MRes Publication

Candidate: Zunhao Zhang

Journal: Environment and Planning B: Urban Analytics and City Science

Word Count for Main Body: 5758

Word Count including Abstract and References: 7894

Submission Date: September 14th 2020

Module Name: MRes Dissertation

Module ID: CASA0004

Supervisor: Prof. Michael Batty

Note: The *Supplementary Materials* should originally be a separate word file with the paper based on the requirements of Environment and Planning B: Urban Analytics and City Science, but I put it after reference in order to keep the submission of assessment as a single PDF file.

Simulating the Spatial Interaction of Internal Migration in Polycentric Mega-city Regions based on Gravity Models: An Application in the Great Bay Area, China

Zunhao Zhang

Abstract

Since the rapid development of meg-city regions around the world, the closely spatial interaction within those urban agglomerations has been a popular agenda in urban studies. The Great Bay Area (GBA) is one of the most important polycentric mega-city regions in China and its coordinated development has also been constantly emphasized as a national planning strategy in response to global competition. Therefore, simulating the spatial interaction within the GBA will contribute to making tailored planning decisions to this area. However, due to the limited availability of Origin-Destination (O-D) data in China, most studies on spatial interaction within spatial systems in China based on gravity model - one of the most widely employed methods for modelling spatial interaction patterns - have remained at the theoretical level or have used assumed parameters values to estimate the flows. In order to contribute to the research in this field, this paper focuses on simulating the migration flows, one of the most common form of spatial interaction, within the GBA that includes the municipalities in Guangdong province. The daily movement data during the return-home travel rush before the Spring Festival from Baidu Map is utilized to interpret migration flows, combining with the socio-economic indicators of each cities and O-D travel times.

Keywords

Spatial interaction, Internal migration in the Great Bay Area (China), Gravity model

Introduction

In recent decades, with the vigorous emergence and development of urban agglomerations around the world, more and more scholars began to conduct research on those large regions containing a cluster of cities with high-close internal spatial interactions between them (Hanssens et al., 2013; He et al., 2017; Yeh and Chen, 2020), and put forward many similar descriptive concepts, such as Polycentric Mega-city Region (PM-CR) (Hall and Pain, 2006). This kind of effective spatial interaction is contributing to the utilisation of different specializations of the nodes that comprise spatial networks (De Goei et al., 2010). In this context, a better understanding of the spatial interaction patterns of these regions is particularly important, rather than conventionally exploring the location attributes of their internal nodes (Batty, 2013).

The concept of the Greater Bay Area (GBA) in the Pearl River Delta, China, was first proposed by national government in 2017 as an important regional development strategy, and the past few years have witnessed a succession of planning policies for facilitating the coordinated development of the GBA, such as the *Overall Planning Outline* in 2019 and the *Inter-city Rail Transit Planning* in 2020. Covering nine cities in Guangdong (Gd) Province as well as Hong Kong and Macau, the GBA is normally considered as one of the most important PM-CRs in China which is on a par with the other three famous international bay areas in New York, Tokyo and San Francisco in terms of different socio-economic scales (Hall, 1999; Li et al., 2018). Therefore, it is of great significance to explore current spatial interaction patterns within the GBA and effectively simulate the potential influence of the implementation of the planning on it.

As the most common form of spatial interaction (Norris, 1972), the internal migration flows within the GBA will be the research object of this paper. The rapid urbanization and the continuous construction of transportation networks in China provide motivations and opportunities for long-term migration (Li et al., 2019), especially in the mega-city regions like the GBA. However, this trend of frequent migration also puts some pressure on housing, public infrastructure and transportation, and is also accompanied by the deterioration of some social problems, such as social inequality and segregation (Li et al., 2019). Measuring and simulating the migration patterns might be the basis for the further exploration of these challenges.

However, as one of the most commonly methods for modelling spatial interaction (Batty, 1972; