–temporal modeling of air pollution data with dynamic variance and leptokurtosis

Amanda M.Y. Chu [a], Raymond W.M. Li [b], Mike K.P. So [b,*]

[a] Department of Mathematics and Statistics, Hang Seng Management College, Hang Shin Link, Siu Lek Yuen, Shatin, N.T., Hong Kong

[b] Department of Information Systems, Business Statistics and Operations Management, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 20 October 2017
Accepted 15 May 2018
Available online 29 May 2018

Keywords:
Bayesian estimation
Dynamic variance
Leptokurtosis
NO_2
Spatial–temporal model

ABSTRACT

Spatial–temporal modeling is commonly used to explain the dependence of environmental and socio-economic variables over space and time. Early published works usually assumed constant second and fourth moments. In this paper, we propose a new spatial time series model with dynamic variance and kurtosis. A distinctive feature of our proposed model is that for variables of interest, the model allows the variability and tail heaviness (which usually are indicated by the level of leptokurtosis) to change over spatial location and time. We establish Bayesian inference for the proposed model and conduct a simulation study to showcase the model's effectiveness compared with that of a baseline model. Air pollution data from Hong Kong and China's delta region are analyzed to further illustrate the dynamic variance behavior over time and the heavy-tailed characteristics of observations.

1. Introduction

Data related to climate or the environment often exhibit spatial and temporal dependence. That dependence has prompted the development of spatial–temporal models to efficiently account for both effects, such as the work of Sahu et al. (2006) modeling spatial–temporal dependence of fine particulate matter, that of Sahu et al. (2007) on ozone modeling, that of Wang et al. (2011) studying

* Corresponding author.
E-mail addresses: amandachu@hsmc.edu.hk (A.M.Y. Chu), wmli@connect.ust.hk (R.W.M. Li), immkpso@ust.hk (M.K.P. So).

https://doi.org/10.1016/j.spasta.2018.05.002
2211-6753/

spatial–temporal characteristics of precipitation in China, and that of Torabi (2013) on spatial–temporal modeling for disease mapping. A common approach in capturing the space–time dynamic of variables is based on the conditional autoregressive model introduced by Besag (1974). Wikle and Cressie (1999) proposed a space–time Kalman filter for prediction over space and time. Another common approach is Bayesian hierarchical modeling (Wikle et al., 1998, 2001; Sengupta and Cressie, 2013; Banerjee et al., 2014). The idea of Bayesian hierarchical modeling (Sang and Gelfand, 2009) can be seen in the modeling of heavy precipitation (Gaetan and Grigoletto, 2007; Ghosh and Mallick, 2011; Wang and So, 2016), pollutant exposure (Bobb et al., 2013; Wilson et al., 2014), storms (Economou et al., 2014), and sea surface temperatures (Lemos and Sansó, 2009). When one is capturing temporal dependence in spatial data, latent time series processes are naturally involved. The time series can have time-dependent parameters at the data level (Banerjee et al., 2014) or spatial–temporal random effects (Sang and Gelfand, 2009). An important question to ask, when analyzing climate or environmental data, is whether spatial–temporal dependence exists not just at the mean level but also at the variance level. The primary focus of previous environmental research has usually been on the mean level of data (Banerjee et al., 2008; Sahu and Mardia, 2005; Sang and Huang, 2012). Inclusion of conditional heteroskedasticity at the spatial–temporal level has an interesting potential to enhance our understanding of changing environmental phenomena. A spatial or temporal clustering of the variance of observations can be vital for practitioners wishing to more accurately assess risk and the duration of abnormal levels of data.

Another important question in analyzing space–time data is how best to model spatial extremes. Previous studies have tended to develop models for extreme climate or environmental events (e.g., extreme rainfalls, heavy storms, pollutants, and temperatures) with a hierarchical Gaussian process or constant variance. Gelfand et al. (2005) introduced Dirichlet process mixing for extreme rainfall spatial data. Cooley et al. (2007) introduced a hierarchical model for data on extreme intensity and frequency of rain. They used a latent spatial process in hierarchical Bayesian settings to model the number of exceedances of rainfall observations among Colorado stations. Research suggests that there will be significant trends of increasingly frequent extreme climate and environmental events, such as rainfall, due to the warming environment (Goswami et al., 2006). Other work on spatial extreme modeling includes that of Davison and Smith (1990) on models for exceedances over high thresholds, that of Davison et al. (2012), who reviewed statistical modeling methods for spatial extremes, that of Huser and Davison (2014) and of Genton et al. (2015), who used max-stable processes to model space–time and multivariate extremes, respectively, and that of Xu and Genton (2017), who proposed the Tukey g-and-h random field. The commonly used Gaussian process formulation is easy to implement. However, the normality assumption often may be too restrictive in real data modeling. Therefore, researchers have developed several non-Gaussian techniques to relax the normality assumptions, such as spatial quantile regression (Reich et al., 2012), mixture representation (Fonseca and Steel, 2011a,b; Kaiser et al., 2002; Zhang and El-Shaarawi, 2010), and the spatial linear model (Palacios and Steel, 2006). In addition, Krupskii and Genton (2017) proposed a factor copula model for analyzing spatial–temporal data. Huser et al. (2017) modeled spatial extremes using Gaussian scale mixtures. Opitz et al. (2018) considered Bayesian tail regression, which can handle extreme spatial–temporal quantiles. However, the issue of time-varying variance and tail heaviness still has not been addressed. This paper aims to fill that gap by proposing a spatial–temporal model with dynamic variance and leptokurtosis (i.e., with a fat- or heavy-tailed distribution relative to a normal distribution).

In this paper, we introduce a general hierarchical framework for a non-Gaussian spatial–temporal model with space–time-dependent variance. We assume that the spatial time series $Y_t(s)$ follows a Student's t distribution, given the mean $\mu_t(s)$, variance $\sigma_t^2(s)$, and degrees of freedom $\nu_t(s)$. The Student's t distribution has a heavier tail than the normal distribution does (i.e., it is leptokurtic), and that can account for severe outliers or extreme events. In addition to the spatial–temporal dependence in $\mu_t(s)$, we incorporate the spatial–temporal dependence for $\sigma_t(s)^2$ and kurtosis $\gamma_t(s)$ by including their respective proxies $(Y_{t-1}(s) - \mu_{t-1}(s))^2$ and $((Y_{t-1}(s) - \mu_{t-1}(s))/\sigma_{t-1}(s))^4$ at time $t-1$ in the formulation. This inclusion can capture the different behaviors of the observation $Y_t(s)$ over spatial location s and time t in terms of dynamic variance and changing tail property. In other words, our main contribution is to add flexibility into the modeling approach by incorporating fat tails with time-varying variance and kurtosis, which are characteristics of many real datasets.