Handbook of Spatial Epidemiology

Paula MORAGA

King Abdullah University of Science and Technology (KAUST)

Computer, Electrical and Mathematical Sciences and Engineering Division

4700 King Abdullah University of Science and Technology

Thuwal 23955-6900, Saudi Arabia

+966 12 808 0900

paula.moraga@kaust.edu.sa

Andrew B. Lawson, Sudipto Banerjee, Robert P. Haining, and María Dolores Ugarte, eds. Boca Raton, FL: Chapman & Hall/CRC Press, 2016, xviii+684 pp., $180.00(H), ISBN: 978-1-48-225301-6.

The Handbook of Spatial Epidemiology is an edited anthology of chapters written by experts in the field that cover a wide range of methods used to understand the geographic and temporal patterns of health events, assess their relationship with potential risk factors, detect clusters, measure inequalities, and evaluate the impact of interventions. It can serve as an excellent reference for statisticians interested in the development of spatial epidemiology methods and their application. The handbook provides a thorough catalog of important theoretical models and a number of well-chosen applications, as well as a detailed discussion of the specific issues that need to be taken into account in spatial epidemiology investigations. The models are presented at an advanced level, so in general the handbook will be most useful to those with considerable statistical training. However, some of the introductory chapters are broad enough to be approachable to students, researchers and public health specialists.

The topics covered by the handbook are vast. A range of Bayesian hierarchical disease mapping models are clearly presented, as well as kernel methods and smoothing techniques based on splines. These are deployed to analyze both case event data that represent residential locations of cases, and count data that arise when case events are aggregated into a finite number of areal units such as counties or municipalities. Moreover, a number of important extensions are presented including survival models to analyze time-to-event data, longitudinal models to analyze the variation of individual responses over time, and multivariate models to jointly analyze multiple health outcomes.

The handbook also provides a thorough treament of clustering methods, which can be divided in global methods that assess the existence of excess risk in the whole study area, local methods such as scan statistics that identify the spatial extent of the clusters and the populations affected, and focused clustering methods that assess whether there is elevated risk near of a suspected source. A comprehensive overview of issues central to the interpretation of clustering studies is provided emphasizing that before determining a statistically significant cluster found is real, a plausible biological explanation needs to be found, and a number of issues such as incomplete reporting and geocoding location errors need to be accounted for as they may reduce the power to detect true clusters or yield spurious cluster findings.

Geostatistical models use disease data that have been collected at a set of specific locations together with risk factors such as environmental and climatic variables to predict disease risk at unobserved locations. The book describes Bayesian variable selection methods to choose an appropriate subset of covariates among many potential predictors as well as determining the functional forms and interactions among them to capture the relationships between the response and the covariates in geostatistical models (Chapter 23). A useful addition to the book could have been a presentation of machine learning

approaches that more fully exploit the information contained in the covariates and may yield improved prediction accuracy.

Early detection of disease outbreaks is essential to facilitate timely public health response. The book describes a number of important spatio-temporal transmission models for the person-to-person spread of infectious diseases that enable to understand the dynamics of disease transmission, predict epidemics, assess the risk of emerging pathogens, and evaluate control measures. These models are crucial to inform the response to diseases such as influenza, SARS and more recently COVID-19.

Like any edited anthology, the benefit of reading from a wide range of authors is weighed against the lack of a single narrative voice. The book would have been an easier read if authors covering the same topic (e.g., disease mapping, clustering) had been synthesized into a single chapter. As is, there is a fair amount of repetition: Bayesian disease risk models, for example, are explained several times. The anthology form also means that there is little discussion of how the presented methods compare when tasked with solving a specific problem.

An interesting problem that arises in many spatial applications is how to combine data that are available at different spatial resolutions to obtain better inferences. This book presents several approaches to deal with this problem in the context of air pollution. Specifically, Bayesian melding assumes that data obtained at different spatial resolutions arise from a common latent process, and in statistical calibration one of the variables is viewed as a predictor for the other measurements. The authors note that the implementation of these approaches using MCMC may involve considerable computational effort. An alternative approach to obtain faster inferences is to use INLA and SPDE (Moraga et al. 2017).

Spatial epidemiological analyses are subject to a number of issues that may impact the interpretation of results and this book contains strategies to avoid making misleading conclusions. Ecological bias is a common challenge in studies that investigate the individual-level association between health outcomes and exposures using area-level data. The book discusses the limitations of such ecological data and provides a number of proposals to address this bias. Another common issue arises when residential locations at the time of diagnosis are used as a surrogate for unmeasured environmental exposures. For diseases with a long latency, these locations have minimal relevance. The book presents approaches that account for this latency through residential histories, a strategy that can also address population mobility.

The spatial epidemiology methods presented in the book can be applied using data obtained from traditional data sources such as registries and surveys. Currently, new data sources are available that can be used to complement data obtained from traditionally sources. For example, health-related social media posts and search engine query logs can be used to understand disease activity levels in real-time. These new data sources and associated methods have the potential to provide improved insights for decision-making in many settings.

In recent years, the spatial capability of the statistical software R has dramatically increased. The book presents the powerful R package `sp` as the primary tool to manipulate, analyze and visualize spatial data. For fitting Bayesian hierarchical models using MCMC the book turns to WinBUGS, a well-established piece of software, for the integrated nested Laplace approximation (INLA) uses `R-INLA`. In the years since the book's publication, the computational toolbox has continued to expand. The `sp` package has been augmented with the `sf` package, which provides a simple and flexible interface that will be familiar to users of the `tidyverse`. Mapping in R has continued to improve through the development of `tmap` and `leaflet`. The probabilistic programming language Stan and the associated R package `rstan` allows more efficient Bayesian inferences using

Hamiltonian Monte Carlo. These packages would be a welcome addition to a second edition of this book.

Although the methodology presented in the book is well-explained, the code to reproduce most of the approaches is not provided. This is an unfortunate omission, as it would have provided means for researchers and practitioners to apply the methods presented to their own investigations easily and unambiguously. It also would have aided instructors who wish to pull chapters from this book for use in an applied course.

Overall, this book provides an overview of the rich variety of spatial epidemiology methods and their application. For readers coming from general statistics, the attention paid to issues such as misalignment, ecological bias, and the interpretation of results will be particularly useful. In the years since the publication of this handbook, the field has continued to evolve. A present survey of the discipline would include a discussion of new data sources, machine learning approaches, and modern software packages for spatial analysis, as well as a supplement containing the code necessary to apply the methods presented. With these additions, this handbook stands as a valuable resource for researchers and professionals interested in solving new challenges in spatial epidemiology and public health surveillance.

Paula Moraga

King Abdullah University of Science and Technology (KAUST)

REFERENCES

Moraga, P., Cramb, S.M., Mengersen, K.L., and Pagano, M. (2017), “A Geostatistical Model for Combined Analysis of Point-Level and Area-Level Data using INLA and SPDE,” *Spatial Statistics*, vol. 21A, 27–41.