Review of Spatial Microsimulation with R

Trond Husby

Netherlands Environmental Assessment Agency (PBL)
trond.husby@pbl.nl

Spatial Microsimulation with R by Robin Lovelace and Morgane Dumont (2016) is a practical guide to spatial microsimulation in the programming language R. The emphasis of the book is clearly on the practical implementation in a modern programming language rather than on the theoretical aspects of the field. The book is meant for anyone who wants to not only understand but to 'use' spatial microsimulation. Readers can implement the numerous worked examples in the book using supporting code and data. Book chapters which include embedded R code can be downloaded from a GitHub repository. Readers can either compile the entire book or copy-paste relevant parts of R code to follow the worked examples.

What is spatial microsimulation? According to the authors, spatial microsimulation can be understood either as a technique or as an approach. In this book, spatial microsimulation is primarily interpreted as a synonyme to population synthesis. The technique entails the creation of a synthetic micro population for small administrative areas by combining a-spatial microdata with geographically aggregated zone-level data.

The book spans 12 chapters, divided into three parts. Part I gives an overview of the field of spatial microsimulation, its applications and its concepts; Part II discusses the practicalities and software decisions involved in generating spatial microdata; and Part III deals with issues related to the modelling of spatial microdata. Individual chapters can be read separately, however, in order to accompany the worked examples readers are advised to follow the sequence of the book. This review primarily deals with Part II - admittedly the core of the book.

Part II begins with a chapter on data preparation (Chapter 4). The focus of this chapter is on loading, cleaning and reshaping aggregate and individual level data used in population synthesis. Practical issues such as subsetting, re-categorising and matching variables are carefully explained with examples. The chapter shows one of the strengths of R in this context, namely the interactive work with the data.

The longest chapter in the book (Chapter 5) explains population synthesis. Starting with an overview on weighting algorithms, it moves on to explain the IPF technique and discusses two different implentations of IPF in R. The authors run examples of both R implementation, discussing similarities between, and advantages and

drawbacks of both approaches. The R implementations of IPF share one shortcoming: individual weights are fractional. To address this shortcoming, the authors provide an easy-to-use function which transforms the fractional weights into integers. Integerised weights can subsequently be used for expansion: the replication of individuals in the micro data. The book provides easy-to-use functions for expansion as well. In my opinion, Chapters 4 and 5 are best read while running the examples in an R console.

Chapter 6 gives an overview of alternative approaches to population synthesis, including regression- and optimization-based algorithms. This chapter also provides codes and data so that readers can play with the alternative algorithms themselves. The chapter provides links to a number of R packages and Python tools built for population synthesis.

In Chapter 7 the reader is introduced to population synthesis 'in the wild', i.e., on a real data set. For a newcomer to spatial microsimulation, this chapter is perhaps the most useful. Real data sets are likely to exhibit problematic features such as non-matching marginals and two-dimensional constraints. The chapter provides solutions to such problems and it uses the IPF procedure discussed in Chapter 5 to perform the population synthesis. Chapter 7 can be read alone but it is definitely more accessible if read with Chapters 4 and 5.

Validation is an important part of any type of simulation, and model checking and evaluation is the topic of Chapter 8. The chapter deals with both internal and external validation, although it is fair to say it devotes most attention to the former. A number of different model-checking tests are presented and the authors give an example using the data set from Chapter 7. Chapters 9 and 10 deal with more specific cases of population synthesis - namely spatial microsimulation without microdata and allocation of individuals into households.

Part III of the book focuses on how a synthetic population can be used in other models. Chapter 11 discusses the TRESIS approach to spatial microsimulation. TRESIS is an integrated approach to modelling transport decisions, where microdata on households and individuals are used to determine residential location and destination and mode choice. The TRESIS approach differs somewhat from the population synthesis presented in Chapter 7. The differences are illustrated with examples run on the data set used in Chapter 7. Finally Chapter 12 shows how the synthetic population from Chapter 4 can be used as input in an Agent-based model implemented in Netlogo.

As an experienced R user but a newcomer to spatial microsimulation, I found the book a great do-it-yourself resource. Using my own data sets I quickly managed to create a spatial micropopulation with the approaches presented in Part II. The book is not written with a single target audience in mind, however, in my opinion this book is best suited for newcomers to spatial microsimulation who are somewhat fluent in R. Spatial Microsimulation in R devotes most of its attention to population synthesis using IPF. Other recent reference guides provide more comprehensive reviews of the field than this book. However, in my view the limited scope is an advantage. The natural flow of the chapters and the worked examples effectively bring a, somewhat devoted, newcomer to spatial microsimulation quickly up to speed on the field. In my view this is a major strength of the Spatial Microsimulation in R.

An R user would notice that the authors primarily stick to functions from base R in the core part of the book. Many of these functions allow the user to replace for-loops with vectorised alternatives, thereby increasing computational efficency. However, some of these functions can seem unintuitive to readers who are not experienced

R users or readers who are not familiar with matrix algebra. Much of the work with the data in Chapters 4 and 5 can be accomplished with functions from either the *tidyverse* or from *data.table*. These packages aim at providing user-friendly functions for data manipulation, although computationally efficient operations are going on under the hood. The syntax involved with these functions is therefore often more intuitive than functions from base R.

In my opinion this book is an excellent resource for everyone who want to learn how to do spatial microsimulation. The possibility to download the contents of the book, compile it and work interactively with the code also makes it a great example of dynamic documents and reproducible research. I would encourage interested readers to check out the GitHub repository [1] and buy the book.

REFERENCES

Lovelace, R., & Dumont, M. (2016). *Spatial microsimulation with R.* CRC Press.

NOTES

[1] `https://github.com/Robinlovelace/spatial-microsim-book`