



香港城市大學
City University of Hong Kong

專業 創新 胸懷全球
Professional · Creative
For The World

CityU Scholars

An attractiveness-based model for human mobility in all spatial ranges

Zhou, Jianfeng; Fan, Zhong-yan; Ng, Kai-Tat; Tang, Wallace K S

Published in:

New Journal of Physics

Published: 01/12/2019

Document Version:

Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

License:

CC BY

Publication record in CityU Scholars:

[Go to record](#)

Published version (DOI):

[10.1088/1367-2630/ab5da4](https://doi.org/10.1088/1367-2630/ab5da4)

Publication details:

Zhou, J., Fan, Z., Ng, K.-T., & Tang, W. K. S. (2019). An attractiveness-based model for human mobility in all spatial ranges. *New Journal of Physics*, 21, Article 123043. <https://doi.org/10.1088/1367-2630/ab5da4>

Citing this paper

Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights

Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission

Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy

Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

PAPER • OPEN ACCESS

An attractiveness-based model for human mobility in all spatial ranges

To cite this article: Jianfeng Zhou *et al* 2019 *New J. Phys.* **21** 123043

View the [article online](#) for updates and enhancements.

**PAPER**

An attractiveness-based model for human mobility in all spatial ranges

OPEN ACCESS**RECEIVED**

16 September 2019

REVISED

17 November 2019

ACCEPTED FOR PUBLICATION

29 November 2019

PUBLISHED

18 December 2019

Jianfeng Zhou , Zhong-yan Fan, Kai-Tat Ng and Wallace K S Tang 

Department of Electrical Engineering, City University of Hong Kong, Hong Kong, People's Republic of China

E-mail: EEKSTANG@CITYU.EDU.HK**Keywords:** human mobility, attractiveness, population-based, optimization, socio-economic indexesSupplementary material for this article is available [online](#)

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Abstract**

In the past decade, various aspects of human mobility, from individual to population levels in both spatial and time scales, have been studied. However, existing human mobility models still fail to describe activities in all spatial ranges, while in particular, local activities like self-loop fluxes are generally ignored. Moreover, the regional attractiveness, a basic but useful concept in mobility model design, remains difficult to be quantified due to its intricacies. To resolve these two fundamental issues, we introduce a trip competition mechanism to empower a mobility model to estimate population fluxes in all spatial ranges. The model includes attraction scores of regions in concern, obtained via optimizing the working population distributions. Its capability of predicting a variety of mobility patterns is verified by empirical data from three different countries, and its accuracy outperforms those of existing models. The quantified attractiveness is also found to be highly correlated with common socioeconomic indicators, and is able to act as a distinct metric to characterize a region.

1. Introduction

The insights into the motivation of human movement and the availability of massive demographic data have facilitated the evolution of human mobility modeling. Based on existing models, the statistical human mobility patterns ranged from individual level [1, 2] to population level, in diverse spatial [2–4] and time scales [5–7], have gradually been unraveled. Furthermore, the characteristics of different classes of individuals, for example, the returners and the explorers [8], have also been studied. These findings not only deepen our understanding in human behaviours but also provide carriers to other momentous issues, such as epidemic spreading [9–11], traffic engineering [12, 13], urban planning [14], just to name a few.

Human mobility is an aggregated outcome of trip selections in a population. As reflected in people's intuition and assumed in most mobility models, distance is one of the critical factors in this decision process. In the seminal gravity model [15], population flux is assumed to be inversely proportional to a function of distance. In opportunities-based models [16–19], the radius of the circular competition region is specified by the distance between the origin and the destination, while a distance ranking approach is adopted in the rank-based model [20].

Due to their formulation, the above-mentioned models focus on coming-in/going-out activities of an area, while local activities, i.e. self-loop fluxes within an area, are excluded. However, self-loop fluxes actually account for a large proportion of the total population fluxes, and in many scenarios, these local activities should be considered. A typical example is that, for workers' population, those who commute in/out for work, and those live and work in the same area are of equal importance. Indeed, in most census data, the percentages of populations who work in their resident regions are huge.

Another key factor in trip selections is the attractiveness of the destination. The simplest and widely adopted approach for representing attractiveness of an area is based on its population size. This method has been adopted

in many mobility models, including the gravity model [15, 21], the intervening-opportunities (IO) models [16, 17], the radiation model [18], the population-weighted opportunities model [19], etc. In addition to population size, some other factors have been suggested. For example, in random utility models [22, 23], a function combining traveling time, traveling distance, and job opportunities of a destination is assigned as the total attractiveness. The models proposed in [2] and [3] assume that the attractiveness is proportional to both population size and the memory effect [24, 25], while the number of nearby settlements is taken into account in a rank-based model [20]. Recently, machine learning has also been applied to explore the influence of various urban indicators to commuting flows, such as distance, GDP and unemployment rate [26].

Unfortunately, due to the diversity of motivations and complex relation in between, a clear definition of attractiveness is still in vain. Therefore, we consider that using a single factor, such as the population size, or an arithmetic combination of several factors to represent such an abstract attribute could be inadequate. In contrast, we here take attractiveness as a collective score, according to personal preference and other factors, reflected by the outcome of trip selections. It should be emphasized that an inappropriate approximation of attractiveness would affect the validation of human mobility laws, resulting in a large error in mobility pattern prediction.

In this paper, a new mobility model, called attraction model, is proposed. Comparing with existing models, our attraction model can predict the flux between any pair of subareas, achieving an excellent agreement with the census data. Moreover, the obtained attraction score well characterizes a subarea and also highly correlates with common socioeconomic metrics, providing a comprehensive indicator to show how competitive an area is. Last but not least, our model is superior in data efficiency. As compared to other parameter-based models, such as [2, 15, 22, 23], which usually require a large amount of data to tune their parameters, our attraction model only requires population distribution of the subareas, which is quantitatively same as popular parameterless models, such as the radiation model. Also, thanks to the population optimization process, our model is able to provide a better estimation in all scenarios.

2. Methods

2.1. Attraction model of human mobility

Similar to other studies in human mobility, we focus on commuting that governs most of the regular mobility patterns, for which the occurrence time, the quantity of fluxes, the origins and destinations are relatively fixed.

