RedFame

International Journal of Social Science Studies
Vol. 12, No. 6; November 2024
ISSN 2324-8033 E-ISSN 2324-8041
Published by Redfame Publishing
URL: http://ijsss.redfame.com

Confirming the Third Law of Geography: Evidence in China's Spring Festival Travel Rush

Kai Liu[1,2], Pengjun Zhao[2,3], Xiaodong Hai[4]

[1] School of Architecture & Urban Planning, Shenzhen University, Shenzhen, Guangdong, China

[2] Key Laboratory of Earth Surface System and Human-Earth Relations of Ministry of Natural Resources of China, School of Urban Planning & Design, Peking University Shenzhen Graduate School, Shenzhen, Guangdong, China

[3] College of Urban and Environmental Sciences, Peking University, Beijing, China

[4] Smart Steps Digital Technology Co., Ltd, Beijing, China

Correspondence: Pengjun Zhao, School of Urban Planning & Design, Peking University Shenzhen Graduate School, China.

Received: October 23, 2024 Accepted: November 23, 2024 Available online: November 27, 2024

doi:10.11114/ijsss.v12i6.7283 URL: https://doi.org/10.11114/ijsss.v12i6.7283

Abstract

The Third Law of Geography had been proposed by researchers for several years, yet empirical evidence within the field of human geography remained scarce. This study focused on the tidal characteristics of China's Spring Festival Travel Rush (*Chunyun*) across 365 cities in 2019, utilizing day-by-day time-series big data. It confirmed the presence and validity of the law, which posited that the more similar geographic configurations of two cities, the more similar the tidal features of human mobility between these two cities. Results showed that the logistic regression model had acceptable goodness of fit (0.421) and impacts of related indexes (city scale, labor supply, city level, traffic hub features, and dominant flow) had a statistically significant performance. Our research not only bolsters the use of data-driven policy decisions based on large-scale human mobility data but also offers a novel perspective to enhance the study of this law, serving as a pioneering effort to advance previous work.

Keywords: Geographic similarity; Logistic regression method; The Third Law of Geography

1. Introduction

From the perspectives of broadness, independence and applicability, researchers summarized the existing three Laws of Geography as three principles: spatial autocorrelation (Tobler, 2004), heterogeneity (Goodchild, 2004) and similarity (Zhu et al., 2018). Tobler (2004) stated that spatial variation of geographic features follows a general principle of 'near things are more related than distant things', which was named as the First Law of Geography. Goodchild (2004) mentioned that the spatial variation of geographic attributes displays 'uncontrolled variance' and noticed that this principle (spatial heterogeneity) could be named as the Second Law of Geography. Zhu et al. (2018) mentioned that the similarity in geographic configuration between two locations could be linked with the existing guiding principles in spatial prediction and named this geographic similarity principle as the Third Law of Geography. Following that study, the Third Law of Geography was herein defined as 'the more similar geographic configurations of two points (areas), the more similar the values (processes) of the target variable at these two points (areas)'.

Some scholars, such as Brunsdon et al. (1996), Li (2007), Wang & Wu (2011), Sun et al. (2012), Zhu et al. (2018), and Zhu et al. (2020), were devoted to theoretical improvement and had sup-ported the validity of three Laws of Geography. The Third Law of Geography has been proposed by researchers for several years, yet empirical evidence within the field of human geography remains scarce, owing to its inadequate application of social science (cf. Zhu et al., 2018) and philosophical criticism (e.g., metaphysics, exceptionalism) (cf. Andrew 1985 & Sun et al. 2012).

Efforts have been made in providing empirical evidence on the validity of the Third Law of Geography (e.g., Zhang & Zhu, 2019; Zhu et al., 2019; Ma et al., 2020; Zhu et al., 2022; & Zhao et al., 2023). For instance, Zhu & Turner (2022) explained that the differences be-tween the Third Law of Geography from the first two were displayed from three aspects, i.e., principle expressed, form of expression, and role of geographic samples. Zhang & Zhu (2019) performed a representativeness-directed approach to mitigate spatial bias for the predictive mapping of geographic phenomena.

Furthermore, issues of landslide susceptibility detection, soil sampling, and groundwater level monitoring could also be connected to the Third Law of Geography (e.g., Zhu et al., 2019, Ma et al., 2020, and Zhao et al., 2023).

Those studies had devoted the promotion of applied research related to the Third Law of Geography. However, these efforts so far have been on the physical geography side. In human geography side, although some applied research focusing on the pattern of intraurban similarity/heterogeneity had touched on this issue and obtained some forward-looking results in terms of spatial prediction (Zhu et al., 2020), feature area recognition (Veneri, 2013), and human-land relationship interpretation (Liu et al., 2012 and Liu et al., 2021), these findings only rest on the intraurban analysis within a limited area (intraurban). Moreover, promoting the application and understanding of the Third Law of Geography at a wider scale and its proof of universality on intercity relations were still a blank-filling work.

Intercity human mobility has multidirectional tidal features (Tan et al., 2021). As a long-term, comprehensive, and complex interaction phenomenon, intercity human mobility re-shaped the patterns of a national-scale urban system, accompanying a nationwide population redistribution (Wei et al., 2018; Liu et al., 2022) and spatiotemporal interaction of intercity materials (Tao & Thill, 2019). The arrival of the big data era required the development of a unified human-land relation system of multiple dimension-scale areas for more data-intensive scientific discoveries (Sun et al., 2012). Furthermore, with the development of geographic information system (GIS) and data mining techniques in the field of computational geography, analytical dimensions of geospatial science should be extended to more comprehensively under interdisciplinary frameworks with the use of big data (Wang & Wu, 2011; Chai & Zhang, 2020).

Previous studies investigated the miscellaneous geographical phenomena within the space of flow by using spatial interactive models (including multiple formulas) (Murayama, 2013). However, definitely speaking, those contribution were based on the context of the First and Second Law of Geography (i.e., spatial proximity and spatial heterogeneity). Few under-standings were on the issue of the similarity recognition between dynamic flow patterns and geographic features within the context of the Third Law of Geography, regarded as one of the motivations of this study. Furthermore, grasping the patterns of dominant flow and the role of traffic hub across the national urban system was the requirement of population-development oriented traffic-strengthening strategy (Wang, 2010; Zhao et al., 2020). Thus, dataset with a high degree of GIS-enable aggregation and spatiotemporal fineness is needed. China's Spring Festival Travel Rush (*Chunyun*), as the world's largest people migrations (Liu et al., 2022), can provide typical sufficient materials in this context.

Against these above-mentioned issues, this study designed a self-constructed indica-tor-based assessment and exploratory analytical framework and aimed to address the hypothesis: the more similar geographic configurations of two cities, the more similar the tidal features of human mobility between these two cities. Specifically, this is a typical topic of binary logistic regression analysis: whether the tidal features of human mobility at two cities were similar or not depended on their geographic configurations (such as labor supply, scale, level, traffic hub role, and flow pattern).

We also proposed another hypothesis that the reason why this law has not been ap-plied or related to social science studies was owing to the "exceptionalism" due to the challenges in defining the "geographic similarity" and "geographical configurations" in social science (Andrew 1985 and Zhu et al. 2018). As the first efforts in this realm, the authors argue that performing a carefully designed binary logistic regression model on this issue and observing the specific quantitative relationships between the dependent variable (i.e., geographic similarity between two cities) and independent variable (i.e., geographical configurations) could bring new insights in this context.

