Should we use Multilevel Regression and Post-stratification when simulating area-level population outcomes?

Roger Beecham*[1,2] and Stephen Clark†[1]

[1]School of Geography, University of Leeds
[2]Leeds Institute for Data Analytics, University of Leeds

GISRUK 2025

Summary

Estimating unknown outcomes at small-area population level is a routine task in GIScience. We ask whether Multilevel Regression and Poststratification (MRP), an approach to simulating public opinion, might overcome deficiencies in spatial microsimulation (SPM), the *de facto* approach in applied spatial analysis. Using microdata from Health Survey for England and 2021 Census tables, we evaluate MRP and SPM at estimating a known, groundtruthed, health outcome. When parameterised with few constraints, there are only slight differences in estimation between the two approaches. With more, and geographic, constraints, often desired by spatially-inclined researchers, errors begin to accumulate in SPM estimates that do not appear in those arrived at via MRP.

KEYWORDS: spatial microsimulation, multilevel modelling and poststratification, small-area estimation, non-representative samples.

1 Introduction

There are many situations in applied policy analysis where it is useful to know about a population that lives in an area. Although Censuses are the canonical population-level dataset, they record only high-level characteristics of a population. Social scientists therefore rely on other data sources, often public opinion surveys, when researching individual attitudes and behaviours. Spatial microsimulation (SPM) is a technique that allows some outcome measured in a survey to be estimated at the population-level. It takes individuals responding to a survey and re-weights or allocates them to small-area spatial units, constrained by the known (Census) characteristics of those units (Lovelace and Dumont, 2016).

Used widely in GIScience and adjacent domains, the assumptions and limitations of SPM are often acknowledged, but seldom addressed: that survey data are representative of the study area, or

*r.j.beecham@leeds.ac.uk
†s.d.clark@leeds.ac.uk

at least are equally well miss-represented across that area; that the survey is sufficiently rich and detailed to capture variation in individuals and outcomes over the study area; that the target variable is dependent on the constraint variable and that this dependency does not vary over the spatial units of the study area.

Readers familiar with political polling might see parallels with Multilevel Regression and Poststratification (MRP). MRP is also a technique for simulating an unknown outcome at small-area level. The outcome of interest in most applications of MRP is population-level voting (Park et al., 2004) – the number of votes political parties might achieve in an upcoming election, estimated separately by voting constituency – the small-area element. Party voting is first *modelled* from the survey data (MR), and to estimate the population-level outcome modelled probabilities are weighted (P) based on the known population-level composition of reporting areas. Individual-level variables are standard demographics collected via the survey known to be correlated with voting intention; area-level variables are commonly lagged vote shares or some other area-level context. At *poststratification*, separate predicted probabilities of voting intention are derived from the regression model for different types of people, or strata. This requires, for each small-area, different cross-tabs describing the known demographic characteristics such as sex-age-education-ethnicity. Local forecasts are made by multiplying predicted probabilities for different strata by the known number of that strata living in a small-area (Hanretty, 2019).

Although rarely applied in other fields, at face-value MRP seems to address several limitations of SPM that should bother us as geographers:

1. **MRP adjusts estimates where target spatial units are poorly represented by survey data**: Since MRP uses multilevel model designs, estimated regression coefficients, and thus predicted probabilities, are partially pooled (shrunk to the global mean) for demographic subsets that are less well-represented in the survey. This helps to address the risk often signalled in SPM primers of assigning high weights to certain individuals in the survey or copying the same individual across many times to the same spatial unit.

2. **MRP adjusts for geographic dependency in the outcome**: It is usual that some subnational context, extra to demographics, will systematically affect the baseline outcome of interest; when studying voting outcomes, Scotland is an entirely different polity. In MRP this sort of dependency can be easily accounted for, without losing data, by including regional context variables as random intercepts. Although in SPM it is possible to add geographic constraints (Birkin and Clarke, 2012) – for example requiring that survey respondents living in a particular region are allocated by the SPM to spatial units in that region only – doing so risks diminishing the diversity of the microdata (as demonstrated in our analysis).

