Spatial Interaction Modelling for Large-Scale Infrastructure Projects

Andrew P Smith*[1] and Nik Lomax†[2]

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

January 29, 2019

Summary

We propose a modelling methodology for exploring and quantifying the impact of large-scale infrastructure changes on internal migration. Choosing measures of the potential emissiveness and attractiveness of geographical units and how the proposed infrastructure changes these values (e.g. housing stock), a spatial interaction model can be used to compute changes to future population distributions. We use the prototype model to investigate the impact on population of the proposed CaMKOx corridor and show that the results are qualitatively in line with expectations. We outline how the model is being developed further in order to produce more quantitatively reliable outputs.

KEYWORDS: Spatial Interaction Models, Population Projections, Internal Migration, Infrastructure.

Introduction

In order to understand the impact of - and perhaps also justify - proposed large-scale infrastructure projects, it is necessary to quantitatively assess the various impacts that the infrastructure changes may have.

The UK government is in an advanced stage of planning or commissioning a number of linked major infrastructure projects along the 'East-West Arc': a region bounded by Oxfordshire in the west, running though Milton Keynes to Cambridgeshire in the east, also sometimes referred to as the **CaMKOx** corridor. This region already contains a number of high-tech and high growth industries plus the prestigious Universities of Oxford and Cambridge, but growth is considered to be hampered by a lack of direct transport links and a lack of affordable of housing. Major new road and rail links are planned, along with large-scale housebuilding, including a number of new towns or villages along the corridor (National Infrastructure Commission, 2017).

*a.p.smith@leeds.ac.uk
†nik.lomax@leeds.ac.uk

We present a modelling methodology which we hope will enable policymakers, implementers and researchers explore and quantify the potential impact of large-scale infrastructure changes on internal migration, at a national level but with a finer geographical resolution. These changed migration flows can then be used to alter official population projections, such as those produced by the ONS.

Some qualitative results are presented, using scenarios inspired by the planned CaMKOx corridor infrastructure project.

Background

Spatial interaction models are a class of models that parameterise flows between geographical units as a function of one or more variables, always including some measure of the separation between geographical units. Originally inspired by an analogy with Newtonian gravity they were popularised by researchers such as Wilson (1971).

Only a brief summary of the mathematical theory behind the model is given here for brevity, a more detailed mathematical treatment can be found in Oshan (2016).

The basic form is that the intensity of an arrival process (e.g. Poisson) is governed by three factors: the strength of the emissiveness of the origin, the attractiveness of the destination, and some measure of the separation between origin and destination. In more precise terms:

$$\bar{F}_{ij} = k\frac{V(i)W(j)}{f(d(i,j))} \tag{1}$$

where $\bar{F}_{ij}$ is the expectation of the number of flows, k is some constant, $V(i)$ is an accumulated measure of the emissiveness of an origin i, $W(j)$ is an accumulated measure of attractiveness of a destination j, and $f(d(i,j))$ the measure of separation, or cost function.

In the 'gravity' variant of the model the overall emissiveness is the product of the m individual (and independent) emissiveness factors

$$V(i) = \prod_m v_m(i)^{\mu_m} \tag{2}$$

and similarly for attractiveness, with n individual factors.

$$W(j) = \prod_n w_n(j)^{\alpha_k} \tag{3}$$

and separation

$$f(d(i,j)) = d(i,j)^{\beta} \tag{4}$$

In logarithmic form, the equation is linear in the parameters (k, μ, α, β):

$$\ln \bar{F}_{ij} = \ln k + \sum_m \mu_m \ln v_m(i) + \sum_n \alpha_n \ln w_n(j) - \beta \ln d(i,j) \tag{5}$$

which is why this class of models are sometimes referred to as *log-linear*. The parameters (k, μ, α, β) can be fitted to observations using a multivariate linear regression, and goodness-of-fit statistics such as R^2 and the root mean squared error computed.

The gravity model is constrained only in the total flows over the entire region. For more constrained models, either production (emission), attraction, or both are constrained to the appropriate marginal totals. Thus the factors in equations 2 and/or 3 are replaced by constant balancing factors that constrain the flows from and/or to a region to the observed values:

$$V(i) = A_i = 1/\sum_j \frac{W(j)}{f(d(i,j))} \tag{6}$$

removing the μ dependency from equation 1, and likewise

$$W(j) = B_j = 1/\sum_i \frac{V(i)}{f(d(i,j))} \tag{7}$$

removing the α dependency.

It should be noted that whilst the constrained models tend to fit observed data more closely than unconstrained models, due to having more parameters (the balancing factors), in practice they have fewer degrees of freedom. The balancing factors are fixed by equations 6 and 7. So, for production constrained models there is/are no μ parameter(s) to fit, for attraction-constrained models there is likewise no α parameter, and for doubly constrained models only β exists as a free parameter (other than the intercept k).

Thus, if we want to gauge the impact of a change to the e.g. housing stock at the destinations, an attraction- or doubly-constrained model would not be appropriate as the flows are not a function of *any* attractiveness value.

Methodology

In this section we outline the methodology by way of an example scenario based on the CaMKOx proposals.

