Human mobility models reveal the underlying mechanism of seasonal movements across China

Bing Song

College of Systems Engineering
National University of Defense Technology
Changsha 410073, P. R. China

Xiao-Yong Yan

Institute of Transportation Systems Science and Engineering
Beijing Jiaotong University
Beijing 100091, P. R. China

Suoyi Tan and Bin Sai

College of Systems Engineering
National University of Defense Technology
Changsha 410073, P. R. China

Shengjie Lai

WorldPop, Department of Geography and Environment
University of Southampton
University Road, Southampton SO171BJ, UK

Hongjie Yu

Key Laboratory of Public Health Safety
Ministry of Education
School of Public Health, Fudan University
Shanghai 200032, P. R. China

Chaomin Ou[*,‡] and Xin Lu[†,‡]

College of Systems Engineering
National University of Defense Technology
Changsha 410073, P. R. China
[*]*ouchm@163.com*
[†]*xin_lyu@sina.com*

[‡]Corresponding authors.

Understanding the spatial interactions of human mobility is crucial for urban planning, traffic engineering, as well as for the prevention and control of infectious diseases. Although many models have been developed to model human mobility, it is not clear whether such models could also capture the traveling mechanisms across different time periods (e.g. workdays, weekends or holidays). With one-year long nationwide location-based service (LBS) data in China, we investigate the spatiotemporal characteristics of population movements during different time periods, and make thorough comparisons for the applicability of five state-of-the-art human mobility models. We find that population flows show significant periodicity and strong inequality across temporal and spatial distribution. A strong "backflow" effect is found for cross-city movements before and after holidays. Parameter fitting of gravity models reveals that travels in different type of days consider the attractiveness of destinations and cost of distance differently. Surprisingly, the comparison indicates that the parameter-free opportunity priority selection (OPS) model outperforms other models and is the best to characterize human mobility in China across all six different types of days. However, there is still an urgent need for development of more dedicated models for human mobility on weekends and different types of holidays.

Keywords: Population movement; seasonal migration; human mobility; human mobility models.

1. Introduction

On the one hand, uncovering the human mobility patterns that characterize the trajectories followed by humans during their daily activities is significant for infectious disease prevention and control,[1–8] city planning,[9,10] traffic engineering,[11] crisis management[12,13] and economic forecasting.[14,15] On the other hand, human mobility is the driving force for many important societal dynamics, such as the spread of infectious diseases,[3,13,16–18,8,5] economic development,[19,20] cultural exchange,[21,22] environmental change[23,24] and traffic demand.[25]

Many researchers have tried to understand human mobility patterns[26–30] and challenges in order to solve the problem of human mobility predicting[29,31–36] by mathematical modeling. Gonzalez *et al.*[26] and many of the other researchers[27–30,37] have found that human trajectories show a high degree of temporal and spatial regularity, which has prompted the proposal of some representative models and prediction approaches for modeling and predicting human mobility. Inspired by Newton's law of gravitation, Zipf has proposed the gravity model[31,38] to predict mobility flow. Along with the gravity model, Stouffer[39] has developed the intervention opportunity (IO) model to describe human mobility. Stouffer suggests that the key role in determining migration comprises both the destination's opportunities (an analogy of something that can fulfill the individual's travel purpose, such as working opportunities, and higher salaries) and the intervening opportunities (see Fig. 1(a)) between the origin and the destination but not the distance. The two models have sometimes been compared in different studies,[40–46] but there is no unified conclusion on the best model for a variety of flows; for example, sometimes the two have similar levels of performance,[40,41] or the gravity model outperforms the IO model[42,44,47] and vice versa.[45,46]

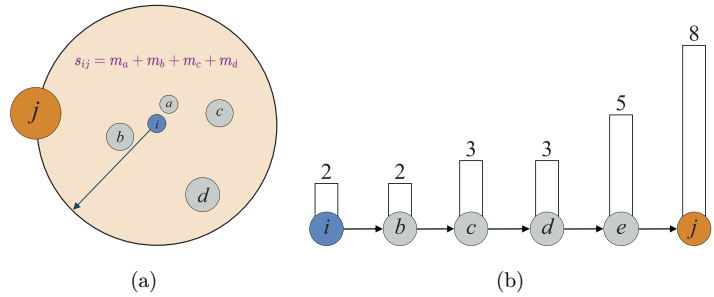


Fig. 1. Illustration of opportunity models. (a) Distance and intervening opportunity. Each circle represents a location with different benefits m_l (represented by the size of the circle), location i is the origin and the other locations are potential destinations. Then, the intervening opportunities between origin i and destination j are equal to the sum of all opportunities of the locations between i and j, that is, $s_{ij} = m_a + m_b + m_c + m_d$. (b) Opportunity ranking and destination selection. We use m_l to represent the maximum opportunity benefit at location l, and the height of the bar in the figure represents the value m at each location. If the maximum opportunity benefits of the locations are $m_i = m_a = 1, m_b = m_c = 3, m_c = 3, m_d = 5$ and $m_j = 8$, then the individual at location i will only choose location b as the destination in the radiation model. Location j will be chosen in the OO model, no matter how far from i it is. However, the individual will choose d or j with different probabilities in the OPS model, which is decided by distance d_{ij}.

However, the applications of the variants of the IO model are limited due to the lack of effective methods to measure the ranking of IOs[48] and to estimate parameters, owing to the nonlinear formulation. To overcome the limitation of the IO model, Simini *et al.*[49] has proposed the radiation model, which is parameter-free and has rigorous derivation based on the assumption that people prefer to select the closest location with higher benefits than those of the origin. The radiation model shows strong predictive performance for a wide range of flows, such as commuting, migration and commodity in the United States. However, some of the later studies have shown that the radiation model cannot be applied to spatial scales for all.[32–34,50–54] Further research has found that the radiation model tends to overestimate short-distance flow and underestimate long-distance flow on the city scale,[50,55] indicating that more universal models are needed to characterize the mobility patterns on different spatial scales.[33,34,50,52,56–58] For this purpose, Yan *et al.*[52] established the population-weighted opportunities (PWOs) model for the prediction of mobility patterns in cities. Inspired by the PWO model and the prudent social model,[58] Liu and Yan[50] introduced the opportunity priority selection (OPS) model, which assumes that the destination selected by the individual is the location that presents higher benefits (usually replaced by the local population) than those of the origin and of the intervening opportunities. Additionally, Yan *et al.*[34] proposed a more universal model for the prediction of intracity and intercity mobility patterns at both the individual and the population levels by combining the PWO model with the continuous-time random walks model.[59] In general, individuals will compare the benefits at different locations when the variants of the IO model are used for human mobility prediction, but the rules for comparing the benefits of different models differ.

