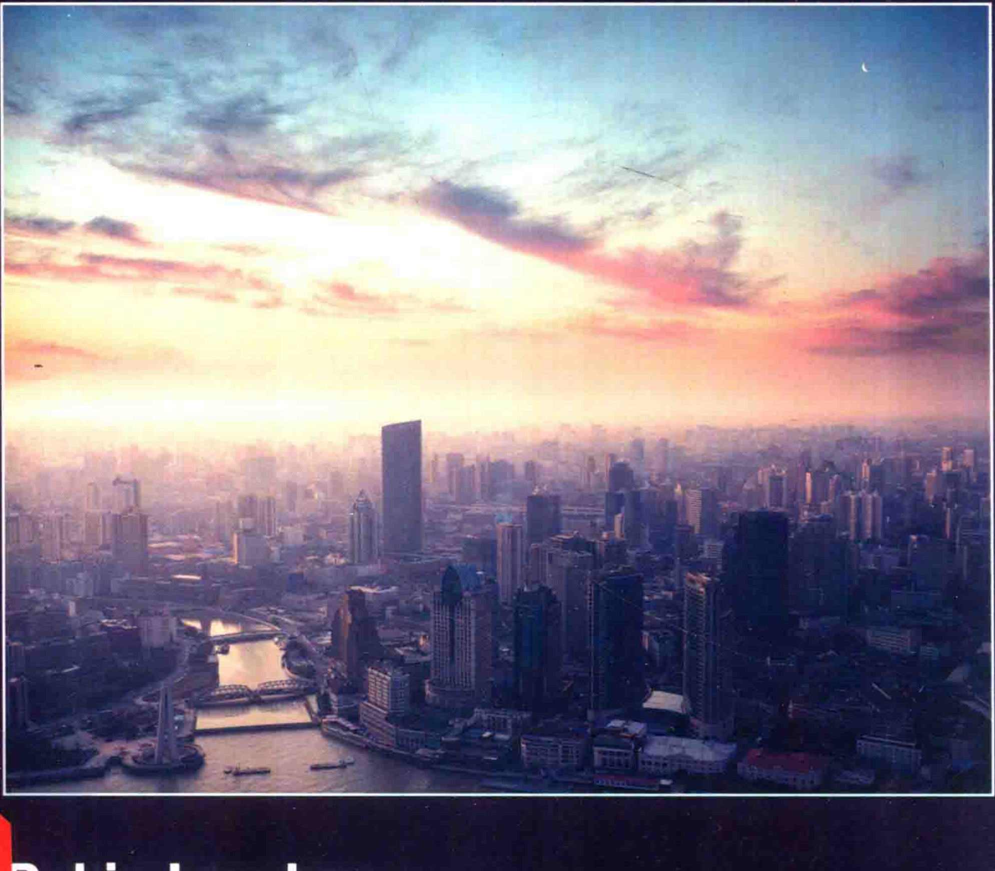
The R Series

Spatial Microsimulation with R



Robin Lovelace
Morgane Dumont

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Spatial Microsimulation with R

Robin Lovelace

Morgane Dumont

With the assistance of
Richard Ellison and Maja Založnik

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20160125

International Standard Book Number-13: 978-1-4987-1154-8 (Paperback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Spatial Microsimulation with R

Chapman & Hall/CRC
The R Series

Aims and Scope

This book series reflects the recent rapid growth in the development and application of R, the programming language and software environment for statistical computing and graphics. R is now widely used in academic research, education, and industry. It is constantly growing, with new versions of the core software released regularly and more than 7,000 packages available. It is difficult for the documentation to keep pace with the expansion of the software, and this vital book series provides a forum for the publication of books covering many aspects of the development and application of R.

The scope of the series is wide, covering three main threads:

- Applications of R to specific disciplines such as biology, epidemiology, genetics, engineering, finance, and the social sciences.
- Using R for the study of topics of statistical methodology, such as linear and mixed modeling, time series, Bayesian methods, and missing data.
- The development of R, including programming, building packages, and graphics.

The books will appeal to programmers and developers of R software, as well as applied statisticians and data analysts in many fields. The books will feature detailed worked examples and R code fully integrated into the text, ensuring their usefulness to researchers, practitioners and students.

Published Titles

Spatial Microsimulation with R, *Robin Lovelace and Morgane Dumont*

Statistics in Toxicology Using R, *Ludwig A. Hothorn*

Stated Preference Methods Using R, *Hideo Aizaki, Tomoaki Nakatani, and Kazuo Sato*

Using R for Numerical Analysis in Science and Engineering, *Victor A. Bloomfield*

Event History Analysis with R, *Göran Broström*

Computational Actuarial Science with R, *Arthur Charpentier*

Statistical Computing in C++ and R, *Randall L. Eubank and Ana Kupresanin*

Basics of Matrix Algebra for Statistics with R, *Nick Fieller*

Reproducible Research with R and RStudio, Second Edition, *Christopher Gandrud*

R and MATLAB®, *David E. Hiebeler*

Nonparametric Statistical Methods Using R, *John Kloke and Joseph McKean*

Displaying Time Series, Spatial, and Space-Time Data with R, *Oscar Perpiñán Lamigueiro*

Programming Graphical User Interfaces with R, *Michael F. Lawrence and John Verzani*

Analyzing Sensory Data with R, *Sébastien Lê and Theirry Worch*

Parallel Computing for Data Science: With Examples in R, C++ and CUDA, *Norman Matloff*

Analyzing Baseball Data with R, *Max Marchi and Jim Albert*

Growth Curve Analysis and Visualization Using R, *Daniel Mirman*

R Graphics, Second Edition, *Paul Murrell*

Introductory Fisheries Analyses with R, *Derek H. Ogle*

Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving, *Deborah Nolan and Duncan Temple Lang*

