Spatial Statistics 1 (2012) 30–39

Contents lists available at SciVerse ScienceDirect

Spatial Statistics

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

Hierarchical modeling for spatial data problems

Alan E. Gelfand

Department of Statistical Science, Duke University, Durham, NC, 27708-0251, USA

ARTICLE INFO

Article history:
Received 3 January 2012
Accepted 27 February 2012
Available online 10 March 2012

Keywords:
Data fusion
Directional data
Dirichlet processes
Extreme values
Kernel predictors
Species distributions

ABSTRACT

This short paper is centered on hierarchical modeling for problems in spatial and spatio-temporal statistics. It draws its motivation from the interdisciplinary research work of the author in terms of applications in the environmental sciences—ecological processes, environmental exposure, and weather modeling. The paper briefly reviews hierarchical modeling specification, adopting a Bayesian perspective with full inference and associated uncertainty within the specification, while achieving exact inference to avoid what may be uncomfortable asymptotics. It focuses on point-referenced (geo-statistical) and point pattern spatial settings. It looks in some detail at problems involving data fusion, species distributions, and large spatial datasets. It also briefly describes four further examples arising from the author's recent research projects.

1. Introduction

At the outset, I am delighted to be invited to contribute this short paper profiling my research interests for this promising new journal. The time is right for a journal devoted to spatial statistics. I note that spatial statistics has experienced an unusual evolution as a field within the discipline of Statistics. The stochastic process theory that underlies much of the field was developed within the mathematical sciences by probabilists, whereas, early on, much of the statistical methodology was developed quite independently. In fact, this methodology grew primarily from the different areas of application: mining engineering leading to the development of geostatistics by Matheron and colleagues, agriculture with spatial considerations owing to the thinking of Fisher on randomization and blocking, and forestry which motivated the seminal Ph.D. Thesis of Matérn. As a result, for many years, spatial statistics labored on the fringe of mainstream statistics. However, the past twenty years have seen an explosion of interest in space and space–time problems. This has been largely fueled by

E-mail address: alan@stat.duke.edu.

2211-6753/$ – see front matter
doi:10.1016/j.spasta.2012.02.005

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, has facilitated the widespread usage of sophisticated geographic information systems (GIS) software to create attractive displays, and has enabled inferential investigation of challenging, evermore appropriate and realistic models.

In the process, spatial statistics has been brought into the mainstream of statistical research with a proliferation of books, conferences and workshops, courses and short courses, and an exciting new journal! Moreover, while there has been a body of strong theoretical work developed since the 1950s, it is safe to say that, broadly, spatial statistics has changed from a somewhat ad hoc field to a more model-driven one. Hence stochastic modeling, in particular, hierarchical modeling is the primary focus of my article, as well as a reflection of my contributions to the field. It is the opportunity to (i) frame flexible stochastic models, (ii) specify models that capture important features of complex processes, and (iii) add a full inference engine to the lovely displays that can be produced with GIS software, that has driven my research in spatial statistics. It is my objective here to give you a bit of the flavor of this work. As requested by the editor, in profiling my research activity, the associated references will primarily supply papers I have been involved with.

1.1. *A paradigm shift*

As we move into the second decade of the 21st century, we are witnessing a dramatic paradigm shift in the way that statisticians collaborate with researchers from other disciplines. Disappearing are the days when the statistician was called in at the end of a project to provide some routine data analysis and some summary displays. Now the statistician is an integral player in a research team, helping to formulate hypotheses, identify data needs, develop suitable stochastic models, and implement fitting of and inference from the resulting challenging models. Altogether, the statistician becomes sufficiently knowledgeable in the subject matter to "walk the walk" and "talk the talk", adding another scientific dimension to her/his skill set.

As part of this shift, there is an increasing attention paid to bigger picture science, to looking at complex processes with an integrative perspective, and to bringing a range of knowledge to this effort. Increasingly, we find researchers working with observational data, less with designed experiments, recognizing that the latter can help inform about the former but the gathering of such experiments provides only one source of data for learning about the complex process. Other information sources, empirical, theoretical, physical, etc., will also be included in the synthesis.

The primary result of all of this is the development of a multi-level stochastic model. Such models are well-suited for incorporating the foregoing knowledge, allowing it to be inserted at various levels of the modeling, as appropriate. Following the vision of Berliner (1996), we imagine a three stage hierarchical specification:

$$\text{First stage}: [data|process, parameters]$$
$$\text{Second stage}: [process|parameters]$$
$$\text{Third stage}: [(hyper)parameters].$$

The simple form of this specification belies its breadth. The process component can include multiple levels. For our interests here, it can be spatial and it can be dynamic. The data can be conditioned on whatever aspects of the process are appropriate. The stochastic forms can be multivariate, perhaps infinite dimensional with parametric and/or nonparametric specifications.

In principle, a hierarchical model can be *flattened* by suitable marginalization/integration. However, the advantage of the hierarchical form lies in convenience of specification, ease of interpretation and, often, in facilitation of model fitting. Furthermore, by recognizing the uncertainty in the model unknowns, uncertainty is properly propagated to inference arising from the model.

In view of the above, hierarchical modeling has taken over the landscape in contemporary stochastic modeling. Though analysis of such modeling can be attempted through non-Bayesian approaches, working within the Bayesian paradigm enables exact inference and proper uncertainty assessment within the given specification.