Motivated by Ohm's law of electricity,[60] Sallah *et al.*[32] introduced a new parameter-free human mobility model called the impedance model for population flow prediction. Analogizing population flow with electricity, the impedance model likens the number of trips per day to an electric current, which is driven by the voltage (the population difference in the model) and blocked by the resistor (the distance between two places in the model). The impedance model has been successfully used in human mobility prediction during the cholera epidemic in Haiti. Additionally, Li *et al.*[61] proposed a concept of "active population", based on which they proposed a unified model that can reproduce the spatial scalings for population, total road length and total number of socioeconomic interactions.

Despite the success of those models in specific spatial scales, there is still a lack of explicit and comprehensive understanding of the underlying mechanism accounting for the effect of time and holiday effects on human mobility patterns. Meanwhile, accurate, complete and high-resolution data across different time periods are still scarce in the field of study. To overcome these limitations, in this work, we analyzed and summarized comprehensively the underlying spatiotemporal mechanism of mobility patterns during workdays, weekends, holidays and the Chunyun migration period (the Spring Festival travel season may be the largest periodic human migration in the world, when usually begins 15 days before Spring Festival and lasts for around 40 days). Using nationwide location-based service (LBS) data, we measured population flow and aggregate national migrants by the prefectures of destination. On this basis, we selected several mobility models to evaluate the validity of assumptions on the underlying mechanisms of population flow during different period. The remainder of the paper is organized as follows: Sec. 2 introduces the dataset and the methods used in this study. Section 3 analyzes the spatial–temporal characteristic of population flow in China and compares five models on the performance of fitting human mobility patterns across different periods. Finally, Sec. 4 concludes and discusses the main findings and limitations.

2. Data and Methods

2.1. *Datasets*

LBS mobility data. A large-scale national-level mobility dataset, extracted from Baidu's LBS, was used for the analysis of human mobility patterns in China. The location of a device (typically a mobile phone) can be determined when it uses Baidu's LBS, provided that the user has agreed to its data collection policy. Baidu is the dominant internet service provider in China, which offers a range of consumer features, including maps, news, videos, an encyclopedia, antivirus software and internet TV for mobile devices and personal computers.

We extracted data from April 23, 2013 to April 29, 2014. The raw data were anonymous, aggregated daily flows of population data collected at the county level (2959 counties). We then aggregated the intracounty flows with those at the prefecture level (358 prefectures) for further analysis. The data from different dates were

divided into six categories according to the type of day: workdays, weekends, three-day holidays (legal holidays other than New Year's Day and the Spring Festival in China), New Year's Day (only one day off on January 1, 2014), National Day (October 1–7, 2013) and the Chunyun migration times (January 16–February 24, 2014). The representativeness and the validity of the LBS data have been verified in previous studies [13] such that the data highly conform with the mobility matrix produced by complete taxi-based global positioning system (GPS) locations.

Population data. The population data used in models for predicting the mobility of people across locations come from the sixth national census[62] (the most recent data are from the 2013 census), which are the most detailed, systematic and authoritative data on the population in China. The study covered 358 administrative units, including four municipalities and 354 prefecture-level cities in the mainland, excluding Hong Kong, Macao and Taiwan.

2.2. *Mobility models*

2.2.1. *Gravity model*

Proposed by Zipf in 1946, the gravity model[31] is analogous to Newton's law of gravitation. This model assumes that the mobility flow between locations i and j is proportional to the populations m_i and m_j in the two locations and decays with a power function[63] of the travel distance d_{ij} between i and j. The probability P_{ij} that individuals move from location i to location j can be approximated by

$$P_{ij} = \alpha \frac{m_i m_j}{d_{ij}^{\beta}}, \tag{1}$$

where α and β are adjustable exponents. To ensure that the predicted flow matrix T satisfies $O_i = \sum_j T_{ij}$, we use an origin-constrained gravity model,

$$P_{ij} = \frac{m_j^{\alpha} d_{ij}^{-\beta}}{\sum\limits_{i \neq j} m_j^{\alpha} d_{ij}^{-\beta}}, \tag{2}$$

where m_j represents the attractiveness of destination j, α is an evaluation indicator of the attractiveness of destination j, β is the evaluation indicator of the travel cost, and the larger the index α/β, the more attractive the destination, the higher the travel cost.

2.2.2. *Intervention opportunity models*

Radiation model. The radiation model[49] assumes that the travelers will select a location as their destination through three steps. It first assigns a fitness (chosen from some distribution, $p(z)$) to every location according to its benefits. Second, it ranks all benefits according to the distance between the origin and the destination. Third, the closest location with higher benefits than those of the origin (location b in Fig. 1(b)) is selected as the travel destination. As a result, the probability of travelers from

location i to location j takes the form

$$P_{ij} = \frac{1}{1-\frac{m_i}{M}} \frac{m_i m_j}{\left(m_i + s_{ij}\right) \cdot \left(m_i + s_{ij} + m_j\right)}, \tag{3}$$

where m_i and m_j represent the opportunities of origin i and destination j, respectively, denotes the total opportunities of all the locations between i and j; and S_{ij} is the number of intervening opportunities [40] (see Fig. 1(a)), which is the total number of opportunities in the circle of radius d_{ij} centered on i (excluding the origin and the destination).

OPS model. The OPS model[50] assumes that the probability of a destination being selected by an individual is proportional to the benefits at the destination and inversely proportional to the total number of intervening opportunities between the origin and the destination (the locations b, d or j in Fig. 1(b)). Then, the probability of the individual at location i choosing location j as the destination is

$$P_{ij} = \frac{m_j / \left(m_i + s_{ij} + m_j\right)}{\sum\limits_{i \neq j} m_j / \left(m_i + s_{ij} + m_j\right)}. \tag{4}$$

Opportunity only model. The opportunity only (OO) model[33] is an extreme form of the IO model, in which the individuals take the benefits as the only consideration. Therefore, travelers tend to select the location with the highest benefits (the location j in Fig. 1(b)), accounting for the probability that a destination chosen by individuals is proportional to the benefits at the destination, where the probability of the mobility flow between two communities i and j can be approximated by

$$P_{ij} = \frac{m_j / \left(m_i + m_j\right)}{\sum\limits_{i \neq j} m_j / \left(m_i + m_j\right)}. \tag{5}$$

2.2.3. *Impedance model*

The impedance model[32] is proposed using an analogy with electricity such that it assimilates the electric resistance to the distance, the electric current to the number of trips per day, and the electric potential to the mobility potential per day on a given trajectory. Therefore, the probability of an individual at location i traveling to j is given by

$$P_{ij} = \frac{\left(m_i + m_j\right)/d_{ij}}{\sum\limits_{i \neq j} \left(m_i + m_j\right)/d_{ij}}, \tag{6}$$

where m_i and m_j denote the populations of locations i and j, respectively.