Example Scenario

In this (extreme) hypothetical scenario, we envisage that over the period 2020-2024, there will be:

- 25000 new jobs created per year in Oxford, Milton Keynes and Cambridge (a total of 375000 new jobs).
- 4000 new houshold spaces built per year in each of the 13 local authorities within the corridor (a total of 260000). **This is additional to the ONS household projections.**

The geographical scope is local authority districts (LADs) in Great Britain, as (principal) subnational and household population projections are available up to 2041 (2039 for Wales). The observed flows are 2011 census migrations rescaled appropriately, as depicted below. (Note that despite its appearance this matrix is not symmetric.)

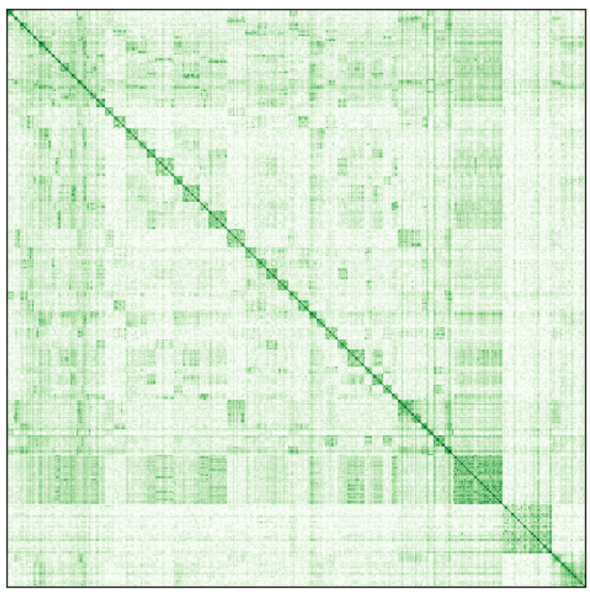

Figure 1: Depiction of the origin-destination matrix for the local authority districts in Great Britain (displaced log scale).

We chose a production constrained model, with the attractiveness parameters being households (as a proxy for household spaces) and jobs. The model is then fitted to the *status quo* projections for flows, households and jobs, resulting in a parameterisation fixing the values of $(\alpha_{houses}, \alpha_{jobs}, \beta)$. The next step is to perturb the attractiveness within the region, as outlined above. Thus, for example, the value for households in each of the 13 CaMKOx LADs is increased by 4000 and the number of jobs in the three centres increased by 25000. The flows are then recomputed, resulting in a new O-D matrix, from which the overall change in population of *all* LADs can be computed. The resulting changes are accumulated year on year.

The algorithm can be run over the entire time horizon of the population projection (2016-2041) and then used to construct a custom variant projection at subnational (i.e. LAD level).

Implementation

The model implementation is in python, at its core using the `SpInt` module (Oshan, 2016), part of the `pysal` spatial analysis package.

The `ukcensusapi`(Smith, 2017) and `ukpopulation` (Smith and Russell, 2018) packages are used for automating access to 2011 census data and subnational population projections respectively.

This model workflow is being implemented as a Data Analytics Facility for National Infrastructure (DAFNI) pilot project with the help of the Science & Technology Facilities Council (STFC): the idea is to make the model user-friendly, easily configurable, and available to non-technical users such as policymakers and industry insiders.

Preliminary Results

The results obtained by the hypothetical scenario and algorithm described in the previous section. The production-constrained model fits well to the observed migration flows, as illustrated in the scatterplot, with an R^2 of 0.94 and and RMS error of approximately 2.7.

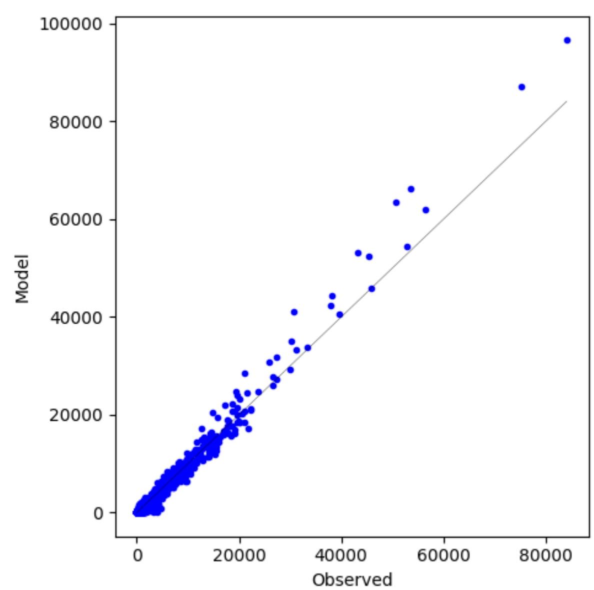

Figure 2: Scatterplot of model migrations versus observed.

The table below shows some typical model parameters. (There will be a slight variation from year to year.)

Table 1: Fitted model parameters

Name	Value
k	1.46
α_{houses}	0.55
α_{jobs}	0.13
β	-1.34

The change in the projected population of Oxford predicted by the model is shown in graph below, with the baseline being the official (principal variant) subnational population projection.

