Preface

Spatial analysis methods have seen a rapid rise in popularity due to demand from a wide range of fields. These include, among others, biology, spatial economics, image processing, environmental and earth science, ecology, geography, epidemiology, agronomy, forestry and mineral prospection.

In spatial problems, observations come from a spatial process $X = \{X_s, s \in S\}$ indexed by a spatial set S, with X_s taking values in a state space E. The positions of observation sites $s \in S$ are either fixed in advance or random. Classically, S is a 2-dimensional subset, $S \subseteq \mathbb{R}^2$. However, it could also be 1-dimensional (chromatography, crop trials along rows) or a subset of $\mathbb{R}^3$ (mineral prospection, earth science, 3D imaging). Other fields such as Bayesian statistics and simulation may even require spaces S of dimension $d \geq 3$. The study of spatial dynamics adds a temporal dimension, for example $(s,t) \in \mathbb{R}^2 \times \mathbb{R}^+$ in the 2-dimensional case.

This multitude of situations and applications makes for a very rich subject. To illustrate, let us give a few examples of the three types of spatial data that will be studied in the book.

Geostatistical data

Here, S is a *continuous* subspace of $\mathbb{R}^d$ and the random field $\{X_s, s \in S\}$ observed at n fixed sites $\{s_1, \ldots, s_n\} \subset S$ takes values in a real-valued state space E. The rainfall data in Figure 0.1-a and soil porosity data in Fig. 0.1-b fall into this category. Observation sites may or may not be regularly spaced. Geostatistics tries to answer questions about modeling, identification and separation of small and large scale variations, prediction (or kriging) at unobserved sites and reconstruction of X across the whole space S.

Lattice data and data on fixed networks

Here, S is a *fixed discrete* non-random set, usually $S \subset \mathbb{R}^d$ and X is observed at points in S. Points s might be geographical regions represented as a network with
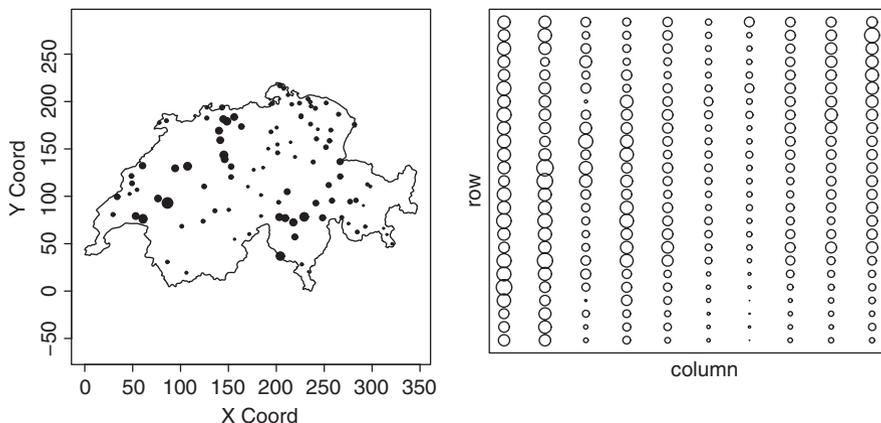
Fig. 0.1 (a) Rainfall over the Swiss meteorological network on May 8, 1986 (during the passage of Chernobyl's radioactive cloud. This is the `sic` dataset from the `geoR` package of *R* (178)); (b) Soil porosity (`soil` dataset from the `geoR` package). For both (a) and (b), the size of symbols are proportional to the value of X_s.
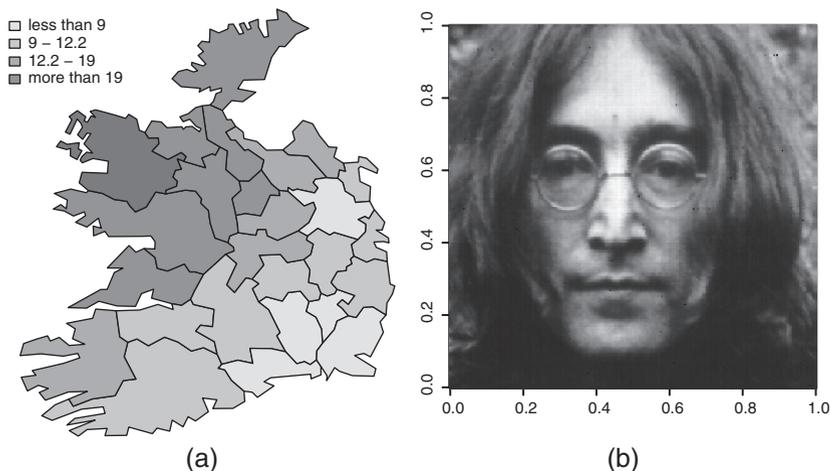

Fig. 0.2 (a) Percentage of people with blood group A in the 26 counties of Ireland (`eire` dataset from the `spdep` package); (b) Image of John Lennon (256×256 pixels in a 193-level grayscale, `lennon` dataset from the `fields` package).

given adjacency graph $\mathscr{G}$ (cf. the 26 counties of Ireland, Fig. 0.2-a) and X_s some value of interest measured at *s*. The state space *E* may or may not be real-valued. In image analysis, *S* is a regularly spaced set of pixels (cf. Fig. 0.2-b). Goals for these types of data include constructing and analyzing explicative models, quantifying spatial correlations, prediction and image restoration.

