Predictive limitations of spatial interaction models: a non-Gaussian analysis

B. Hilton*, A. P. Sood*, T. S. Evans,

(*) Equal first authors

Centre for Complexity Science, and Theoretical Physics Group,
Physics Dept., Imperial College London, SW7 2AZ, U.K.

16th September 2019

arXiv:1909.07194v1 [physics.soc-ph] 16 Sep 2019

Abstract

We present a method to compare spatial interaction models against data based on well known statistical measures which are appropriate for such models and data. We illustrate our approach using an widely used example: commuting data, specifically from the US Census 2000. We find that the radiation model performs significantly worse than an appropriately chosen simple gravity model. Various conclusions are made regarding the development and use of spatial interaction models, including: that spatial interaction models fit badly to data in an absolute sense, that therefore the risk of over-fitting is small and adding additional fitted parameters improves the predictive power of models, and that appropriate choices of input data can improve model fit.

1 Introduction

The ability to predict the number of vehicles, the amount of goods, or the spread of disease between two locations, using only limited data about each location, is important in a variety of academic disciplines. Problems of this nature can be studied using 'spatial interaction models'. Given some measures of the importance of each site i, and the distance d_{ij} between two sites i and j, these models predict the flow from site i to site j, denoted F_{ij}. The distance d_{ij} need not be a geographical distance; it could reflect the cost of travel or other socio-economic measures of separation. These models only predict flows between distinct sites, and so $i \neq j$.

The nature of spatial interaction models and the associated data means that the Gaussian statistical techniques cannot always be used, though they are often found in the literature. Our primary goal is to improve upon the statistical analysis commonly carried out in the literature and apply this improved analysis to determine the relative effectiveness of key examples from two popular families of models: gravity models and radiation models. Additionally, our methods are used to identify which features of these models give the greatest improvement in results.

We will start by reviewing the data used in our work. In Section 3, we will look at the various spatial interaction models we consider. The statistical methods used are described in Section 4 and the results shown in Section 5. We will conclude with a discussion of our work.

2 Data

It is inherent to the nature of statistical analysis that models must be compared against data. In order to minimise the introduction of bias resulting from the selection of a particular data set, we chose to use a freely available and widely used data set: the county-to-county worker flow data from the US Census 2000 [2]. For instance, this data was used in the paper in which the radiation model was introduced [1]. Using this data, the radiation model was compared favourably against the gravity model in [1] . To verify that our conclusions are a result of the models and the US commuter flow system rather than merely a feature of the specific data

set, we also used the parallel data set from the American Community Survey 2009–2013 [3]. Further, we obtained the populations of the counties at the census dates of 2000 [4] and 2010 [5]. Though we often use the language of commuting to describe our approach, our methods are data set agnostic, and therefore our results have wider applicability.

In the US Census 2000, there are 3109 counties or their equivalents within the 48 contiguous United States. These form the sites used by our spatial models. The US Census 2000 asked, for each person listed: "at what location did this person work *last week*?" Respondents were further instructed "if this person worked at more than one location, [to] print where he or she worked most last week." This means that our figures for commuting will include data from those who occasionally work at other locations for a few days and these are likely to inflate the number of long distance trips recorded relative to data representing where a person worked for most of a year. Information on the distribution of flows is shown in Fig. 1 and Table D2.

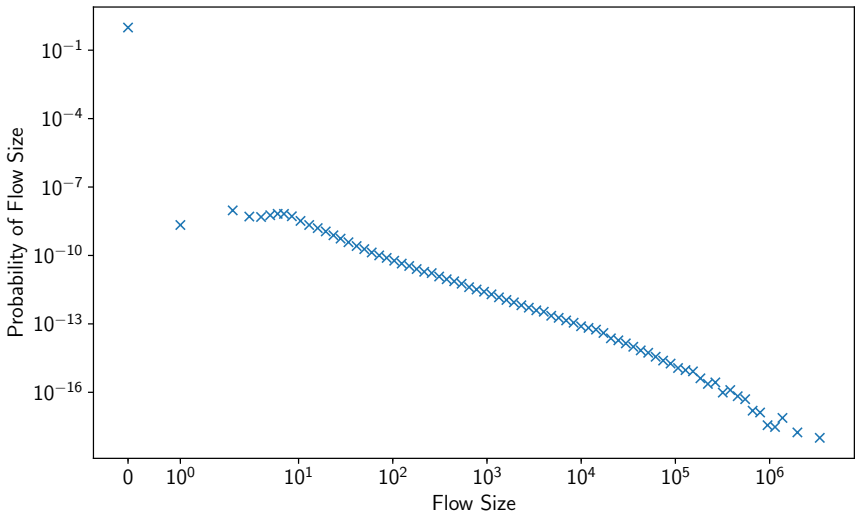


Figure 1: The distribution of commuter flow sizes in the US Census 2000 data [2].

From this data we define three values associated with each site i which are generic to many spatial interaction contexts: the site population P_i, the flow into a site I_i, and the flow out O_i. While these three values are likely to be correlated at each site for our commuting data, there are large individual differences as sites may have developed specialised functions. For instance in the US Census 2002 data [2], many people work in San Francisco county who commute in from other counties (265,291 people), but fewer live in San Francisco county and commute elsewhere (130,036 people).

We use this data on sites, the population and the number of commuters arriving and leaving a site, to determine model parameters associated with site importance. These are used to set our site model parameters. We use w_i (site weight) as a generic site importance model parameter but, depending on the model, we can use up to three more specific site parameters to characterise a site: a repulsiveness parameter t_i controlling the total flow out of a site, an attractiveness parameter n_i which controls the flow into a site, and in some cases an 'aspiration' parameter m_i which controls how far a commuter will travel.

The distances needed for the models were great-circle distances between the geographical centres of each pair of US counties. These data were obtained from the National Bureau of Economic Research [6].

3 Models

3.1 Gravity Models

One of the most widely used spatial interaction models is a class of models known as 'gravity models', which have been used in a variety of socio-economic contexts since the 19th Century but have seen much development since the 1950s (see [7, 8] for general reviews).