A targeted area is divided into N subareas according to its natural partitioning (for example, a state in the United States is divided into counties). Every subarea i has three attributes:

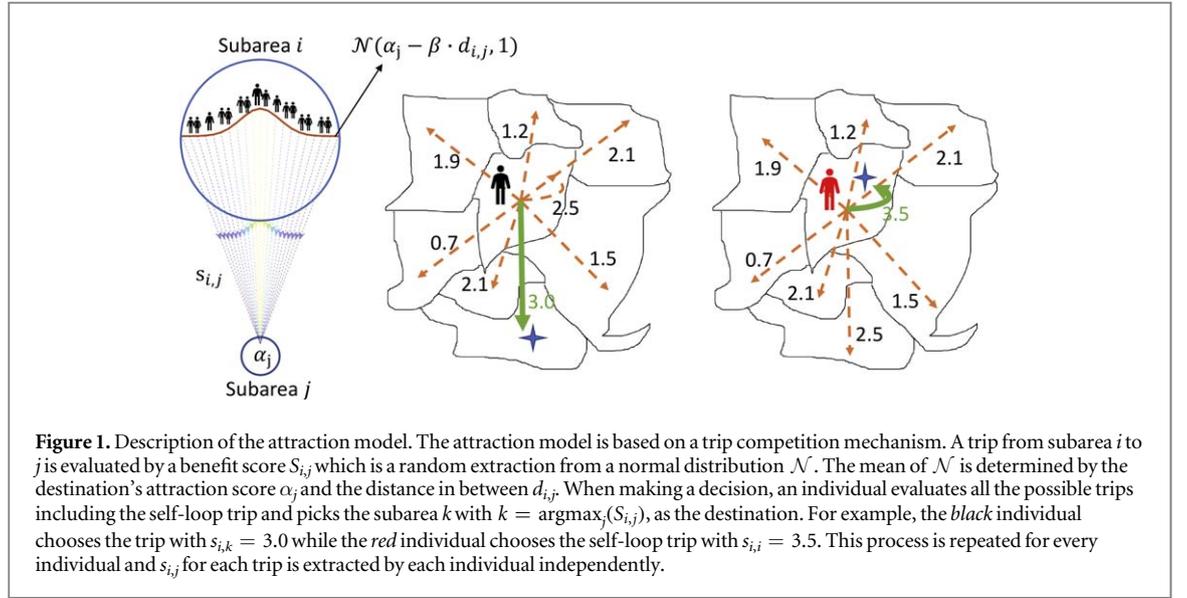
- resident working population pop_i^r : it refers to the group of persons who live in subarea i and have their jobs in any subarea;
- working population pop_i^w : it refers to the group of persons who work in subarea i but can live in any subarea;
- attraction score α_i : it reflects the attractiveness of subarea i .

Denoting the population flux from subarea i to j as $f_{i,j}$, the following assumptions are given:

1. $\text{pop}_i^r = \sum_{j=1}^N f_{i,j}$
2. $\text{pop}_i^w = \sum_{j=1}^N f_{j,i}$
3. $\sum_{i=1}^N \text{pop}_i^r = \sum_{i=1}^N \text{pop}_i^w$.

The basic idea of the proposed attraction model is depicted in figure 1. The selection process of a trip is conducted by a competition mechanism, evaluated by a benefit score. Generally, human subjective scores can be regarded as weighted averages of many independent small effects, and thus, according to the Central Limit Theorem, normal distribution is frequently adopted. Therefore, the benefit score is assumed to follow a normal distribution $\mathcal{N}(\mu, \sigma)$ with mean μ and variance σ^2 , so as to capsule the diversity and variations in the decision-making process of individuals. It is remarked that, similar ways of modeling have also been used in other human perception related studies, such as quantification of human response [27], emotion modeling [28], neural coding [29], and so on. Also, some further study about using normal distribution is presented in section 4 of the supplementary information.

Concretely, let the random variable $S_{i,j}$ be the benefit score of working at subarea j evaluated by the resident working population living in subarea i , one has



$$S_{i,j} \sim \mathcal{N}(\mu_{i,j}, \sigma_{i,j}) \equiv \mathcal{N}_{ij}, \quad (1)$$

with

$$\begin{aligned} \mu_{i,j} &= \alpha_j - \beta \times d_{i,j}, \\ \sigma_{i,j} &= 1, \end{aligned}$$

where α_i is the attraction score of subarea i , $d_{i,j}$ represents the distance between i and j , and β is a distance penalty constant. Here, $\mu_{i,j}$ is assumed to be linearly proportional to the distance $d_{i,j}$ and this assumption will be further discussed in section 3.1.

The trip competition mechanism is operated as follows. For each individual m who lives in subarea i , the benefit score $s_{i,j}^m$ of working at subarea j is a random extraction from $\mathcal{N}_{i,j}$. He/she would then choose to work at the subarea with the highest benefit score. Therefore, the transition probability of commuting from subarea i to j is:

$$p_{i,j} = \int_{-\infty}^{\infty} P(S_{i,j} = z) \times \prod_{k \neq j} P(S_{i,k} < z) dz, \quad (2)$$

where $P(\cdot)$ is the probability density function.

In terms of weak law of large number, the expectation of the flux $f_{i,j}$ can be estimated by:

$$E(f_{i,j}) = \text{pop}_i^r \times p_{i,j}. \quad (3)$$

2.2. Parameter identification via population optimization

Most of the parameter-based mobility models estimate their parameters by fitting the model with empirical commuting fluxes data. This method can be regarded as a flux optimization method because its goal is to minimize the errors between the fluxes obtained from the models and the empirical data. On the one hand, a high precision may be achieved via the optimization. On the other hand, it takes much time and efforts to collect such a large amount of data (for N subareas, the number of data is $N(N - 1)$), and it is difficult to ensure their accuracy. Also, it is challenging to tackle with the high-dimensional optimization.

To overcome these drawbacks, we introduce another method that only requires pop_i^r and pop_i^w of each subarea (for N subareas, the number of data is $2N$). We refer to it as the population optimization method. By matching the predicted and actual population distributions (or census data in practical applications), attraction score of every subarea and a weighted distance term, i.e. α_i and β , can be obtained.

We consider the population estimation error, which can be defined as

$$J_{\text{pop}}(\alpha, \beta) = \frac{\sum_{i=1}^N \left| \text{pop}_i^w - \sum_{j=1}^N E(f_{j,i}) \right|}{2 \times \sum_{i=1}^N \text{pop}_i^w}, \quad (4)$$

where $\alpha \equiv [\alpha_1, \dots, \alpha_N]$ are the attraction scores and β is the distance penalty factor as defined in equation (1).

By minimizing J_{pop} , α and β are given as

$$\underset{\alpha \in \mathbb{R}^N, \beta > 0}{\text{argmin}} J_{\text{pop}}(\alpha, \beta). \quad (5)$$

The minimization can be easily accomplished by designing a simple greedy algorithm, while the details of the algorithm can be found in supplementary information, algorithm S1.