Multiple Factor Analysis by Example Using R, *Jérôme Pagès*

Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R, *Daniel S. Putler and Robert E. Krider*

Implementing Reproducible Research, *Victoria Stodden, Friedrich Leisch, and Roger D. Peng*

Graphical Data Analysis with R, *Antony Unwin*

Using R for Introductory Statistics, Second Edition, *John Verzani*

Advanced R, *Hadley Wickham*

Dynamic Documents with R and knitr, Second Edition, *Yihui Xie*

Preface

Spatial microsimulation is a set of methods for modelling phenomena which operate at individual and geographical levels simultaneously. For example the functioning of a modern city (Shanghai is illustrated on the front cover) involves an overwhelmingly complex web of human interactions. Simulating such complexity may seem impossible. Yet, by breaking the problem up into its constituent parts — discrete geographic areas and a sample of the population — spatial microsimulation can be used to model key aspects of the city system on an everyday laptop computer. There are dangers associated with reductionism, but condensing a problem down to its fundamentals has many advantages. Using the techniques described we can simulate scenarios such as population growth, increased energy efficiency and major shifts in transport technologies and modes. By linking the synthetic population to an agent-based model, in which individuals interact over time with each other and their environment, complex behaviours such as social segregation could also be simulated, as illustrated in Chapter 12.

This book is for anyone who wants to not only understand but to *use* spatial microsimulation. The emphasis is on the practical rather than theoretical aspects of the field. R packages such as **mipfp**, for enabling the generation of synthetic populations, are described in detail, with reference to practical examples and reproducible code. The aim is to enable you to implement the methods on your own data.

By explaining how to use tools for modelling phenomena that vary over space, this book should help enhance your knowledge of complex systems. We hope the book empowers the reader with the confidence and know-how needed to provide evidence-based policy guidance.

The origins of this book are more prosaic: during my PhD at the University of Sheffield I was tasked with using spatial microsimulation to model transport energy use. Despite the growing academic literature on the subject, there was little information that explained *how* to do spatial microsimulation, using a modern programming language. It was informal communication and code-sharing with a colleague, Malcolm Campbell, that led to the development of my models in R. This experience demonstrated the importance of reproducible research. Following this 'open science' ethic, readers are encouraged to comment on and contribute to the book's continued development via the code sharing site GitHub (see `https://github.com/Robinlovelace/spatial-microsim-book`).

The opportunity to turn the idea into reality came in the spring of 2014, when I developed notes for an 'Introduction to Spatial Microsimulation' course at the University of Leeds. The high demand for and positive feedback after the course suggested the need for practical teaching materials in the area. Four months later I delivered another course on spatial microsimulation at the University of Cambridge. The materials had been greatly updated and, thanks to the involvement of CRC Press, these provided the foundation for a book on the subject.

Morgane Dumont (NaXys, University of Namur), who attended the Cambridge course, became involved shortly after and has greatly improved the work. Morgane's background in Mathematics and Statistics made her the ideal co-author, complementing the focus on practical examples and code.

Maja Zaloznik (University of Oxford) and Richard Ellison (University of Sydney) have also greatly improved the book through their contributed chapters. Richard's chapter (11) illustrates how R can be used as the basis for transport demand modelling, using an approach known as TRESIS. Maja's chapter (12) is the most advanced in the book and demonstrates how spatial microsimulation can be used in parallel with agent-based modelling, with an implementation in the NetLogo language.

Spatial microsimulation with R is therefore the result of international teamwork. It is, to the best of our knowledge, the only practical book on the subject. We hope it is useful in your work. More widely, we hope it provides a solid foundation for advancement in the field and a toolkit for solving real-world problems.

If you have any feedback on the book please do get in touch via the book's online repository, hosted on the code sharing platform GitHub: `https://github.com/Robinlovelace/spatial-microsim-book`.

Robin Lovelace, February 2016.

Acknowledgements

As with any worthwhile textbook, this was not a solo effort. We benefited immensely from teaching spatial microsimulation to diverse audiences, the formal and informal feedback they provided, and correspondence with a number of people using spatial microsimulation 'in the wild'. Of these, the following deserve special mention:

- James Gleeson, from the Greater London Authority (GLA), provided insight into how spatial microsimulation can be used in local government and made several improvements to the book.
- Ulrike Rauer, from the University of Oxford, commented on early drafts of the book and showed how it could be made more relevant to PhD students new to the approach.
- Stephen Clarke at the University of Leeds demonstrated the benefits of the Flexible Modelling Framework and encouraged testing of the R code on much larger datasets than had previously been used, encouraging optimisation of the code.
- Johan Barthélemy, from the SMART Infrastructure (Wollongong), helped in understanding his methods and R package (mipfp).
- Lex Comber, who provided crucial comments on the structure of the first part of the book and a great insight into how to make it more useful for teaching.
- Malcolm Campbell, my predecessor in the PhD. Malcolm provided a huge amount of support during the early phase of my PhD and shared all the R code he developed. He's been a great support of the book from the beginning.
- Everyone who provided input from the University of Leeds, including Mark Birkin, Nick Malleson and Andy Evans.

Morgane Dumont's research is funded by the Wallonia Region (Belgium) and she is a member of NaXys (University of Namur). She thanks these two affiliations in particular. We also thank the FNRS (Belgium) for the funding that allowed Morgane to meet Robin in Leeds in 2015 to work together on the book. Finally, computational resources (for Chapter 10) have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11.

Thanks also to all the people who provided the wider resources for this project to happen.

List of Figures

List of Tables

Contents