The purpose of this study is to discover the tidal features of *Chunyun* mobility across China's 365 cities in 2019 based on a day-by-day time-series big data, and then to explore the relationship with multiple hidden relevant variables and finally to confirm the existence of the Third Law of Geography. Two research questions were addressed in this study: 1) how is the tidal features of *Chunyun* mobility across China's 365 cities and is there any geographic similarity between them? 2) how can we confirm the Third Law of Geography and provide empirical evidence in support of this law in social science?

2. Materials and Methods

2.1 Empirical Dataset

Figure 1 illustrated the spatial distribution, grade classification, urban planning blueprint within the layout of China's master plan, and geographical division of 365 cities across the mainland of China. The 14th Five-Year Plan is the guiding document of China's urban development (The State Council of the People's Republic of China 2021). Following the previous studies (Wang, 2010; Murayama, 1982), this study would introduce a new parameter in the sense of urban planning (cf. the next section), to quantify the patterns of dominant flow with the consideration of differentiating the external and internal influence of 19 urban agglomerations (UAs) (cf. Figure 1a).

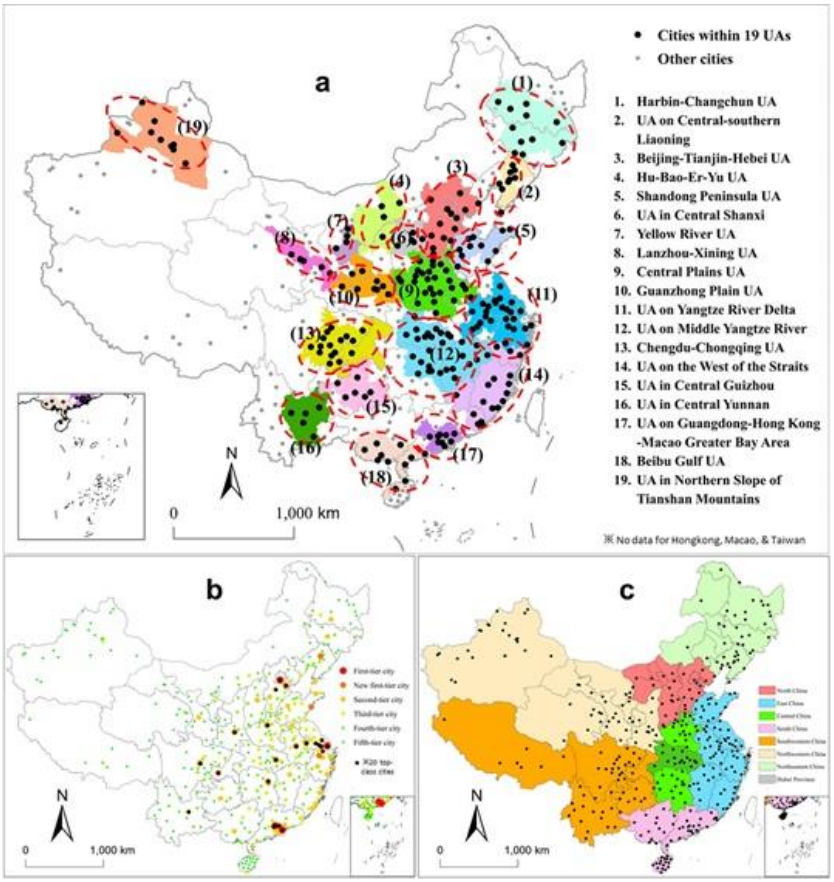

Figure 1. 365 cities across the mainland of China. (a) 19 major urban agglomeration designed by the China's master plan. (b) Grade classification of China's 365 cities. (c) Geographical division of China's 365 cities.

This study selected the city scale (i.e., $\sqrt{(GDP \times POP)}$, labor supply, city level, traffic hub role, and dominant flow feature as the independent variables correlated with the human mobility patterns (dependent variable) before the construction of logistic regression model. Table 1 summarized the descriptions for each parameter. A GIS-enabled environment was available for our in-depth analyses as shown in Figure 1.

Table 1. Data description of independent variable

independent variable	remark	description	source
city scale	G	$\sqrt{\mathrm{GDP} * \mathrm{POP}}$	National Bureau of Statistics of China (2021); China City Statistical Yearbook (2020)
labor supply	K	Proportion of the total number of rural migrant workers (RMWs) to the permanent population (National Bureau of Statistics of China 2021)	Obtained from mobile signaling data
city level	L1	Dummy variable; 1 for a super-tier city, otherwise 0	Research Institute of New First-tier City (2020); Liu et al., (2022)
	LN1	Dummy variable; 1 for a new-tier-1 city, otherwise 0	
	L2	Dummy variable; 1 for a tier-2 city, otherwise 0	
	L3	Dummy variable; 1 for a tier-3 city, otherwise 0	
	L4	Dummy variable; 1 for a tier-4 city, otherwise 0	
	L5	Dummy variable; 1 for a tier-5 city, otherwise 0	

traffic hub role	H1	Dummy variable; 1 for a large-scale integrated transport hub city, otherwise 0	Zhao et al., (2020)
	H2	Dummy variable; 1 for a national transport hub city, otherwise 0	
	H3	Dummy variable; 1 for a local transport hub city, otherwise 0	
dominant flow index	D1	Defined as the formula: $\sum_{r\in[1,10]}^{UA_\Omega}[(11-r)*d_{i\to j_\Omega}]$, where city *i* belongs to any of 19 UAs shown in Figure 1a. Ω is the city code vector of this UA. r indicated the rank number of top-10 target cities that have a higher pointing probability of coming from city *i* (top-10 dominant flows). $d_{i\to j}$ means the pointing probability from city *i* to city *j*, calculated from the OD matrix. j_Ω is the code of top-10-ranked linkage city tied to city i; in addition, city j_Ω must belong to the same UA. If city i does not belong to any of 19 UAs, Ω is the city code vector of the nearest UA from city i.	Calculated by the authors from the OD data.
	D2	Defined as the formula: $\sum_{r\in[1,10]}^{UA_{j\notin\Omega}}[(11-r)*d_{i\to j_{\neg\Omega}}]$, where city *i* belongs to any of 19 UAs shown in Figure 1a. In this case, the constraint conditions of D2 were basically same as D1, but top-10-ranked city *j* does not belong to UA Ω. If city *i* does not belong to any of 19 UAs, Ω is the city code vector of the nearest UA from city *i*.	

A good grade classification can not only simplify the difficulty of identifying the patterns of the complex urban network (Tan et al. 2021), but also provide tracing clues for investigating large-scale population migration (Liu et al., 2022). The city-level indicator we used can describe a kind of comprehensive strength that considered business resource aggregation, urban resident vitality, traffic accessibility, lifestyle diversity, and future plasticity (Research Institute of New First-tier City, 2020), which was the newest standard with reliability and scientific value (Tan et al., 2021; Liu et al., 2022). This newest grade classification of China's 365 cities was provided by the Research Institute of New First-Tier City, Shanghai Media Group.

As the independent variables, three kinds of the traffic hub roles were considered to join in our model (cf. Table 1). Based on abundant yearbook data and Zhao et al. (2020)'s judgement, 13 large-scale global-national integrated transport hub cities, 21 national transport hub cities, and 49 local transport hub cities were recognized (cf. Figure 2).

Appendix 1 provided the descriptive statistics of independent variables (365 city samples) shown in Table 1 (cf. Table A1). In particular, this set of independent variables mentioned here corresponds to the mathematical mapping onto a single city, differing from the independent variables (geographic similarity between a two-city-pair) that would be input into the logistic regression model (cf. Section 2.4).

