Predicting highway freight transportation networks using radiation models

Wei-Xing Zhou,[1,2,3] Li Wang,[1] Wen-Jie Xie,[1,3,*] and Wanfeng Yan[3,4]

[1]*School of Business, East China University of Science and Technology, Shanghai 200237, China*
[2]*Department of Mathematics, East China University of Science and Technology, Shanghai 200237, China*
[3]*Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China*
[4]*Zhicang Technologies, Beijing 100016, China*

(Received 28 June 2020; accepted 1 November 2020; published 24 November 2020)

Highway freight transportation (HFT) plays an important role in the economic activities. Predicting HFT networks is not only scientifically significant in the understanding of the mechanism governing the formation and dynamics of these networks, but also of practical significance in highway planning and design for policymakers and truck allocation and route planning for logistic companies. In this work we apply parameter-free radiation models to predict the HFT network in mainland China and assess their predictive performance using metrics based on links and fluxes, which can be done in reference to the real directed and weighted HFT network between 338 Chinese cities constructed from about 15.06 million truck transportation records in five months. It is found that the radiation models exhibit relatively high accuracy in predicting links but low accuracy in predicting fluxes on links. Similar to gravity models, radiation models also suffer difficulty in predicting long-distance links and the fluxes on them. Nevertheless, the radiation models perform well in reproducing several scaling laws of the HFT network. The adoption of population or gross domestic product in the model has a minor impact on the results, and replacing the geographic distance by the path length taken by most truck drivers does not improve the results.

DOI: 10.1103/PhysRevE.102.052314

I. INTRODUCTION

Freight transportation is vital in the daily activities of an economy, which transfers raw materials, intermediate products, and final products from origin locations to destination locations. Freight transportation can be accomplished through different modes such as airplanes, trains, ships, and trucks. Unlike long-distance human trips which are usually through well-scheduled public transportation systems with prefixed timetables, most freight transportation activities occur when a transportation demand appears and is matched by a transporter. Usually, data records for such kinds of freight transportation are unavailable to researchers. Fortunately, we obtained a database of the highway freight transportation (HFT) network between 338 cities in mainland China [1], which records the transportation dates, origin cities, and destination cities associated with trucks and is a directed, weighted network. This database allows us to study the structure, function and dynamics of the HFT network.

In this work we investigate the problem of reconstructing the HFT network by predicting the directed links between cities and the fluxes (or weights) on the links. This problem is in essence that of *trip distribution* in the field of human mobility or immigration, and many spatial interaction models have been developed [2–6]. The problem can have different settings by considering multiple user classes [7], different trip modes [8,9], and different constraints on the origins or destinations [5,10].

There are different spatial interaction models [6,11], mainly based on the intervening opportunities model [12], the gravity model [13], the fitness model [14], and the radiation model [15–17]. The gravity model and its variants are among the most studied and most widely applied [18–22]. Empirical evidence has accumulated from the studies of human migration and mobility [15,17,23–27], transportation networks [1,28–34], intranational and international trades [13,35–48], and mobile phone communication [49–54], to mention a few.

However, gravity models and many other models have a few limitations or shortcomings [15,55,56]. Wang *et al.* have studied the directed and undirected Chinese HFT networks using four gravity models [1]. However, these gravity models are symmetric and are not very suitable to handle the directions of truck flows. In contrast, the radiation model proposed by Simini *et al.* can fix these problems and is parameter free [15]. One does not need to calibrate the radiation model and can use only the "masses" (usually population or gross domestic product) of the cities and the distance matrix between the cities to predict the truck fluxes between pairs of cities. Ren *et al.* further proposed the cost-based radiation model by considering the cost needed to take to transport between two cities instead of the geographic distance and found that the new model outperforms the raw radiation model in predicting the highway traffic in the United States [16]. Xia *et al.* proposed the gravity-radiation model to estimate population flows and measure the overall accessibility at multiple scales in Australia [57].

*wjxie@ecust.edu.cn

There are a number of works in the field of complex transportation networks that focus on the physics of the problem [58]. Such complex transportation networks are spatially embedded networks that may exhibit phase transitions [59]. It is found that long-range shortcuts have important impacts on the navigation in transportation networks [60,61]. Researchers also proposed design principles for optimal transport networks [62,63]. These numerical and theoretical studies not only deepen our understanding about the behavior of complex transportation networks but also have important practical implications.

In this work we apply four parameter-free radiation models (the raw radiation model with population, the raw radiation model with gross domestic product, the cost-based radiation model with population, and the cost-based radiation model with gross domestic product) to predict the HFT network in mainland China. We assess and compare their predictive performance using metrics based on links and fluxes, which can be done in reference to the real directed and weighted HFT network. We find that the radiation models exhibit relatively high accuracy in predicting the links but low accuracy in predicting the fluxes on the links. Similar to gravity models, these radiation models also perform badly in predicting long-distance links and the fluxes on them. Interestingly, we find that the radiation models perform well in reproducing several scaling laws of the HFT network. Our results also show that the four models have comparative performance, that is, using population or gross domestic product in the model has a minor impact on the results, and using the path length as the distance cannot improve the results either.

The remainder of this work is organized as follows. Section II explains briefly the radiation models adopted. Section III describes the data we investigate and the basic properties they possess. Section IV presents the main results of our analysis. Section V summarizes this work.

II. THE RADIATION MODELS

Consider two cities i and j with population P_i and P_j and gross domestic product G_i and G_j, respectively, at distance d_{ij}^{geo} from each other. We denote F_{ij} and $\tilde{F}_{ij}$ the real truck flux and the predicted flux from i to j. According to Simini *et al.* [15], the radiation model predicts the flux from city i to city j in the following form:

$$\tilde{F}_{ij} = F_i^{\text{out}} \frac{P_i P_j}{(P_i + S_{ij})(P_i + P_j + S_{ij})}, \tag{1}$$

where S_{ij} is the total population in the circle of radius d_{ij}^{geo} centered at i but excluding the source and destination population, and

$$F_i^{\text{out}} = \sum_{j \neq i} F_{ij} \tag{2}$$

is the total number of departing trucks from city i. Alternatively, when we use the gross domestic product as the proxy, we have

$$\tilde{F}_{ij} = F_i^{\text{out}} \frac{G_i G_j}{(G_i + S_{ij})(G_i + G_j + S_{ij})}, \tag{3}$$

where S_{ij} is the total GDP in the circle of radius d_{ij}^{geo} centered at i but excluding the source and destination GDP. We can see that the transportation probability $\tilde{F}_{ij}/F_i^{\text{out}}$ from i to j depends only on their population or GDP and the distribution of population and GDP around i. Note that S_{ij} does not include the population or GDP outside of mainland China, as in the U.S. case [15].

Using the commuting data between the cities of England and Wales at macroscales and between the wards of London at microscales, Masucci *et al.* found that the raw radiation model significantly underestimates the commuting flows for large cities and generalized the radiation model by introducing a normalization factor for finite systems [64]:

$$\tilde{F}_{ij} = \frac{F_i^{\text{out}}}{1 - \frac{M_i}{M}} \frac{M_i M_j}{(M_i + S_{ij})(M_i + M_j + S_{ij})}, \tag{4}$$

where M_i could be the population P_i or the GDP G_i, and $M = \sum_i M_i$ is the total population or GDP. However, our analysis shows that this factor only has a negligible impact on the predicted fluxes.

Ren *et al.* proposed the cost-based radiation model, which improves the prediction of highway traffic in the United States [16]. We also adopt this model in our analysis. The formulas are the same as the raw radiation models presented in Eqs. (1) and (3), except that d_{ij}^{geo} is replaced by the length d_{ij}^{cost} of the path taken by most truck drivers, the driving distance from city i to city j.

III. DATA DESCRIPTION

A. Highway transportation flux

The highway truck transportation data we analyze were provided by a leading truck logistics company in China. The data contain 15.06 million truck transportation records between pairs of 338 cities in mainland China during the period from 1 January 2019 to 31 May 2019 [1]. Each record contains the origin and destination cities. We construct the flux matrix $\mathbf{F} = [F_{ij}]_{338\times338}$, where F_{ij} stands for the number of trucks with freights driven from city i to city j, and unloaded trucks are not counted in. Because radiation models do not consider intracity transportation, we pose

$$F_{ii} = 0. \tag{5}$$

Note that F_{ij} is not necessary to be equal to F_{ji} for $i \neq j$.