The simplest gravity model is given by

$$\widehat{F}_{ij} = w_i w_j f(d_{ij}) \,, \tag{3.1}$$

where $\widehat{F}_{ij}$ is the model's estimate of the flow F_{ij} from i to j. The w_i and w_j parameters are the weights of sites i and j respectively, some measure of the importance of sites. The function $f(d_{ij})$ is some monotonically decreasing function of (generalised) distance: the 'deterrence function'. This function is often chosen without theoretical motivation and typically includes additional parameters which must be determined using previously known data. Such flexibility in the form of the deterrence function can be regarded as a key limitation of the gravity model [1]. However, in practice simple forms are often found to be effective. Common deterrence functions include exponentials ($f(x) = e^{-\beta x}$ for some $\beta > 0$) [9] and power laws ($f(x) = x^{-\beta}$ for some $\beta > 0$) [10, 11]. The deterrence functions invariably include a global parameter, β in our examples, which is the same for all pairs of sites. This might be set from data, for instance β^{-1} represents a typical length scale for the exponential form. However, such global model parameters are often determined by varying their values until the model has the best possible fit to the data.

In order to accurately test the extent of the difference in predictive power between models, they must share any feature which is not being explicitly compared. All the models considered here are 'production constrained' models in which the output of each site is fixed by a model parameter for that site. So rather than the simplest gravity model of (3.1), we will use a production constrained gravity model [7, 12, 13] of the form

$$\widehat{F}_{ij} = \frac{t_i n_j d_{ij}^{-\beta}}{\sum_k n_k d_{ik}^{-\beta}} \,. \quad (i \neq j) \tag{3.2}$$

This obeys $\sum_j \widehat{F}_{ij} = t_i$, the production constraint making the site model parameter t_i equal to the total flow leaving site i. The n_j parameter is some measure of the 'attractiveness' of site j which controls the flow into each site, though this is not necessarily equal to the flow into site j. Even if $t_i = n_i$ (as is often assumed), it is worth noting that this model already describes an asymmetric flow with $\widehat{F}_{ij} \neq \widehat{F}_{ji}$ in general. Thus, unlike the simple gravity model, this production constrained gravity model can reproduce flow asymmetries present in real data, as illustrated by the example of San Francisco county considered in Section 2.

For our work with the gravity model (3.2), we will set the output site parameter equal to the number of commuters leaving a site, $t_i = O_i$, while the site attractiveness parameter will be set equal to the number of commuters arriving $n_i = I_i$. We will choose the single global model parameter β in (3.2) to be the value that gives the best fit to our data as explained below. For comparison, the gravity model against which the radiation model is compared in [1] also used a power law deterrence function, but had no constraints on inputs or outputs, and used nine fitted parameters (see Section C).

Other forms for the deterrence function in our gravity model were also investigated, but the power law in (3.2) proved the fairest comparison [14, 15].

3.2 The Radiation Model

The radiation model was derived in the context of commuter flows, using the underlying assumption that a worker seeking employment will accept the most proximate job offer that meets their requirements. The most general form of the radiation model in [1] is

$$\widehat{F}_{ij} = t_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \,. \tag{3.3}$$

The model parameter t_i controls the total flow leaving each site i and we have that $\sum_j \widehat{F}_{ij} \approx t_i$ making this radiation model a production constrained model. We will return to this approximation below. The n_i model parameter is the number of opportunities drawing commuters into site i, the site attractiveness parameter in this model. The s_{ij} is given by the sum of all opportunities of sites closer to i than j, the intervening opportunities measure of [16]

$$s_{ij} = \sum_{k|k\neq i} n_k \theta(d_{ij} - d_{ik}) \,. \tag{3.4}$$

Here $\theta(x)$ is one for $x > 0$ and zero otherwise so the sum does not include n_i or n_j. The last model parameter m_i is a measure of the aspiration of commuters leaving site i. That is, the larger the value of m_i, the more commuters leaving site i want and so the further they have to travel to achieve their aspirations. Thus, m_i does not alter the total flow leaving site i, but m_i controls the distribution of the flow leaving site i.

We noted above that the flow leaving each site i is not exactly equal to the t_i model parameter. This is easily corrected [17] to give us a normalised form of the radiation model

$$\widehat{F}_{ij} = \left(\frac{N_c + m_i - n_i}{N_c - n_i} \right) t_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \,. \tag{3.5}$$

Here $N_c = \sum_i n_i$ is the total number of opportunities in the system. With this normalisation, the production constraint is perfectly enforced in the normalised radiation model, $\sum_j \widehat{F}_{ij} = t_i$. If $N_c \gg n_i, m_i$ then this normalised radiation model form is almost the same as (3.3) showing this correction (the factor in brackets) is often small.

One of the important features of the radiation model is that the form is fixed; there is no equivalent here to the choice of deterrence function seen in gravity models. This means there are no explicit global model parameters in the radiation model, such as the β in (3.2). The lack of such global model parameters (as opposed to those parameters linked to site properties) leads to the description in [1] of the radiation model as having a "parameter-free nature".

However, to use the radiation model, or indeed any spatial interaction models, we must first relate the site model parameters to values in our data. Mapping these site model parameters to data values can be done in many ways and this leads to a family of radiation models. The versions of the radiation model analysed here are summarised in Table 1, with more details given in Appendix B. In particular, the original radiation model [1] used the total population P_i of site i to set the three site model parameters with $m_i = n_i = P_i$ and $t_i = \alpha P_i$: model F in Table 1 (see also (B.6)). Note that α is a single fitted global model parameter which illustrates how such parameters can be introduced to spatial interaction models through the mapping of data to model parameters. In such a case, even the radiation model is no longer parameter free in the sense defined above. In our examples only our radiation models A to E are parameter free, the remaining radiation models and our gravity model both have one fitted global model parameter.

	Name	$\mathbf{m_i}$	$\mathbf{n_i}$	$\mathbf{t_i}$	Normalised?	Eq.
A.	Total population	P_i	P_i	P_i	×	(B.1)
B.	Departing commuters	O_i	O_i	O_i	×	(B.2)
C.	Departing commuters, Normalised	O_i	O_i	O_i	✓	(B.3)
D.	Arriving & Departing, Naïve split	O_i	I_i	O_i	×	(B.4)
E.	Arriving & Departing, Revised split	I_i	I_i	O_i	✓	(B.5)
F.	Total population, Fitted factor	P_i	P_i	αP_i	×	(B.6)
G.	Departing commuters, Fitted factor	O_i	O_i	αO_i	×	(B.7)
H.	Arriving & Departing, Revised, Fit factor	I_i	I_i	αO_i	✓	(B.8)

Table 1: A summary of the different versions of the radiation model. The tick in the 'Normalised?' column indicates that a model uses a normalisation that enforces the production constraint exactly (3.5), while a cross in that column indicates that the original form (3.3) is used for that model. In each case we specify which of the site data values, (P_i population, I_i commuters arriving, O_i commuters leaving) is used for the model site parameters (aspirations m_i, opportunities n_i, out flow t_i). See Table A1 for a summary of the notation. The single global model parameter α is found by optimising the fit of the model to the data. The model used in [1] is equivalent to our model F. The full equations are given in Appendix B as indicated in the final column.