Similarly, the flux estimation errors can be defined as:

$$J_{\text{flux}}(\alpha, \beta) = \frac{\sum_{i=1}^N \sum_{j=1}^N |f_{i,j} - E(f_{i,j})|}{2 \times \sum_{i=1}^N \text{pop}_i^w}. \quad (6)$$

However, it should be remarked that the flux information is not required in our model, and J_{flux} is solely for the sake of comparison and discussion.

2.3. Overview of the dataset

In this study, the commuting data of three countries, including the United States (US), Canada, and the United Kingdom (UK), are adopted. The commuting data of the United States is the county-to-county working flow from the US Census (2009–2013) [30]. This dataset contains all the states except for the State of Hawaii and Alaska. Each state is defined as a closed area in our study, while the counties of a state are the subareas. The Canada data [31] is from Statistics Canada (2016). It includes all the Provinces and each of them is defined as a targeted area, while the census subdivision regions in the province are regarded as subareas. Data of the Greater London area is from the UK Census (2011) [32, 33]. We regard the whole Greater London as an area and extract each borough as a subarea.

For the distance between two subareas, we adopted a practical way and used the ‘traveling time by car’ as calculated by Google Map.

3. Results

3.1. Model validation

As discussed in section 1, distance is a critical factor in the decision of trip selection. Therefore, a linear relation between the mean of benefit scores $\mu_{i,j}$ and the distance $d_{i,j}$ is assumed in (1). This relationship is further investigated by obtaining $\mu_{i,j}$ directly from the data (i.e. without using equation of $\mu_{i,j}$ in (1)) via minimizing flux estimation error (6) and $d_{i,j}$ from Google Map as described in section 2.3. Firstly, the Pearson correlation coefficients ($\rho(\mu, d)$) for each subarea j are calculated. Then, the coefficients of determination (R^2) and mean square errors (MSE) are measured by applying linear regression to $\mu_{i,j}$ and $d_{i,j}$ for each j . Figure 2 shows the boxplot of the results based on the data of New Jersey (with 21 counties). The large absolute value of $\rho(\mu, d)$ (close to 1), large R^2 and small MSE indicate a high linear correlation between μ and d . As a result, the use of linearity assumption in (1) is justified.

The attraction model is firstly validated based on the commuting data of the three countries, as described in section 2.3. The optimal $\alpha \equiv [\alpha_1, \alpha_2, \dots, \alpha_N]$ and β (see equation (1)) are found by minimizing (5). After obtaining α and β , the predicted population fluxes $E(f_{i,j})$ can be easily derived by equation (3). The obtained results are compared with the census data, as shown in figure 3 and referred as Case 1.

To further evaluate its effectiveness, the attraction model is compared with the gravity model [15] and the radiation model [18]. Since the compared models fail to estimate self-loop fluxes $E(f_{i,i})$, for fair comparison, all models are targeted to obtain the out-going fluxes $f_{i,j}$ with $i \neq j$. The results are given in figure 3, referred as Case 2.

For Case 2, the establishment of attraction model is similar to that for Case 1, except that self-loop fluxes $f_{i,i}$ are excluded from pop_i^r and pop_i^w (denoted as $\text{pop}_i^{r \setminus i}$ and $\text{pop}_i^{w \setminus i}$, respectively). The radiation model is parameter-free for which $E(f_{i,j})$ is directly computed by the population size (pop_i), the total out-going fluxes of each sub-area (i.e. $\text{pop}_i^{r \setminus i}$) and $d_{i,j}$. On the other hand, the gravity model, as a parameter-based model, obtains $E(f_{i,j})$ by minimizing its difference with the census data $f_{i,j}$. Table 1 compares the commuting/population data required by each model, while table 2 and table 3 list the overall population estimation error J_{pop} and flux estimation error J_{flux} of each model, respectively.

As shown by Case 2 in figure 3, there are significant deviations in the predicted fluxes from the gravity model, especially for both high and low fluxes. The radiation model generally achieves better prediction than the gravity model. Except for the Greater London, the radiation model performs well with some slight underestimation of high fluxes. The predicted median values align with the census data, but most of the mean values are overestimated. The deficiency of the radiation model for studying the Greater London is similar to the phenomenon reported in [19], which is due to the small scale prediction.