In Fig. 1 we illustrate two examples of a flux map with different transportation fluxes. The arrows of the segments are omitted. In Fig. 1(a) we show the origin-destination pairs with $F_{ij} > 2000$. In Fig. 1(b) we show the origin-destination pairs with $F_{ij} = 100$ or $F_{ij} = 101$. It is rational that the transportation network is denser in the east than in the west. In addition, there is a network community of cities in the east that has dense interconnections but not in the west. This is due to the fact that cities close to the coast have a more diverse and higher developing economy compared to cities in the west.

B. Distance between different cities

In our analysis, we use the geographic distance d_{ij}^{geo} measured in kilometers in the raw radiation model and the driving

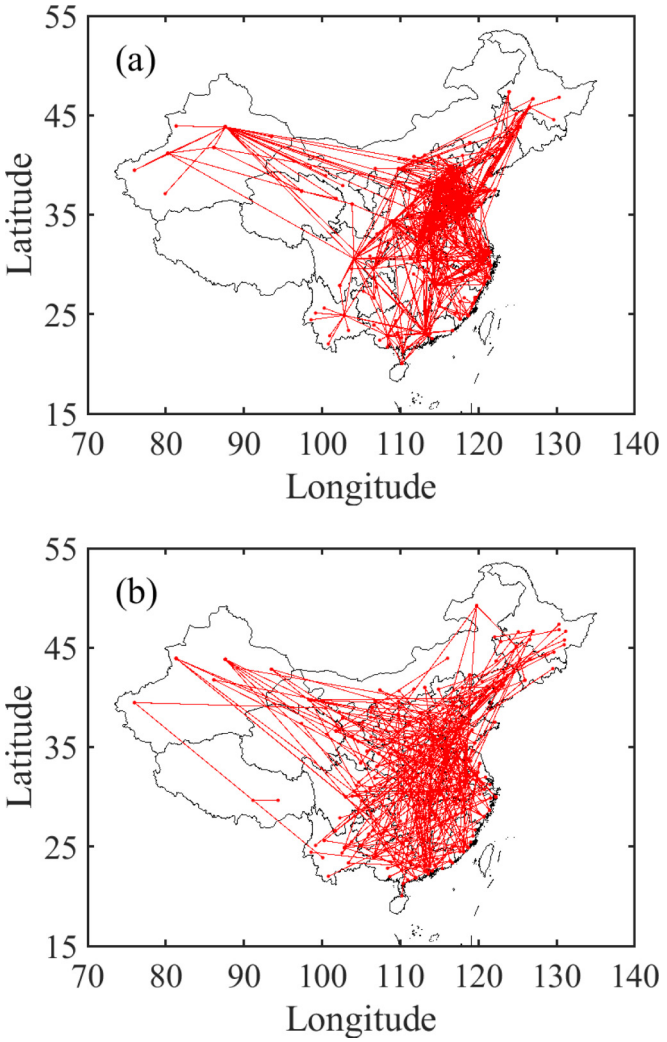

FIG. 1. Examples of flux map with different transportation fluxes. The arrows of the segments are omitted. In plot (a) we show the origin-destination pairs with $F_{ij} > 2000$. In plot (b) we show the origin-destination pairs with $F_{ij} = 100$ or $F_{ij} = 101$.

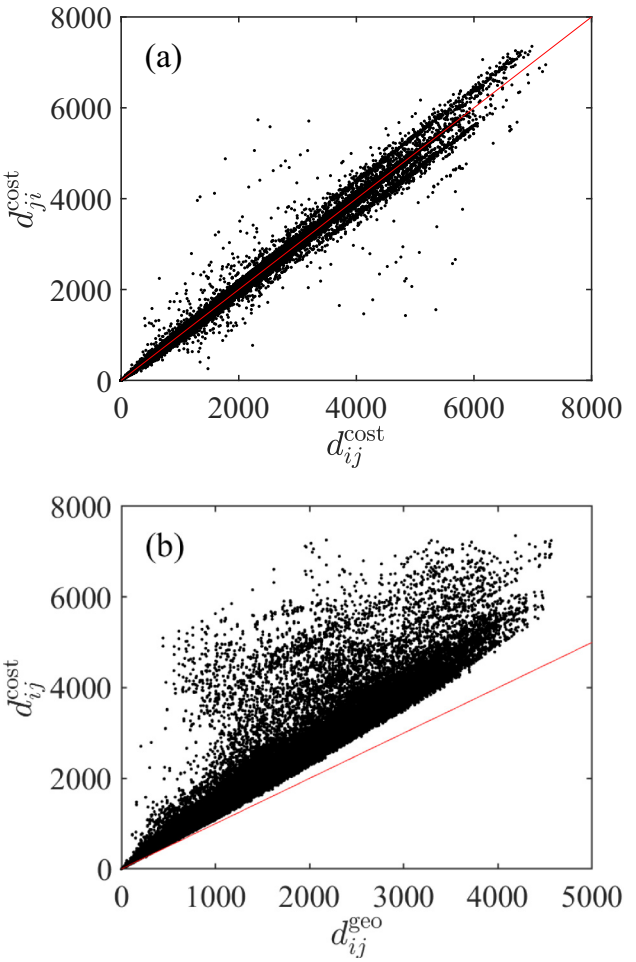

FIG. 2. Important traits of the two distance measures. (a) Asymmetry of the driving distance matrix between different cities. (b) The relationship between driving distance and geographic distance.

distance d_{ij}^{cost} measured in kilometers in the cost-based radiation model. The geographic distance d_{ij}^{geo} is the shortest surface distance between two cities i and j whose locations are determined by their longitudes and latitudes, which is the length of the great circle arc connecting these two cities on the surface of the earth. Obviously, the geographic distance matrix is symmetric such that

$$d_{ij}^{\mathrm{geo}} = d_{ji}^{\mathrm{geo}}. \tag{6}$$

The driving distance d_{ij}^{cost} between two cities i and j is not calculated due to the information of longitude and latitude. Rather, we use the path length between the two cities that is usually optimal since the path is chosen by truck drivers. As shown in Fig. 2(a), the driving distance matrix is asymmetric, that is,

$$d_{ij}^{\mathrm{cost}} \neq d_{ji}^{\mathrm{cost}}, \tag{7}$$

which shows that due to local roads, the driving distance or path length d_{ij}^{cost} from i to j is not necessarily identical to the driving distance d_{ji}^{cost} from j to i.

It is obvious that the driving distance d_{ij}^{cost} is longer than the geographic distance d_{ij}^{geo}:

$$d_{ij}^{\mathrm{geo}} < d_{ij}^{\mathrm{cost}} \tag{8}$$

for all pairs of cities, as illustrated in Fig. 2(b). On average, the difference between these two distance measures increases when the two cities are farther away from each other.

C. Population and gross domestic product

The population and gross domestic product data for the 338 Chinese cities in 2017 were retrieved online from the Complete Collection of World Population [65], which is publicly available except for a few cities. We found the missing data from Baidu Encyclopedias [66]. We then calculate the four matrices of S_{ij} corresponding to the four combinations of M

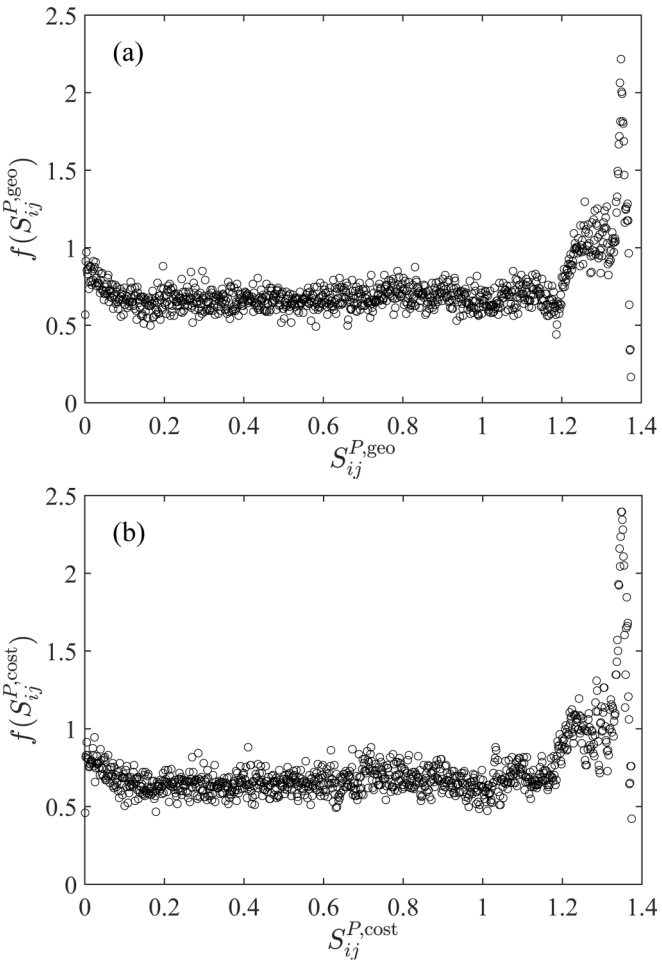

FIG. 3. Empirical distribution of S_{ij} between two cities. (a) The population $S_{ij}^{P,\mathrm{geo}}$ (billion people) is determined by geographic distance d_{ij}^{geo}. (b) The population $S_{ij}^{P,\mathrm{cost}}$ (billion people) is determined by driving distance d_{ij}^{cost}.