To measure each city's dominant flow patterns from the UA's external and UA's internal aspect, a self-constructed indicator assessment was originally built (cf. Table 1). Our formula was designed as a rank-size weighted index; moreover, UA's boundary effect was fully considered. Following previous studies (Murayama, 1983; Murayama, 2000; & Wang, 2010), the basic of dominant flow theory was applicative in this case and considering top-10-ranked dominant flows was enough for correctly simplifying the nodal dominant-dominated patterns, as mentioned in the bottom two lines in Table 1.

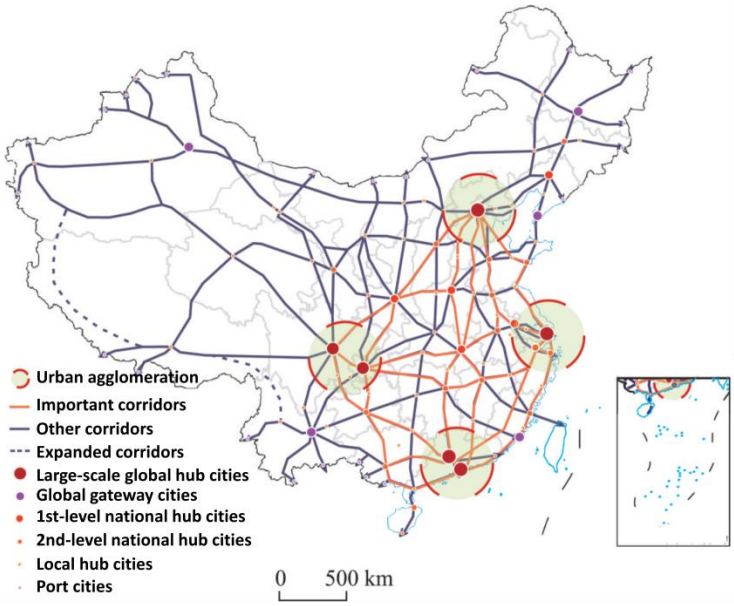

Figure 2. Layout of China's national transportation system

Numbers of rural migrant workers by hometown was obtained from mobile signaling data (MSD). MSD used for this study was provided by a certain telecom operator in China with 445 million users (cf. Liu et al., 2022). When users are making phone calls, sending messages, browsing websites, or switching on/off their devices, mass real-time base-station-dependent location information are recorded, updated, and interchanged between users' mobile phone and mobile base station. To enhance the performance of the data extrapolation, geographical weight calculated from the real county-level census data, user coverage, market share with other operators, as well as a variety of parameters extracted from the data structure of users' age, gender, etc., were aggregated and handled using a machine learning approach by the operator. Under the premise of a definite total amount, the data production team made estimated number of mobility data from the whole network closer to the actual value and helped to extrapolate the data to all users of the entire network, based on these mass samples and their corresponding regional census-based population weights. Figure 3 explained the detailed information of performing the MSD pre-processing, and generating the mobility data, recognizing geo-tagged user portrait information (age, gender, tag of rural migrant worker, etc.) by a series of technical methods.

To explore the tidal features of 365 cities' human mobility during the *Chunyun* cycles, this study used the day-by-day origin-destination (OD) data between China's 365 cities covering 50 days before/after the Spring Festival Travel Rush of 2019, which was of high quality, reliability and fineness (Geographical Society of China 2021). A series of exploratory statistical analysis would be performed (see the following sections). *Chunyun* (China's Spring Festival Travel Rush) mobility data was created as a time-series city-level national-scale origin-destination (OD) matrices between China's 365 cities. Based on the authors' calculation, the fitting accuracy of total population movement can reach about 80%, between the MSD and published official figure by the Ministry of Transport of the People's Republic of China (Geographical Society of China, 2021). Particularly, data of 167 million rural migrant workers' hometown (cross-city) divided by 365 cities across China was also available for this study. As an important variable measuring cross-domain labor supply level, this study employed and calculated the city-level proportion of the total number of rural migrant workers (RMWs) to the permanent population from census data (National Bureau of Statistics of China 2021). This figure had a coincidence rate of 98.98% with the official figure (Akskci Coperation, 2021), which implied that the accuracy of data extrapolation method was acceptable in this context.

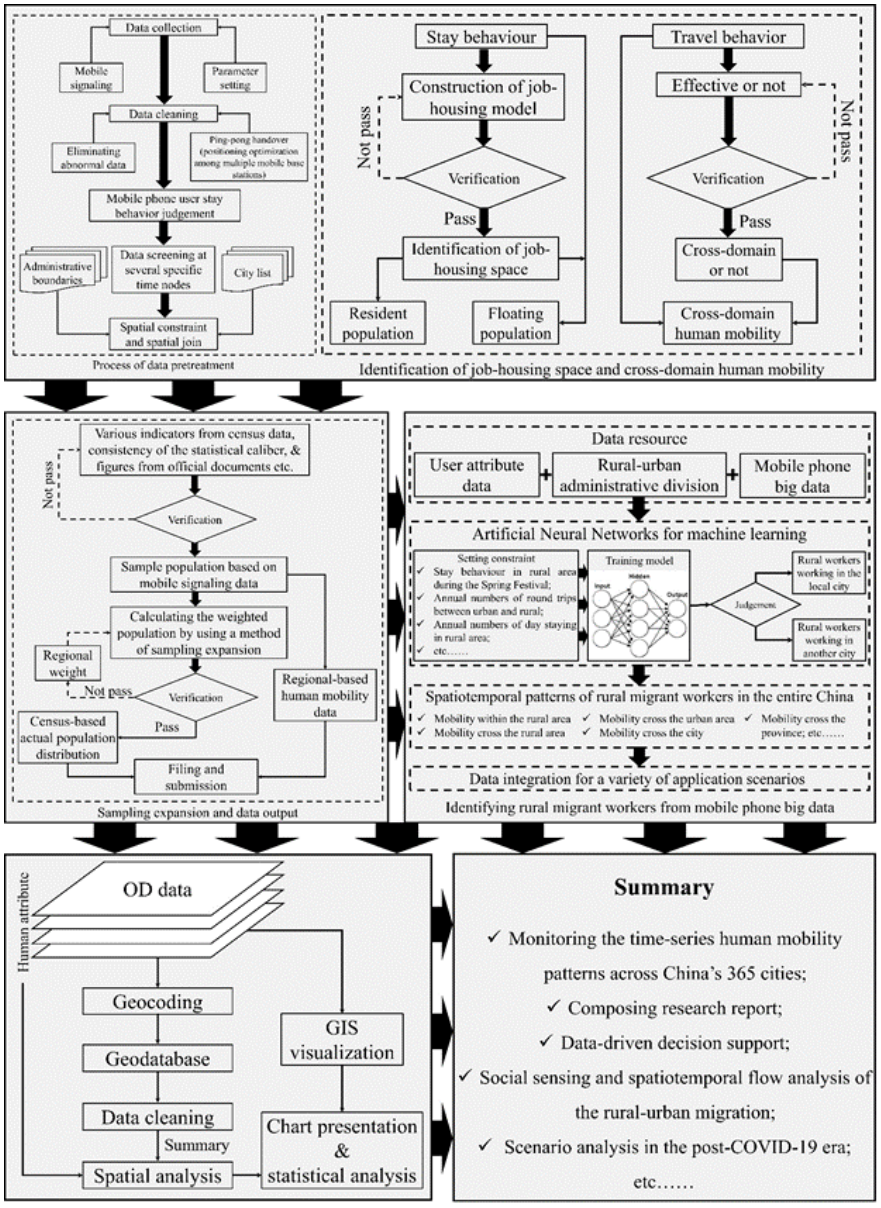

Figure 3. Framework of data processing and production.

