Handbook of Environmental and Ecological Statistics

Alan Gelfand, Montse Fuentes, Jennifer A. Hoeting, Richard L. Smith

Introduction

Publication details


Alan E. Gelfand, Montserrat Fuentes, Jennifer Hoeting, Richard L. Smith

Published online on: 15 Sep 2017

How to cite: Alan E Gelfand, Montserrat Fuentes, Jennifer Hoeting, Richard L. Smith. 15 Sep 2017, Introduction from: Handbook of Environmental and Ecological Statistics CRC Press

Accessed on: 07 Dec 2023

Introduction

Alan E Gelfand  
Duke University

Montserrat Fuentes  
Virginia Commonwealth University

Jennifer Hoeting  
Colorado State University

Richard L. Smith  
University of North Carolina Chapel Hill

The papers in Section I focus on methodology. The intent is to present a collection of approaches that span a wide area of statistical thinking but also a broad range of methodology needed for the challenges that arise in working with environmental and ecological data. Each of these chapters is fairly formal, somewhat technical, elaborating properties and behaviors but also suggestive and, in some cases, inclusive of application. The level of the papers varies technically but an effort has been made to make them quite accessible and focused.

Chapter 2 (Gelfand) is a modeling chapter, discussing the basics of modeling for environmental processes, broadly reviewing stochastic modeling, hierarchical modeling, capturing uncertainty, model adequacy, and model comparison. Chapter 3 (Craigmile) takes up time series methodology, in particular, classical methods, long-range dependence and change points. Chapter 4 (Schmidt and Lopes) turns to dynamic models, opening up state space ideas, stochastic partial differential equations, integro-difference equations, projection ideas, along with discretization for implementation.

Next, we move to spatial approaches. Chapter 5 (Banerjee) takes up geostatistical modeling for environmental processes, reviewing the basic tools, e.g., variograms, covariance functions, kriging, particularly working with Gaussian processes. Chapter 6 (Illian) considers point patterns, offering basic theory and inference beginning with one dimensional models. However, primarily, it focuses on point patterns in two dimensions, again with associated inference. Chapter 7 (Berrocal) reviews data fusion, summarizing the literature from climatology. It then considers fusing environmental data sources, in particular, combining deterministic models with observations, using, e.g., melding and downscaling.

Chapter 8 (Cooley, Hunter, and Smith) turns to the analysis of extremes. Here, we start with basic theory but then move to multivariate data and spatially referenced extremes, citing illustrations for climate and for environmental exposure. Chapter 9 (Zimmerman and Buckland) focuses on techniques for environmental sampling. Topics include sampling for locations, e.g., for monitoring sites, geometric/space filling strategies, model-based strategies, sampling for field sites and transects with distance sampling. Chapter 10 (Clark and Gelfand) takes up a frequent challenge with ecological data, the problem of zeros. Zeros require care in terms of accommodating so-called zero-inflation settings but, more generally,
need attention in the context of multinomial trials and joint species distribution modeling.
Chapter 11 (Palmer), the last chapter of this section, takes up ordination methods, also
referred to as gradient analysis. Tools like principal components analysis (PCA), canonical
correlation analysis (CCA), and factor analysis are the bread and butter for clustering and
dimension reduction approaches in ecological data analysis.

Turning to Section II, by definition, ecology is the study of how organisms interact with
one another and with their surroundings. Ecologists seek to understand factors that impact
population size, the distribution of populations, and interactions with other species and
the environment. All of these factors can vary over space and time. There is a long history
of statisticians developing new methodology to address challenging problems in ecology.
Likewise, ecologists often use sophisticated statistical methods. The chapters in this section
provide a comprehensive overview of the most important and currently relevant topics in
ecological statistics.

There are many reasons for the close relationship between statisticians and ecologists.
Data collected to study ecological processes are often observational, messy, and can involve
very large data sets. Thus basic statistical approaches like ANOVA models are typically
inadequate. In addition, the field of theoretical ecology is closely linked to mathematical
biology—a field in which applied mathematical tools are used to address problems in biology.
Statistical models for ecological processes are often inspired by or directly incorporate these
theoretical models allowing for statistical inference on the parameters of the theoretical
models. Altogether, in order to develop new and useful statistical methods, a statistician
working on ecological problems needs to bring substantial understanding of the ecological
process under study as well as a broad knowledge of statistical methodology.