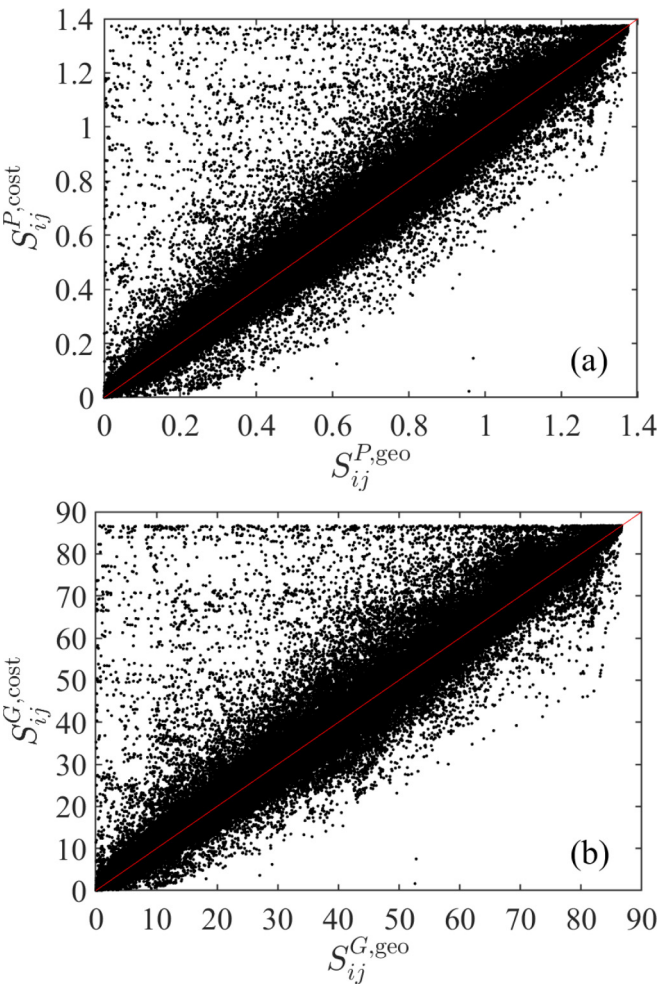

FIG. 4. Comparison of $S_{ij}^{M,\mathrm{geo}}$ and $S_{ij}^{M,\mathrm{cost}}$ to investigate the influence of driving distance and geographic distance. (a) The radiation models are based on population with $M = P$ (billion people). (b) The radiation models are based on GDP with $M = G$ (trillion yuan).

(population P or gross domestic product G) and d (geographic distance d_{ij}^{geo} or driving distance d_{ij}^{cost}).

Figure 3 shows the empirical distributions of S_{ij} for $M = P$. It is observed that the two distributions related to the geographic distance d_{ij}^{geo} in Fig. 3(a) and the driving distance d_{ij}^{cost} in Fig. 3(b) are very similar. It is found that the distribution is roughly uniform,

$$f(S_{ij}) \sim \text{constant}, \tag{9}$$

except for very small and very large values of S_{ij}. The majority of small S_{ij} values correspond to the city pairs in the west. The S_{ij} values in the bulk are mainly obtained from those city pairs in the east. The large S_{ij} values reflect city pairs that are far away from each other. The two distributions of S_{ij} for $M = G$ are also very similar to those in Fig. 3, which are thus not shown in this work.

We investigate the impact of distance definition on the value of S_{ij}. The results are depicted in Fig. 4(a) for S_{ij}^{P} and in Fig. 4(b) for S_{ij}^{G}. These two plots exhibit similar patterns as the transportation probabilities, showing that the choice of M as P or G will give quantitatively similar results [34]. It is observed that most of the data points fluctuate around $S_{ij}^{M,\mathrm{cost}} = S_{ij}^{M,\mathrm{geo}}$. There are also many points deviating far away from the diagonals, showing that $S_{ij}^{M,\mathrm{cost}} \gg S_{ij}^{M,\mathrm{geo}}$. It reflects the complex geographic roads and rather heterogeneous distribution of population and GDP in mainland China. It also implies that the predicted fluxes from the raw radiation model and the cost-based radiation model will be very different for many city pairs.

The asymmetric relationship between $S_{ij}^{M,\mathrm{geo}}$ and $S_{ji}^{M,\mathrm{geo}}$ for $M = P$ is illustrated in Fig. 5. The plots exhibit similar patterns: The $S_{ij}^{M,d}$ values span four orders of magnitude, the points fluctuate around the diagonal, and the highest density is around $(10^5, 10^5)$, as shown in Fig. 3. The figure also shows that the $S_{ij}^{M,d}$ matrices are asymmetric, such that

$$S_{ij}^{M,d} \neq S_{ji}^{M,d} \tag{10}$$

for most origin-destination (OD) city pairs. This is not surprising, because $S_{ij}^{M,\mathrm{geo}}$ and $S_{ji}^{M,\mathrm{geo}}$ correspond to two different circles with different circle centers at i and j.

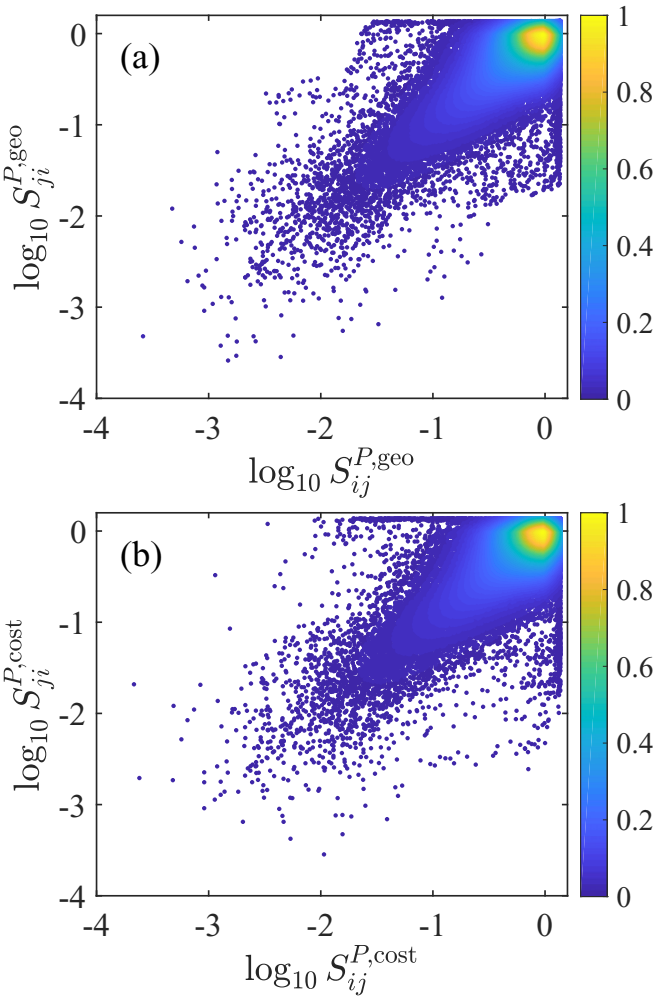

FIG. 5. Asymmetric relationship between S_{ij}^{P} and S_{ji}^{P}. (a) The population $S_{ij}^{P,\mathrm{geo}}$ is determined by driving distance d_{ij}^{geo}. (b) The population $S_{ij}^{P,\mathrm{cost}}$ is determined by driving distance d_{ij}^{cost}.