4 Statistical Methods

There are two statistical challenges when dealing with spatial interaction models and data. Suitable statistical measures must be chosen to evaluate how well the models' parameters (where present) give the best fit to data, and secondly some metric must be selected to establish which model is 'best'. However, the choice of this metric is not obvious. For example, one may decide to prioritise accurately predicting which pairs of sites will have zero flow ($F_{ij} = 0$) over gaining accurate estimates of the sizes of large flows. We attempt to sidestep such issues by asking in an unbiased statistical sense how *probable* the models are. In order to achieve this, it is worth first considering some of the techniques found in the existing literature.

4.1 Common techniques for comparing models

The Sørensen-Dice coefficient is often used to compare models against real data, for instance [18, 19, 17, 20, 21], but it is measure which lacks a rigorous statistical basis. The Kolmogorov-Smirnov test is also seen in spatial modelling [22]. However, the test requires the two input functions to be independent, and is therefore invalid when a model has fitted parameters estimated using the same data from which the weights are drawn [23].

Sometimes a comparison is made using statistics which assume underlying Gaussian distributions: i.e. where it is assumed that the distribution $p(F_{ij}|\widehat{F}_{ij})$ (the probability that the flow is found to be F_{ij} given a predicted flow $\widehat{F}_{ij}$) is Gaussian for any i, j. Examples include mean squared errors [24], the coefficient of determination R^2 values [17, 24], and Pearson correlation coefficients [25]. However, real data sets feature no negative flows and usually a high proportion of very small flows, meaning that the distributions of flows between any one pair of sites cannot be assumed to be Gaussian.

Additionally, none of these tests measure the effects of fitting parameters: varying a model parameter to fit data can improve the accuracy of the model for that data set, but at the expense of reducing the model's predictive power on other data sets.

4.2 Poisson regression

The limitations of these techniques motivate the application of alternative statistical methods [26]. Our starting point is the determination of the error distribution $p(F_{ij}|\hat{F}_{ij})$. Were there data on commuting for every day over a few years, we could look at the actual fluctuations in flows and examine the validity of this statistical model. However, without this data, and given that the chosen data sets (see section 2) contain discrete count data, the simplest assumption we can make is to assume that the flow F_{ij} between any one pair of sites is Poisson distributed: that for any given pair of sites, we model the probability of finding flow F_{ij} in the data as $p(F_{ij}|\widehat{F}_{ij}) = \exp(-\widehat{F}_{ij})(\widehat{F}_{ij})^{F_{ij}}/(F_{ij}!)$, where we have taken the model estimate $\widehat{F}_{ij}$ to be the mean of our distribution. For small flows, the majority of values in our data, this is significantly different from a Gaussian distribution.

Using these assumptions, we can now ask how probable it is that that data would be observed given the distribution predicted by the model. This is known as 'Poisson regression'. Using Poisson regression, we calculate the log-likelihood $\ln L$ for model values $\widehat{F}_{ij}$, given some flow data F_{ij}, where we retain the option to work only with flows above a minimum value $F_{\min}$, namely

$$\ln L(F_{\min}) = \sum_{\substack{i,j \\ i\neq j,}} \left(-\widehat{F}_{ij} + F_{ij}\ln(\widehat{F}_{ij}) - \ln(F_{ij}!)\right)\theta(F_{ij} - F_{\min})\,. \tag{4.1}$$

Note that the predicted flow is never zero: we always get a finite result. Log-likelihood functions and maximum likelihood estimations provide a rigorous way to estimate fitted parameters, and

to quantitatively compare how well models fit data. While adding more fitted parameters will always improve the fit of the model to the data, this risks over-fitting to the particular data set used, reducing models' general predictive power. Thus log-likelihood values cannot tell us whether or not these fitted parameters have truly improved the model, and we need a different measure of model effectiveness.

Ideally, in order to test model effectiveness, a form of cross-validation would be used, wherein the model is fit to some data and then tested against a second data set drawn from the same distribution [27]. However, the difficulty in obtaining multiple real data sets drawn from the same distribution means that some other model selection criterion must be used. One widely-used method is the Bayesian information criterion [28, 29] given by

$$\mathrm{BIC}(F_{\min}) = k\ln(n) - 2\ln(L(F_{\min})), \tag{4.2}$$

where k is the number of fitted parameters, n is the number of data points, and L is the likelihood. The Bayesian information criterion can be used to compare models against a single common data set. It has a robust statistical basis [30], introducing a penalty that increases with the number of fitted parameters. This penalty is sometimes considered too harsh [31].

Finally, it would be useful to have a measure of goodness-of-fit. Log-likelihood (and therefore Bayesian information criterion) values can only be used to compare models. They allow us to say one model matches real data more closely than another, but do not conclude that they resemble real data well in any absolute sense. For this, we need some value against which likelihood values can be compared. One method is to use the saturated likelihood L_s: the value that the likelihood would take if the predictions from the model exactly matched the data. The ratio of the actual likelihood to this saturated value must be between zero and one and can be used to define the deviance $D \geq 0$ through $L/L_s = \exp(-D/2)$. In our case we have that

$$D(F_{\min}) = 2\sum_{\substack{i,j \\ i\neq j,}} \left((\widehat{F}_{ij} - F_{ij}) + F_{ij}\ln(F_{ij}/\widehat{F}_{ij}) \right) \theta(F_{ij} - F_{\min})\,. \tag{4.3}$$

For all three of these statistics (log-likelihood, BIC and deviance), the lower the magnitude, the better the model fits the data.

5 Results

Figure 2 shows the log-likelihoods for the various versions of the radiation model in Table 1, and for the production constrained gravity model of (3.2), calculated using the commuting data of the US census 2000 [2]. The exact values of the log-likelihoods and associated standard errors are shown in Table 2. Radiation model D (see Table 1) has been omitted from the figures in this section because of its extremely large log-likelihood — it is far worse than any other model. This is unsurprising since this model uses out-flow data O_i for the 'aspiration' parameter m_i: this result thus acts as a simple check of our approach in dealing with radiation model parameters.