2.3. *Metrics of evaluation*

Sørensenrensen similarity index. The Sørensen similarity index (SSI)[64] is frequently used to measure the similarity between two samples. Here, we apply a

modified version of the index, which has been proposed by Liu and Yan[33] to measure the accuracy of the predictive flow reproduced by theoretical models, defined as

$$\mathrm{SSI} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\min\left(T_{ij}, \hat{T}_{ij}\right)}{T_{ij} + \hat{T}_{ij}}, \tag{7}$$

where N is the number of locations, T_{ij} denotes the predicted population flow from location i to location j and $\hat{T}_{ij}$ represents the empirical flow. Apparently, if each T_{ij} is equal to $\hat{T}_{ij}$, SSI = 1, and if all $\hat{T}_{ij}$ are far from the real values, then SSI ≈ 0.

Root mean square error. The root mean square error (RMSE) or the root mean square deviation (RMSD) is a measure of the differences between the values predicted by a model or an estimator and the values observed.[65] The RMSE is calculated as

$$\mathrm{RMSE} = \frac{1}{N} \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} \left(\hat{T}_{ij} - T_{ij}\right)^2}. \tag{8}$$

3. Results

3.1. *Temporal characteristics of population flows in China*

With the LBS data, we first calculate the total number of individuals moving into and out of each of the 358 prefectures on a daily basis. Then, we calculate the net population flow, that is, the total inflow minus the total outflow, in each city. The netflows for all 358 cities, as well as the top 10 cities with the highest and the lowest netflows before the Spring Festival, are shown in Fig. 2. We can observe a strong weekly pattern for the overall netflows during normal periods, that is, different cities may receive additional travelers or lose visitors on workdays or weekends. Another strong signal is the sharp variation of netflows on days before and after major holidays, such as Labor Day, National Day and traditional Chinese festivals, including the Mid-Autumn Festival, the Dragon Boat Festival, the Spring Festival, among others, and the change is almost symmetrical around holidays. Additionally, the magnitude of the flow varies a lot, on account of the differences in the cities' populations, economies, and geographic and sociodemographic factors such that a few big prefectures with large values of netflow are more visible in Fig. 2, however, there are more prefectures with small netflows in opposite signs, the conservation of flows always holds.

The symmetry of the variation of the netflows before and after the major holidays can be explained by the likelihood that people might visit their hometowns or travel when the holiday season begins and then return to their prior locations afterward, producing the return trips that roughly match the departures. We refer to this as the backflow effect. However, the backflow effect varies from region to region, such that cities of economic and strategic importance (Fig. 2(b)) have a large population

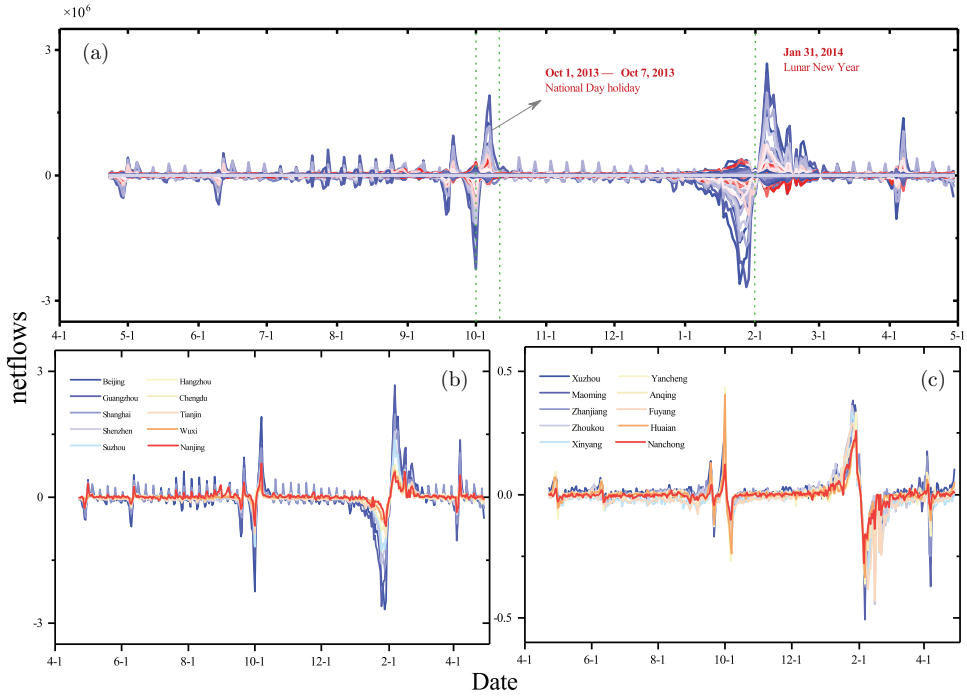


Fig. 2. Temporal patterns of population flows in China in 2013. (a) The netflows of population mobility at the prefecture level from April 23, 2013 to April 30, 2014. Each curve represents the change of the net population flow in a prefecture. (b) Top 10 cities with the largest net population outflow before the holidays. (c) Top 10 cities with the largest net population inflow before the holidays.

exodus before the holidays and arrivals at the end of the holidays, in reverse to the cities that host labor-export services (Fig. 2(c)).