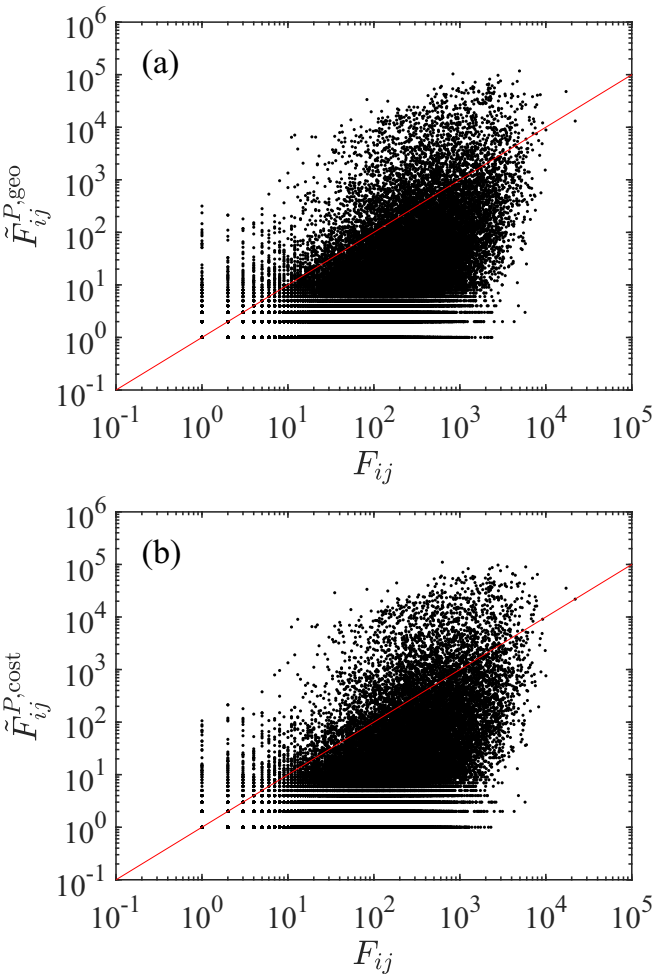

FIG. 6. Comparing the predicted flux, $\tilde{F}_{ij}^{P,d}$, with the measured flux, F_{ij}, for each pair of cities. We compare the real data with two formulations of the radiation model, $d_{ij} = d_{ij}^{\mathrm{geo}}$ (a) and $d_{ij} = d_{ij}^{\mathrm{cost}}$ (b). Data points are scatter plot for each pair of cities.

IV. RESULTS

A. Predicting truck fluxes

We predict the fluxes $\tilde{F}_{ij}^{P,\mathrm{geo}}$ and $\tilde{F}_{ij}^{P,\mathrm{cost}}$ using Eq. (1) and $\tilde{F}_{ij}^{G,\mathrm{geo}}$ and $\tilde{F}_{ij}^{G,\mathrm{cost}}$ using Eq. (3). To have a qualitative impression of the prediction performance, we compare the predicted flux $\tilde{F}_{ij}^{P,\mathrm{geo}}$ from the raw radiation model with the measured flux F_{ij} for each pair of cities in Fig. 6(a) and compare the predicted flux $\tilde{F}_{ij}^{P,\mathrm{geo}}$ from the cost-based radiation model with the measured flux F_{ij} for each pair of cities in Fig. 6(b). The two scatter plots exhibit very similar patterns, and so do the results for $\tilde{F}_{ij}^{G,\mathrm{geo}}$ and $\tilde{F}_{ij}^{G,\mathrm{cost}}$.

Overall, the predicted fluxes $\tilde{F}_{ij}^{P,d}$ fluctuate around the measured fluxes F_{ij}. However, for many OD city pairs, the predicted values could deviate remarkably, as observed in other systems (see Ref. [15] and related works). For the highway freight transportation network, it shows that these models are more likely to underestimate the flux than to overestimate it, because Fig. 6 shows that there are more points below the diagonal $\tilde{F}_{ij}^{P,d} = F_{ij}$ than above it.

In order to quantify the performance of the models in predicting fluxes, we adopt two classes of indicators. The first class of indicators considers the link weights or fluxes. We adopt an analogy of the concept of common part of commuters (CPC) [55,56,67,68], which is similar to the Sørensen similarity index with a slight difference in the denominator [69]. Following Refs. [55,56], we call it the common part of fluxes CPF,

$$\mathrm{CPF}(F,\tilde{F}) = \frac{\sum_i^n \sum_{j=1}^n \min(F_{ij},\tilde{F}_{ij})}{\sum_i^n \sum_{j=1}^n F_{ij}}, \tag{11}$$

where the numerator is the number of common trucks in the real and predicted transportation networks, and the denominator is the number of total trucks in the real transportation network. We also adopt two more indicators called the normalized mean absolute error (NMAE),

$$\mathrm{NMAE}(F,\tilde{F}) = \frac{\sum_{i=1}^n \sum_{j=1}^n |F_{ij}-\tilde{F}_{ij}|}{\sum_{i=1}^n \sum_{j=1}^n F_{ij}}, \tag{12}$$

TABLE I. Performance of the raw radiation model and the cost-based radiation model in predicting highway freight transportation network.

	$M=P$ $d=d^{\text{geo}}$	$M=P$ $d=d^{\text{cost}}$	$M=G$ $d=d^{\text{geo}}$	$M=G$ $d=d^{\text{cost}}$
CPF	0.2312	0.2330	0.2281	0.2267
NRMSE	0.0371	0.0374	0.0386	0.0384
NMAE	1.5337	1.5330	1.5392	1.5460
CPL	0.7673	0.7648	0.7038	0.7009
PCPEL	0.6317	0.6276	0.5487	0.5444
PCPML	0.9313	0.9375	0.9510	0.9573
PTIE	0.3683	0.3724	0.4513	0.4556
PTIIE	0.0687	0.0625	0.0490	0.0427

and the normalized root-mean-square error (NRMSE) [68],

$$\text{NRMSE}(F,\tilde{F}) = \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(F_{ij}-\tilde{F}_{ij})^2}}{\sum_{i=1}^{n}\sum_{j=1}^{n}F_{ij}}. \tag{13}$$

The second class of indicators considers how good the models correctly predict existing links. Reference [68] introduced the common part of links (CPL),

$$\text{CPL}(F,\tilde{F}) = \frac{2\sum_{i}^{n}\sum_{j=1}^{n}\mathbf{1}_{F_{ij}>0}\times\mathbf{1}_{\tilde{F}_{ij}>0}}{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}_{F_{ij}>0}+\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}_{\tilde{F}_{ij}>0}}, \tag{14}$$

where $\mathbf{1}_X = 1$ if the condition X is fulfilled and $\mathbf{1}_X = 0$ otherwise. Hence $\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}_{F_{ij}>0}$ is the number of edges in the raw network and $\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}_{\tilde{F}_{ij}>0}$ is the number of edges in the predicted network. Alternatively, we propose to check the proportion of correctly predicted existing links (PCPEL),

$$\text{PCPEL}(F,\tilde{F}) = \frac{\sum_{i}^{n}\sum_{j=1}^{n}\mathbf{1}_{F_{ij}>0}\times\mathbf{1}_{\tilde{F}_{ij}>0}}{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}_{F_{ij}>0}}, \tag{15}$$

and the proportion of correctly predicted missing links (PCPML),

$$\text{PCPML}(F,\tilde{F}) = \frac{\sum_{i}^{n}\sum_{j=1}^{n}\mathbf{1}_{F_{ij}=0}\times\mathbf{1}_{\tilde{F}_{ij}=0}}{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}_{F_{ij}=0}}, \tag{16}$$

where "missing links" refers to the links that do not exist in the real transportation network. Hence, the proportion of type-I error (PTIE) is

$$\text{PTIE}(F,\tilde{F}) = 1 - \text{PCPEL}(F,\tilde{F}), \tag{17}$$

which is the proportion of missed prediction for existing links, and the proportion of type-II error (PTIIE) is

$$\text{PTIIE}(F,\tilde{F}) = 1 - \text{PCPML}(F,\tilde{F}), \tag{18}$$

which is the proportion of false alarms for nonexisting links.

The performance of the raw radiation model and the cost-based radiation model in predicting a highway freight transportation network is presented in Table I. First of all, the adoption of driving distance in the cost-based radiation model does not improve the predictive power. The first class of indicators about the network weights in Table I also show that the radiation models with $M = P$ give a larger common part of fluxes (CPF) and smaller errors (NRMSE and NMAE). However, the difference is marginal. The second class of indicators about links give mixed results. The radiation models with $M = P$ result in larger values of CPL and PCPEL and thus give better prediction of existing links. In contrast, models with $M = G$ result in larger values of PCPML and thus give slightly better prediction of nonexisting links. Speaking differently, the models with $M = P$ have a lower proportion of type-I errors but a higher proportion of type-II errors. It shows that the raw radiation model and the cost-based radiation model have comparative predictive power.