The method of tagging a user as a rural migrant worker from MSD was explained (cf. Figure 3). First, some specific rules for identifying the RMWs were set before the machine learning process. The typical samples which have a specific kind of features and constraints, such as staying behavior and periodic rural-urban migration before/after the Spring Festival, lower income, difference of hometown and workplace were carefully trained (cf. Figure 3). Additionally, a college student is not regarded as a RMW but must be firstly excluded before the identification of RMWs began, as one of the premise constraints. Second, a neural network model was used to training the fittest model to judge and identify a RMW which better meet the following conditions: such as staying in his/her hometown during the Spring Festival but working in another city, specific user attributes, typical urban-rural migration pattern, etc. Finally, the data producer (a third party) employed a method involving spatiotemporal constraints and human behaviour identification and a range of machine learning and data mining technology to handle the raw data (including a desensitization and encryption process), while setting specific rules to distinguish the RMWs from the total population. Besides the patent of the method of identifying RMWs using mobile signaling big data, the data producer had also used the total population extrapolation algorithm, spatiotemporal human mobility and labor migration models, and the return-to-city/return-to-work model to derive the data available for this study.

Overall, the variables selected in this study were already rich enough for descripting a national urban system displayed from the *Chunyun* mobility, which was regarded as a comprehensive collection of cities which are interdependent through economic fluctuations, diffusion and exchange of population flow (Murayama, 2000; Liu & Dauda, 2020),

while too many variables might bring terrible problems of collinearity and chronological dislocation (e.g., yearbook data with different published time) between variables (Liu & Dauda, 2020).

2.2 Analytical Framework

Figure 4 showed the analytical framework as a guideline for understanding the following processes of cluster analysis and logistic regression analysis. The day-by-day time-series OD data were reconstructed into 5 matrixes, corresponding to 5 *Chunyun* cycles (cf. the middle part of Figure 4). The *Chunyun* (40 days) starts on the 16th day of the 12th Lunar Month and ends on the 25th day of the 1st Lunar Month. Another 10-day cycle before the *Chunyun*'s start was chosen as an initial state of the tidal feature.

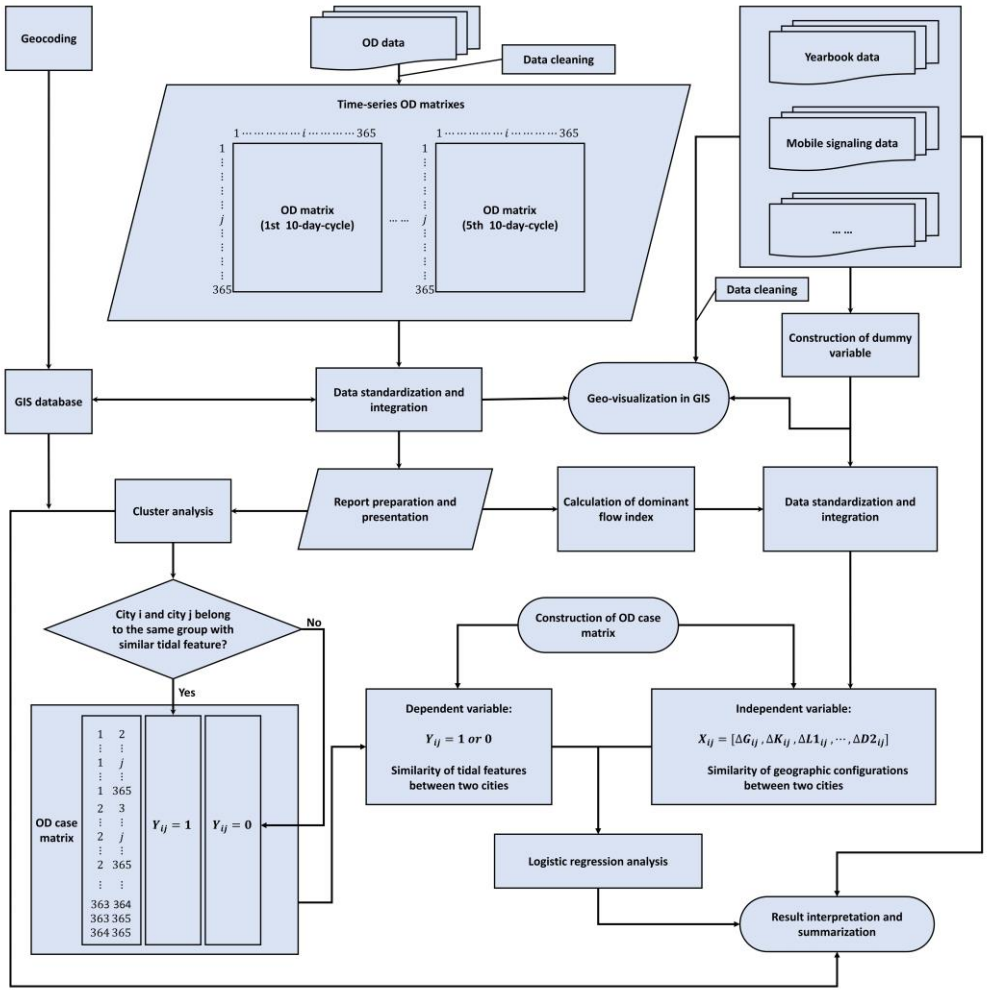

Figure 4. Analytical framework

The reason why we divided 40-day *Chunyun* timeline into four 10-day segments would be explained. We would try to reveal the tidal feature of human mobility arising in a city by investigating 5 sub-cycles during the *Chunyun* timeline. Considering that when dividing *Chunyun* timeline to 4 10-days cycles the cosine similarity measurement between the time-series trend line has the highest between-group variation (based on the authors' calculation). Furthermore, when dividing 40 days into 4 sub-cycles, the data compatibility (time-axis disruption and completeness of one sub-cycle) tended to be better comparing with other division methods, i.e., dividing by 7, 8, or 9 days, because that 10 is divisible by 40 and the dislocation problem between Lunar calendar and Gregorian calendar could be weakened.

As mentioned in Section 1, validating the Third Law of Geography in the case of *Chunyun* is a typical topic of binary logistic regression analysis: whether the tidal features of human mobility at two cities were similar or not is the dependent variable. It depended on their geographic configurations (such as labor supply, scale, level, traffic hub role, and flow pattern), which were set as independent variables (cf. Figure 4). More detailed information would be touched in Section 2.4.

On the other hand, 13 independent variables were calculated by 365 cities based on the raw data (cf. the right part of

Figure 4). All data were geocoded, standardized, and in-putted to a unified GIS database. Finally, the cluster analysis and logistic regression analysis were performed.