3. **MRP can capture spatial and categorical heterogeneity in process**: In GIScience, we often acknowledge geographic heterogeneity in process: that the extent and direction of associations between context variables and the outcome of interest can vary systematically. While in SPM it is possible to have two-way constraints that may indirectly capture these heterogeneities, again this is at the cost of less heterogeneous microdata. Moreover, MRP *models* for this process, or other interactions between constraint variables and the outcome of

interest.

2 Data and Methods

We compare small-area-level predictions of an outcome estimated via SPM versus those generated via an MRP model. To do this, we identify a variable that is known and already collected via the 2021 UK Census: respondents' self-reported health status. This variable is also collected in the 2021 release of Health Survey for England (HSE) and framed in the same way as in the Census: *How is your health in general?*, with repondents self-reporting on a five-point scale from very-good to very-bad.

The 2021 HSE achieved a random sample of over 7100 respondents. While an undoubtedly large sample, the HES cannot be used to study area-level outcomes at the sub-regional level and so we use it as microdata and simulate the outcome (health-status) for each of the c.7000 middle super output areas (MSOA) in England.

In order to make a fair comparison across the simulated outcomes, we use equivalent constraints or covariates (see Figure 1). The stochastic flexible modelling framework was used to generate microsimulated outputs; the `rstanarm` R package to estimate the multilevel models and draw from the posteriors to postratify on Census data.

3 Analysis

Our models estimate self-reported health status as a binary outcome: *good* or *not good*. Seventy-eight percent of the (unweighted) HSE respondents self-reported that their general health was good; for the Census that proportion was 82.5%. This underestimate in the HSE might be due to the context under which the Census and HSE surveys are conducted or some form of non-response or differential response bias.

Presented in Figure 1 are plots summarising how well the SPM and MRP models simulate the *good* | *not-good* health outcome; separate models are constructed using constraints shared between the HSE and Census judged to be discriminating of health status. The histograms show the distribution in Pearson's residuals comparing predicted MSOA-level counts of individuals with *good* self-reported health against the *known* counts – or groundtruth – as per the Census. The maps also show these values for the MRP and SPM models respectively, along with summary statistics of mean absolute error in predicting the *good* | *not-good* counts and mean percentage point error in predicting the percentage good outcome.

While there are only minor differences in the predictive performance of outcomes simulated using SPM and MRP, as we start to introduce more constraints, errors in the SPM begin to accumulate. This is evidenced in the global error statistics, but most notably in the residuals maps for the most constrained SPM models, which display extreme residuals that are *spatially* concentrated. Most likely these extreme values are due to diminishing effective sample size when the SPM is constrained on subnational and geodemographic context: in rows five and six of Figure 1 by region and Index of Multiple Deprivation quintile. To explore this, we plot Shannon entropy for each MSOA. This