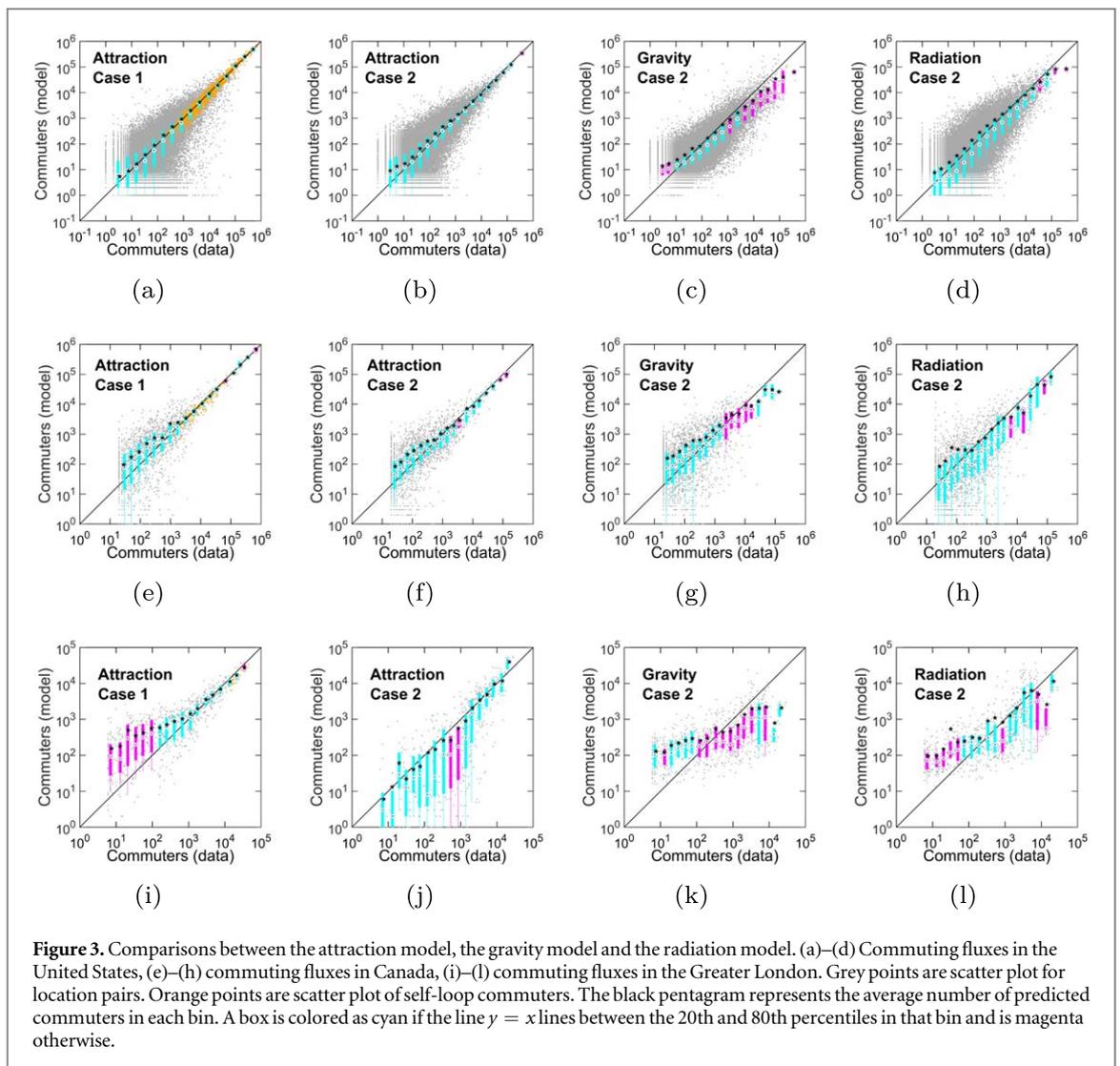
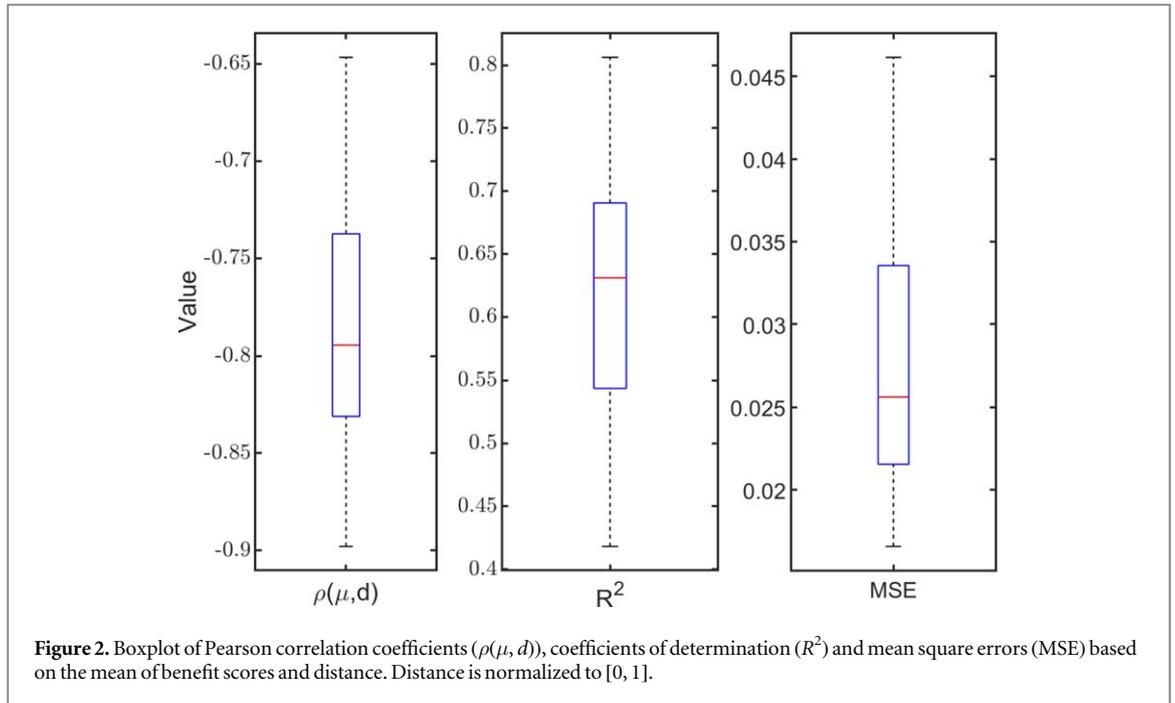


Table 1. Comparisons of data requirements of different models in Case 2.

Models	Data required	No. of data
Attraction model	$\text{pop}_i^{r\setminus i}, \text{pop}_i^{w\setminus i}$	$2N$
Gravity model	pop_i, f_{ij}	N^2
Radiation model	$\text{pop}_i, \text{pop}_i^{r\setminus i}$	$2N$

Note. N is the number of subareas. Also, distance information $d_{i,j}$ are required for all three models.

Table 2. Comparison of population estimation errors of different models.

	$J_{\text{pop}}^{A,C1a}$	$J_{\text{pop}}^{G,C2}$	$J_{\text{pop}}^{R,C2}$	$J_{\text{pop}}^{A,C2}$
States in US	0.0018	0.4235	0.3863	0.0322
Provinces in Canada	0.0050	0.4214	0.4187	0.0306
Greater London in UK	0.0188	0.3611	0.2649	0.0002

^a G, R, A denote the gravity, radiation and attraction model, respectively. $C1$ and $C2$ represent Case 1 and Case 2, respectively. These notations are also used in table 3.

Table 3. Comparison of flux estimation errors of different models.

	$J_{\text{flux}}^{A,C1}$	$J_{\text{flux}}^{G,C2}$	$J_{\text{flux}}^{R,C2}$	$J_{\text{flux}}^{A,C2}$
States in US	0.1158	0.4203	0.4003	0.2860
Provinces in Canada	0.1048	0.5038	0.4360	0.3152
Greater London in UK	0.2198	0.4340	0.5328	0.3000

Table 4. Comparison of Pearson correlation coefficient (ρ_{flux}) between the census commuting fluxes and those predicted by different models.

	$\rho_{\text{flux}}^{A,C1}$	$\rho_{\text{flux}}^{G,C2}$	$\rho_{\text{flux}}^{R,C2}$	$\rho_{\text{flux}}^{A,C2}$
States in US	0.8718	0.4774	0.6833	0.9666
Provinces in Canada	0.8890	0.6349	0.8112	0.9583
Greater London in UK	0.9081	0.2866	0.3598	0.8152

In contrast, the attraction model achieves more accurate prediction (also see tables 2 and 3). It not only obtains the predicted working populations with negligible deviations from the empirical data, but also achieves a small flux estimation error. As observed, both the predicted mean and median fluxes tightly align with the census data. This thus confirms the effectiveness of the optimization mechanism and also supports the validity of the attraction model. Moreover, as shown in Case 1, it is able to achieve high precision when estimating both the out-going and self-loop fluxes, even though only working population distributions pop_i^r and pop_i^w are available.

