Systematic testing of microsimulation methods

Robin Lovelace (robinlovelace.net)

24/10/2014

Motivation

- Dozens of methods available for (spatial) microsimulation
- Difficult to choose from options
- Testing can be time consuming and tricky (Harland et al. 2012)

Need for fast and consistent testing framework

Broader motivations

Problem: each researcher has their own 'horse' in the race

Past testing efforts in the literature

The ‘model experiment’ genre

Results from past work

- Many useful findings - often researcher's own model 'best'
- No conclusive results - not reproducible - comparing different things

Table 1: Summary comparison of the three algorithms

	Deterministic Reweighting	**Conditional Probabilities**	**Simulated Annealing**
Easy setup (is there much pre-processing)?	Yes	Yes	No
Sensitive to specification of constraint order?	Yes	Yes	No
Limit to number of constraints that can be used?	Yes	Yes	No
Requires a sample population?	Yes	No	Yes
Can take forward and backward steps to find an appropriate solution?	No	No	Yes
Stochastic?	No	Yes	Yes
Speed of execution	Fastest	Middle	Slowest

Microsimulation as an experimental procedure

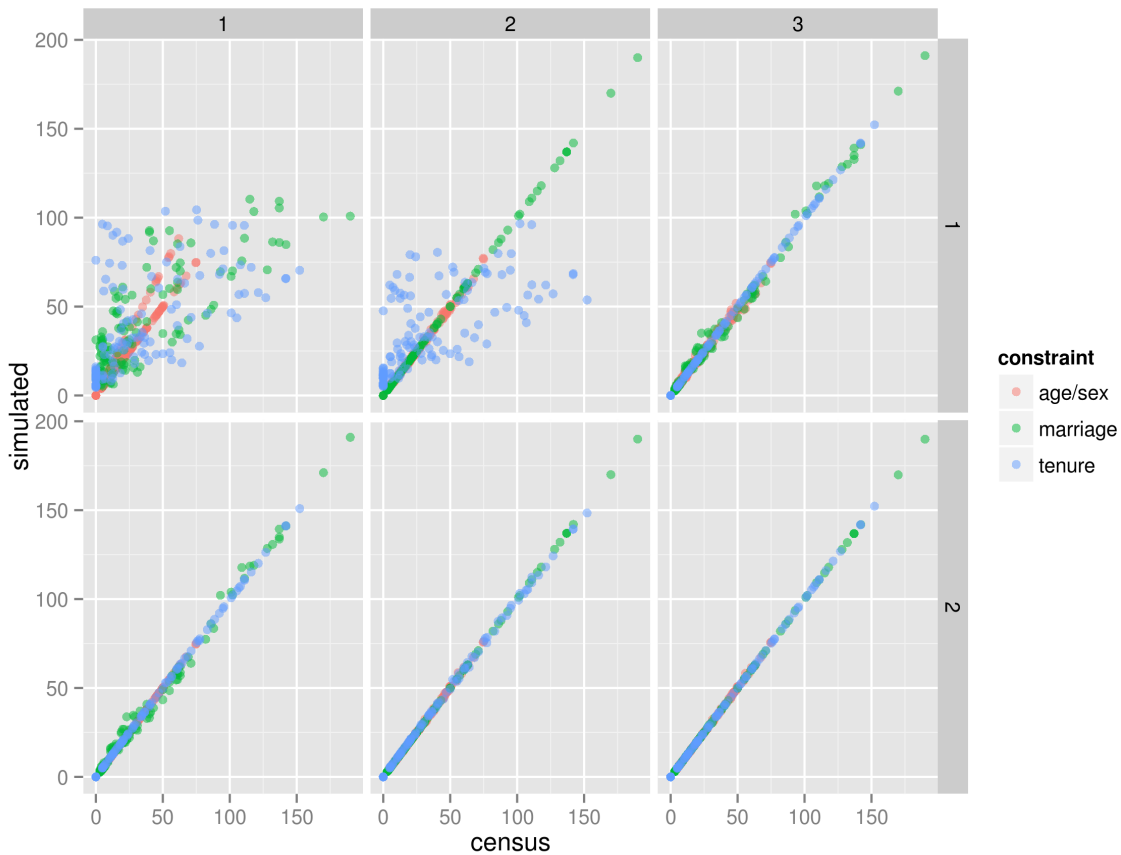
- Controlled experiments are the foundation of science
- Real-world experiments impossible
- Simulation allows range of alternatives to be tested safely