B. Scaling behavior of city's influx

In our analysis, the flux matrix $\tilde{F}_{ij}$ is predicted using the population or GDP information and the measured outflux data F_i^{out}. We can obtain the predicted influx of a city:

$$\tilde{F}_i^{\text{in}} = \sum_{j\neq i}\tilde{F}_{ji}. \tag{19}$$

We test if the predicted influxes are able to reproduce the scaling behavior of the city's influx observed in the real transportation network. Specifically, we investigate node's in-strength and out-strength correlation and the power-law dependence between the node's in-strength and node's traits that are widely observed in directed networks [15]. Note that the concept of flux in this work is actually the node strength in the general language of network science.

We illustrate in Fig. 7 the dependence of influx with respect to outflux in log-log scales. This figure shows that the data points locate around the diagonal in each plot and the plots share similar patterns. Cities with high influxes and outfluxes are more concentrated on the diagonal, while cities with low influxes and outfluxes have higher influxes than outfluxes. Wang *et al.* find that a city with higher transportation diversity usually has larger population, higher GDP, larger influx, and larger outflux [34]. It is rational that less developed cities need to buy more freight than to sell freight.

If we regress $\log_{10}(F_i^{\text{in}})$ with respect to $\log_{10}(F_i^{\text{out}})$, we obtain the power-law dependence

$$F_i^{\text{in}} \sim (F_i^{\text{out}})^{\gamma}, \tag{20}$$

where $\gamma^{\text{Data}} = 0.665$ in Fig. 7(a) with the adj-R^2 being 0.776, $\gamma^{P,\text{geo}} = 0.742$ in Fig. 7(b) with the adj-R^2 being 0.672, $\gamma^{P,\text{cost}} = 0.626$ in Fig. 7(c) with the adj-R^2 being 0.641, $\gamma^{G,\text{geo}} = 0.795$ in Fig. 7(d) with the adj-R^2 being 0.687, and $\gamma^{G,\text{cost}} = 0.664$ in Fig. 7(e) with the adj-R^2 being 0.652. This finding is consistent with the results that the transportation diversity scales as a power law with influx and outflux [34]. It indicates that the predicted transportation network using the radiation models can well capture the correlation behavior between a city's influx and outflux.

For the measured transportation network, Fig. 8(a) shows that the influx F_i^{in} scales as a power law with respect to the population P,

$$F_i^{\text{in}} \sim P^{\beta_P}, \tag{21}$$

where the power-law exponent is $\beta_P = 1.08$, and Fig. 8(b) shows that the influx F_i^{in} scales as a power law with respect to

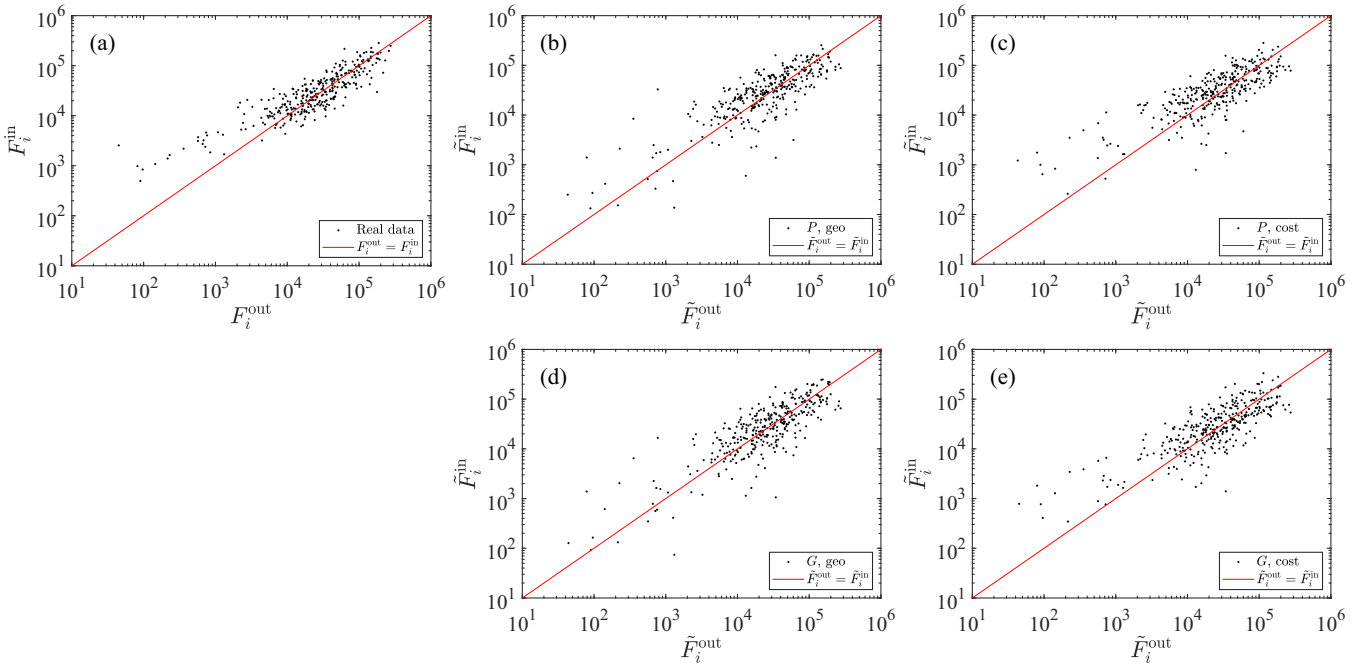

FIG. 7. Correlation between influx and outflux of a city. (a) Real data. (b) Obtained from the raw radiation model with $M = P$ and $d = d^{\rm geo}$. (c) Obtained with the cost-based radiation model with $M = P$ and $d = d^{\rm cost}$. (d) Obtained from the raw radiation model with $M = G$ and $d = d^{\rm geo}$. (e) Obtained with the cost-based radiation model with $M = G$ and $d = d^{\rm cost}$.

the GDP G,

$$F_i^{\rm in} \sim G^{\beta_G}, \tag{22}$$

where the power-law exponent is $\beta_G = 0.84$. The power-law scaling is well reproduced in Figs. 8(c) and 8(d) by the raw radiation model with $M = P$ and in Figs. 8(e) and 8(f) by the cost-based radiation model with $M = P$. Moreover, the raw radiation model with $M = G$ and the cost-based radiation model with $M = G$ can also reproduce this power-law scaling dependence behavior. When we consider the predicted transportation networks, the variable $F_i^{\rm in}$ should be changed to $\tilde{F}_i^{\rm in}$ in Eqs. (21) and (22).