3.2. *Spatial distribution of population flows in China*

The geographic distribution of population flows exhibits an extremely uneven pattern, with the large-scale mobility occurring in the more economically developed eastern side of the Hu Huanyong Line[66] (dashed line in Fig. 3), also called the Heihe–Tengchong Line, which is the well-recognized line for the east–west divide of the population density in China. Provincial capitals (e.g. Chengdu in Sichuan Province, Zhengzhou in Henan Province, etc.), as well as the more economically advanced cities (e.g. Shenzhen in Guangdong Province, Suzhou in Jiangsu Province, etc.), are mostly observed to be the centers of population flows regionally. These local networks are scattered across various regions. Figure 3 also shows a high intensity of population flows between provincial capital cities and regional central municipalities. Figures 4(a) and 4(b) show the geographic distributions of the population flows between provinces and cities, respectively, through the chord diagram. We find that the population movement in each province is mainly concentrated in its cities

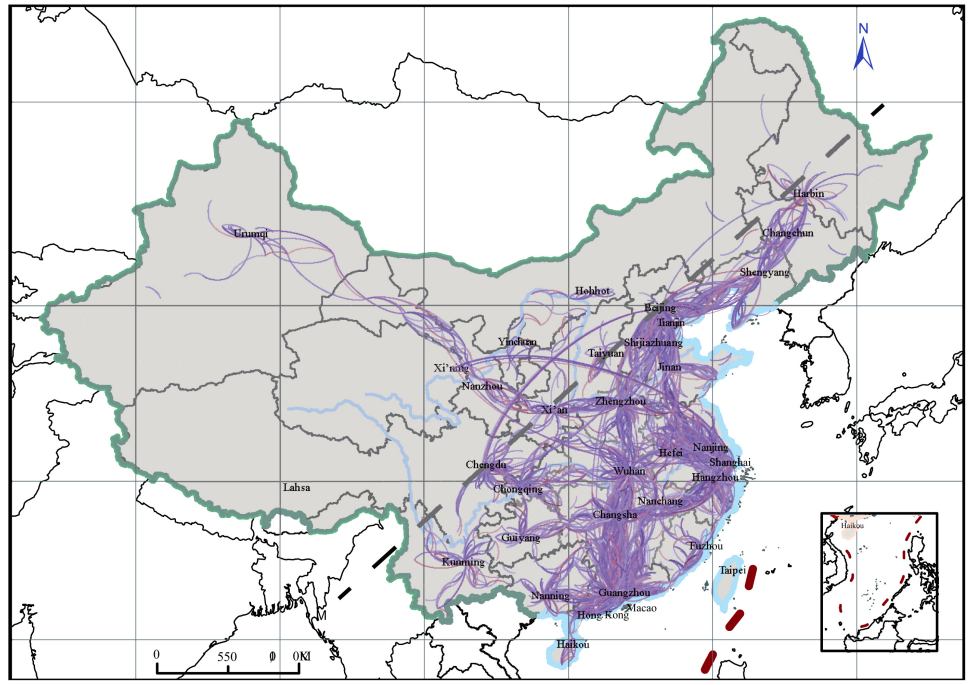
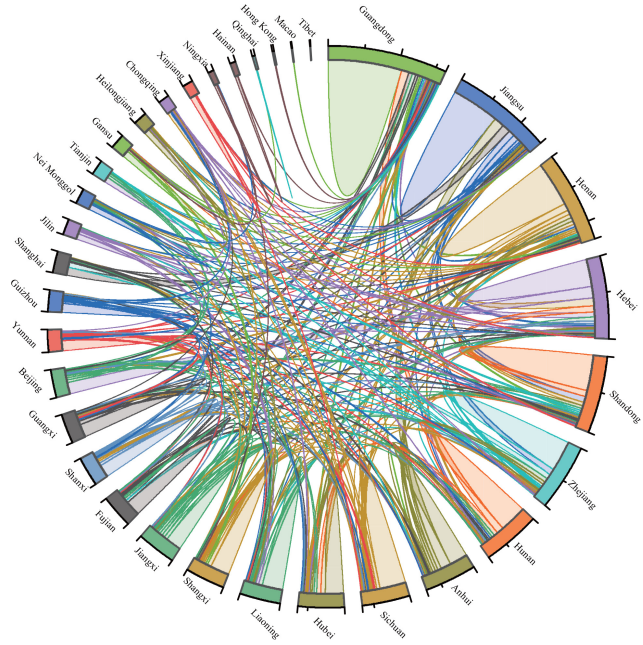


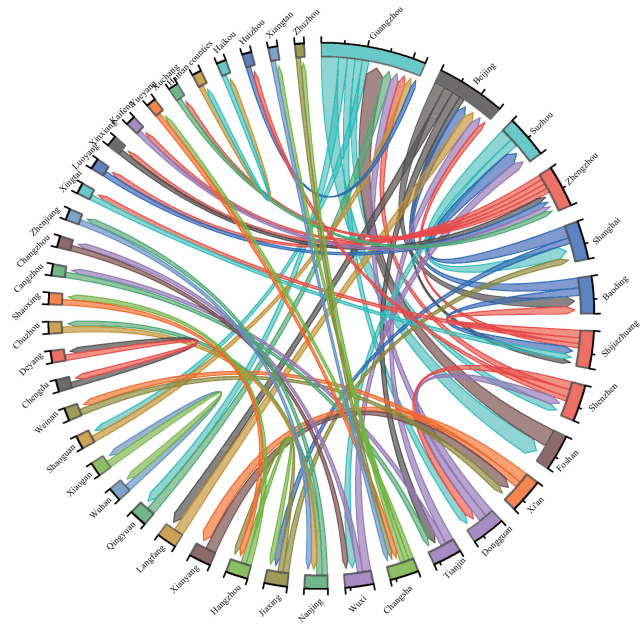
Fig. 3. The geographic distribution of human mobility. The size of point represents the average inflow in the period of China Chunyun. Plots interprovince and intercity Top 42 cites with the highest average daily flow, respectively.

(this can also be observed in Fig. 2(b)). The average rate of intraprovincial population migration in China is 45.22%, with the lowest at approximately 30% in Tibet and the highest at 64.67% in Guangdong. We can also observe a high frequency of population flows across urban agglomerations, such as the Jing-Jin-Ji (Beijing–Tianjin–Hebei) Region, the Yangtze River Delta Economic Zone (economic region including Shanghai, Zhejiang and Jiangsu) and the Pearl River Delta Metropolitan Region (a dense network of cities that covers nine prefectures of Guangdong Province, namely Dongguan, Foshan, Guangzhou, Huizhou, Jiangmen, Shenzhen, Zhaoqing, Zhongshan and Zhuhai and the Special Administrative Regions (SARs) of Hong Kong and Macau). These regions host many of the most high-tech and capital-intensive manufacturing industries in China, which provide many job openings and competitive salaries.