Simulation, then is the process of imitating the behavior of system patterns. Simulation as one method of problem-solving becomes attractive when conventional analytic, numeric or physical experimental methods would be too time-consuming, expensive, difficult, hazardous and/or irreversible or even impossible as real world experiments intended to solve a problem. (Merz, 1991). **International Journal of Forecasting 7 (1991) 77-104 77**

IPF performance testing

Setting-up model the experiments

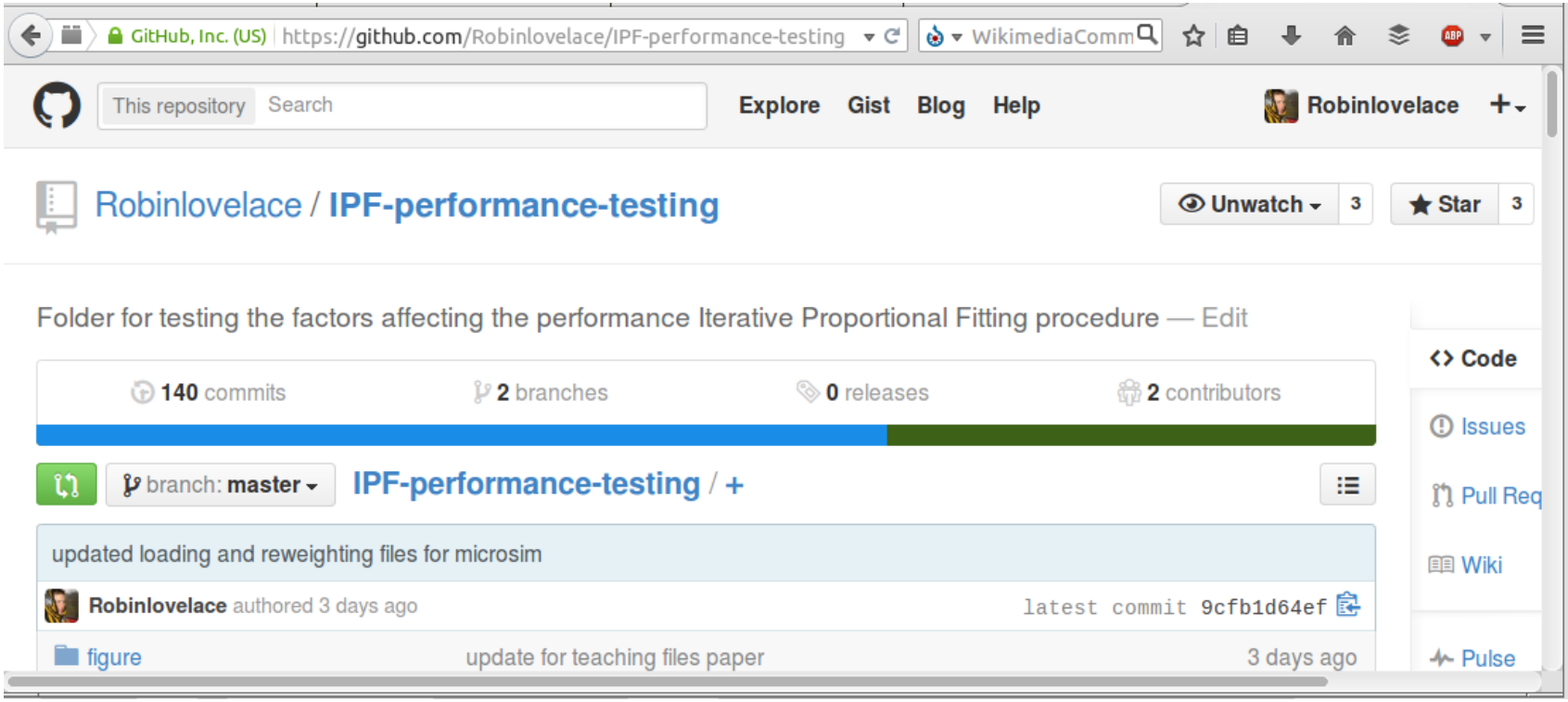
- ‘Scrambled’ versions of official datasets used
- Work ongoing on larger examples