These log-likelihoods allow for an initial comparison between models. Radiation model A ('Populations') is the worst model other than radiation model D. The total flow out of each site in radiation model A is generally significantly larger than real flows, leading to its poor performance. Changing the site model parameters to be equal to the departing commuters data value O_i (radiation model B - 'Departing commuters') improves the model significantly, as expected. Adding in a normalisation (radiation model C - 'Departing commuters, Normalised') only results in a slight improvement. This is because of the large number of commuters in the USA; the largest possible value of the normalisation factor is 1.0168 and the mean value is

Model	Log-likelihood $\ln L$	Error due to fit
A	-84209685.1963	N/A
B	-32332647.1270	N/A
C	-32289322.6070	N/A
D	-700368330.4228	N/A
E	-26019946.6913	N/A
F	-27384424.2552	0.0001
G	-25664008.8022	0.0003
H	-19395250.0656	0.0003
gravity Model	-13505652.9746	0.0003

Table 2: The log-likelihood values (4.1) for the various radiation models of Table 1 and the production constrained gravity model of (3.2). The standard error in the log-liklihood comes from the uncertainty in the value of any fitted parameters. Thus we have no estimate of uncertainty for models without a fitted parameter, as indicated by an "N/A" entry.

1.0003. Using a model in which the site model parameters for input and output flow, n_i and t_i respectively, are related to the corresponding data values, I_i and O_i respectively, produces the best results. This is radiation model E — 'Arriving & Departing, Revised'.

Every model with an additional fitted factor works better than its counterpart: F is better than A, G is better than B, and H is better than E. Moreover, even model F ('Populations, Additional Fitted Factor'), which one might expect would overestimate the flows due to its large site parameter values, arrives at a better log-likelihood than model B ('Departing Commuters'). However, model G ('Departing Commuters, Additional Fitted Factor') is more successful than model F ('Populations, Additional Fitted Factor'), indicating that the matching model site parameters to appropriate site data values still has merit.

The explanation for the particularly strong improvement resulting from fitting lies in the idea, corroborated below, that none of these models fit real data particularly well. Consequently, allowing a parameter to vary until the best possible value is found optimises the models' effectiveness far more than ensuring model site values are well matched to data when the overall model only approximates reality very roughly. Intriguingly, our gravity model (3.2), whose form was chosen so as to be comparable to our radiation models, matches our real data more closely than any of our radiation models.

Log-likelihoods alone do not tell the full story. Fig. 3 shows the BIC values for the models. Despite the BIC often being regarded as overly harsh with regards to additional parameters [31], the trend shown is exactly the same as in Fig. 2. This is because the penalty applied by the BIC is $k \ln(n)$, and $\ln(n)$ is only 8.04. This is much smaller than the log-likelihood values of order 10^7. We can therefore conclude that there is very little risk of over-fitting, and that adding relevant additional fitted parameters significantly improves the models.

Fig. 4 shows the deviance values for each model. The blue bars are almost identical in appearance to Fig. 3 because the magnitude of the actual log-likelihood ($\sim 10^7$) far exceeds that of the saturated log-likelihood (~ 40000). This comparison underscores how poorly these models fit real data in an absolute sense.

Given that most of the data is zero, we might wonder to what extent these trends are an artefact of how well the zero-flows are predicted rather than how well the models predict the exact sizes of the other flows. Fig. 4 addresses this by considering the deviance values for the models compared against truncated data sets, in which only flows above a certain $F_{\min}$ are considered. The figure shows that the trends are almost completely as above. The only exception is for flows greater than 10,000 predicted by model A ('Populations'). This model uses the largest weights and therefore overestimates most flows, but predicts more reasonable values

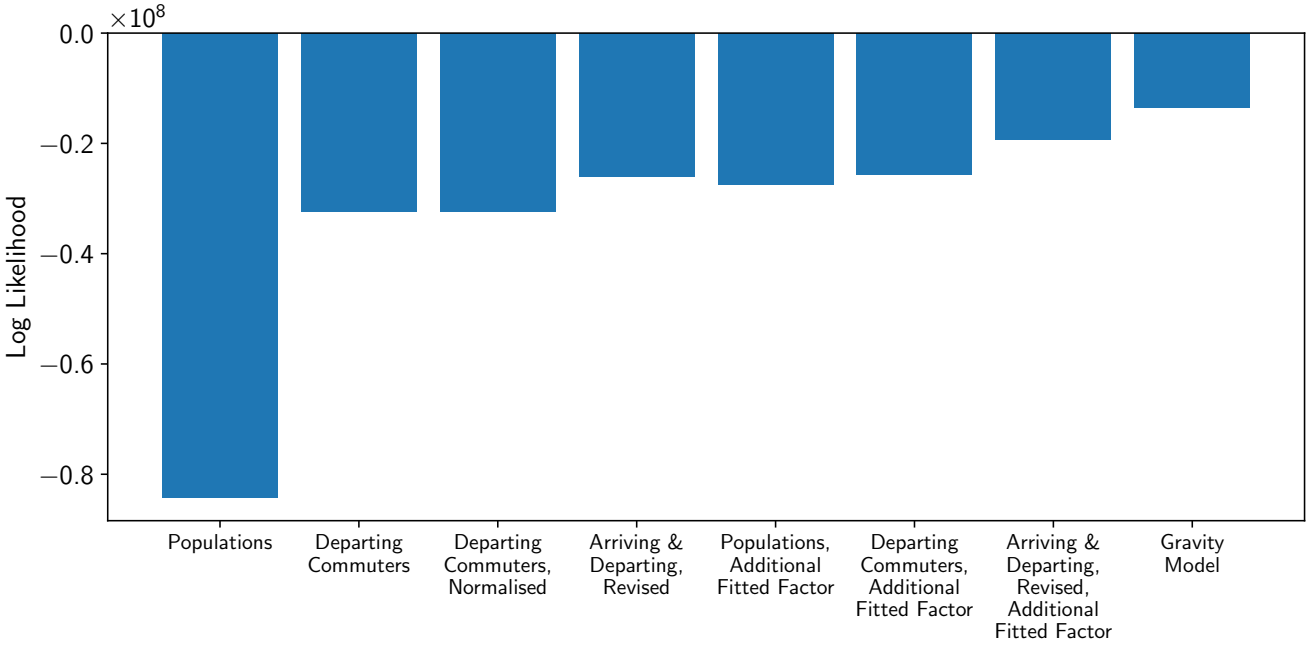


Figure 2: The log-likelihood values (4.1) for radiation models A, B, C, E, F, G, H (from left to right) described in Table 1, alongside the production constrained gravity model of (3.2). Less negative values represent better models. These data are from the US Census 2000 [2]. The uncertainty in the value of any fitted parameter led to a negligible change in these results.