2.3 Analytical Framework

This study used a 1×5 vector to record the tidal feature for a better understanding of the time-series broken line (cf. Figure 6 in Section 3.1) for the cycle status of each city. The Z-score ($(x-\mu)/\sigma$) of the net amount of flow (inflow - outflow) for each *Chunyun* cycle was calculated, because the authors found that focusing on the net flow amount can better re-veal how the rules of festival periodicity of intercity human mobility change. Then, we used the method of Jenks Natural Breaks Classification to obtain 5 groups: 5 for recording the status of large-scale outflow, 4 for recording the status of middle-scale outflow, 3 for recording the relatively balanced situation, 2 for recording the status of middle-scale in-flow, and 1 for recording the status of large-scale inflow (cf. Table 2). Finally, cluster analysis was performed to reveal the hidden characteristics of *Chunyun* mobility's tidal feature. Based on the Elbow Principle and dendrogram of the clustering process shown in Figure 4, this study found that the tidal features of spatiotemporal interaction in terms of *Chunyun* human mobility displayed 7 kinds of typical patterns in the pre-pandemic scenario (2019), mainly because the information loss is larger (-0.68) when 7 clusters were recognized. Be-sides, the figure of the Cophenetic Correlation Coefficient reached 0.9242 (based on the authors' calculation), which meant that the clustering result had a high level of statistical significance. We found that the tidal features of *Chunyun* mobility arising in 365 cities dis-played 7 kinds of typical patterns, as summarized in Table 2.

Table 2. Tidal features of *Chunyun* mobility pattern arising in 365 cities.

Remark	Periodicity type	Count	Tidal feature
Type_1	302030403	17	medium-scale inflow in the early and middle stage; medium-scale outflow in the later stages
	302030404	13	
	303020403	3	
	202020404	2	
	202030404	1	
Type_2	304040202	14	medium-scale outflow in the early and middle stages; medium-scale inflow in the later stages
	304030202	6	
	403040202	1	
	304040102	1	
	304030203	1	
	303040203	1	
	303030202	1	
	304030303	2	
	303040303	1	
	303030302	3	
	303030203	1	
Type_3	302020404	14	medium-scale inflow in the early and middle stages; medium-scale outflow in the later stages; but with a relatively larger drop
	302020403	7	
Type_4	303030403	12	small-scale inflow in the early and middle stages; small-scale outflow in the later stages; but with a relatively larger drop
	303030304	3	
Type_5	302030303	5	small-scale inflow in the early and middle stages; small-scale outflow in the later stages
	303020303	3	
Type_6	405050101	4	large-scale outflow in the early and middle stages; large-scale inflow in the later stages
	404040101	2	
	404040102	4	
	405040102	3	
	404050102	1	
Type_7	303030303	239	Weak fluctuations in all stages

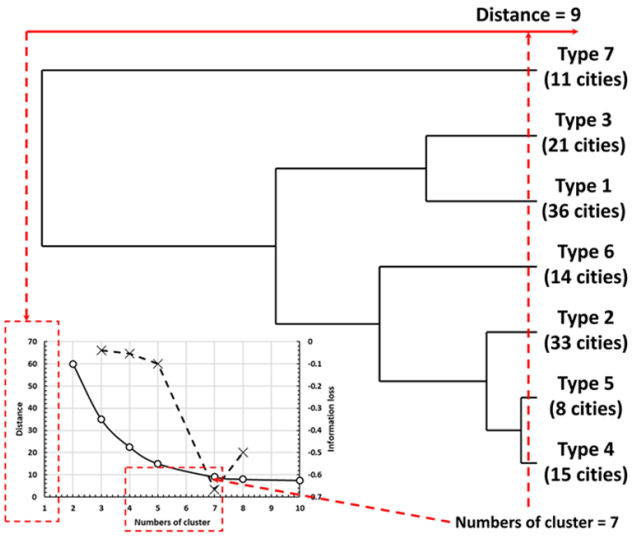

Figure 5. Clustering process. Note: the process of type 7's agglomeration was not fully shown because of its untypicality, only high-ranked 138 cities (not every 365 cities) were picked up for the clustering model (Top 90% threshold).

2.4 Logistic Regression Analysis

The logistic regression model was defined as the following formula:

$$Z = logit(y) = \beta_0 + \beta_1 \Delta G_{ij} + \beta_2 \Delta K_{ij} + \beta_3 \Delta LN1_{ij} + \beta_4 \Delta L2_{ij} + \beta_5 \Delta L3_{ij} + \beta_6 \Delta L4_{ij} + \beta_7 \Delta L5_{ij} + \beta_8 \Delta H2_{ij} + \beta_9 \Delta H3_{ij} + \beta_{10} \Delta D1_{ij} + \beta_{11} \Delta D2_{ij},$$

$$y = 1/(1 + e^{-Z}),$$

$$\Delta X_{ij} = 1 - |x_i - x_j| \ \wedge \ i \neq j$$

where the β. meant regression coefficient. ΔX_{ij} indicated the geographic similarity between city i and city j. Z was the transformation expression of logistic regression.

Descriptive statistics of all variables used in the model (n=66430, i.e., 365×364÷2) could be addressed in Table A2 (cf. Appendix 2). We have normalized all independent variables by using the Min-Max Normalization Method.

The dependent variable (y) meant whether the tidal features of human mobility in two cities were similar or not. 1 indicated that tidal features of *Chunyun* mobility arising in city i and city j were grouped into the same type after the cluster analysis, and otherwise 0. The set of independent variables was a 1 × 12 vector: [1, ΔG_{ij}, ΔK_{ij}, $\Delta LN1_{ij}$, $\Delta L2_{ij}$, $\Delta L3_{ij}$, $\Delta L4_{ij}$, $\Delta L5_{ij}$, $\Delta H2_{ij}$, $\Delta H3_{ij}$, $\Delta D1_{ij}$, $\Delta D2_{ij}$]. The Δ here meant a process of subtraction operation, i.e., the similarity of geographic configurations (such as labor supply, scale, level, traffic hub role, and flow pattern) between city i and city j.

We conducted the collinearity diagnosis process by using SPSS software. If putting variables L1 and H1 into the logistic regression model, we might bring the collinearity problem (based on their values of tolerance and variance inflation factor), that is to say, retaining the variable G (city scale) was enough. Retaining the variables L1 and H1 would make their variance inflation factors (VIF) close to threshold 5, which were 4.137 and 5.824, respectively. After removing these two variables, all VIFs could be below 2.2 (cf. Table 3), which meant that the collinearity problem could be ignored at this moment. Therefore, these two variables were rejected from the model.

3. Result

3.1 The Tidal Features of Chunyun Mobility Across China's 365 Cities

Figure 6 showed the spatial distribution of 365 cities with 7 types of tidal features (cf. Figure 6a) and their corresponding time-series broken line along 5 cycles during 2019 *Chunyun* (cf. Figure 6b-h). The description of tidal waves with different types had been explained in Table 2. Type 1 (pink dots) was distributed in the gateway location undertaking the overlap-ping area from local cities to the periphery of the super-tier cities (core cities). Type 2 (blue dots) was mostly distributed in the provincial capital city, regional sub-central city, and satellite cities surrounding the core city of three major UAs: Beijing-Tianjin-Hebei, Yangtze River Delta, and Greater Bay Area as shown in Figure 1a. Their location, planning role, and traffic hub role made them had a transit nodal function for gathering intra-provincial

flows and projecting extra-provincial flows (such as trans-provincial labor supply).