Project organisation

```
|-- data-big (just README links)
|-- figure
|-- input-data
|   |-- sheffield
|   |-- simple
|   `-- small-area-eg
|-- literature
|-- models
|   |-- ipfinr
|   |-- FMF
|   |-- simSALUD
|   `-- GREGWT
`-- output
```

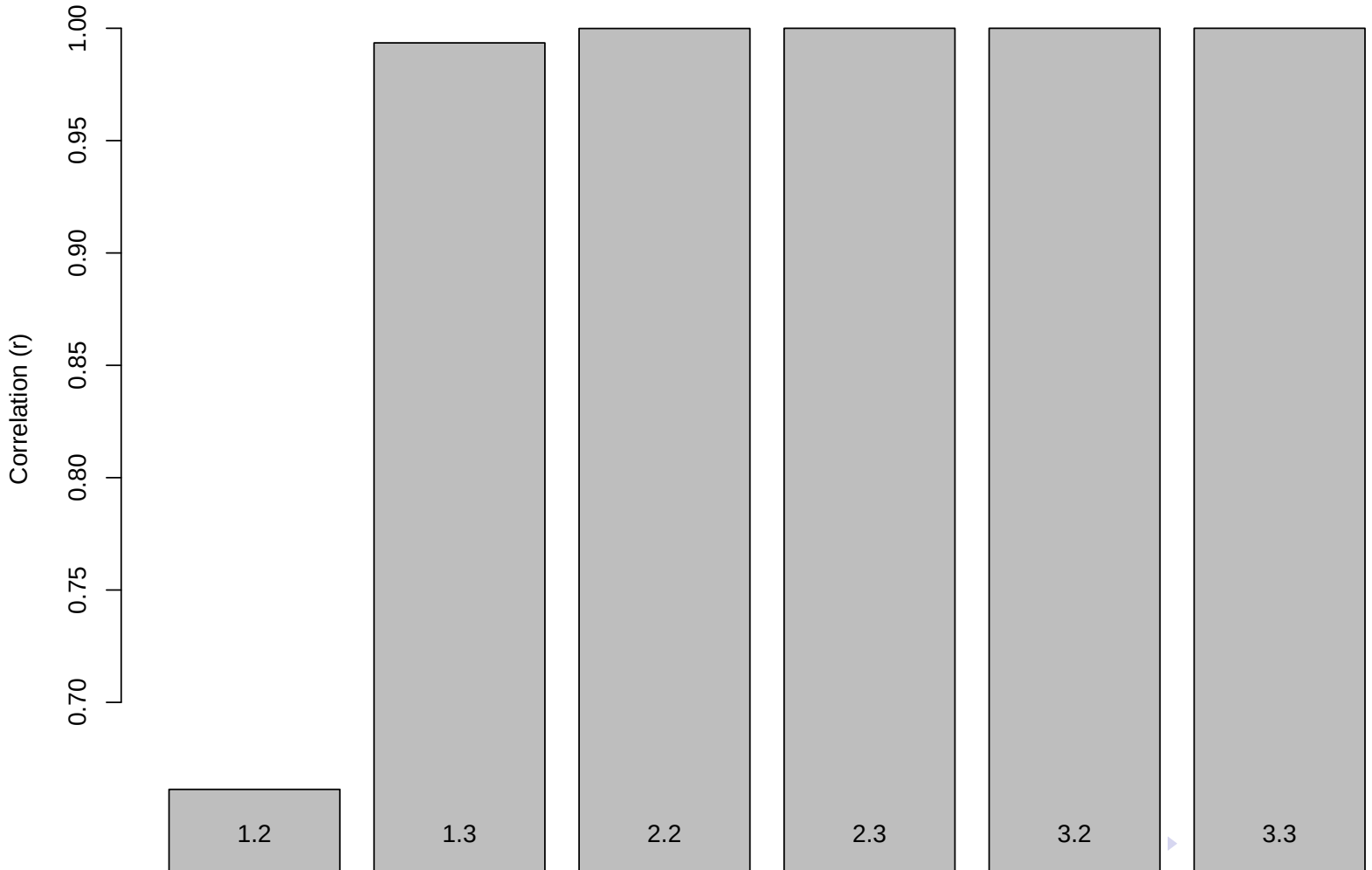
Try it yourself!