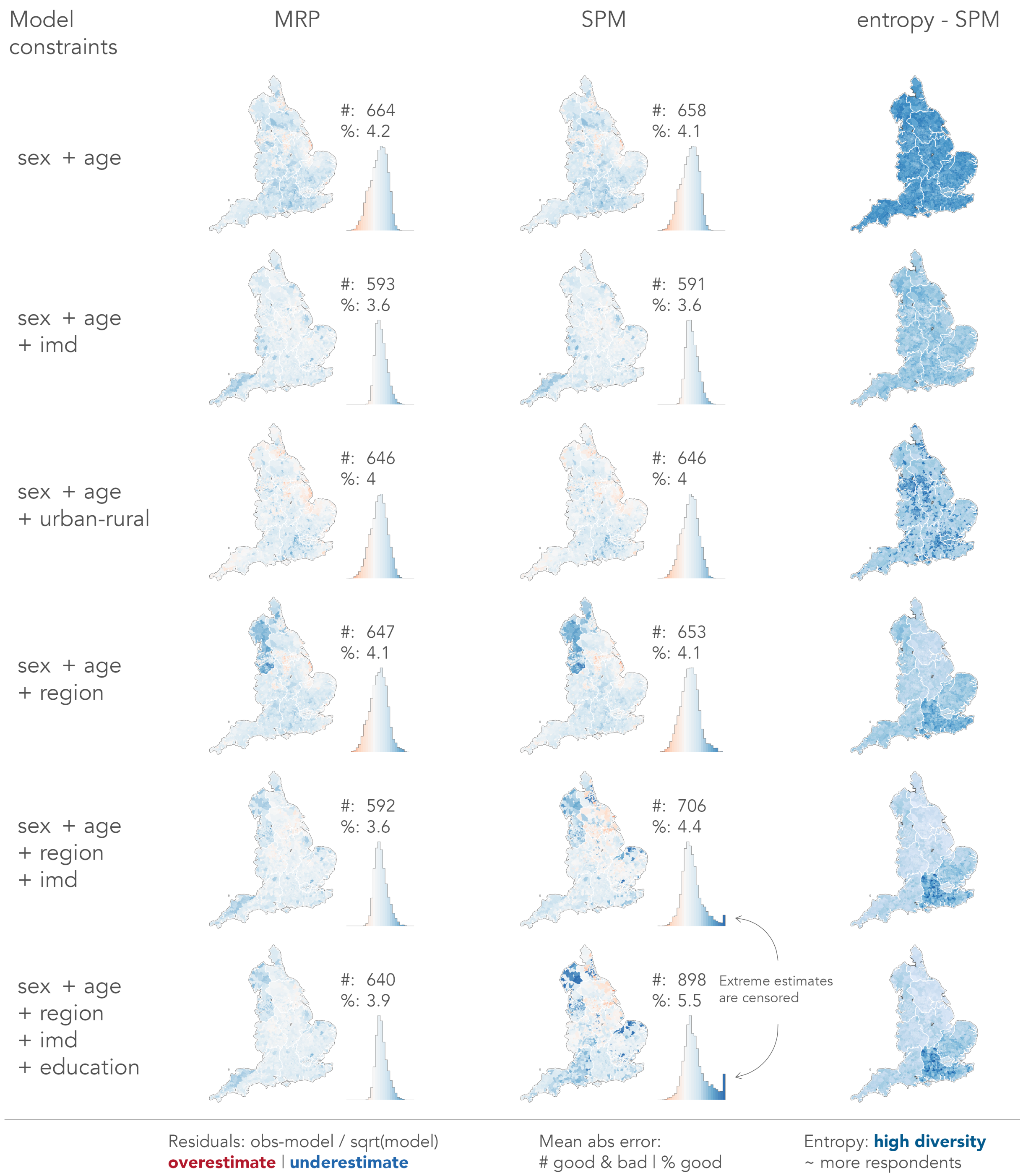

Figure 1: Summary of model performance in SPM and MRP for simulated estimates self-reported good health at MSOA level. Plot row labels identify the variables used as constraints in the SPM and as covariates in the MRP.

describes the number of unique respondents from the HSE used by the SPM to estimate the health outcome of an MSOA. Darker colours are associated with higher diversity and therefore where the MSOA is, in relative terms, better represented by the sample in the HSE. As more constraints are used, entropy necessarily decreases, and so a local colour scale is used to emphasise relative entropy within each model solution.

4 Conclusion

We present a principled case for how MRP might overcome some oft-cited limitations of SPM and demonstrate evidence to support this case. Our initial comparisons are made on a very paired back set of constraints and an outcome variable for which a high-level demographic – age – is by far the biggest determinant. An immediate analysis activity is to explore and evaluate the performance of SPM and MRP on other outcomes. Unlike in political polling, the target outcome in applied GIScience usually remains unknown. Perhaps for this reason, area-level estimates generated from SPMs tend not to be scrutinised to the same intensity as those generated from MRPs. With this paper we hope to initiate greater methodological introspection into small-area estimation in GIScience.

5 Biography

Roger Beecham is Associate Professor in Geographic Data Science at School of Geography, University of Leeds.

Stephen Clark is Research Fellow at School of Geography, University of Leeds.

References

Birkin, M. & Clarke, G. (2012). The enhancement of spatial microsimulation models using geodemographics. *The Annals of Regional Science*, *49*, 515–532.

Hanretty, C. (2019). An introduction to multilevel regression and post-stratification for estimating constituency opinion. *Political Studies Review*, (pp. 1–16).

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

Park, D., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State–level estimates from national polls. *Political Analysis*, *12*, 375–85.