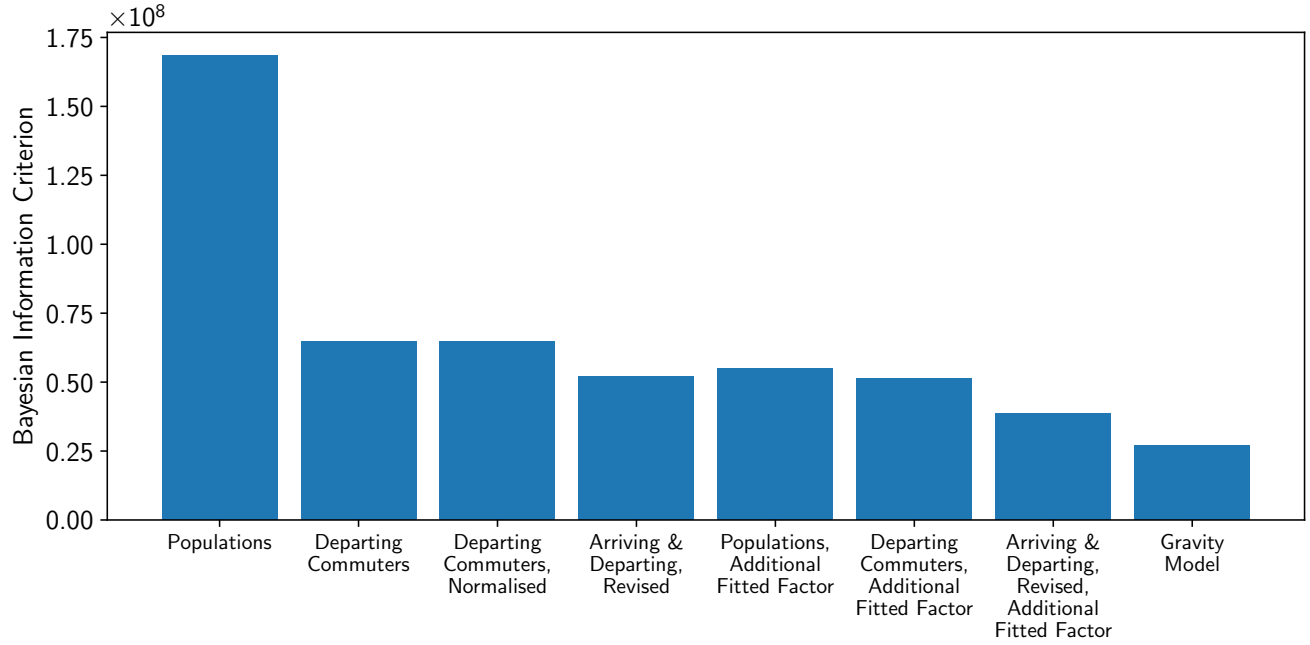


Figure 3: Bayesian information criterion values (4.2) for radiation models A, B, C, E, F, G, H (from left to right) described in Table 1, alongside the production constrained gravity model of (3.2). Lower values represent better models. These data are from the US Census 2000 [2].

for the larger flows. This suggests larger flows are therefore systematically underestimated by the other models. However, only 0.022% of flows predicted by model A are greater than 10,000, so this trend does not significantly affect the validity of our overall conclusions.

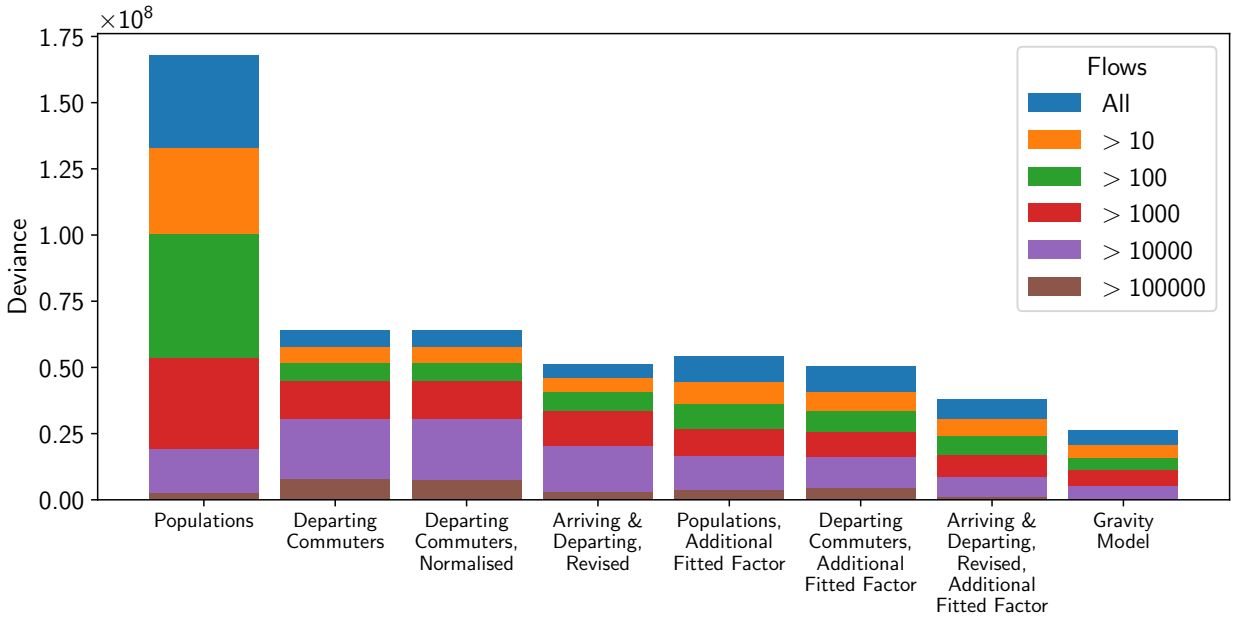


Figure 4: Deviance values (4.3) for radiation models A, B, C, E, F, G, H (from left to right) described in Table 1, alongside the production constrained gravity model of (3.2), with data sets which are truncated using the minimum values shown in the legend. Lower values represent better models. These data are from the US Census 2000 [2]. For each model the top of a coloured bar represents the deviance value for that model when the data is limited to flows above the value indicated in the legend.

Finally, we consider the second data set (the American Commuter Survey 2009-2013 [3]). In Fig. 5 we show the deviance values for these data, though the trends are the same in all our measures. The results for this data reinforce all of our conclusions.

6 Conclusions and Discussion

For our data on modern US commuter flows, the most accurate flow predictions came from the production-constrained gravity model. Looking at the truly "parameter free" radiation models, that is models with no fitted global parameters, radiation model E ('Arriving & Departing, Revised') was most successful. The set of parameter free radiation models A–E showed that matching each model parameter to an appropriate data value improves the model performance as we should expect. This radiation model E benefits from a number of improvements over the original radiation model: choosing an appropriate input data set (e.g. the number of individuals who leave each site rather than population); correctly adjusting the model to include measures of both attractiveness and repulsiveness for each site; and introducing the correct normalisation.