Replicable results

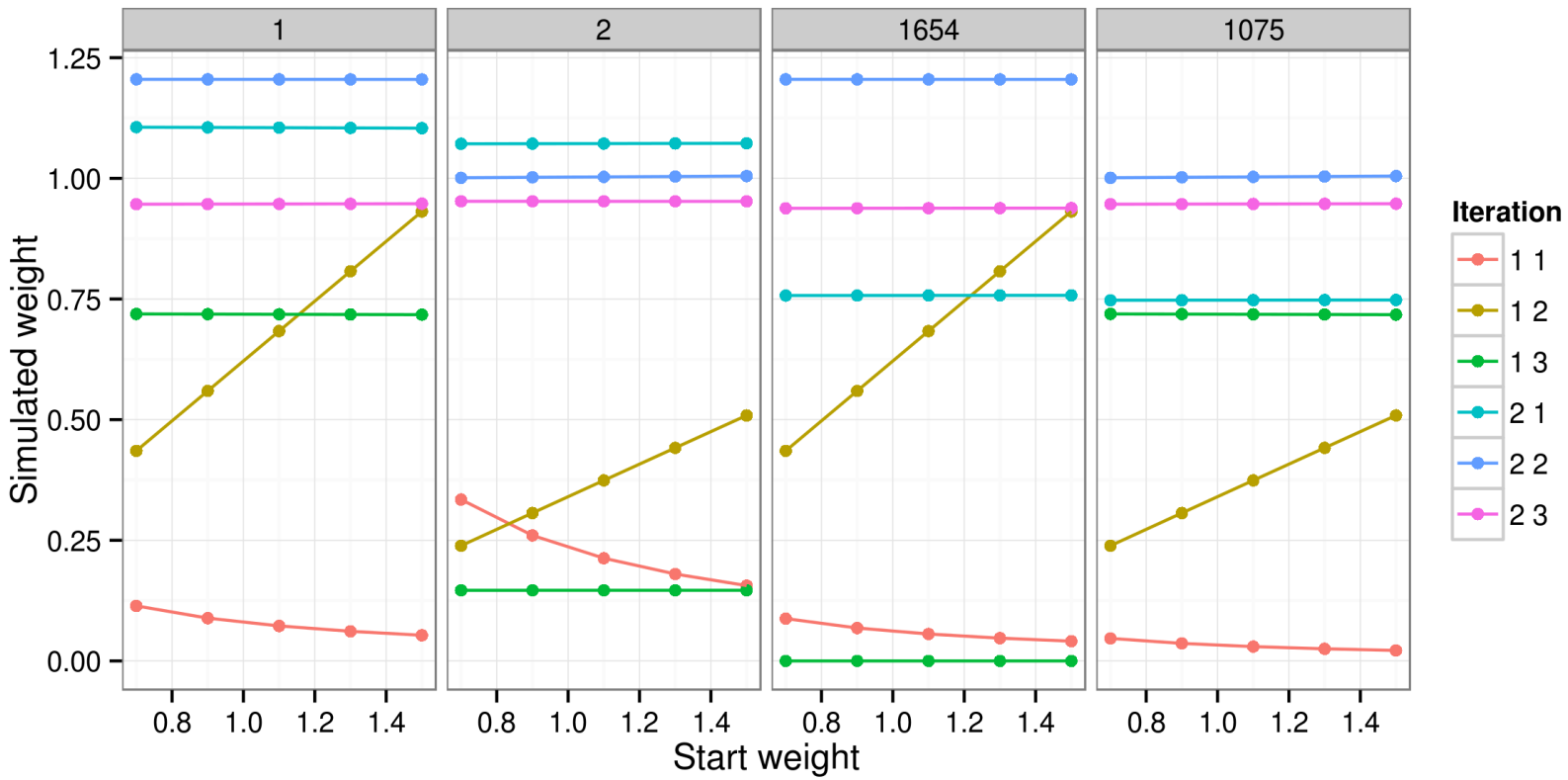
Reproducible example:

```
source("models/etsim.R")
```