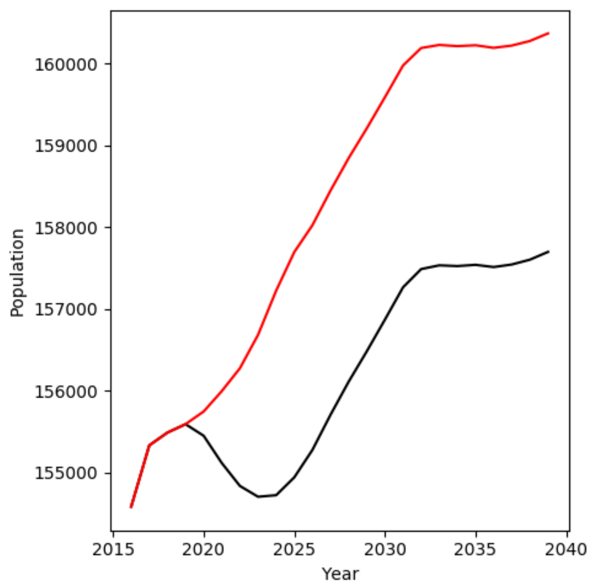

Figure 3: Population of Oxford under scenario (red) and baseline projection (black).

Finally, the map below shows the wider impact of the scenario, showing that there is an increase in population across the CaMKOx corridor which is greatest in the centres, and that the population decreases most in areas close to the corridor.

Preliminary Analysis

The most obvious issue is that the scenario is extreme in terms of the increase in housing stock and jobs, yet the change in population is very modest in comparison. Figure 3 shows an approximate

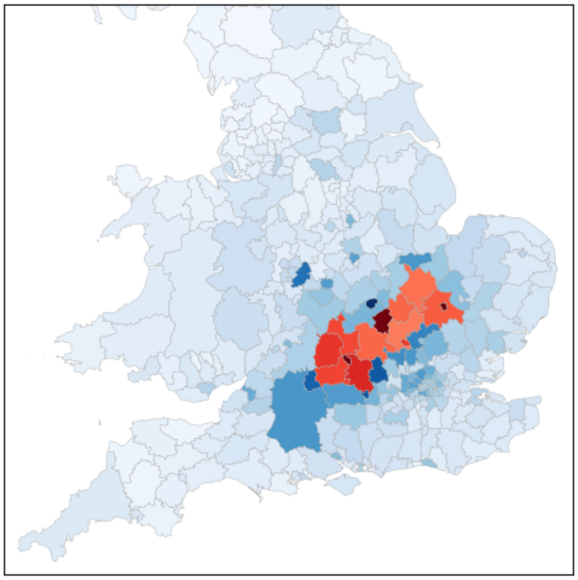

Figure 4: Net change in LAD population under scenario. Red is increase, blue is decrease, the darker the greater.

increase in the population of Oxford of only 3000 people as a result of a scenario in which the number of jobs increases by 20000 and the number of housing units increases by 125000.

Whilst the model results appear to be sensible from a qualitative perspective, there is clearly something awry with the model or the assumptions when it comes to quantitative predictions, which needs to be addressed.

The most obvious reason for this is that, by considering only migration flows from the 2011 census, we are restricting the population to only those that did move house, rather than the population at large. We must consider people who would otherwise not move at all without the infrastructure changes. The magnitude of this oversight is in line with the disparity between scenario and results.

Conclusion

Ongoing work will focus on resolving the disparity between the magnitude of the scenario changes and that of the changed migration flows. We shall then experiment with different quantities - and combinations thereof - for attraction, additionally devising more realistic scenarios over realistic timelines.

Acknowledgements

This work is being carried out as part of EPSRC-funded MISTRAL programme, part of the Infrastructure Transitions Research Consortium (https://www.itrc.org.uk/).

The authors would like to thank the STFC DAFNI team for providing the resources for an ongoing pilot scheme which implements this work on the DAFNI platform.

Biographies

Andrew Smith is a Research Fellow at the University of Leeds. After a Ph.D. in computational physics he spent nearly 20 years working as a Quantitative Analyst. He returned to academia to pursue a more research-based career. He is a keen advocate of software development practices for reproducible research.

Nik Lomax is a University Academic Fellow in Data Analytics, based in the School of Geography, University of Leeds. Niks research interests are in population estimation and projection. Nik is a Co-Investigator of the ESRC funded Consumer Data Research Centre and is a funded Turing Fellow.

References

National Infrastructure Commission (2017). Partnering For Prosperity: A new deal for the Cambridge-Milton Keynes-Oxford Arc.

Oshan, T. (2016). A primer for working with the spatial interaction modeling (SpInt) module in the python spatial analysis library. 3(2):R11–R23.

Smith, A. P. (2017). UKCensusAPI: python and R interfaces to the nomisweb UK census data API. *Journal of Open Source Software*, 19(2):408.

Smith, A. P. and Russell, T. (2018). ukpopulation: unified national and subnational population estimates and projections, including variants. *Journal of Open Source Software*, 28(3):803.

Wilson, A. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A*, 3:1–32.