Another conclusion was that adding an additional global parameter, and setting that parameter by finding the best fit, improves the performance of any model. The penalty of having an extra parameter is negligible for our data sets while there is vast room for improvement in what are poor fits in statistical terms. This is why the radiation model which best fits both data sets is radiation model H. This is the same as radiation model E, but with a single additional fitted parameter.

Despite these improvements, and in direct contrast with the results of [1], our statistical measures show that for these US commuting data sets the radiation model is vastly inferior to an appropriately chosen gravity model for most realistic purposes, i.e. where there is data that

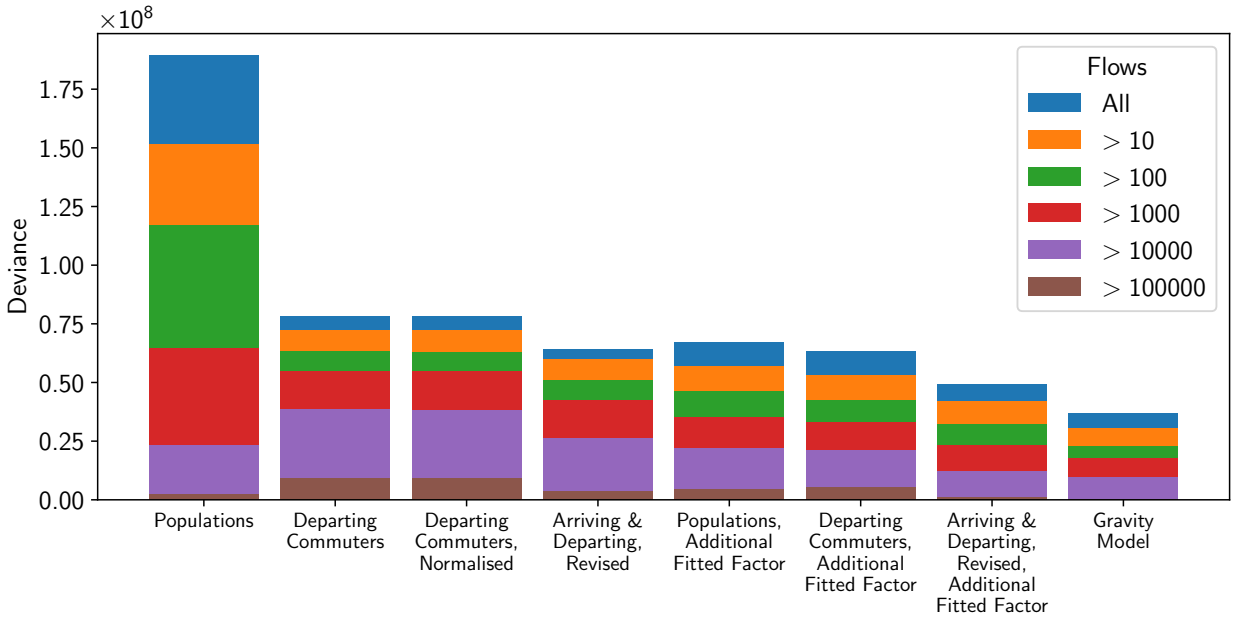


Figure 5: Deviance values (4.3) for radiation models A, B, C, E, F, G, H (from left to right) described in Table 1, alongside the production constrained gravity model of (3.2), with data sets which are truncated using the minimum values shown in the legend. Lower values represent better models. These data are from the American Commuter Survey 2009-2013 [3]. For each model the top of a coloured bar represents the deviance value for that model when the data is limited to flows above the value indicated in the legend.

can be used to fit parameters – what appears to be a small visual difference between models in our plots represents a large numerical difference.

The relative success of our chosen gravity model highlights another result. The gravity model used in [1] made less successful predictions than the radiation model in spite of its having nine fitted parameters to the latter's zero. This underscores the importance of constraints, and the requirement that only models with corresponding constraints be compared against each other when the impact of these constraints is not the topic of investigation. This is why in this work all our models are production constrained in order to make our comparisons fair.

By examining the deviance values, we further established that none of these models fit our data well in an absolute sense. This is unsurprising: the large number of factors affecting commuter flows – geographical and socio-economic – limit the extent to which a simple model with very few parameters could make accurate predictions.

Our work leads us to make recommendations for spatial interaction modelling in general. First, we suggest that non-Gaussian regression (in particular Poisson regression) as applied to log-likelihood, Bayesian information criterion and deviance, are good statistical methods to use when analysing spatial interaction models. These have a firm theoretical grounding and provide an unbiased statistical approach. Second, we should make sure any feature which is not being explicitly tested is controlled for. Here, this means all our models enforce the production constraint. In fact, it would be trivial to add the input constraint into all these models, as is standard for gravity models [7]. Such an improvement requires no additional parameters. Third, the small penalty in the Bayesian information criterion arising from additional parameters, as well as the lower deviance values of models with fitted parameters, attest to the fact that if data exist that can be used for fitting, then a model with many physically relevant parameters can be improved by fitting to this data. Having such fitted model parameters is an advantage, not

a disadvantage. Fourth, models should make use of as much available information as possible. We found that if we used the actual commuter flows in and out of sites in a way that matched that narrative behind a model, then results were better than trying to use the population as some proxy for the actual flows. Lastly, these simple spatial interaction models should be used only to provide an outline of real-world processes, with fitted parameter values giving general insights into spatially-constrained processes. These models are only ever crude approximations of reality.

Acknowledgements

TSE thanks E. Bamis, P. Expert, M. T. Gastner, and R. J. Rivers for useful conversations.

Availability of data and materials

All the data used in this work is publicly available as cited within the text, see [2, 3, 6, 5, 4, 32].

Author's contributions

B. H. and A. P. S. conducted the numerical simulations and data analysis. All authors analysed and interpreted the results, and wrote the manuscript.

References

[1] F. Simini, M. C. Gonzalez, A. Maritan, and A.-L. Barabasi, "A universal model for mobility and migration patterns," *Nature*, 484, pp. 96–100, 2012.

[2] United States Census Bureau, "County-to-county worker flow files," 2001.

[3] United States Census Bureau, "2009-2013 5-year american community survey commuting flows," 2013.

[4] United States Census Bureau, "State and county intercensal tables: 1990-2000," 2016.