Results

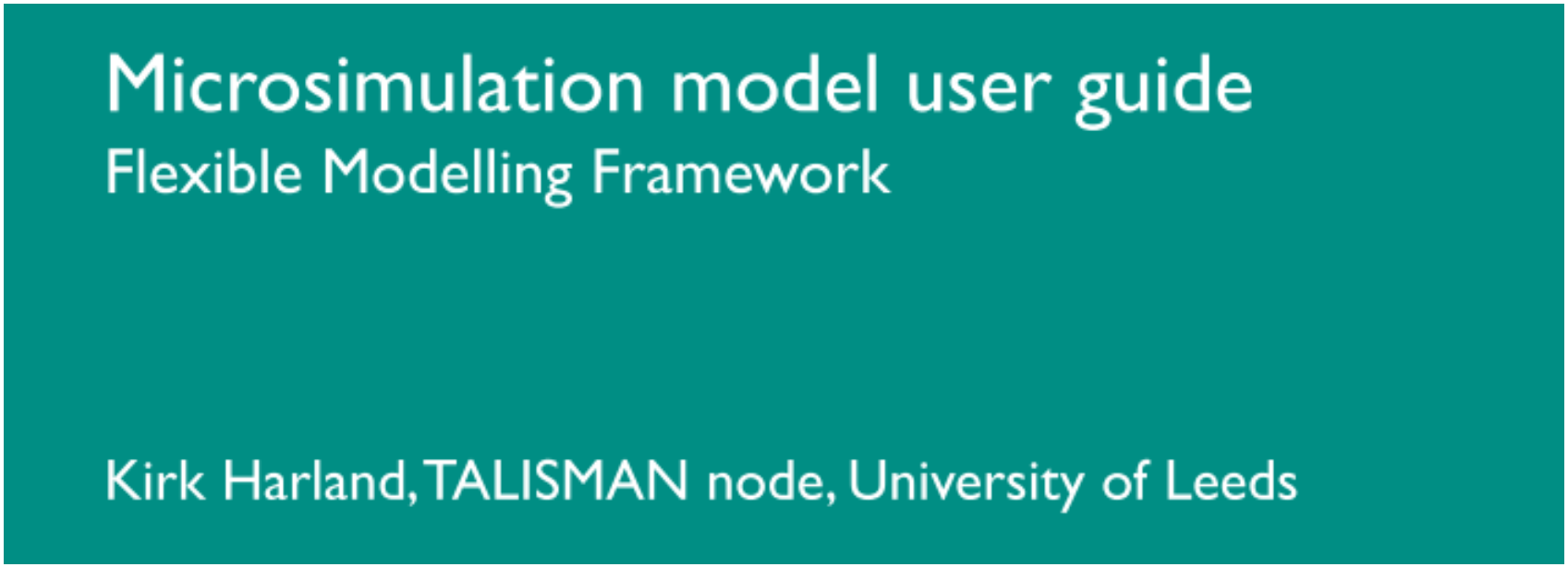
- 'Empty cells' found to have largest impact on fit
- **Initial weights had very little impact**
- C code (**ipfp** package): **50 fold** speed increase



Broadening the tests

CO in FMF vs IPF in R

- New project to test techniques on very large microdatasets
- Challenge: allocate 569,741 individuals to 7,787 zones
- Almost 60 million people in output spatial microdata!
- New methodology for IPF developed



External validation

- More important that ‘internal validation’ is how well results fit reality
- Opportunity provided by Census variable on census well-being
- Simulated at small area level with FMF and R

Work in progress

- Compare different approaches in terms of timing, model fit and ease of use
- External validation
- Use alternative methods to generate same output: GREGWT? SimObesity? simSALUD?