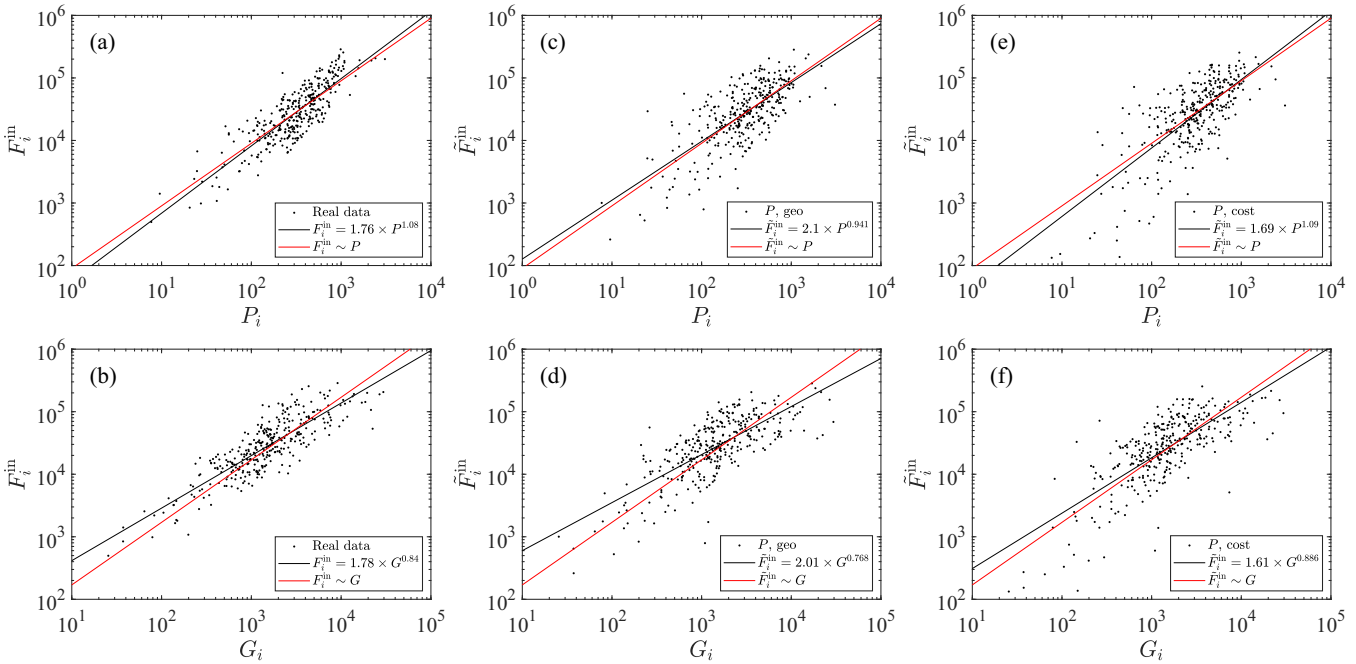

FIG. 8. Comparison of the power-law dependence of influx with respect to population and GDP in the raw and cost-based radiation models. (a, b) Real data from the measured transportation network. (c, d) Predicted influxes from the raw radiation model with $M = P$. (e, f) Predicted influxes from the cost-based radiation model with $M = P$.

TABLE II. Estimated exponents of the power-law dependence.

	Real data	P, d^{geo}	P, d^{cost}	G, d^{geo}	G, d^{cost}
β_G	0.84	0.77	0.89	0.85	0.99
β_P	1.08	0.94	1.09	0.91	1.08
$\kappa\beta_G$	0.94	0.86	0.99	0.95	1.11

Table II presents the power-law exponents β_P and β_G obtained from the measured network and the predicted networks. It is found that

$$\beta_P > \beta_G \tag{23}$$

for all the five networks. Indeed, there is a more accurate relationship between β_P and β_G for each network. It is well documented that there is a power-law dependence between population and GDP for the Chinese cities [1,70]:

$$G \sim P^{\kappa}, \tag{24}$$

where $\kappa = 1.116$ [1]. Combining Eq. (24) with Eqs. (21) and (22), we have

$$\beta_P = \kappa\beta_G. \tag{25}$$

We present the values of $\kappa\beta_G$ in the third row of Table II and compare them with β_P. We find that relation (25) is well validated, especially for the two models with $M = G$.

C. Out-links and in-links of a representative city

Using a representative city, Chengdu, we compare the predicted out-links and in-links with the out-links and in-links in the real transportation network when the corresponding outfluxes and influxes exceed certain levels. The findings seem qualitatively similar for other cities.

Figure 9 illustrates the results for out-links. We show only the results of the raw radiation model and the cost-based radiation model with $M = P$, because using $M = G$ gives similar results. Basically, the models are able to predict out-links with long path lengths when the outflux thresholds are low, as

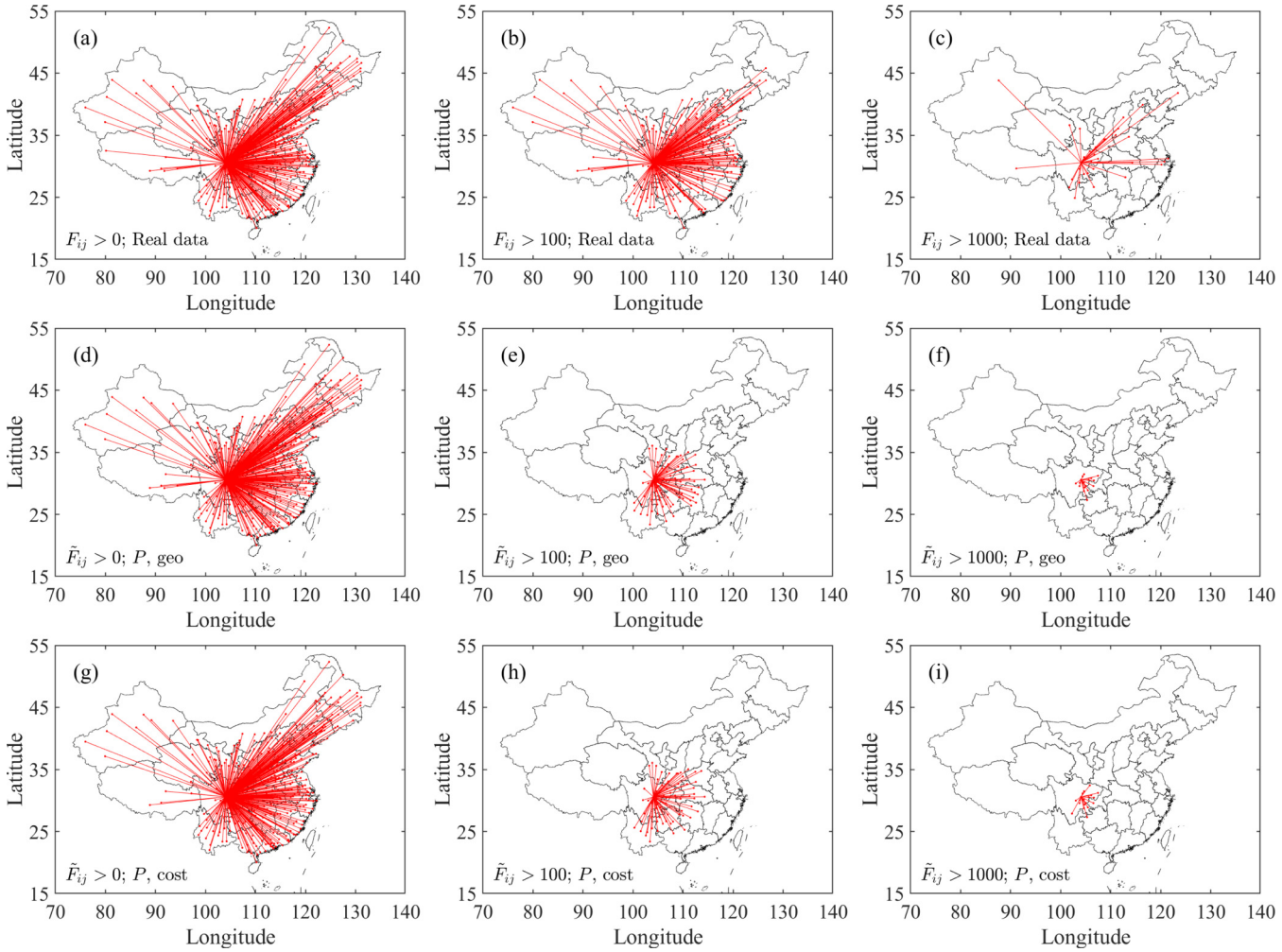

FIG. 9. Comparing the predictions of out-links started from Chengdu for the four radiation models with the real data. The first column shows the out-links with more than one truck from Chengdu ($F_{ij} > 0, \tilde{F}_{ij} > 0$). The second column shows the out-links with more than ten trucks from Chengdu ($F_{ij} > 100, \tilde{F}_{ij} > 100$). The third column shows the out-links with more than one hundred trucks from Chengdu ($F_{ij} > 1000, \tilde{F}_{ij} > 1000$). The panels in different rows display in turn the real data (a, b, c), the predicted out-links from the raw radiation model with $M = P$ (d, e, f), and out-links from the cost-based radiation model with $M = P$ (g, h, i).