[5] United States Census Bureau, "County intercensal tables: 2000-2010," 2017.

[6] National Bureau of Economic Research, "County distance database," 2016.

[7] S. Erlander and N. Stewart, *The Gravity Model in Transportation Analysis*. VSP, 1990.

[8] P. Nijkamp and A. Reggiani, *Interaction, evolution and chaos in space*. Springer Science & Business Media, 2012.

[9] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences*, 106, pp. 21484–21489, 2009.

[10] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, "The complex network of global cargo ship movements," *Journal of The Royal Society Interface*, 7, pp. 1093–1103, 2010.

[11] C. Viboud, O. Bjørnstad, D. Smith, L. Simonsen, M. Miller, and B. Grenfell, "Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza," *Science*, 312, pp. 447–451, 2006.

[12] A. G. Wilson, "A statistical theory of spatial distribution models," *Transportation Research*, 1, pp. 253–269, 1967.

[13] A. G. Wilson, "A Family of Spatial Interaction Models, and Associated Developments," *Environment and Planning A: Economy and Space*, 3, pp. 1–32, 1971.

[14] B. Hilton, "Investigations into the accuracy of spatial interaction models," Master's thesis, Imperial College London, 2019
[`https://doi.org/10.6084/m9.figshare.9751919`.

[15] A. Sood, "An investigation of models of flow in complete spatially embedded networks," Master's thesis, Imperial College London, 2019
[`https://doi.org/10.6084/m9.figshare.9751919`].

[16] S. A. Stouffer, "Intervening opportunities: A theory relating to mobility and distance," *American Sociological Review*, 5, pp. 845–867, 1940.

[17] A. P. Masucci, J. Serras, A. Johansson, and M. Batty, "Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows," *Physical Review E*, 88, p. 022812, 2013.

[18] F. Gargiulo, M. Lenormand, S. Huet, and O. Baqueiro Espinosa, "Commuting network models: Getting the essentials," *Journal of Artificial Societies and Social Simulation*, 15, pp. 6–, 2012.

[19] M. Lenormand, S. Huet, F. Gargiulo, and G. Deffuant, "A universal model of commuting networks," *PLOS ONE*, 7, p. e45985, 2012.

[20] Y. Yang, C. Herrera, N. Eagle, and M. C. González, "Limits of predictability in commuting flows in the absence of data for calibration," *Scientific Reports*, 4, 5662, 2014.

[21] M. Lenormand, A. Bassolas, and J. J. Ramasco, "Systematic comparison of trip distribution laws and models," *Journal of Transport Geography*, 51, pp. 158–169, 2016.

[22] C. Kang, Y. Liu, D. Guo, and K. Qin, "A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint," *PLOS One*, 10, 2015.

[23] D. J. Steinskog, D. B. Tjøstheim, and N. G. Kvamstø, "A cautionary note on the use of the kolmogorov–smirnov test for normality," *Monthly Weather Review*, 135, pp. 1151–1157, 2007.

[24] R. P. Curiel, L. Pappalardo, L. Gabrielli, and S. R. Bishop, "Gravity and scaling laws of city to city migration," *PLOS ONE*, 13, p. e0199892, 2018.

[25] J. Liu, K. Zhao, S. Khan, M. Cameron, and R. Jurdak, "Multi-scale population and mobility estimation with geo-tagged tweets," in *2015 31st IEEE International Conference on Data Engineering Workshops*, IEEE, 2015.

[26] I. Bamis, "Constrained Gravity Models for Network Flows," MSc Thesis, Imperial College, London, 2012.

[27] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *International Joint Conference on Artificial Intelligence (IJCAI)*, 14, pp. 1137–1145, Montreal, Canada, 1995.

[28] A. E. Raftery, "Choosing models for cross-classifications," *American Sociological Review*, 51, pp. 145–146, 1986.

[29] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag New York, 2002.

[30] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, 6, pp. 461–464, 1978.

[31] S. I. Vrieze, "Model selection and psychological theory: A discussion of the differences between the akaike information criterion (AIC) and the bayesian information criterion (BIC)," *Psychological Methods*, 17, pp. 228–243, 2012.

[32] Office For National Statistics, National Records Of Scotland, and Northern Ireland Statistics And Research Agency, "2011 Census Aggregate Data," 2016.

Appendices

A Summary of Notation

A summary of the notation used in this work is given in Table A1.

Notation	Meaning
i, j	Indices of sites.
P_i	The population of site i.
O_i	The number of commuters leaving site i.
I_i	The number of commuters arriving at site i.
N_c	The total number of commuters in the data. This satisfies $N_c = \sum_i O_i = \sum_j I_j$.
F_{ij}	The actual flow from a source site i to a target site j as found in the data.
d_{ij}	A measure of the distance from site i to site j.
$\widehat{F}_{ij}$	The estimated flow from a source site i to a target site j as predicted by some model.
w_i	The site 'weight' model parameter. Controls the flow into and out of a site.
m_i	The site 'aspiration' model parameter. Controls the distribution of flows from site i.
n_i	The site 'attractiveness' model parameter, the number of 'opportunities' . Controls flow into site i.
t_i	Model parameter controlling the total flow leaving site i (site 'repulsiveness').

Table A1: A summary of the different data values (top half) and the different model parameters (bottom half) used in this paper.

B Versions of the Radiation model

In this section we give explicit forms for the Radiation models used in our work written in terms of the actual data values used.

B.1 The Populations Radiation model

The 'Populations' model (model A) is a standard radiation model (3.3) which sets all input parameters equal to the population ($m_i = n_i = t_i = P_i$). This gives us that

$$\widehat{F}_{ij} = P_i \frac{P_i P_j}{(P_i + s_{ij})(P_i + s_{ij} + P_j)} \,. \tag{B.1}$$

Here the intervening opportunities measure s_{ij} is the total population of sites lying closer to site i than site j (excluding site i itself).

B.2 The Departing Commuters Radiation model

The 'Departing Commuters' model (model B) is a standard radiation model (3.3) defined as

$$\widehat{F}_{ij} = O_i \frac{O_i O_j}{(O_i + s_{ij})(O_i + s_{ij} + O_j)} \,, \tag{B.2}$$

Here all input parameters are set equal to the number of commuters who depart from site i, $m_i = n_i = t_i = O_i$. The intervening opportunities measure s_{ij} in this model is the total number of commuters leaving all sites which are closer to site i than site j (excluding the commuters leaving site i itself). Note that in this model the total flow leaving site i is *not* equal to the number of commuters leaving site i, $\sum_{j\in\mathcal{T}_i} \widehat{F}_{ij} \neq O_i$. This model has failed this normalisation criteria but in many cases this can be a small effect so this is not an unreasonable model to use.