To further compare the prediction accuracy of different models, the Pearson correlation coefficients ρ_{flux} between the predicted and census commuting fluxes for different models are listed in table 4 while the coefficients of determination R^2_{flux} based on linear regression are tabulated in table 5. The larger coefficients in using attraction model for both Case 1 and Case 2 clearly confirm its superiority.

Another important metric for mobility is the distance probability $p_{\text{dist}}(d)$ which stands for the probability of a trip between two subareas with distance d . Figure 4 shows these probabilities of each model, while figure 5 focuses on cases with $d = 0$, i.e. the self-loop fluxes, as a supplementary. Except for some slight underestimation of long-distance fluxes, the $p_{\text{dist}}(d)$ of the attraction model matches the census data very well, demonstrating its high robustness on diverse spatial scales. On the contrary, both the gravity model and radiation model have difficulty in predicting short and long-distance trips. In particular for the Greater London area, both compared models fail to follow the trend of the census data, resulting in large prediction errors.

The performance of a population-level model also depends on the discretization resolution, which specifies the size of subareas. Usually, when a population-level model is concerned, the resolution is low and thus the size

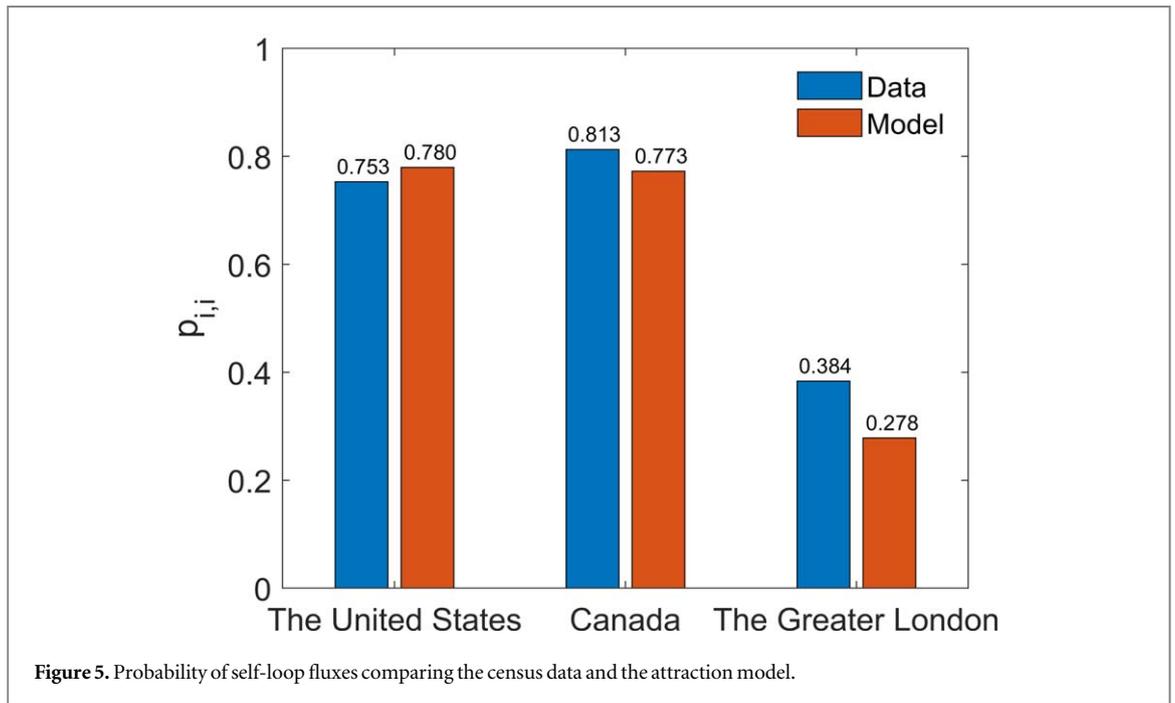
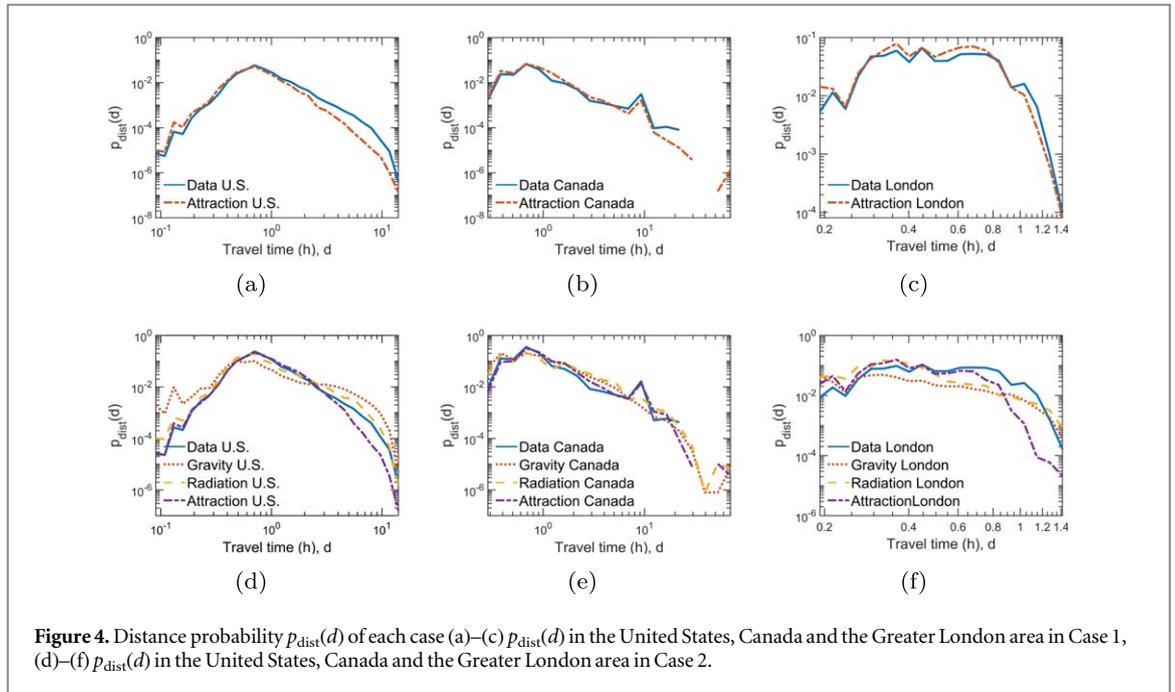


Table 5. Comparison of coefficient of determination (R^2_{flux}) between the census commuting fluxes and those predicted by different models.