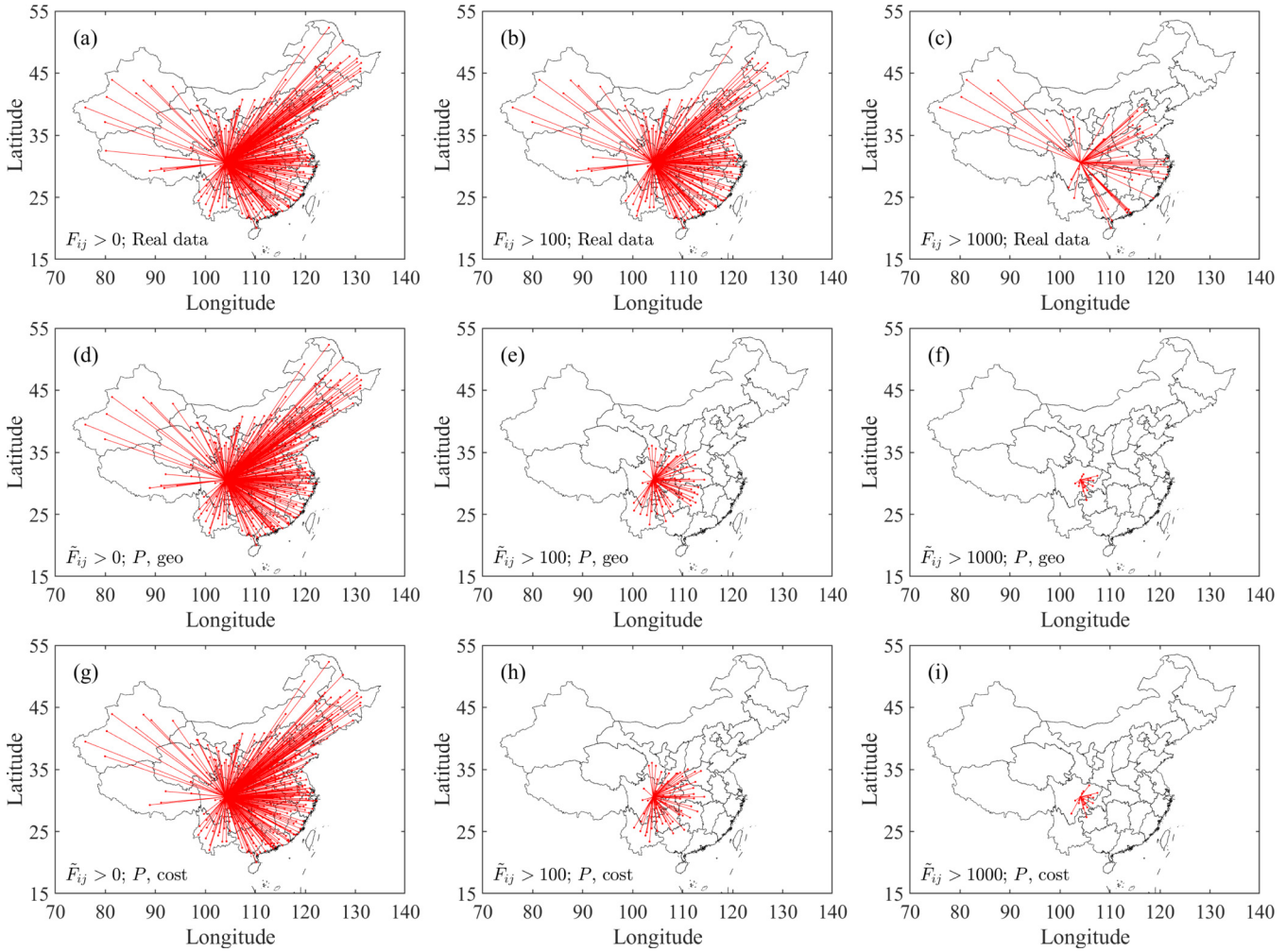

FIG. 10. Comparing the predictions of in-links started from Chengdu for the four radiation models with the real data. The first column shows the in-links with more than one truck from Chengdu ($F_{ij} > 0, \tilde{F}_{ij} > 0$). The second column shows the in-links with more than ten trucks from Chengdu ($F_{ij} > 100, \tilde{F}_{ij} > 100$). The third column shows the in-links with more than one hundred trucks from Chengdu ($F_{ij} > 1000, \tilde{F}_{ij} > 1000$). The panels in different rows display in turn the real data (a, b, c), the predicted out-links from the raw radiation model with $M = P$ (d, e, f), and out-links from the cost-based radiation model with $M = P$ (g, h, i).

shown in the first column of the figure. When we look at the out-links with large outfluxes, for instance, when we require that $F_{ij} > 100$ and $\tilde{F}_{ij} > 100$, the out-links with long path lengths cannot be predicted. Both the raw radiation model and the cost-based radiation model show comparable performance. Although radiation models usually perform better in predicting long-distance links than gravity models [15], their performance of predicting the fluxes is not satisfying. This finding is consistent with the result that the radiation models underpredict the truck fluxes, as shown in Fig. 6.

Figure 10 illustrates the results for in-links. We show only the results of the raw radiation model and the cost-based radiation model with $M = P$, because using $M = G$ gives similar results. Basically, the models are able to predict in-links with long path lengths when the influx thresholds are small, as shown in the first column of the figure. When we look at the in-links with large influxes, for instance, when we require that $F_{ij} > 100$ and $\tilde{F}_{ij} > 100$, the in-links with long path lengths cannot be predicted. Both the raw radiation model and the cost-based radiation model show comparable performance. Although radiation models usually perform better in predicting long-distance links than gravity models [15], their performance of predicting the fluxes is not satisfying. This finding is also consistent with the result that the radiation models underpredict the fluxes, as show in Fig. 6.

Comparing Figs. 9(a) and 10(a), we find that the out-link pattern and the in-link pattern are indistinguishable by eye-balling, especially for those long-distance links. This is rational since the real transportation network is highly interconnected with the network density being 0.8225, while the predicted transportation networks are theoretically complete graphs. Certainly, the predicted fluxes are rounded, that is, when the predicted flux is less than 0.5, we treat it as zero. Therefore the network density is 0.5322 when $M = P$ and $d_{ij} = d_{ij}^{\text{geo}}$, 0.4605 when $M = G$ and $d_{ij} = d_{ij}^{\text{geo}}$, 0.5277 when $M = P$ and $d_{ij} = d_{ij}^{\text{cost}}$, and 0.4556 when $M = G$ and $d_{ij} = d_{ij}^{\text{cost}}$. The predicted networks are still able to correctly predict most of the long-distance links, although their fluxes are underestimated. However, when we filter out low-flux links, most of the long-distance links in the predicted HFT networks

disappear, as shown in Figs. 9(b) and 9(c), and Figs. 10(b) and 10(c).

V. DISCUSSION AND SUMMARY

In this work we applied the raw radiation model and the cost-based radiation model to predict the highway freight transportation network between 338 cities in mainland China. These models are adopted because they are parameter-free and require minimal information, which is widely and publicly available. The predicted network is weighted and directed. We then assessed the predictive performance of these models using metrics based on links and fluxes, which can be accomplished by comparing the predicted HFT networks to the real HFT network.

We found that the radiation models have relatively high accuracy in predicting links but low accuracy in predicting fluxes on links. In particular, the radiation models do not perform well in predicting long-distance links and the fluxes on them, just like the gravity models behave. We note that each trip (link) is not a combination of multiple trips with breaks to deliver or collect goods (not for rest) along the path. Hence the discrepancies between data and predictions of long-distance fluxes are partially caused by the high heterogeneous distribution of different industries and population in mainland China. Nevertheless, the radiation models perform well in reproducing several scaling laws of the HFT network, including the pair between influx and outflux, between the influx and population, and between the influx and GDP.

In our analysis we adopted four parameter-free radiation models: the raw radiation model with population, the raw radiation model with gross domestic product, the cost-based radiation model with population, and the cost-based radiation model with gross domestic product. These models perform similarly in predicting the HFT network. The adoption of population or gross domestic product in the radiation model has a minor impact on the results, and replacing the geographic distance by the driving distance does not improve the results either.

Therefore when we want to predict or reconstruct the highway freight transportation network, there is still much room to improve the performance of the parameter-free radiation models. First of all, GDP or population are calculated based on residents, not on the people living there. However, there are huge population movements between cities, which usually flow out from less developed cities to more developed cities. The population mobility causes the Spring Festival travel rush in China, which also influences the HFT network pattern and causes fluctuations in the network at smaller timescales [1].

There is an important issue about the impact of missing nodes outside the investigated region (here it is mainland China) on the predicted transportation network. In the classical four-step travel demand model, one usually adds virtual nodes at all important infrastructure elements (in this case highways), crossing the boundaries of the investigated region (mainland China) [55,56]. However, in the present case, the so-called total number of departing trucks from city i, F_i^{out}, is actually the sum of F_{ij}, where the j's represent all other Chinese cities. In other words, F_i^{out} does not include those trucks for international transportation. In addition, there is only one border-crossing highway between Kunming (China) to Bangkok (Thailand). Therefore we do not consider virtual nodes in our analysis.

There are other extensions of the radiation model [64,71] and related models [17,69]. However, we do not expect to see much improvement of the results if we apply these models. When applied to the highway freight transportation network, the main missing ingredients in these models are the diversity and heterogeneity of industrial structures. Different cities have different comparative advantages and no city is autarkic. Specific models that integrate these factors into the radiation model would be more suitable and perform better in predicting the links and fluxes of the highway freight transportation network.