For further understanding of the variation of human mobility networks across periods, we mapped the population mobility network with an average daily flow $\geq$ 1000 and a distance >50 km at the county level during workdays, weekends, three-day holidays (e.g. Labor Day and Dragon Boat Festival) and others, as shown in Fig. 5. We find a significant increase in the volume of movements and a more diverse mobility during the holidays, except New Year's Day, which is mainly caused by the vacation policy (only one day off on January 1, 2014). Notably, during National Day,



(a)



(b)

Fig. 4. Chord diagram for human mobility. (a) and (b) Plots interprovince and intercity Top 42 cites with the highest average daily flow, respectively.

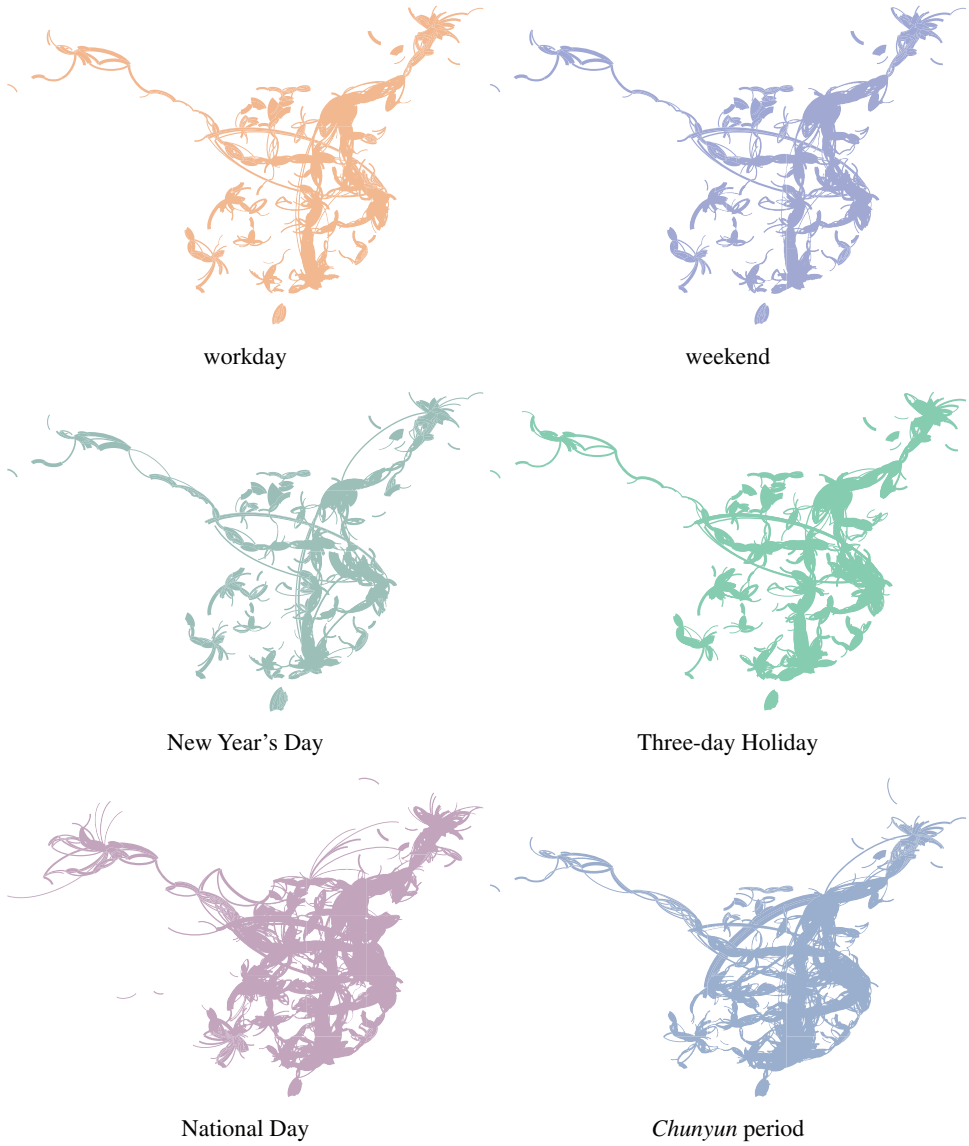


Fig. 5. A comparison of spatial patterns of mobility at difference periods.The human mobility networks are visualized with average daily flow on links ≥ 1000 and distance > 50 km at county level, the size of the edge is proportional to the average daily flow from origin to destination.

the mobility network patterns are diverse, such that links with heavier flows above a certain amount increase substantially, even more than during the Spring Festival, indicating that people travel for more diverse purposes with a variety of choices during the seven-day holiday. In contrast, peoples' main reason for traveling during the Chunyun period is to go home for the Spring Festival, which also happens during the other traditional Chinese festivals (e.g. the Dragon Boat Festival and the

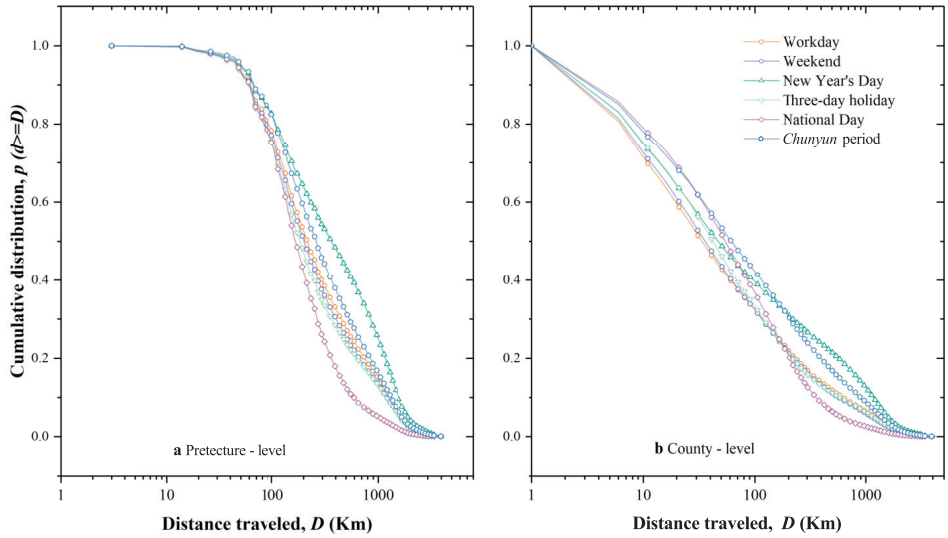


Fig. 6. Cumulative distribution of travel distances at (a) prefectural level and (b) county level.

Mid-Autumn Festival), such that the shape of the network connected by edges in other periods is more monotonous in comparison with National Day. However, the travel distance during the Chunyun period is longer than in any other period. In Sec. 3.2, we will illustrate the reason behind it in conjunction with the results of the gravity model.