B.3 The Normalised Departing Commuters Radiation model

The 'Departing Commuters, Normalised' model (model C) is a normalised Radiation model (3.5) defined as

$$\widehat{F}_{ij} = \left(\frac{O_i}{1 - O_i/N_c}\right) \frac{O_i O_j}{(O_i + s_{ij})(O_i + s_{ij} + O_j)}\,, \qquad N_c = \sum_i O_i \tag{B.3}$$

Again all input parameters are set equal to the number of commuters who depart from site i, $m_i = n_i = t_i = O_i$. The intervening opportunities measure s_{ij} is given in terms of the outputs of intervening sites, exactly as in the Departing Commuters Radiation model (model B) (B.2). Unlike that model, this version is normalised properly so the total flow out of the model equals the associated data value exactly, $\sum_{j\in\mathcal{T}_i} \widehat{F}_{ij} = O_i$ (see (3.5)).

B.4 The Arriving & Departing, Naïve Split, Radiation Model

The 'Arriving & Departing, Naïve Split' radiation model (model D) is based on (3.3), and defined as

$$\widehat{F}_{ij} = O_i \frac{O_i I_j}{(O_i + s_{ij})(O_i + s_{ij} + I_j)}\,. \tag{B.4}$$

Here both m_i and t_i (see (3.3)) are set equal to the number of commuters who depart from site i so $m_i = t_i = O_i$. We set the attractiveness model parameter, the number of opportunities at site j, to be equal to the total number of commuters found in the data to be arriving at site j, so $n_j = I_j$. This last identification then means that the intervening opportunities measure s_{ij} has to be the cumulative number of commuters arriving at all sites closer to i than j (excluding site i), regardless of their origin.

B.5 The Arriving & Departing, Revised, Radiation Model

The 'Arriving & Departing, Revised' model (model E) is a normalised Radiation model (3.5) defined as

$$\widehat{F}_{ij} = \left(\frac{N_c}{N_c - I_i}\right) O_i \frac{I_i I_j}{(I_i + s_{ij})(I_i + s_{ij} + I_j)}\,, \qquad N_c = \sum_i I_i\,. \tag{B.5}$$

Here we have set the site attractiveness parameter n_i and site aspiration parameter m_i at each site i are set equal to the number of commuters arriving at site i, $m_i = n_i = I_i$. The site repulsiveness parameter t_i is set equal to the number of commuters leaving a site O_i and the normalisation factor here ensures this is equal to the total flow predicted from the model, $\sum_j \widehat{F}_{ij} = O_i$. The intervening opportunities measure s_{ij} in this model is the total number of commuters arriving (I_i) at all sites which are closer to site i than site j (excluding the commuters arriving at site i itself).

B.6 The Populations, Additional Fitted Factor, Radiation model

The 'Populations, Additional Fitted Factor' model (model F) is a standard radiation model (3.3) defined as

$$\widehat{F}_{ij} = \alpha P_i \frac{P_i P_j}{(P_i + s_{ij})(P_i + s_{ij} + P_j)} \,. \tag{B.6}$$

Here we have $m_i = n_i = P_i$, but the flow parameter t_i is set proportional to the total population $t_i = \alpha P_i$. This α is a single additional parameter found by optimising the fit to the data using a maximum likelihood estimation. The intervening opportunities measure s_{ij} is the total population of sites lying closer to site i than site j (excluding site i itself).

B.7 The Departing Commuters, Additional Fitted Factor, Radiation model

The 'Departing Commuters, Additional Fitted Factor' model (model G) is a standard radiation model (3.3) defined as

$$\widehat{F}_{ij} = \alpha O_i \frac{O_i O_j}{(O_i + s_{ij})(O_i + s_{ij} + O_j)} \,. \tag{B.7}$$

Here we have set $m_i = n_i = O_i$, the number of commuters who depart from site i, and $t_i = \alpha O_i$ where α is a single fitted parameter. The intervening opportunities measure s_{ij} is therefore the sum of all the outputs, O_i, of intervening sites. Note that this model is not normalised, so $\sum_j \widehat{F}_{ij} \neq \alpha O_i$ and so in turn $\alpha = 1$ is not to be expected even in 'perfect' data generated from the model itself. In reality, the lack of accuracy in the model predictions is likely to ensure some $\alpha \neq 1$ will provide an optimal fit to the data.

B.8 The Arriving & Departing, Revised, Additional Fitted Factor, Radiation model

The 'Arriving & Departing, Revised, Additional Fitted Factor' model (model H) is a normalised Radiation model (3.5) defined as

$$\widehat{F}_{ij} = \alpha O_i \left(\frac{N_c}{N_c - I_i} \right) \frac{I_i I_j}{(I_i + s_{ij})(I_i + s_{ij} + I_j)} \,, \qquad N_c = \sum_i I_i \tag{B.8}$$

Here we have set $m_i = n_i = I_i$ and $t_i = \alpha O_i$. This model is a normalised radiation (3.5). The intervening opportunities measure s_{ij} is the total number of commuters arriving at all sites closer to i than j.

The model is normalised so in principle we might expect $\alpha = 1$. However we leave α as a single parameter to be found by optimising the fit to the data and so $\alpha \neq 1$ is likely given the inevitable imperfections in this simple model.

C Gravity Models

The Gravity model used in [1] was

$$\widehat{F}_{ij} = \theta(D - d_{ij}) C_1 (P_i)^{\alpha_1} (P_j)^{\beta_1} (d_{ij})^{-\gamma_1} + \theta(d_{ij} - D) C_2 (P_i)^{\alpha_2} (P_j)^{\beta_2} (d_{ij})^{-\gamma_2} \,, \tag{C.1}$$

and similar with the exponential form for the deterrence function. The Heaviside theta functions split the model into short range and long range forms with a global model parameter D determining the distance scale. This is a model with nine global parameters found through

finding a best fit to the data. We have not used this form in our work. We have instead worked with a production constrained gravity model with one parameter (3.5) as this matches the approximate production constraint and the single parameter used in the Radiation model of [1] (the same as our Radiation model F of (B.6), see Table 1).

D Data

The distribution of commuter flows in the US Census 2000 [2] data is shown in Table D2.

Flow	Number
All	9665881
> 0	164764
> 10	77432
> 100	21237
$> 1,000$	7058
$> 10,000$	1814
$> 100,000$	212

Table D2: The number of county-county pairs with flows equal to or greater than the flow minimum given. Data for county-county commuter numbers is as given in the US Census 2000 [2].