ACKNOWLEDGMENTS

This work was partly supported by the Fundamental Research Funds for the Central Universities. We are grateful to Jun-Chao Ma and Zhi-Qiang Jiang for discussions.

[1] L. Wang, J.-C. Ma, Z.-Q. Jiang, W. Yan, and W.-X. Zhou, EPJ Data Sci. **8**, 37 (2019).
[2] A. Wilson, Transp. Res. **1**, 253 (1967).
[3] A. G. Wilson, J. Transp. Econ. Policy **3**, 108 (1969).
[4] A. G. Wilson, Transp. Res. **4**, 1 (1970).
[5] J. Roy and J. Thill, Pap. Reg. Sci. **83**, 339 (2004).
[6] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, Phys. Rep. **734**, 1 (2018).
[7] W. H. K. Lam and H. J. Huang, Transp. Res. B **26**, 275 (1992).
[8] M. Florian and S. Nguyen, Transp. Res. **12**, 241 (1978).
[9] E. Fernandez, J. Decea, M. Florian, and E. Cabrera, Transp. Sci. **28**, 182 (1994).
[10] H. Bar-Gera and D. Boyce, Transp. Res. B **37**, 405 (2003).
[11] L. de Grange, E. Fernandez, and J. de Cea, Transp. Res. E **46**, 61 (2010).
[12] S. A. Stouffer, Am. Sociol. Rev. **5**, 845 (1940).
[13] W. Isard, Q. J. Econ. **68**, 305 (1954).
[14] D. Garlaschelli and M. I. Loffredo, Phys. Rev. Lett. **93**, 188701 (2004).
[15] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, Nature (London) **484**, 96 (2012).
[16] Y. Ren, M. Ercsey-Ravasz, P. Wang, M. C. Gonzalez, and Z. Toroczkai, Nat. Commun. **5**, 5347 (2014).
[17] X.-Y. Yan, W.-X. Wang, Z.-Y. Gao, and Y.-C. Lai, Nat. Commun. **8**, 1639 (2017).
[18] A. W. Evans, Transp. Res. **4**, 19 (1970).
[19] A. W. Evans, Transp. Res. **5**, 15 (1971).
[20] S. P. Evans, Transp. Res. **7**, 39 (1973).
[21] S. P. Evans, Transp. Res. **10**, 37 (1976).
[22] G. Hyman, Environ. Plan. **1**, 105 (1969).
[23] E. G. Ravenstein, J. R. Stat. Soc. **52**, 241 (1889).

[24] D. Karemera, V. I. Oguledo, and B. Davis, Appl. Econ. **32**, 1745 (2000).
[25] M. Beine, S. Bertoli, and J. Fernandez-Huertas Moraga, World Econ. **39**, 496 (2016).
[26] G. Fagiolo and G. Santoni, Appl. Econ. Lett. **23**, 188 (2016).
[27] I. Hong, W.-S. Jung, and H.-H. Jo, PLoS One **14**, e0218028 (2019).
[28] W.-S. Jung, F.-Z. Wang, and H. E. Stanley, Europhys. Lett. **81**, 48005 (2008).
[29] O. Kwon and W.-S. Jung, Physica A **391**, 4261 (2012).
[30] I. Hong and W.-S. Jung, Physica A **462**, 48 (2016).
[31] N. Tu, D. Adiputranto, X. Fu, and Z.-C. Li, Transp. Res. E **117**, 108 (2018).
[32] Q. Zeng, G. W. Y. Wang, C. Qu, and K. X. Li, Transp. Res. E **117**, 96 (2018).
[33] T. V. Nguyen, J. Zhang, L. Zhou, M. Meng, and Y. He, Transp. Res. E **134**, 101816 (2020).
[34] L. Wang, J.-C. Ma, Z.-Q. Jiang, W. Yan, and W.-X. Zhou, arXiv:2011.08468.
[35] J. Tinbergen, *Shaping the World Economy: Suggestions for an International Economic Policy* (Twentieth Century Fund, New York, 1962).
[36] J. E. Anderson, Am. Econ. Rev. **69**, 106 (1979).
[37] M. L. Senior, Prog. Human Geogr. **3**, 175 (1979).
[38] J. H. Bergstrand, Rev. Econ. Stat. **71**, 143 (1989).
[39] J. E. Anderson and E. van Wincoop, Am. Econ. Rev. **93**, 170 (2003).
[40] K. Yamaguchi, Transp. Res. Part E, Logist. Transp. Rev. **44**, 653 (2008).
[41] B. Levine, L. Nozick, and D. Jones, Transp. Res. Part E, Logist. Transp. Rev. **45**, 611 (2009).
[42] G. Fagiolo, J. Econ. Interact. Coord. **5**, 1 (2010).
[43] J. E. Anderson, Ann. Econ. Rev. **3**, 133 (2011).
[44] M. Duenas and G. Fagiolo, J. Econ. Interact. Coord. **8**, 155 (2013).
[45] C. Llano, T. De la Mata, J. Diaz-Lanchas, and N. Gallego, Transp. Res. Part A, Policy Pract. **95**, 334 (2017).
[46] A. Bottasso, M. Conti, P. C. de Sa Porto, C. Ferrari, and A. Tei, Transp. Res. Part A, Policy Pract. **107**, 126 (2018).
[47] A. Brugnoli, A. Dal Bianco, G. Martini, and D. Scotti, Transp. Res. A **112**, 95 (2018).
[48] A. Almog, R. Bird, and D. Garlaschelli, Front. Phys. **7**, 55 (2019).
[49] R. Lambiotte, B. V. D., C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren, Physica A **387**, 5317 (2008).
[50] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel, J. Stat. Mech.: Theory Exp. (2009) L07003.
[51] J.-P. Onnela, S. Arbesman, M. C. González, A.-L. Barabási, and N. A. Christakis, PLoS One **6**, e16939 (2011).
[52] V. Palchykov, M. Mitrovic, H.-H. Jo, J. Saramäki, and R. K. Pan, Sci. Rep. **4**, 6174 (2014).
[53] V. D. Blondel, A. Decuyper, and G. Krings, EPJ Data Sci. **4**, 10 (2015).
[54] M. G. Beiro, L. Bravo, D. Caro, C. Cattuto, L. Ferres, and E. Graells-Garrido, EPJ Data Sci. **7**, 28 (2018).
[55] M. Lenormand, S. Huet, F. Gargiulo, and G. Deffuant, PLoS One **7**, e45985 (2012).
[56] F. Gargiulo, M. Lenormand, S. Huet, and O. B. Espinosa, J. Artif. Soc. Soc. Simul. **15**, 6 (2012).
[57] N. Xia, L. Cheng, S. Chen, X. Wei, W. Zong, and M. Li, J. Transp. Geogr. **72**, 178 (2018).
[58] M. Barthelemy, Phys. Rep. **499**, 1 (2011).
[59] K. Kosmidis, S. Havlin, and A. Bunde, Europhys. Lett. **82**, 48005 (2008).
[60] J. M. Kleinberg, Nature (London) **406**, 845 (2000).
[61] S. M. da Silva, S. D. S. Reis, A. D. Araujo, and J. S. Andrade, Jr., Phys. Rev. E **98**, 032306 (2018).
[62] G. Li, S. D. S. Reis, A. A. Moreira, S. Havlin, H. E. Stanley, and J. S. Andrade, Jr., Phys. Rev. Lett. **104**, 018701 (2010).
[63] G. Li, S. D. S. Reis, A. A. Moreira, S. Havlin, H. E. Stanley, and J. S. Andrade, Jr., Phys. Rev. E **87**, 042810 (2013).
[64] A. P. Masucci, J. Serras, A. Johansson, and M. Batty, Phys. Rev. E **88**, 022812 (2013).
[65] Complete Collection of World Population, http://www.chamiji.com.
[66] Baidu Encyclopedias, https://baike.baidu.com.
[67] M. Lenormand, S. Huet, and F. Gargiulo, J. Transp. Land Use **7**, 43 (2014).
[68] M. Lenormand, A. Bassolas, and J. J. Ramasco, J. Transp. Geogr. **51**, 158 (2016).
[69] E.-J. Liu and X.-Y. Yan, Sci. Rep. **10**, 4657 (2020).
[70] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, and G. B. West, Proc. Natl. Acad. Sci. USA **104**, 7301 (2007).
[71] Y. Yang, C. Herrera, N. Eagle, and M. C. Gonzalez, Sci. Rep. **4**, 5662 (2014).