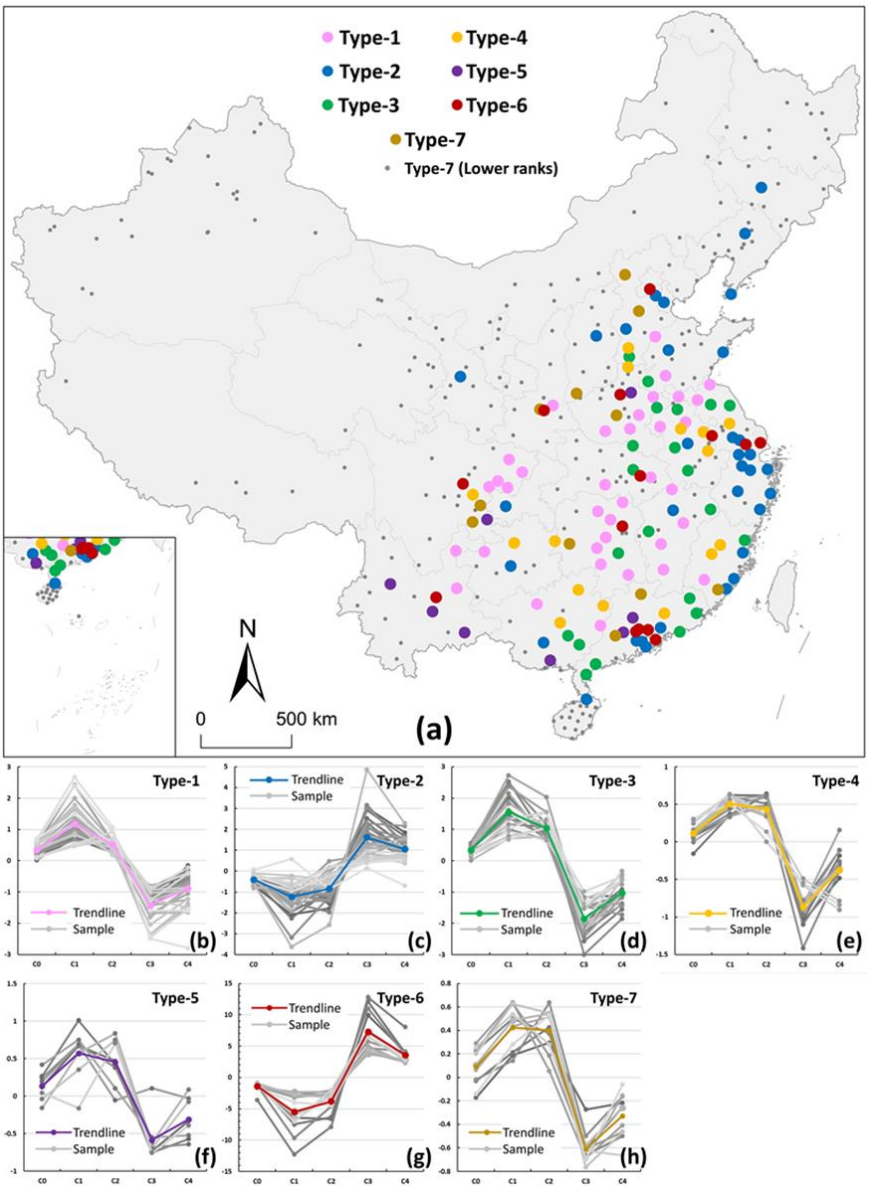

Figure 6. Results of tidal features of 2019 Chunyun mobility across China's 365 cities. (a) Spatial distribution of the types of tidal features. (b-h) 7 types of tidal wave for 2019 *Chunyun* mobility. Note: only top-138-ranked cities' tidal broken line were shown (Top 90% threshold)

Cities belongs to Type 3 were mostly located at the highway or railway cross of regional corridor (cf. Figure 6 and Figure 2). Cities belongs to Types 4 and 5 were usually centers for labor export (with higher K value), while cities of type 5 had a relatively inconvenient condition of transportation. Type 6, covering several super-tier cities and new first-tier cities (cf. Fig-ure 1b) can be read, such as Beijing, Shanghai, Guangzhou, Shenzhen, Suzhou, Zhengzhou, Chengdu, and Changsha. Its tidal waves look similar with the ones of Type 2 but accompanied by intense changes in the variable of net flow (cf. Figures 6c & 6g). The rest cities, with weak fluctuations in all stages of their tidal waves, were grouped in the Type 7.

In one word, the tidal features of *Chunyun* mobility across China's 365 cities had 7 typical kinds of type, and their spatial proximity and heterogeneity were mixed and complex but displayed a certain degree of regional similarity/differentiation. To a certain ex-tent, all the above-mentioned come from the First and Second Laws of Geography, because spatial transmission effects can be perceived by data.

3.2 Corroboration of the Third Law of Geography

This study used the stepwise method to fit the regression coefficients (β) of inde-pendent variables (cf. Section 2.4). As shown in the lower-left portion of Figure 4, we de-fined the dependent variable (using dummy variable) as the similar mobility pattern be-tween two cities, that is, whether the tidal features of human mobility in two cities were similar or not, as mentioned in Section 2.4. The set of independent variables were regarded as the similarity of geographic configurations (such as labor supply, scale, level, traffic hub role, and flow pattern) between city i and city j, which were calculated by subtraction operation, as noted in the right-lower portion of Figure 4. By investigating the results obtained from this designed binary logistic regression model (cf. Section 2.4), we could evaluate specific quantitative relationships between the dependent variable (i.e., geographic similarity) and independent variable (i.e., geographical configurations). In this sense, the proposition of confirming the Third law of Geography can be interpreted using a binary logistic regression model.

Results were summarized in Table 3. Our model had acceptable goodness of fit (Nagelkerke R2 = 0.421) and impacts of 10 related indexes (labor supply, scale, level, traffic hub features of a city, and dominant flow) had a statistically significant performance. The city scale has the highest coefficient, which implies that GDP and population were the two most important factors that affected the tidal feature of Chunyun mobility. The second higher index is the coefficient of K, i.e., the proportion of the total number of RMWs to the permanent population. That is to say, the more power of trans-provincial labor supply of a city, the more likely that it might be classified as Type 4 (cf. Section 3.1). Indexes of city level and traffic hub role had positive values, except for the L5 and H3. This implied that local transport hub cities or tie-5 cities had mismatched tidal features between the predictive value and true value, owing to their relatively lagging development level. The dominant flow index of UA's internal influence had a positive impact (0.3), while the dominant flow index of UA's external influence has a negative impact. This showed that linkages between different UAs cannot significantly affect the tidal patterns' similarity.

Table 3. Results of the logistic regression model (n=66430). Note: Dependent variable (i.e., geographic similarity between two cities) and independent variable (i.e., geographical configurations)

Model summary	Nagelkerke R^2		Odds ratio	Y=0: 75.5% Y=1: 75.1%
	0.421			75.1%
Independent variable	VIF	Coefficients	Standard Error	Significance
b_0		-14.203	0.164	0.000
G	1.338	8.418	0.053	0.000
K	1.012	3.934	0.187	0.000
LN1	1.560	0.407	0.069	0.000
L2	1.236	1.160	0.035	0.000
L3	1.079	0.987	0.051	0.000
L4	1.060	0.212	0.020	0.000
L5	1.059	-0.194	0.021	0.000
H2	1.388	2.104	0.062	0.000
H3	1.052	3.039	0.023	*0.081*
D1	2.107	0.300	0.072	0.000
D2	2.130	-1.245	0.066	0.000

In addition, scale limitation effect and geographic proximity effect in the geographic environment were always connected with collinearity, endogeneity, and uncertainty among variables (such as space distance, flow transitivity, and inhabitant factors), which might be the major obstacles to conduct confirmatory research on the Third Law of Geography in this context. Based on the result shown in Table 3, the authors could herein confirm that the reason why this law has not been applied or related to social science studies was owing to the "exceptionalism" due to the challenges in defining the "geographic similarity" and "geographical configurations" in social science (Andrew, 1985; Zhu et al., 2018), as noted in Section 1. Even regional socio-economic characteristics that are strongly affected by human subjective initiative might be possible to be tied with an "objective" law. When the explanatory variables were gradually added on the premise of eliminating the influence of collinearity, the explanatory power of the logistic regression model might be improved, i.e., R2 raised from 0.115 to 0.421, based on the authors' calculation.

