

Preface

We live in a complex world, and clever people are continually coming up with new ways to observe and record increasingly large parts of it so we can comprehend it better (warts and all!). We are squarely in the midst of a “big data” era, and it seems that every day new methodologies and algorithms emerge that are designed to deal with the ever-increasing size of these data streams.

It so happens that the “big data” available to us are often *spatio-temporal data*. That is, they can be indexed by spatial locations and time stamps. The space might be geographic space, or socio-economic space, or more generally network space, and the time scales might range from microseconds to millennia. Although scientists have long been interested in spatio-temporal data (e.g., Kepler’s studies based on planetary observations several centuries ago), it is only relatively recently that statisticians have taken a keen interest in the topic. At the risk of two of us being found guilty of self-promotion, we believe that the book *Statistics for Spatio-Temporal Data* by Cressie and Wikle (2011) was perhaps the first dedicated and comprehensive statistical monograph on the topic. In the decade (almost) since the publication of that book, there has been an exponential increase in the number of papers dealing with spatio-temporal data analysis – not only in statistics, but also in many other branches of science. Although Cressie and Wikle (2011) is still extremely relevant, it was intended for a fairly advanced, technically trained audience, and it did not include software or coding examples. In contrast, the present book provides a more accessible introduction, with hands-on applications of the methods through the use of R Labs at the end of each chapter. At the time of writing, this unique aspect of the book fills a void in the literature that can provide a bridge for students and researchers alike who wish to learn the basics of spatio-temporal statistics.

What level is expected of readers of this book? First, although each chapter is fairly self-contained and they can be read in any order, we ordered the book deliberately to “ease” the reader into more technical material in later chapters. Spatio-temporal data can be complex, and their representations in terms of mathematical and statistical models can be complex as well. They require a number of indices (e.g., for space, for time, for multiple variables). In addition, being able to account for dependent random processes requires a bit of statistical sophistication that cannot be completely avoided, even in an applications-based introductory book. We believe that a reader who has taken a class or two in calculus-based prob-

ability and inference, and who is comfortable with basic matrix-algebra representations of statistical models (e.g., a multiple regression or a multivariate time-series representation), could comfortably get through this book. For those who would like a brief refresher on matrix algebra, we provide an overview of the components that we use in an appendix. To make this a bit easier on readers with just a few statistics courses on their transcript, we have interspersed “technical notes” throughout the book that provide short, gentle reviews of methods and ideas from the broader statistical literature.

Chapter 1 is the place to start, to get you intrigued and perhaps even excited about what is to come. We organized the rest of the book to follow what we believe to be good statistical practice. First, look at your data and do exploratory analyses (Chapter 2), then fit simple statistical models to the data to indicate possible patterns and see if assumptions are violated (Chapter 3), and then use what you learned in these analyses to build a spatio-temporal model that allows valid inferences (Chapters 4 and 5). The end of the cycle is to evaluate your model formally to find areas of improvement and to help choose the best model possible (Chapter 6). Then, if needed, repeat with a better-informed spatio-temporal model.

The bulk of the material on spatio-temporal modeling appears in Chapters 4 and 5. Chapter 4 covers descriptive (*marginal*) models formed by characterizing the spatio-temporal dependence structure (mainly through spatio-temporal covariances), which in turn leads to models that are analogous to the ubiquitous geostatistical models used in kriging. Chapter 5 focuses on dynamic (*conditional*) models that characterize the dynamic evolution of spatial processes through time, analogous to multivariate time-series models. Like Cressie and Wikle (2011), both Chapters 4 and 5 are firmly rooted in the notion of *hierarchical thinking* (i.e., hierarchical statistical modeling), which makes a clear distinction between the data and the underlying latent process of interest. This is based on the very practical notion that “[w]hat you see (data) is not always what you want to get (process)” (Cressie and Wikle, 2011, p. xvi).

Spatio-temporal statistics is such a vast field and this modestly sized book is necessarily not comprehensive. For example, we focus primarily on data whose spatial reference is a point, and we do not explore issues related to the “change-of-support” problem, nor do we deal with spatio-temporal point processes. Further, we mostly limit our discussion to models and methodologies that are relatively mature, understood, and widely used. Some of the applications our readers are confronted with will undoubtedly require cutting-edge methods beyond the scope of this book. In that regard, the book provides a down-to-earth introduction. We hope you find that the path is wide and the slope is gentle, ultimately giving you the confidence to explore the literature for new developments. For this reason, we have named our epilogical chapter *Pergimus*, Latin for “let us continue to progress.”

A substantial portion of this book is devoted to “Labs,” which enable the reader to put his or her understanding into practice using the programming language R. There are several reasons why we chose R: it is one of the most versatile languages designed for statistics; it is open source; it enjoys a vibrant online community whose members post

solutions to virtually any problem you will encounter when coding; and, most importantly, a large number of packages that can be used for spatio-temporal modeling, exploratory data analysis, and statistical inference (estimation, prediction, uncertainty quantification, and so forth) are written in R. The last point is crucial, as it was our aim right from the beginning to make use of as much tried-and-tested code as possible to reduce the analyst’s barrier to entry. Indeed, it is fair to say that this book would not have been possible without the excellent work, openness, and generosity of the R community as a whole.

In presenting the Labs, we intentionally use a “code-after-methodology” approach, since we firmly believe that the reader should have an understanding of the statistical methods being used before delving into the computational details. To facilitate the connections between methodology and computation, we have added “R Tips” where needed. The Labs themselves assume some prior knowledge of R and, in particular, of the *tidyverse*, which is built on an underlying philosophy of how to deal with data and graphics. Readers who would like to know more can consult the excellent book by Wickham and Grolemund (2016) for background reading (freely available online).

Finally, our goal when we started this project was to help as many people as we could to start analyzing spatio-temporal data. Consequently, with the generous support of our editors at Chapman & Hall/CRC, we have made the .pdf file of this book and the accompanying R package, **STRbook**, freely available for download from the website listed below. In addition, this website is a place where users can post *errata*, comment on the code examples, post their own code for different problems, their own spatio-temporal data sets, and articles on spatio-temporal statistics. You are invited to go to:

<https://spacetimewithr.org>

We hope you find this book useful for your endeavors as you begin to explore the complexities of the spatio-temporal world around us – and within us! Let’s get started . . .

Christopher K. Wikle
Columbia, Missouri, USA
Andrew Zammit-Mangion
Wollongong, NSW, Australia
Noel Cressie
Sydney, NSW, Australia