Wider context of spatial microsimulation

Issues within the field

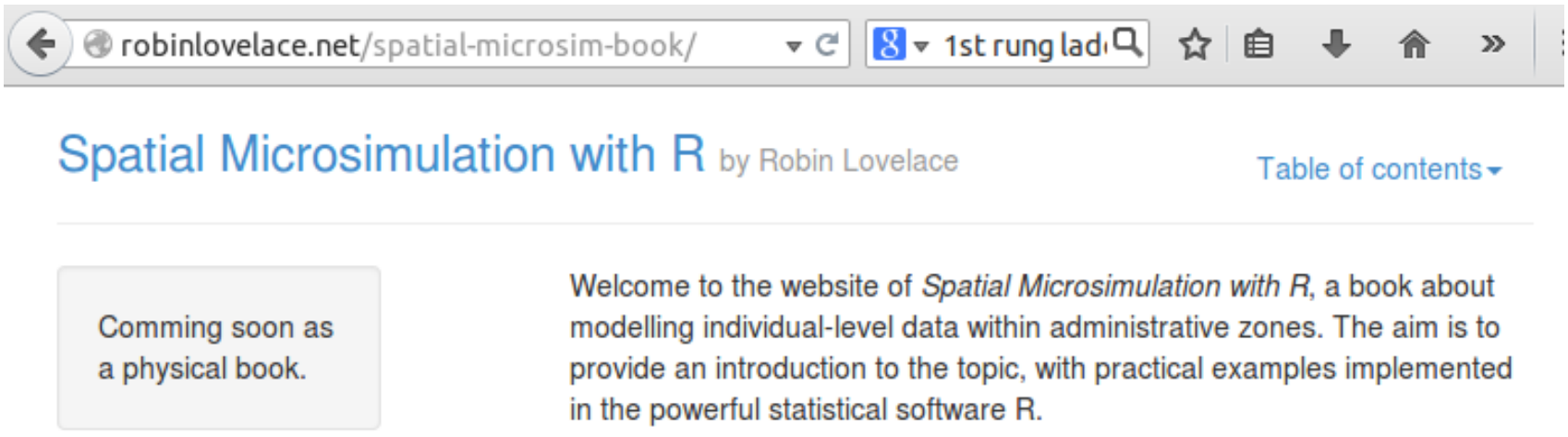
- "Little attention is paid to the choice of programming language used" for microsimulation (Clarke and Holm 1987)
- Lack of reproducibility (Lovelace and Ballas 2013)
- Hard to get started
- Few simple examples - uses tend to be big and complicated
- Few introductory teaching resources

Teaching spatial microsimulation

- Two courses in May (Leeds) and August (Cambridge)
- Taught basic principles of spatial microsimulation
- And implementation in R
- Feedback: students grateful for first rung on ladder
- More success with latter course focussing on applications

Spatial microsimulation introductory textbook

- Contract with CRC Press as part of their R Series
- Draft of book available online in its entirety
- Open 'wiki' style allows anyone to contribute
- Any feedback/input gratefully received
- Check it out here: robinlovelace.net/spatial-microsim-book/



Key References

Clarke, Martin, and Einar Holm. 1987. "Microsimulation Methods in Spatial Analysis and Planning." *Geografiska Annaler. Series B. Human Geography* 69 (2): 145–164. http://www.jstor.org/stable/10.2307/490448.

Harland, Kirk, Alison Heppenstall, Dianna Smith, and Mark Birkin. 2012. "Creating Realistic Synthetic Populations at Varying Spatial Scales: A Comparative Critique of Population Synthesis Techniques." *Journal of Artificial Societies and Social Simulation* 15 (1): 1. http://jasss.soc.surrey.ac.uk/15/1/1.html.

Lovelace, Robin, and Dimitris Ballas. 2013. "'Truncate, Replicate, Sample': A Method for Creating Integer Weights for Spatial Microsimulation." *Computers, Environment and Urban Systems* 41 (September): 1–11. doi:10.1016/j.compenvurbsys.2013.03.004. http://dx.doi.org/10.1016/j.compenvurbsys.2013.03.004.