4. Discussion

In the aspect of methodology, unlike Shen (2004), Zhu et al. (2018), Huang et al. (2020)., Zhu et al. (2020), Mu et al. (2021), and Zhao et al. (2022), that focused on model examination or spatial prediction, this study designed a

self-constructed indicator-based assessment and exploratory analytical framework. Some scholars had mentioned that regional socio-economic characteristics and inter-regional interactions are mutually associated, and they are basically in a heterogeneous but analogous relationship (Murayama, 2000; Liu & Dauda, 2020). Unfortunately, they did not touch the proof of this law. Our findings can provide new thinking about the application of the logistic regression model and exploratory analysis.

From the perspective of urban planning significance, our findings fit the theme of urban planning optimization or policy evaluation by considering the human-land inter-action arising from human mobility. For example, good data-driven thinking can support policymakers in formulating related policies to assessing the influence of the lockdown measures in the post-epidemic era and achieving sustainable resilient cities and seeking a population-development-oriented comprehensive modern transport system (Zhao et al., 2020). Specific suggestions can be put forward for the national-level master plan. For example, policymakers can make more precise decisions (from the city level to the national level) based on our analyses. Local authorities can know the specific geographical configurations of a certain city and better understand the city's role-playing in this *Chunyun* network.

Indeed, some limitations were existing in this study. First, using the mobile signaling data to measure the population movement between cities had uncontrollable uncertainty and bias. More verification should be done by using statistical data or census data. Second, this study was a case study by using *Chunyun* data to replenish new insights into the Third Law of Geography. Future studies should focus on more cases of other countries or datasets with other forms, except for China's *Chunyun*. Furthermore, through a tentative approach, the authors tried to provide a kind of thinking that combined the application of the logistic regression method and the beingness of the Third Law of Geography, more evidence is needed to affirm the establishment of this law.

5. Conclusion

Overall, this study constructed a national-scale time-series analytical framework. Based on a self-constructed indicator-based assessment, this study could confirm that: the more similar geographic configurations of two cities, the more similar the tidal features of human mobility between these two cities.

Within a unified urban system, nodal influence and tidal features along the timeline were strongly tied to its city scale and labor supply-demand balance. Patterns of city levels, traffic hub roles, and dominant flow indexes also impacted the similarity of tidal features. As a conclusion, on the premise that the sample/variable was sufficient enough and the collinearity problem was solved, whether the tidal features of human mobility at two cities were similar or not depended on their geographic configurations (such as labor supply, scale, level, traffic hub role, and flow pattern). In this sense, it is possible to address the existence and legitimacy of the Third Law of Geography.

As mentioned in Section 1, the authors argued that interpreting the geographic similarity between dynamic patterns of space of flow (from the human side) and geographic configuration (from the land side) in the spatiotemporal dimension requires updating traditional spatial interaction models, which could be closely associated with the theoretical extension of the Third Law of geography and the interdisciplinary development of human geography (Chai & Zhang 2020). Future research should focus on the construction of new theoretical frameworks in this context. Furthermore, other geographic contexts, such as USA's Thanksgiving break, China's Mid-Autumn Festival, etc. should be considered for enriching the achievements and conducting comparative research, which might facilitate the discovery and induction of potential universal laws related to the Third Law of Geography.

Therefore, our research not only bolsters the use of data-driven policy decisions based on large-scale human mobility data but also offers a novel perspective to enhance the study of this law, serving as a pioneering effort to advance previous work. Our achievements aimed to align with the demands of a humanistic research paradigm and to invigorate the interdisciplinary field.

Acknowledgments

We are grateful to Smart Steps Digital Technology CO., LTD for providing us the raw data.

Authors' contributions

Methodology, K.L; writing—original draft preparation, K.L; review & supervision, P.Z.; resources, X.H.

Funding

This research was funded by National Natural Science Foundation of China (41925003; 42130402).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Redfame Publishing.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Akskci Coperation. (2021). *Analysis of the number and income of migrant workers in China: 2015-2020*. Retrieved from: https://www.askci.com/news/data/hongguan/20210119/0908281332761.shtml (accessed on June 1st, 2023).

Andrew, S. (1985). *The future of geography*. Methuen, London. pp 159-173.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationary. *Geographical Analysis, 28*, 281-298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Chai, Y., & Zhang, Y. (2020). *Time Geography*. Southeast University Press, Nanjing.

Geographical Society of China. (2021). *Application of national-scale spatiotemporal pattern of population flow based on long time-series signaling big data (project), 2020 Geographic Information Technology Progress Award*. Retrieved from: https://urban.pkusz.edu.cn/info/1007/2788.htm (accessed on June 1st, 2023)

Goodchild, M. F. (2004). The Validity and Usefulness of Laws in Geographic Information Science and Geography. *Annals of the Association of American Geographers, 94*, 300-303. https://doi.org/10.1111/j.1467-8306.2004.09402008.x

Huang, B., Zhou, Y., Li, Z., Song, Y., Cai, J., & Tu, W. (2020). Evaluating and characterizing urban vibrancy using spatial big data: Shanghai as a case study. *Environment and Planning B: Urban Analytics and City Science, 47*(9), 1543-1559. https://doi.org/10.1177/2399808319828730

Li, X. W. (2007). The First Law of Geography and spatial-temporal proximity. *Chinese Journal of Nature, 29*(2), 69-71. https://doi.org/10.3969/j.ssn.0253-9608.2007.29.002

Liu, K., & Dauda, S. T. (2020). Structural changes in Japan's urban system from 1990 to 2010. *Journal of Urban and Regional Analysis, 12*(2), 117-143. https://doi.org/10.37043/JURA.2020.12.2.1

Liu, K., Murayama, Y., & Ichinose, T. (2021). Exploring the relationship between functional urban polycentricity and regional characteristics of human mobility: a multi-view analysis under the layout of the Tokyo metropolitan area. *Cities, 111*, 103109. https://doi.org/10.1016/j.cities.2021.103109

Liu, K., Zhao, P. J., Wan, D., Hai, X. D., He, Z. Y., Liu, Q. Y., … Yu, L. (2022). Using mobile phone big data to discover the spatial patterns of rural migrant workers' return to work in China's three urban agglomerations in the post-COVID-19 era. *Environment and Planning B Urban Analytics and City Science, 0*, 1-17. https://doi.org/10.1177/23998083211069375

Liu, Y., Wang, F. H., Xiao, Y., & Gao, S. (2012). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning, 106*(1), 73-87. https://doi.org/10.1016/j.landurbplan.2012.02.012

Ma, T. W., Wei, T. F., Qin, C. Z., Zhu, A. X., Qi, F., Liu, J. Z., … Pan, H. B. (2020). In-situ recommendation of alternative soil samples during field sampling based on environmental similarity. *Earth Science Informatics, 13*(1), 39-53. https://doi.org/10.1007/s12145-019-00407-x

Mu, X., Yeh, A. G. O., & Zhang, X. (2021). The interplay of spatial spread of COVID-19 and human mobility in the urban system of China during the Chinese New Year. *Environment and Planning B: Urban Analytics and City Science, 48*(7), 1955-1971. https://doi.org/10.1007/s12145-019-00407-x

Murayama, Y. (1982). A re-examination of Nystuen-Dacey model in terms of nodal regionalization. *Geographical Sciences, 37(2),* 73-84. https://doi.org/10.20630/chirikagaku.37.2_73

Murayama, Y. (2000). *Japanese urban system*. GeoJournal Library, Springer 2000. https://doi.org/10.1007/978-94-017-2006-9

Murayama, Y., & Komaki, N. (2013). *Regional analysis: data acquisition, analysis, and evaluation*. Kokonshouin, Tokyo.