Point data

Figure 0.3-a shows the location of cell centers in a histological section seen under a microscope and Figure 0.3-b the location and size of pine trees in a forest. Here, the set of observation sites $x = \{x_1, x_2, \ldots, x_n\}$, $x_i \in S \subset \mathbb{R}^d$ is *random*, along with the number $n = n(x)$ of observation sites; x is the outcome of a spatial point process (PP) observed in window S. The process X is said to be *marked* if at each x_i we record a value, for example the diameter of the pine trees found at x_i. A central question in the statistical analysis of PPs is to know if the distribution of points is essentially regular (Figure 0.3-a), completely random (Poisson PP) or aggregated (Figure 0.3-b).
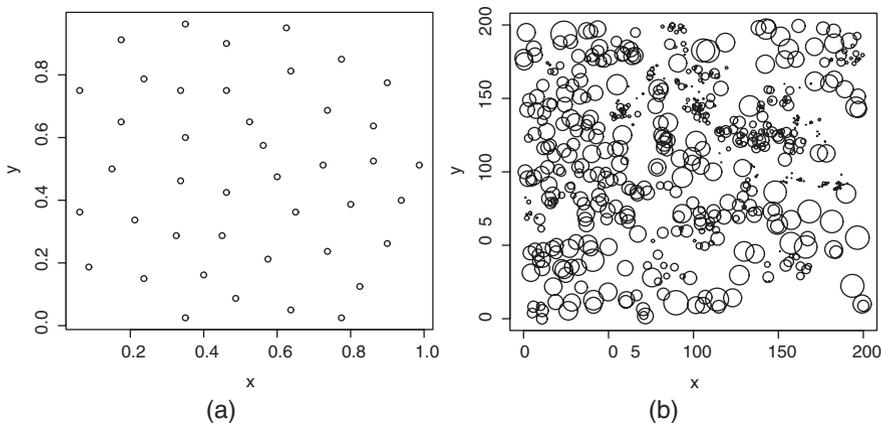

(a)   (b)

Fig. 0.3 (a) 42 cell centers of a histological section seen under a microscope (`cells` dataset from the `spatstat` package); (b) Position and diameter of 584 pine trees in a forest (`longleaf` dataset from the `spatstat` package).

As is the case for time series, spatial statistics differ from classical statistics due to non-independence of observations; throughout this book, we will generally call X a spatial process or *random field*.

This dependency structure means there is redundancy in available information that can be exploited when making predictions, though it also modifies statistical behavior. Unbiasedness, consistency, efficiency and convergence in distribution of estimators all have to be reexamined in this context. The originality of spatial statistics is to make use of *non-causal modeling*; in this sense, spatial statistics is radically different to time series statistics where causal models use the passage of time and a notion of the "past" (modeling river flows, stock prices, evolution of unemployment rates, etc.). Markov spatial modeling works with the idea of the spatial neighborhood of site s "in all directions." This includes dimension $d = 1$: for example, if $S \subseteq \mathbb{Z}^1$ and X_s is the quantity of corn harvested from each corn stalk along a row, a reasonable model would compare X_s with its *two neighbors*, the stalks to the "left" X_{s-1} and "right" X_{s+1}. We see that causal autoregressive modeling of X_s based on

X_{s-1} has no obvious meaning. If the crop is in a field, we could let the harvested quantity $X_{s,t}$ at site (s,t) depend on that of its 4 nearest neighbors $X_{s-1,t}$, $X_{s+1,t}$, $X_{s,t-1}$ and $X_{s,t+1}$, or even perhaps its 8 nearest neighbors.

These three types of spatial structure (cf. Cressie, (48)) provide the framework to this book. The first three chapters are devoted to modeling each in turn (Chapter 1: Second-order models, geostatistics, intrinsic models and autoregressive models; Chapter 2: Gibbs-Markov random fields over networks; Chapter 3: Spatial point processes). Due to the importance of simulation in spatial statistics, Chapter 4 presents *Monte Carlo Markov Chain* (MCMC) methods for spatial statistics. Chapter 5 then brings together the most important statistical methods for the various models and data types and investigates their properties. Four appendices round things off with a presentation of the most useful probabilistic and statistical tools in spatial statistics (simulation, limit theorems and minimum contrast estimation) as well as software packages for performing analyses presented in the book.

Numerous examples, most of them treated with the *R* software package (178), shed light on the topics being examined. When the data being studied are not directly available in *R* or from some other specified location, descriptions, relevant program scripts and links can be found at the website of the book:

$$\texttt{www.dst.unive.it/}\sim\texttt{gaetan/ModStatSpat}.$$

Each chapter ends with a set of exercises.

The bibliography gives the reader the chance to enrich their knowledge of notions only briefly presented here as well as several technical results whose proofs have been omitted. We also list reference books that fill gaps remaining after our intentionally reduced and non-exhaustive treatment of this multi-faceted subject undergoing great development (69).

Our thanks go to all our colleagues who have given us a taste for spatial analysis, for their ideas, remarks, contributions and those who have allowed us to use data collected from their own work. We would equally like to thank the *R Development Core Team* and authors of spatial packages for *R* (178) who have made their powerful and efficient software freely available to the public, indispensable when working with methods and tools described here. We thank reviewers for their careful rereading of the first draft; their remarks have helped to significantly improve the present version. Thanks to Bernard Ycart for encouraging us to expand an initially more modest project. Of course, we could never have undertaken this work without the patience and support of our families and the backing of our respective research teams, Dipartimento di Statistica - Università Ca' Foscari Venezia and Laboratoire SAMOS - Université Paris 1. Lastly, many thanks to Kevin Bleakley for the translation and English adaptation, done with much competence. Any remaining errors are ours.

Venice and Paris,   *Carlo Gaetan*
August 2009   *Xavier Guyon*

Springer