	$R^2_{\text{flux}}^{A,C1}$	$R^2_{\text{flux}}^{G,C2}$	$R^2_{\text{flux}}^{R,C2}$	$R^2_{\text{flux}}^{A,C2}$
States in US	0.7601	0.2279	0.4670	0.9344
Provinces in Canada	0.7904	0.4031	0.6581	0.9185
Greater London in UK	0.8246	0.0821	0.1294	0.6646

of subareas is large. Obviously, any changes in resolution would affect the commuting network and the commuting fluxes.

Generally, the performance of population-level models, in terms of prediction accuracy, becomes worse when the discretization resolution increases (see table S2 of supplementary information available online at stacks.iop.org/NJP/21/123043/mmedia). That’s because the commuting fluxes gradually shift from the

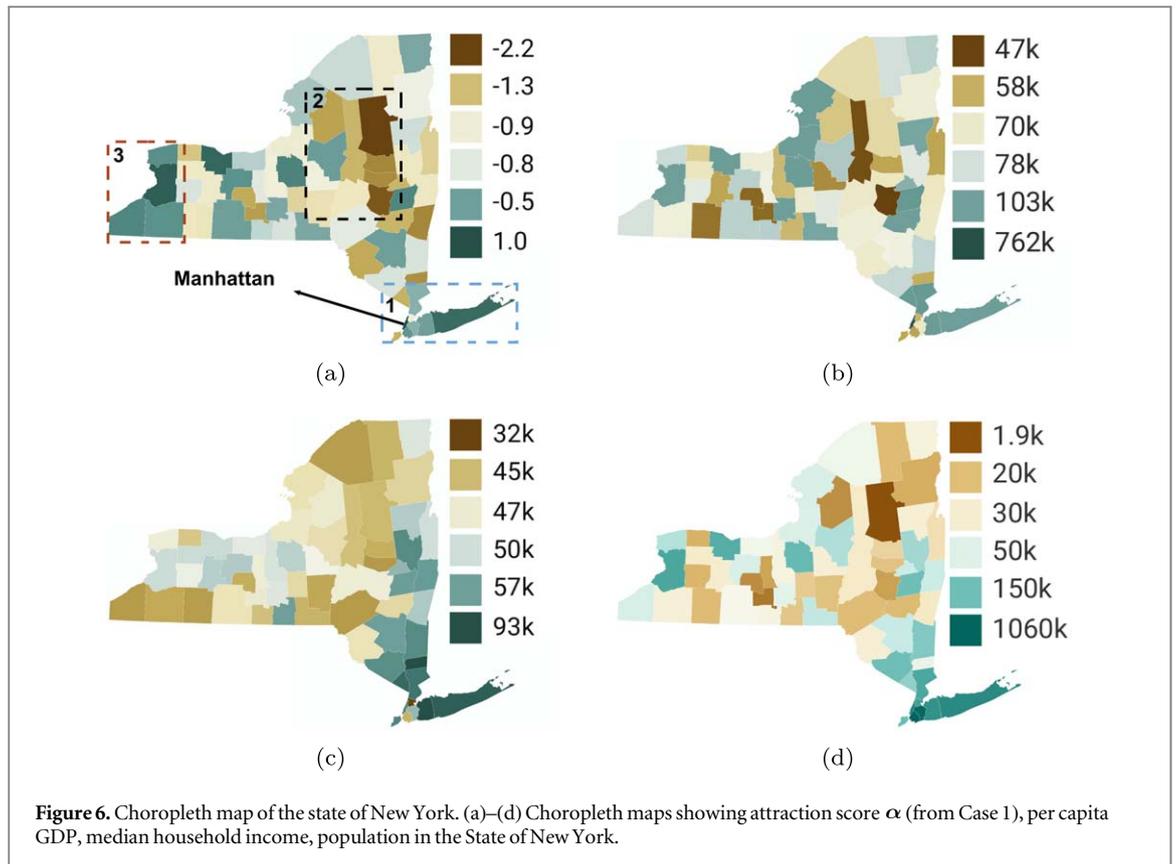


Figure 6. Choropleth map of the state of New York. (a)–(d) Choropleth maps showing attraction score α (from Case 1), per capita GDP, median household income, population in the State of New York.

population's collective behavior to individuals' when resolution increases. (Note: the size of subareas becomes smaller.) Moreover, the more subareas there are, the harder the optimization is. For the attraction model, due to the decrease in the proportion of self-loop fluxes (as subarea is smaller), the advantages of the attraction model would become less significant. However, as observed by the results in table S2, even though the prediction accuracy of the attraction model reduces, it is still much better than the other existing models for both Case 1 and 2. More detailed analysis can be referred to section 5 of the supplementary information.

3.2. Implications of attraction score

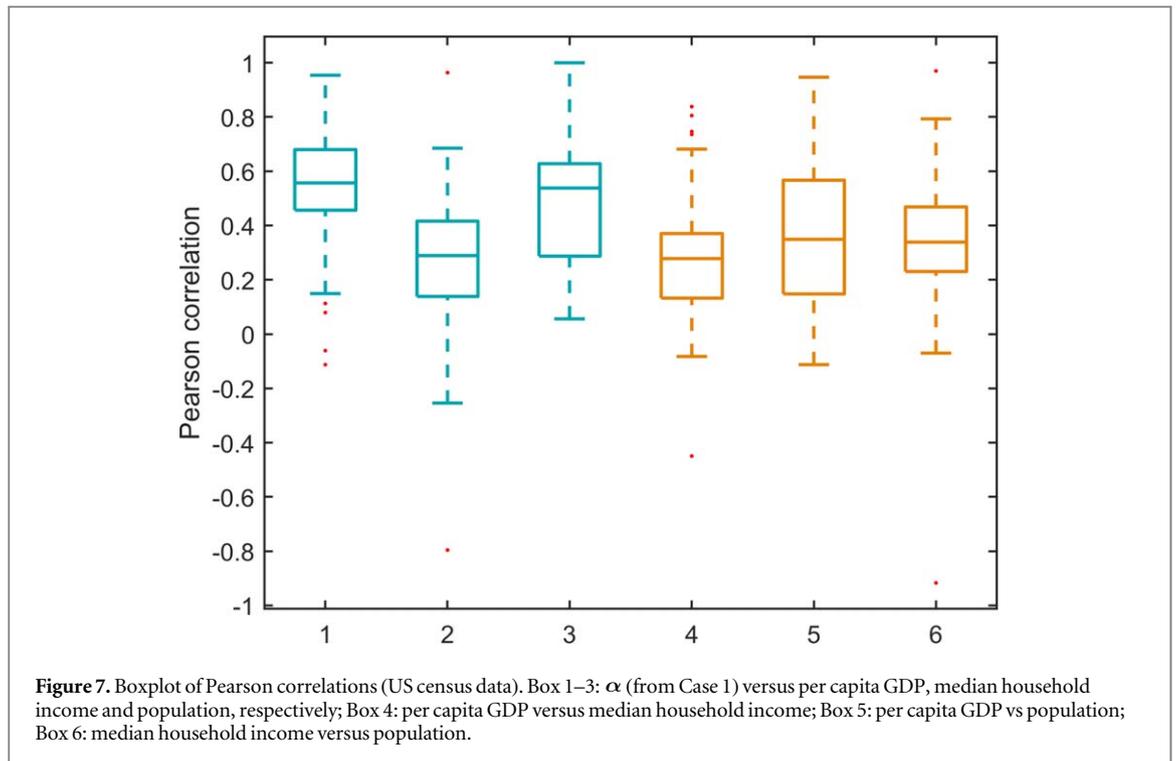
The attraction score α_i is a collective estimation of the attractiveness of a subarea. Take the State of New York as an example, figure 6 depicts the choropleth maps of attraction scores and some socioeconomic metrics. To be precise, we focus our discussion on three regions as numbered in the figure.