National Bureau of Statistics of China. (2021). *The Seventh National Population Census*. Retrieved from http://www.stats.gov.cn/tjsj/zxfb/202105/t20210510_1817176.html (accessed on June 1st, 2023)

Research Institute of New First-tier City. (2020). *Ranking of cities' business attractiveness in China, 2020*. Retrieved from https://www.yicai.com/news/100648666.html (accessed on June 1st, 2023)

Shen, G. (2004). Reverse-fitting the gravity model to inter-city airline passenger flows by an algebraic simplification. *Journal of Transport Geography, 12*, 219-234. https://doi.org/10.1016/j.jtrangeo.2003.12.006

Sun, J., Pan, Y. J., He, R. F., Liu, H. Q., Chang, N. J., Liu, S. F., & Li, H. X. (2012). The enlightenment of geographical theories construction from the First Law of Geography and its debate. *Geographical Research, 31*(10), 1749-1763. https://doi.org/10.11821/yj2012100002

Tan, S., Lai, S., Fang F., Cao, Z., Sai, B., Song, B., … Lu, X. (2021). Mobility in China, 2020: a tale of four phases. *National Science Review, 16 Aug 2021, 8*(11), nwab148. https://doi.org/10.1093/nsr/nwab148

Tao, R., & Thill, J. C. (2019). Flow Cross K-function: a bivariate flow analytical method. *International Journal of Geographical Information Science, 33*(10), 2055-2071. https://doi.org/10.1080/13658816.2019.1608362

The State Council of the People's Republic of China. (2021). The 14th Five-Year Plan for National Economic and Social Development of the People's Republic of China. Retrieved from http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm (accessed on June 1st, 2023)

Tobler, W. R. (2004). On the first law of geography: A reply. *Annals of the Association of American Geographers, 94*(2), 304-310. https://doi.org/10.1111/j.1467-8306.2004.09402009.x

Veneri, P. (2013). The identification of sub-centres in two Italian metropolitan areas: A functional approach. *Cities, 31*, 177-185. https://doi.org/10.1016/j.cities.2012.04.006

Wang, C. J. (2010). Identification of inter-urban container transport hubs and their spatial characteristics: a case study of railway transportation in China. *Acta Geographica Sinica, 65*(10), 1275-1286. https://doi.org/10.11821/xb201010012

Wang, Z., & Wu, J. (2011). *Computational Geography*. Science Press, Beijing.

Wei, Y., Song, W., Xiu, C., & Zhao, Z. (2018). The rich-club phenomenon of China's population flow network during the country's spring festival. *Applied Geography, 96*, 77-85. https://doi.org/10.1016/j.apgeog.2018.05.009

Zhang, G. M., & Zhu, A. X. (2019). A representativeness-directed approach to mitigate spatial bias in VGI for the predictive mapping of geographic phenomena. *International Journal of Geographical Information Science, 33*(9), 1873-1893. https://doi.org/10.1080/13658816.2019.1615071

Zhang, L. P. (2020). *China City Statistical Yearbook*. China Statistics Press, Beijing.

Zhao, F. H., Huang, J. Y., & Zhu, A. X. (2023). Spatial prediction of groundwater level change based on the Third Law of Geography. *International Journal of Geographical Information Science, 37*(10), 2129-2149. https://doi.org/10.1080/13658816.2023.2248215

Zhao, P. J., Hu, H., Zeng, L., Chen, J., & Ye, X. Y. (2022). Revisiting the gravity laws of inter-city mobility in megacity regions. *Sci. China Earth Sci. 66*, 271-281. https://doi.org/10.1007/s11430-022-1022-9

Zhao, P. J., Lyu, D., Hu, H. Y., Cao, Y. S., Xie, J. X., Pang, L., … Yuan D. D. (2020). Population-development oriented comprehensive modern transport system in China. *Acta Geographica Sinica, 75*(12), 2699-2715.

https://doi.org/10.11821/dlxb202012011

Zhu, A. X., & Turner, M. (2022). How is the Third Law of Geography Different? *Annals of GIS, 28*(1), 57-67. https://doi.org/10.1080/19475683.2022.2026467

Zhu, A. X., Lu, G. N., Liu, J., Qin, C. Z., & Zhou, C. H. (2018). Spatial prediction based on Third Law of Geography, *Annals of GIS, 24*(4), 225-240. https://doi.org/10.1080/19475683.2018.1534890

Zhu, A. X., Miao, Y. M., Liu, J. Z., Bai, S. B., Zeng, C. Y., Ma, T. W., & Hong, H. Y. (2019). A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. *Catena, 183*, 104188. https://doi.org/10.1016/j.catena.2019.104188

Zhu, D., Wang, N., Wu, L., & Liu, Y. (2017). Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data. *Applied Geography, 86*, 152-164. https://doi.org/10.1016/j.apgeog.2017.07.001

Zhu, D., Zhang, F., Wang, S. Y., Wang, Y. L., Cheng, X. M., Huang, Z., & Liu, T. (2020). Understanding place characteristics in geographic contexts through graph convolutional neural networks. *Annals of the American Association of Geographers, 110*(2), 408-420. https://doi.org/10.1080/24694452.2019.1694403

Appendix 1

Table A1. Descriptive statistics of the raw data (n=365). *Note: IV: independent variable; SD: standard deviation*.

IV	Remark	Max	Min	Mean	SD
X1	*G*	98.1080	0.0000	10.0259	12.4099
X2	*K*	0.3781	0.0020	0.1228	0.0884
X3	*L1*	1	0	0.0110	0.1043
X4	*LN1*	1	0	0.0411	0.1988
X5	*L2*	1	0	0.0822	0.2750
X6	*L3*	1	0	0.1918	0.3942
X7	*L4*	1	0	0.2466	0.4316
X8	*L5*	1	0	0.4274	0.4954
X9	*H1*	1	0	0.0356	0.1856
X10	*H2*	1	0	0.0575	0.2332
X11	*H3*	1	0	0.1342	0.3414
X12	*D1*	9.8239	0.0110	4.5648	2.2969
X13	*D2*	9.6724	0.0000	2.4722	2.3399

Appendix 2

Table A2. Descriptive statistics of all variables used in the logistic regression model (n=66430). *Note: DV: dependent variable; IV: independent variable; SD: standard deviation*.

Variables	Remark	Max	Min	Mean	SD
DV	Y	1	0	0.4516	0.4977
IV	ΔG_{ij}	1	0	0.7349	0.2006
	ΔK_{ij}	1	0	0.8937	0.1439
	$\Delta L1_{ij}$	1	0	0.9783	0.1458
	$\Delta LN1_{ij}$	1	0	0.9210	0.2698
	$\Delta L2_{ij}$	1	0	0.8487	0.3583
	$\Delta L3_{ij}$	1	0	0.6891	0.4628
	$\Delta L4_{ij}$	1	0	0.6274	0.4835
	$\Delta L5_{ij}$	1	0	0.5092	0.4999
	$\Delta H1_{ij}$	1	0	0.9311	0.2533
	$\Delta H2_{ij}$	1	0	0.8913	0.3113
	$\Delta H3_{ij}$	1	0	0.7669	0.4228
	$\Delta D1_{ij}$	1	0	0.7322	0.1945
	$\Delta D2_{ij}$	1	0	0.7378	0.2198