Comparing the distributions of travel distance during workdays and weekends, we observe that short-distance travel occurs more frequently during weekends, while long-distance travel occurs more frequently during workdays (Fig. 6(a)). We infer that the increase in the proportion of long-distance travel during workdays is mainly caused by business trips. On New Year's Day (with only one day off on January 1, 2014), the number of flows is small (Fig. 2), while long-distance travel occurs more frequently in this period, suggesting that people are more likely to stay home or travel shorter distances if they do not need to work overtime. With additional comparisons, we also find that the size (the average value during different periods) of netflows increases as the days of the holiday decrease (Fig. 2), but the travel distance decreases as the days of the holiday decrease (Fig. 6(a)).

To further investigate the differences of distributions of travel distance in various periods, we draw a cumulative probability distribution against distance at the county level (Fig. 6(b)). Apparently, as county-level mobility networks capture more short-distance travels, the proportion of population movements with longer-distance travels is substantially higher in the prefecture-level distribution than in the county-level distribution. For example, at the prefecture level, 41.04% movements occur within 200 km during the Chunyun period; at the county level, the proportion increases to 69.16%.

3.3. *Comparison of mobility models*

Parameter estimation in gravity model. First, we use the grid search method[67] to estimate parameters α and β of the gravity model. Second, we set the candidate value for both parameters from 0 to 10 and then calculate all the candidate parameter sets to calculate the SSI (see Eq. (7)) of the gravity model. Finally, we select the parameter set that maximizes SSI. The results are shown in Table 1.

We observe that the value of α differs in normal times (workdays and weekends) and on holidays; the former has the minimum value of 0.89, and the latter has close to 1, indicating that the pattern of mobility in normal times is routine travel dominated by a fixed mode of work and life, without much flexibility, such that the destinations' attraction in normal times has a lower effect on human mobility. In contrast, people travel for long-distance tourism during the festivals such that the attraction of the destination comes into play. Notably, there is little difference in the parameters of workdays and weekends, although the value of β on weekends is a little larger than that on workdays. This is because both belong to normal times, and few people choose to travel freely for long distances (crossing city borders).

Additionally, the parameter β value of 1.31 being smaller during the Chunyun period than in any other period (except New Year's Day) suggests that individuals are less concerned about the travel costs to return home for the Spring Festival such that more long-distance travel occurs in this period (Fig. 5). In contrast, the value of β on National Day (also called the golden week for tourism in China) is the highest, which is caused by the increase of long trips.

Human mobility calculation. We then use the models presented in Sec. 2 to calculate the probability P_{ij} of an individual at location i selecting location j as the destination. Then, the population flow T_{ij} between the two communities i and j can be approximated by

$$T_{ij} = O_i \cdot P_{ij}, \tag{9}$$

where O_i represents the total outflow of location i. Therefore, we can compare the predictive accuracy of the mobility flow of each model by calculating the indicators of SSI and RMSE.

Comparison of calculated results. As shown in Tables 2 and 3, in general, the OPS model has a better prediction effect across all six periods, indicating that the travel assumptions of the OPS model among the variants of the IO model correspond to the actual human mobility. The gravity model also performs well during different periods. Surprisingly, the radiation model, which has shown excellent prediction performance in commuting, migration and commodity flows in the United States, is

Table 1. The parameter estimation results of the gravity model.

Workday	Weekend	New Year's Day	Three-day holiday	National Day	Chunyun period
0.89	0.89	1.04	1.05	0.98	0.96
1.43	1.50	1.07	1.51	2.19	1.31

Table 2. Comparison of models' prediction accuracy by SSI.

Time	Gravity	Radiation	OPS	OO	Impedance
Workday	0.643	0.203	**0.646**	0.431	0.553
Weekend	0.634	0.214	**0.637**	0.406	0.525
New Year's Day	**0.663**	0.172	0.637	0.521	0.617
Three-day holiday	0.614	0.224	**0.617**	0.386	0.500
National Day	**0.513**	0.358	0.404	0.211	0.275
Chunyun period	**0.683**	0.184	0.672	0.472	0.599

Table 3. Comparison of models' prediction accuracy by RMSE.

Time	Gravity	Radiation	OPS	OO	Impedance
Workday	4226.668	3726.321	**2643.611**	3978.870	3902.999
Weekend	4530.526	3709.948	**2821.425**	4179.692	4000.847
New Year's Day	2907.887	5184.735	**2625.831**	4154.186	4189.328
Three-day holiday	4883.928	4040.197	**3308.976**	4824.597	4586.256
National Day	11241.906	6211.685	**5116.416**	7390.114	6912.823
Chunyun period	5321.168	6760.330	**401.073**	5515.990	5353.729

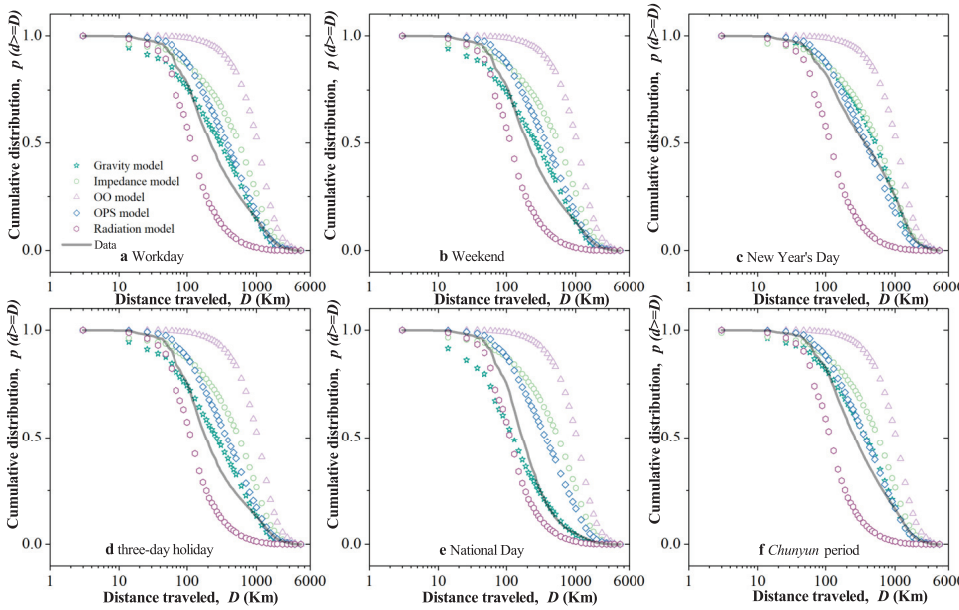


Fig. 7. Comparison between the predictions of the mobility models and the empirical data during across over periods. We calculate the distribution of travel distance by gravity model, radiation model, OO model, OPS model and the impedance model, compared with real data across over periods. p is defined as the probability of travel between locations at a certain distance, travel distance was processed by logarithmic function.

the worst predictor of human mobility in China. This may be because the individual prefers a potential destination near the origin rather than far from the origin in commuting and freight, where travel costs are more important, or China and the United States have differences in their economic development models, cultures, industry categories and so on. In contrast, the radiation model suggests that the destination selected by the individual is the closest location with higher benefits than those of the origin, which may not be in accord with the more complex destination selection rule in China, that is, people compare benefits and costs when choosing a destination.