Region 1 (figure 6(a)) includes the New York City and Long Island. New York City is the most populous city in the US with its global impact in various areas. Long Island owns the largest industrial park in the east coast which provides more than 71 000 job positions. Moreover, it plays a prominent role in scientific research and engineering. As expected, counties in *Region 1* gain high attraction scores while the Manhattan is particularly high as it serves as the economic and administrative center of New York City.

Region 2 is mainly covered by the Mohawk Valley which is a suburban and rural area. It is low in per capita GDP, in median household income, and in population size. Not to our surprise, these subareas have poor attraction scores.

Region 3 is the Western New York. The loss of traditional jobs in manufacturing has led to a decline in economic. Thus, even though it is still a populous area, it suffers from a drop of population. The median household income is relatively low around the year 2010. According to these observations, low attraction scores might be predicted. However, it is the opposite as observed in figure 6(a). The fact is that, its role has been transferred from traditional manufacturing to light manufacturing, high technology, and services. Together with the help of the government's policies (e.g. the casino business in the late 1990s and the investment of a billion dollars under 'Buffalo Billion' project, etc), the economy in this region has been improving significantly since early 2010s and there are many job opportunities. Therefore, the counties in this region gain high attraction scores during the year 2009 and 2013. (Note: data used for the attraction score quantification is from this 5 year survey.)

Based on the very similar patterns of the four choropleth maps in figure 6, a good correlation between the attraction scores and socioeconomic metrics is expected. To quantify the observation, their Pearson correlations



are given in figure 7. The population size of an area does highly correlate with its attraction score but not high enough to be regarded as a linear correlation. In other words, a mobility model may not be precise enough if we simply take the population size to represent the attractiveness. The per capita gross domestic product (GDP) also contributes a lot to the attraction score. It even has a higher correlation than that of the population size and similar observation has also been reported in [26]. On the other hand, the median household income shows a relatively low correlation with the attraction score.

4. Discussion and conclusion

Attractiveness of destinations and distance are two key elements in the formulation of human mobility models. Nowadays, distance information can be easily (and accurately) obtained, e.g. by the use of electronic maps, but it is still hard to quantify attractiveness, in spite of many attempts. In addition, existing mobility models did not address issues of self-loop flux prediction, which is of great importance. From both theoretical and practical points of view, a model reflecting the true physical mechanism behind human mobility should predict fluxes in any spatial range without exceptions.

The above two challenges are resolved by two key mechanisms in the attraction model, namely the trip competition mechanism and the population-based attractiveness quantification method.

In the trip competition mechanism, all possible trips within the studied area would participate in the competition. The competitiveness of a trip is a normal distribution, with its mean determined by a linear combination of the attractiveness of the destination and the distance of the trip. As a result, self-loop flux can be compared in a fair manner. The parameters of the attraction model, i.e. the attraction scores and the weighting factors of distance, are obtained through optimization, minimizing the estimation error of the working populations. Therefore, the inputs of the attraction model are only the distance and the working population distributions of the subareas, in contrast to other parameter-based models, for which costly empirical commuting data is required. Our model has been verified with census data from three different countries through a comparison with the classical gravity model and the radiation model.

Validation of the attraction scores is twofold. Firstly, the commuting fluxes are predicted based on our model after optimization. As shown in figure 3, the predicted fluxes well match with the empirical fluxes. Secondly, the relation between the attraction score and common socioeconomic metrics has been studied. The obtained attraction scores are found to be positively correlated with several socioeconomic measurements including population, per capita GDP and the median household income [34]. These metrics are chosen as they intuitively reflect how attractive an area is, while the results conform with our intuition. Furthermore, the correlation with the population size reflects the limitation of using population size as an approximation of

attractiveness in other models, which is also confirmed by the fair pattern estimation results of these models given in section 3.

Similar to other specific attractiveness such as the tourism attractiveness [35–38] and the investment attractiveness [39, 40], the usage of our attraction score is not limited to mobility prediction. It also serves as a monitoring indicator which properly reflects the healthiness of a regional labor market. As demonstrated in the case of Western New York, it is able to reveal some potential crisis or opportunities of an area, which would be useful for decision making and socioeconomic analyses. For individuals and enterprises, the quantified attraction score suggests a simple and unified way to assess an area. As reported in [41], creative workers would rather be attracted by the overall attractiveness of a city rather than work-related motivations. For governments, the correlations between attraction score and different socioeconomic metrics help to identify areas for improvements, and are also useful for verifying the effectiveness of the governments' policies.

5. Data accessibility

Data and code used in this paper can be downloaded from <https://hippie-caltsby.github.io/attraction-download.html>.

Author contributions

JZ, ZF, KTN and WKST conceived and designed the manuscript. JZ and ZF designed the model, collected data, and performed model validation. JZ and WKST wrote the paper.

Competing interests

The authors declare no competing financial interests.