Species distribution models are a primary focus in the study of animal and plant popu-
lations in order to develop a better understanding of where species live and why. Chapter 12
(Ovaskainen) presents an overview including models for individual species as well as models
for more than one species. The focus in this chapter is on species-distribution models for
presence-absence data via generalized mixed models.

A related problem is population size estimation. Chapter 13 (Barker) surveys the vast
field of capture-recapture and distance sampling methodology to estimate population abun-
dance. Knowledge of the population under study as well as study design influence the choice
of model for abundance estimation. There are separate models for open and closed popu-
lations and for different study designs such as mark-recapture and tag-recovery designs. In
addition, sampling challenges must be addressed in the modeling such as imperfect detec-
tion of animals. Furthermore, these models allow for estimation of other key parameters of
interest such as birth and death rates, probability of survival, and population growth rates.

Technological advances have made it less expensive to collect fine-scale temporal data
on geo-referenced animal locations. In response, a rich set of statistical models for telemetry
data have been proposed to better understand movement of individual animals in space.
Chapter 14 (Hooten and Johnson) surveys parametric spatio-temporal animal movement
models including point-process as well as discrete- and continuous time-models. Models
of animal movement can be used improve understanding of the impact of human-caused
changes in habitat size and availability for animal populations or to better understand
migration patterns for bird, insect or mammal populations.

Understanding the population size and age structure of animal populations can aid in
basic understanding or improve efforts aimed at managing fishery and wildlife populations.
Chapter 15 (Newman) presents traditional as well state-of-the-art statistical methodology
for modeling the structure and dynamics of biological populations. Key models used in an-
imal demography are described including population dynamics, matrix projection, integral
projection, individual-based, and state-space models. This chapter also examines the various
statistical approaches used to estimate parameters for this wide-ranging set of models.
Introduction

In traits-based analyses, organisms are categorized by their biological attributes. For example, in order to understand the impact of stream disturbances, stream macroinvertebrate species may be categorized based on pollution sensitivity or tolerance to water temperature changes due to drought. Traits-based analyses allow ecologists to move beyond a narrow focus on one or two species, particularly useful when many species live in a given habitat type. Trait-based analyses are used by ecologists to better understand patterns of species occurrence, predict changes in community composition, and understand ecosystem function.

Chapter 16 (Aiello-Lammens and Silander) reviews statistical methods for trait-based ecological modeling. The authors describe data structures, exploratory data analyses methods as well as algorithmic approaches and statistical model-based inference for traits data.

The section on models for ecological processes closes with two chapters focusing on statistical models to address specific ecological problems. Chapter 17 (Pereira and Turkman) provides an overview of statistical models for vegetation fires like grass and forest fires. With the changing climate, forest fires have already or are expected to increase in size and intensity in many regions of the world, with the recent large fires in Canada and Australia as two examples. Careful modeling of spatial, temporal and spatio-temporal patterns of vegetation fires can help government planners to predict and manage future fires. The authors provide a comprehensive overview of models for fire size, fire incidence, and fire risk maps which have been used by ecologists, fire scientists and others to advance the field of fire science.

Similarly, the changing climate has impacted precipitation patterns and timing which in turn impacts stream flow patterns. Chapter 18 (Ver Hoef, Peterson, and Isaak) provides an in-depth study of models for stream networks. Stream data requires unique models because streams work differently from other physical processes. Stream data are spatially-referenced but come from physically-connected processes. It is not appropriate to use Euclidean distance to measure stream distance because two locations that are nearby in physical space may not be part of the same stream network. Using hydrological distance in a standard geostatistical model has been proposed as a solution to this problem, but naive use of hydrological distance in geostatistical models can produce non-positive covariance matrices and other mathematical inconsistencies. Thus neither standard time-series nor standard spatial statistical models can be applied to data collected on stream networks. The authors describe models specifically developed to address stream-network construction. They show that such models can produce accurate predictions that are consistent with the physical processes of streams.

For Section III of this handbook we include eight chapters offering a comprehensive coverage of topics in environmental exposure. We include standard and state-of-art methods and models for environmental exposure, while illustrating the methods in the context of different case studies.

Chapter 19 (Fuentes, Reich and Huang) is an introductory chapter that defines exposure and introduces methods for mapping across space and time of ground data, going from inverse distance weighting to Bayesian interpolation. It then follows with approaches for data fusion to combine ground data, satellite data, numerical model output and other sources of relevant information while characterizing errors and uncertainty in all sources of data.

