Journal of Statistical Software

August 2016, Volume 72, Book Review 1. *doi: 10.18637/jss.v072.b01*

Reviewer: Ezra Haber Glenn
Massachusetts Institute of Technology

Spatial Microsimulation with R

Robin Lovelace and Morgane Dumont
Chapman & Hall/CRC, Boca Raton, 2016.
ISBN 978-1-4987-1154-8. 259 pp. USD 89.95 (P), 62.97 (eBook).
https://www.crcpress.com/9781498711548

Starting in the preface, Lovelace and Dumont explain that their book is intended to serve multiple functions: to introduce readers to the theoretical approaches used in spatial microsimulation while simultaneously teaching ways to use R to generate and analyze spatial microdata. Accomplishing either of these tasks alone would be an impressive – and much needed – contribution to the literature, which is sorely lacking in clear introductory materials and hands-on worked examples for students; to accomplish *both* tasks – in a slim 259 pages – would be impressive indeed. Happily, the authors succeed for the most part: the book provides an excellent introduction to the theory and practice of spatial microsimulation, as well as a bridge to working in R to actually *do* the various tasks described; in the estimation of this reviewer, any shortcomings – most notably, some glossing of the more technical material and an over-reliance on highly simplified datasets – seem forgivable, given the desire to keep the volume small and not overwhelm students at the outset.

In passing, note the use of the word "students," instead of readers; the book, based on Lovelace's tested materials from a course on "Introduction to Microsimulation" at the University of Leeds, reads very much like a digested graduate course. As such, the book would serve as an excellent text for a higher-level course on the subject. Readers can even download sample datasets and source code from the book's GitHub site at https://github.com/Robinlovelace/spatial-microsim-book. (That said, the inclusion of other standard textbook features, such as review questions or homework assignments, would make this even more useful in a course setting.)

Before proceeding to describe the book's many strengths, it would be wise to highlight an important caveat regarding the book's topic: readers attracted by the title but unfamiliar with the practice of spatial microsimulation may find themselves disappointed when they learn exactly what the term means. This is no fault of the authors, or of the field *per se*: it's simply that any use of the term "simulation" immediately conjures up images of SimCity, agent-based models, interactive games, virtual reality, and the like. The actual practice of "spatial microsimulation" (in R or anywhere else) is far less sexy a topic.

As defined by the authors in one of the many clarifications near the start of the book:

> *Spatial microsimulation can be understood either as a technique or an approach ... for [t]he creation, analysis and modelling of individual level data allocated to geographic zones.*

Elsewhere the authors clarify issues further, stressing what their topic is *not*, in an attempt to dispel unwarranted notions of simulation: "[s]patial microsimulation is not (quite) agent-based modelling." In short, the book could be more accurately called *"Simulation Techniques (in R) to Generate Spatial Microdata,"* although this doesn't exactly roll off the tongue.... Importantly, while spatial microsimulation can be used to *generate* microdata, *other techniques and approaches* — say, agent based modelling – are needed to actually "play with" it. (And unfortunately, it is precisely the notion of "play" that makes other types of simulation seem so much more fun....)

Regardless, the roots of this particular confusion lie deeper than the current work, and there is little Lovelace and Dumont can do to change the terminology now without creating even more chaos – and luckily, there are plenty of potential readers who are likely to recognize that even if spatial microsimulation is perhaps less flashy than these popular associations, it is still crucial, powerful, and even sort of fun when you put the theories into practice.

And "putting theories into practice" is exactly where the strengths of the book come to light. Starting with a few quick examples in Chapter 3 to showcase the power of spatial microsimulation in the realms of health indicators, economic policy, and transportation data, the book turns to the real work of this sort of analysis.

The heart of the book is an extended, two-chapter discussion of multiple approaches to population simulation, with attention to the use of various weighting algorithms to generate data. Importantly, "discussion" is an appropriate description of the authors' treatment of this material here, given their willingness to wade into the details and trade-offs between different approaches; this is where the value of "learning while doing" really pays off. An impressive inclusion is a section comparing two different R packages to perform the crucial step of iterative proportional fitting (IPF): Alexander Blocker's **ipfp** and Barthelemy, Suesse, and Namazi-Rad's **mipfp** package; a lesser volume would have simply chosen one or the other, or presented both without discussing when, why, and how each would be most useful.

Other helpful additions are chapters on data preparation (Chapter 4) and model checking and evaluation (Chapter 8), two "boring" – but extremely crucial – topics that are too-often ignored. Again, thanks to the book's insistence on learning-while-doing and attending to the "less-flashy" aspects of simulation, the devil in the details can be tamed.

For those readers chomping at the bit, hoping to move beyond IPF and spatial microdata generation and eager to actually "play" with the data, an additional chapter does demonstrate "spatial microsimulation for agent-based models" using the popular **NetLogo** platform. (This chapter, by Maja Založnik, and Chapter 12 on "The TRESIS approach to spatial microsimulation," by contributing writers Richard B. Ellison and David A. Hensher, seem to be tacked on as "bonus tracks," essentially serving as two additional "micro-books" for power-users.)

One final point concerning the book's coverage of the R statistical programming language (the "... with R" promised in the book's title): be warned, the book can hardly be considered a stand-alone introduction to working in R. Time was that users could expect to learn a programming language *along the way* through worked examples in a book such as this, and in the past most authors would have included more introductory material on the syntax and commands of the language. (In essence, a book could be read as *either* "[Learning] Spatial

Microsimulation with R" or "[Learning] R with Spatial Microsimulation.") Here the authors have included only the most basic R help: a five-page "quick start" guide at the start, as well as some additional guidance in a six-page appendix (which really serves as more of an extended sales pitch for the language); R newbies will want to work through this book *along with* (or perhaps even *following*) one of the many good (often free) "Introduction to R" offerings.

That said, R is a powerful tool for this sort of work — all the more so thanks to the ongoing development of special-purpose packages like **ipfp** and **mipfp** by the academic and user communities – and it is heartening to see more and more social science applications migrating to this platform. At least for those interested in working with spatial microdata, Lovelace and Dumont's book will help build this movement.

Reviewer:

Ezra Haber Glenn
Massachusetts Institute of Technology
Department of Urban Studies & Planning
77 Massachusetts Avenue, Room 7-337
Cambridge, MA 02139, United States of America
E-mail: eglenn@mit.edu
URL: http://dusp.mit.edu/faculty/ezra-glenn

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
August 2016, Volume 72, Book Review 1 *Published:* 2016-08-07
doi:10.18637/jss.v072.b01