Spatial Statistics 18 (2016) 86–104

Contents lists available at ScienceDirect

Spatial Statistics

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

Spatial statistics and Gaussian processes: A beautiful marriage

Alan E. Gelfand [a,*], Erin M. Schliep [b,1]

[a] *Department of Statistical Science, Duke University, Box 90251, Durham, NC, 27708, USA*

[b] *Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO, 65211, USA*

ARTICLE INFO

Article history:
Received 30 December 2015
Accepted 30 March 2016
Available online 7 April 2016

Keywords:
Data fusion
Directional data
Hierarchical model
Multivariate processes
Spatial Dirichlet processes
Spatial gradients

ABSTRACT

Spatial analysis has grown at a remarkable rate over the past two decades. Fueled by sophisticated GIS software and inexpensive and fast computation, collection of data with spatially referenced information has increased. Recognizing that such information can improve data analysis has led to an explosion of modeling and model fitting.

The contribution of this paper is to illustrate how Gaussian processes have emerged as, arguably, the most valuable tool in the toolkit for geostatistical modeling. Apart from the simplest versions, geostatistical modeling can be viewed as a hierarchical specification with Gaussian processes introduced appropriately at different levels of the specification. This naturally leads to adopting a Bayesian framework for inference and suitable Gibbs sampling/Markov chain Monte Carlo for model fitting.

Here, we review twenty years of modeling work spanning multivariate spatial analysis, gradient analysis, Bayesian nonparametric spatial ideas, directional data, extremes, data fusion, and large spatial and spatio-temporal datasets. We demonstrate that Gaussian processes are the key ingredients in all of this work. Most of the content is focused on modeling with examples being limited due to length constraints for the article. Altogether, we are able to conclude that spatial statistics and Gaussian processes do, indeed, make a beautiful marriage.

* Corresponding author. Tel.: +1 919 668 5229.
E-mail addresses: alan@stat.duke.edu (A.E. Gelfand), schliepe@missouri.edu (E.M. Schliep).

1 Tel.: +1 573 882 4455.

http://dx.doi.org/10.1016/j.spasta.2016.03.006
2211-6753/

1. Introduction

Spatial statistics has had an unusual history as a field within the discipline of Statistics. The stochastic process theory underlying much of the field was developed by probabilists, whereas, early on, much of the statistical methodology was developed quite independently and informally. In fact, this methodology grew primarily from the different areas of application, e.g., mining engineering, agriculture, and forestry.

As a result, for many years, spatial statistics labored on the fringe of mainstream statistics. However, the past twenty years has seen an explosion of interest in space and space–time problems. This has been largely fueled by the increased availability of inexpensive, high speed computing (as has been the case for many other areas). Such availability has enabled the collection of large spatial and spatio-temporal datasets across many fields, leading to widespread usage of sophisticated geographic information systems (GIS) software to create attractive displays along with the ability to investigate (fit and infer under) challenging, evermore appropriate and realistic models. As a result, spatial statistics has been brought into the mainstream of statistical research, changing from a somewhat ad hoc field to one that is more model-driven.

Full specification of stochastic models for the spatial process being investigated enables full inference and uncertainty assessment regarding the process. Gaussian processes (GPs) on $\mathcal{R}^2$ have become a fundamental specification in such modeling, particularly in settings where prediction is a primary goal.

Here, we focus primarily on geostatistical models, i.e., point-referenced data models. Apart from the simplest versions, such geostatistical modeling can be viewed as a hierarchical specification, with Gaussian processes introduced appropriately at different levels of the specification. Adoption of a Bayesian framework for inference and suitable Gibbs sampling/Markov chain Monte Carlo (MCMC) for model fitting follows. This is not surprising since, more generally, hierarchical modeling has emerged as the modeling paradigm for scientific work in the 21st century (e.g., Gelfand and Ghosh, 2013) and spatial statistics in particular (Gelfand, 2012).

Over the past twenty years there has been an enormous growth in such modeling. The contribution here is to highlight the substantial range of spatial settings where Gaussian processes have enabled rich and flexible specification. While our focus is on the geostatistical spatial setting, we note the importance of Gaussian processes in modeling spatial point patterns, e.g., log Gaussian Cox processes (Banerjee et al., 2014; Moller and Waagepetersen, 2003) or with lattice, grid, and areal data models using Gaussian Markov random fields (Rue and Held, 2005).

Even confining ourselves solely to the geostatistical setting, there is still too much to cover. What we offer is a review of the basic geostatistical model in hierarchical form. This leads to generalized linear spatial regression models and multivariate process models including spatially varying coefficient models. We also briefly mention simple extensions of Gaussian processes that enable more flexible process specifications. Next, we turn to elegant gradient analysis to study directionality in random realizations of spatial surfaces. Then, we consider nonparametric distributional models for spatial data. Here, rather than interpolating realizations at unobserved locations, we interpolate random distributions at unobserved locations. We next discuss spatial extremes, spatial directional data, and data fusion. We conclude with the use of Gaussian processes to accommodate large datasets.

The format of the paper is as follows. In Section 2 we clarify why Gaussian processes prove so attractive. In Section 3 we review the customary geostatistical model, viewing it hierarchically, with either a Gaussian or non-Gaussian first stage specification. In Section 4 we consider several extensions of Gaussian processes to enhance their flexibility. Section 5 considers multivariate Gaussian spatial processes while Section 6 looks at the gradient behavior associated with realizations of Gaussian processes. Section 7 turns to nonparametric extensions of Gaussian processes using Dirichlet processes. Section 8 highlights modeling of spatial extremes while Section 9 examines spatial directional data. Section 10 looks at data fusion or data assimilation and Section 11 concludes with two Gaussian process models for large spatial datasets. We give a brief summary in Section 12.