Cumulative distribution of distance. To show the differences of the mobility model more intuitively, we plot the cumulative distribution of the travel distance calculated using the gravity model, the impedance model, the OO model, the OPS model and the radiation model over all periods (Fig. 7). As shown in Fig. 7, the results predicted by the OO model are far from the origin data, suggesting that the individual selecting only the location with the highest opportunity benefits does not reflect the reality. Additionally, we find that the second half of the curve, calculated using the gravity model, almost coincides with the original data, revealing the gravity model as suitable for predicting human mobility in long distances. In general, the OPS model performs best in human mobility prediction and understanding. This evidence demonstrates that the individual may choose not only the closest location with higher opportunity benefits than those of the origin but also other locations with higher opportunity benefits than the benefits of the origin and intervening opportunities[50] in actual practice.

4. Discussion

For a country as large and populous as China, analyzing the patterns of human mobility during specific periods and comparing models for population migration prediction are of extraordinary significance for transportation planning, urban construction, economic improvement and crisis management. Over the past half-century, from gravity models to IO models and their variants, models for understanding human mobility patterns are constantly being proposed and developed, as well as the acquisition of trajectory data by intelligent terminal equipment, making it possible to study the spatiotemporal characteristics and patterns of human mobility. In this study, we have analyzed large-scale, national mobility data, extracted from China's largest internet service provider, Baidu. The spatial coverage, resolution and time length of these data are superior in comparison to those of previous studies in China. We have used these data to examine the general and the exceptional mobility patterns in China. Through our analysis of temporal patterns in 359 prefectures, we have found that the population netflows show a strong weekly pattern and differences in the migration scale among the regions. Additionally, we have discovered the backflow effect around the holidays, which is driven either by traditional cultural conditioning or by special events occurring in China, and the backflow effect varies

from region to region due to differences in cultures, economies and urban infrastructures. Interestingly, in our further study, we have found that the number of cross-city movements increases with the number of days of the holidays, that is, people tend to travel more often during longer holidays. However, such an increase of traveling frequency is associated with a decrease of traveling distance, indicating that short trips comprise the main type of travel during long holidays. Furthermore, through our analysis of spatial patterns, we have found that the communities form spatially cohesive regions, reflecting the regional division of the geography, in turn indicating that even in the era of convenient transportation and the seeming decrease of the influence of distance, people still prefer to travel locally, visiting neighboring regions more often than those farther away.

Furthermore, we have explored the mechanisms underlying the individual's destination selection behavior in different periods of human mobility using the gravity model and four parameter-free models — three IO models that characterize different types of the individual's destination selection behavior, and the impedance model that has been proven effective in human mobility prediction during the 2010 Haiti cholera epidemic. Notably, the OPS model performs best in predicting intercity population flows in China, suggesting that people may choose not only the closest location with higher opportunity benefits than those of the origin but also other locations with higher opportunity benefits than those of the origin and intervening opportunities in real-life practice. The results confirm other studies' findings[33,50] that the OPS model can better predict intercity population flows. Additionally, the calculated results of the gravity model successfully uncover the travel mechanisms during different periods, that is, the different effects of a destination's attraction and travel costs drive human mobility during different time periods.

However, with economic development, the improvement of the transportation system, especially the launch of high-speed rail and aviation networks, human mobility patterns in China may have changed, such that the data we have used may be somewhat outdated. It is plausible that the limitations may have influenced the applications of some of the conclusions in this paper for the current population mobility. We believe that the analysis presented in this paper is of great significance for the government departments concerned, when they face the problems of transportation planning, urban construction, economic improvement, crisis management and so on. Further research should focus on exploring the applicability of human mobility for finer spatial scales and on exploring the universal human mobility model concerned with time effects.

Acknowledgments

We would like to thank Huihua Zhao for the help in data preparation and analysis. This work was supported by the National Natural Science Foundation of China (Grant Numbers 72025405, 82041020, 71771213, 71774168 and 71822102) and the

Hunan Science and Technology Plan Project (2020TP1013, 2020JJ4673). B.S., X.Y. and X.L. contributed equally to this work.