ORCID iDs

Jianfeng Zhou  <https://orcid.org/0000-0002-5916-6457>

Wallace K S Tang  <https://orcid.org/0000-0002-5786-418X>

References

- [1] Gonzalez M C, Hidalgo C A and Barabasi A L 2008 Understanding individual human mobility patterns *Nature* **453** 779
- [2] Yan X-Y, Wang W-X, Gao Z-Y and Lai Y-C 2017 Universal model of individual and population mobility on diverse spatial scales *Nat. Commun.* **8** 1639
- [3] Song C, Koren T, Wang P and Barabási A-L 2010 Modelling the scaling properties of human mobility *Nat. Phys.* **6** 818
- [4] Zhang H-T, Zhu T, Fu D, Xu B, Han X-P and Chen D 2018 Spatiotemporal property and predictability of large-scale human mobility *Physica A* **495** 40–8
- [5] Castillo-Chavez C, Bichara D and Morin B R 2016 Perspectives on the role of mobility, behavior, and time scales in the spread of diseases *Proc. Natl Acad. Sci.* **113** 14582–8
- [6] Wesolowski A et al 2017 Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics *Nat. Commun.* **8** 2069
- [7] Schneider C M, Belik V, Couronné T, Smoreda Z and González M C 2013 Unravelling daily human mobility motifs *J. R. Soc. Interface* **10** 20130246
- [8] Pappalardo L, Simini F, Rinzivillo S, Pedreschi D, Giannotti F and Barabási A-L 2015 Returners and explorers dichotomy in human mobility *Nat. Commun.* **6** 8166
- [9] Hufnagel L, Brockmann D and Geisel T 2004 Forecast and control of epidemics in a globalized world *Proc. Natl Acad. Sci.* **101** 15124–9
- [10] Viboud C, Bjørnstad O N, Smith D L, Simonsen L, Miller M A and Grenfell B T 2006 Synchrony, waves, and spatial hierarchies in the spread of influenza *Science* **312** 447–51
- [11] Mari L, Bertuzzo E, Righetto L, Casagrandi R, Gatto M, Rodriguez-Iturbe I and Rinaldo A 2011 Modelling cholera epidemics: the role of waterways, human mobility and sanitation *J. R. Soc. Interface* **9** 376–88
- [12] Helbing D 2001 Traffic and related self-driven many-particle systems *Rev. Mod. Phys.* **73** 1067
- [13] de Dios Ortúzar J and Willumsen L G 2002 *Modelling Transport* vol 3 (New York: Wiley) (<https://doi.org/10.1002/9781119993308>)
- [14] Batty M 2008 The size, scale, and shape of cities *Science* **319** 769–71
- [15] Zipf G K 1946 The p 1 p 2/d hypothesis: on the intercity movement of persons *Am. Soc. Rev.* **11** 677–86
- [16] S A Stouffer 1940 Intervening opportunities: a theory relating mobility and distance *Am. Soc. Rev.* **5** 845–67
- [17] Schneider M 1959 Gravity models and trip distribution theory *Papers Reg. Sci.* **5** 51–6
- [18] Simini F, González M C, Maritan A and Barabási A-L 2012 A universal model for mobility and migration patterns *Nature* **484** 96
- [19] Yan X-Y, Zhao C and Fan Y 2014 Universal predictability of mobility patterns in cities *J. R. Soc. Interface* **11** 20140834
- [20] Noulas A, Scellato S, Lambiotte R, Pontil M and Mascolo C 2012 A tale of many cities: universal patterns in human urban mobility *PLoS One* **7** e37027
- [21] Park H J, Jo W S, Lee S H and Kim B J 2018 Generalized gravity model for human migration *New J. Phys.* **20** 093018

- [22] McFadden D 1978 Modeling the choice of residential location *Transp. Res. Rec.* **12** 72–77
- [23] Block H D 1974 Random orderings and stochastic theories of response *Economic Information, Decision, and Prediction* (Dordrecht: Springer) pp 172–217
- [24] Szell M, Sinatra R, Petri G, Thurner S and Latora V 2012 Understanding mobility in a social petri dish *Sci. Rep.* **2** 457
- [25] Zhao Y-M, Zeng A, Yan X-Y, Wang W-X and Lai Y-C 2016 Unified underpinning of human mobility in the real world and cyberspace *New J. Phys.* **18** 053025
- [26] Spadon G, de Carvalho A C, Rodrigues J F Jr and Alves L G A 2019 Reconstructing commuters network using machine learning and urban indicators *Sci. Rep.* **9** 1–13
- [27] Pendrill L R and Fisher W P Jr 2013 Quantifying human response: linking metrological and psychometric characterisations of man as a measurement instrument *J. Phys.: Conf. Ser.* **459** 012057
- [28] Poria S, Cambria E, Winterstein G and Huang G-B 2014 Sentic patterns: dependency-based rules for concept-level sentiment analysis *Knowl.-Based Syst.* **69** 45–63
- [29] Rehn M and Sommer F T 2007 A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields *J. Comput. Neurosci.* **22** 135–46
- [30] U.S. Census Bureau 2019 2009-2013 5-year American community survey commuting flows (<https://census.gov/data/tables.html>)
- [31] Statistics Canada 2019 Commuting flow from geography of residence to geography of work: - census subdivisions, 2016 census (<https://statcan.gc.ca>)
- [32] U.K. Data Service 2019 2011 census: flow data (<https://ukdataservice.ac.uk>)
- [33] Commute-flow 2019 Source data (<http://commute-flow.net/about.html>)
- [34] Carra G, Mulalic I, Fosgerau M and Barthelemy M 2016 Modelling the relation between income and commuting distance *J. R. Soc. Interface* **13** 20160306
- [35] Buhalis D 2000 Marketing the competitive destination of the future *Tourism Manage.* **21** 97–116
- [36] Dwyer L and Kim C 2003 Destination competitiveness: determinants and indicators *Curr. Issues Tourism* **6** 369–414
- [37] Formica S and Uysal M 2006 Destination attractiveness based on supply and demand evaluations: an analytical framework *J. Travel Res.* **44** 418–30
- [38] Cracolici M F and Nijkamp P 2009 The attractiveness and competitiveness of tourist destinations: a study of southern italian regions *Tourism Manage.* **30** 336–44
- [39] Soboleva Y P and Parshutina I G 2016 Management of investment attractiveness of the region by improving company strategic planning *Indian J. Sci. Technol.* **9** 1
- [40] Zheltenkov A, Syuzeva O, Vasilyeva E and Sapozhnikova E 2017 Development of investment infrastructure as the factor of the increase in investment attractiveness of the region *IOP Conf. Ser.: Earth and Environmental Science* Vol 90 (Bristol: IOP Publishing) 012122
- [41] Bontje M et al 2008 *The Amsterdam Region—a Home for Creative Knowledge Workers and Graduates?: Understanding the Attractiveness of the Metropolitan Region for Creative Knowledge Workers* 5.1 (University of Amsterdam, AMIDSt) (<https://hdl.handle.net/11245/1.284693>)