Chapter 20 (Bell and Warren) introduces two alternative models for estimating environmental exposure: land use regression and a stochastic model that simulates human exposure. This chapter discusses limitations and opportunities for enhancement of these alternative methods. Chapter 21 (Diggle and Giorgi) illuminates limitations in our exposure assessment due to sampling bias.

Chapter 22 (Zidek and Zimmerman) introduces a very comprehensive review of the principled statistical methods and approaches for monitoring network design. Many air pollution and environmental stressors are associated to different sources. Source apportionment aims
to decompose the observed mixture of environmental stressors, as represented by measurements of individual constituents, into information about the sources that contribute to the mixture. Chapter 23 (Krall and Chang) describes statistical methods and the challenges for source apportionment. The concept of confounding is critical in providing estimates of health effects that are unbiased and precise. Chapter 24 (Dominici and Wilson) presents a high level overview of the important concept of confounding in environmental health studies.

Chapter 25 (Szpiro) presents an overview of statistical methods applied to exposure assessment, describing statistical theory on the consequences for health effect inferences of measurement error resulting from using predicted rather than true exposures, while presenting methods to optimize health effects inference. Chapter 26 (Sheppard), the final chapter in the environmental exposure section, focuses on study designs for environmental epidemiology, as a fundamental aspect of exposure assessment.

The chapters in Section IV of this Handbook concern statistical methods in climate science. Climate science has long been concerned with large datasets databases of historical weather data now cover tens of thousands of stations that in many cases contain daily data from the mid nineteenth century onwards. Added to that, there are many new sources of remote sensing data such as NASA’s OCO2 satellite, launched in 2014, which measures near-surface carbon dioxide data at resolutions down to a kilometer. On the modeling side, climate model datasets have grown in little more than a decade from 40 terabytes (the CMIP3 dataset, completed in 2006) to 2 petabytes (CMIP5, 2012) to an estimated 10-30 petabytes for the CMIP6 dataset that is projected to become available in 2019. These datasets are compilations of climate model simulations from centers all over the world and play a major role in the successive reports of the Intergovernmental Panel on Climate Change. These large datasets pose major challenges for statistical analysis.

Chapter 27 by Craigmile and Guttorp addresses one of the oldest but still very critical questions in climate science: how does one measure and determine the size of trends? This chapter focuses on classical methods based on time series analysis but gives a particular focus to the problem of measurement error in the observed series. Chapter 28 by Stephenson is a review of climate models and the kind of datasets they produce, and includes a short review of the multi-model ensemble problem, which is about combining the output of different climate models for prediction purposes.

Chapter 29 by Nychka and Wikle discusses spatial statistics given that most climate datasets are indexed in space as well as time, this is a key statistical methodology for anyone hoping to do statistical research in this field. Chapter 30, by Budhiraja and colleagues, is about data assimilation, which is the process of combining climate and weather models with observational data so that the unobserved state of the system is continuously updated. Classical applied mathematics techniques such as 3D-VAR and 4D-VAR are compared with more statistically oriented approaches such as particle filtering. These methods have long been used for numerical weather forecasting but are finding increasing application for long-term climate prediction as well.

Chapter 31, by Davison and colleagues, is a continuation of Chapter 8 on extremes, in this case, focused on spatial extremes. This chapter describes the modern class of stochastic process models known as max-stable processes, including the estimation methods of composite likelihood and exact likelihood analysis, and concludes with two examples applied to large spatiotemporal datasets. Chapter 32 by Wikle discusses the use of statistics in oceanography, covering both time series and spatial statistics methods and illustrating how Bayesian hierarchical models may be used for predicting processes such as El Niño. Chapter 33, by Craigmile and colleagues, also demonstrates the huge power of Bayesian hierarchical models to combine data from different sources, here applied to the problems of paleoclimatology, which is about the use of proxy datasets such as tree rings and ice cores.
to reconstruct historical temperature records before modern measurement techniques were available.

Chapter 34 by Hammerling and colleagues is about detection and attribution, the technique that exists for deciding whether trends in climate data are directly associated with anthropogenic signals (the detection problem) and, if so, how the trend may be apportioned among the different anthropogenic and natural sources of variation (attribution). They also touch upon problems of extreme event attribution, which is concerned with how much specific extreme events (such as the successive hurricanes that devastated parts of Central and North America during the summer of 2017) are associated with anthropogenic sources of climate change. Finally in Chapter 35, Ebi and co-authors have provided an expert review of many different aspects of climate and health, covering both vector-borne diseases such as malaria and the direct health effects of extreme weather such as heatwaves, and address many of the interdisciplinary challenges that this field raises.