References

1. J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia and N. A. Christakis, *Nature* **582**, 389 (2020).
2. M. Chinazzi *et al.*, *Science* **368**, 395 (2020).
3. M. U. Kraemer *et al.*, *Science* **368**, 493 (2020).
4. X. Lu, L. Bengtsson and P. Holme, *Proc. Natl. Acad. Sci. USA* **109**, 11576 (2012).
5. L. Bengtsson, X. Lu, A. Thorson, R. Garfield and J. Von Schreeb, *PLoS Med.* **8**, e1001083 (2011).
6. X. Lu *et al.*, *Glob. Environ. Change* **38**, 1 (2016).
7. S.-Y. Tan *et al.*, *Natl. Sci. Rev.* (2021), doi: 10.1093/nsr/nwab148.
8. B. T. Dai, S. Y. Tan, S. R. Chen, M. S. Cai, S. Qin and X. Lu, *Wuli Xuebao/Acta Phys. Sin.* **70** (2021).
9. A. Bassolas *et al.*, *Nat. Commun.* **10**, 1 (2019).
10. B. Slavko, K. Glavatskiy and M. Prokopenko, *Sci. Rep.* **10**, 1 (2020).
11. F. Becker and K. W. Axhausen, *Transportation* **44**, 1293 (2017).
12. X. Song, Q. Zhang, Y. Sekimoto and R. Shibasaki, Prediction of human emergency behavior and their mobility following large-scale disaster, in *KDD'14: Proc. 20th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, New York, NY, USA, 2014, pp. 5–14.
13. M. U. Kraemer *et al.*, *Nat. Microbiol.* **4**, 854 (2019).
14. T. Venables, *LSE Research Online Documents on Economics*, Vol. **5** (2003), p. 3.
15. R. J. Lalonde and R. H. Topel, *Handbook of Population and Family Economics*, Vol. **1** (1997), p. 799.
16. S. Meloni, N. Perra, A. Arenas, S. Gómez and A. Vespignani, *Sci. Rep.* **1**, 62 (2011).
17. F. Aidan and I. I. Bogoch, *Trends Parasitol.* **34**, 772 (2018).
18. X. Lu *et al.*, *Health Data Sci.* **2021** (2021).
19. L. Pappalardo, D. Pedreschi, Z. Smoreda and F. Giannotti, Using big data to study the link between human mobility and socio-economic development, in *2015 IEEE Int. Conf. Big Data (Big Data)* (IEEE, 2015), pp. 871–878.
20. C. Di Maria and P. Stryszowski, *J. Dev. Econ.* **90**, 306 (2009).
21. W. H. McNeill, *Popul. Dev. Rev.* **10**, 1 (1984).
22. C. Ning *et al.*, *Nat. Commun.* **11**, 1 (2020).
23. R. Black, W. N. Adger, N. W. Arnell, S. Dercon, A. Geddes and D. Thomas, *Glob. Environ. Change* **21**, S3 (2011).
24. G. McNicoll, *Dev. Rev.* (2014).
25. B. Jiang, J. Yin and S. Zhao, *Phys. Rev. E* **80**, 021136 (2009).
26. M. C. Gonzalez, C. A. Hidalgo and A.-L. Barabasi, *Nature* **453**, 779 (2008).
27. Y. Liang, X. Zhou, B. Guo and Z. Yu, **1**, 409 (2012).
28. C. Song, Z. Qu, N. Blumm and A.-L. Barabási, *Science* **327**, 1018 (2010).
29. L. Alessandretti, U. Aslak and S. Lehmann, *Nature* **587**, 402 (2020).
30. M. Schläpfer *et al.*, *Nature* **593**, 522 (2021).
31. G. K. Zipf, *Am. Sociol. Rev.* **11**, 677 (1946).
32. K. Sallah *et al.*, *Int. J. Health Geogr.* **16**, 1 (2017).
33. E.-J. Liu and X.-Y. Yan, *Sci. Rep.* **10**, 1 (2020).
34. X.-Y. Yan, W.-X. Wang, Z.-Y. Gao and Y.-C. Lai, *Nat. Commun.* **8**, 1 (2017).
35. S. Lai *et al.*, *Palgrave Commun.* **5**, 1 (2019).

36. C. Ruktanonchai *et al. Sci. Rep.* **11**, 15389 (2021), https://doi.org/10.1038/s41598-021-94683-7.
37. L. Alessandretti, P. Sapiezynski, S. Lehmann and A. Baronchelli, *PLoS One* **12**, e0171686 (2017).
38. J. Q. Stewart, *Sociometry* **11**, 31 (1948).
39. S. A. Stouffer, *Am. Sociol. Rev.* **5**, 845 (1940).
40. S. Akwawua and J. A. Pooler, *J. Geogr. Syst.* **3**, 69 (2001).
41. T. R. Anderson, *Am. Sociol. Rev.* **20**, 287 (1955).
42. O. R. Galle and K. E. Taeuber, *Am. Sociol. Rev.* **31**, 5 (1966).
43. K. E. Haynes, D. L. Poston Jr. and P. Schnirring, *Econ. Geogr.* **49**, 68 (1973).
44. E. Miller, *J. Reg. Sci.* **12**, 475 (1972).
45. H. Elffers, D. Reynald, M. Averdijk, W. Bernasco and R. Block, *Crime Prev. Community Saf.* **10**, 85 (2008).
46. D. W. Dison and C. W. Hale, *Growth Change* **8**, 15 (1977).
47. X.-Y. Yan and T. Zhou, *Sci. Rep.* **9**, 1 (2019).
48. H. Barbosa *et al.*, *Phys. Rep.* **734**, 1 (2018).
49. F. Simini, M. C. González, A. Maritan and A.-L. Barabási, *Nature* **484**, 96 (2012).
50. E. Liu and X. Yan, *Physica A, Stat. Mech. Appl.* **526**, 121023 (2019).
51. A. P. Masucci, J. Serras, A. Johansson and M. Batty, *Phys. Rev. E* **88**, 022812 (2013).
52. X.-Y. Yan, C. Zhao, Y. Fan, Z. Di and W.-X. Wang, *J. R. Soc. Interface* **11**, 20140834 (2014).
53. M. Lenormand, M. Lenormand, S. Huet, F. Gargiulo, G. Deffuant *et al.*, *PLoS One* **7**, e45985 (2012).
54. B. D. Dalziel, B. Pourbohloul and S. P. Ellner, *Proc. R. Soc. B, Biol. Sci.* **280**, 20130763 (2013).
55. X. Liang, J. Zhao, L. Dong and K. Xu, *Sci. Rep.* **3**, 1 (2013).
56. F. Simini, A. Maritan and Z. Néda, *PLoS One* **8**, e60069 (2013).
57. C. Kang, Y. Liu, D. Guo and K. Qin, *PLoS One* **10**, e0143500 (2015).
58. A. Sim, S. N. Yaliraki, M. Barahona and M. P. Stumpf, *J. R. Soc. Interface* **12**, 20150315 (2015).
59. E. W. Montroll and G. H. Weiss, *J. Math. Phys.* **6**, 167 (1965).
60. R. A. Serway and J. W. Jewett, *Physics for Scientists and Engineers* (Cengage Learning, 2018).
61. R. Li *et al.*, *Nat. Commun.* **8**, 1 (2017).
62. C. N. B. Statistics, The sixth national census bulletin, National Bureau of Statistics (2010).
63. Y. Chen, *Chaos Solitons Fractals* **77**, 174 (2015).
64. D. W. Goodall, *Biometrics* **22**, 882 (1966).
65. R. J. Hyndman and A. B. Koehler, *Int. J. Forecast.* **22**, 679 (2006).
66. Z. Hu, Y. Wang, Y. Liu, H. Long and J. Peng, *ISPRS Int. J. Geo-Inf.* **5**, 24 (2016).
67. P. Lerman, *J. R. Stat. Soc. C, Appl. Stat.* **29**, 77 (1980).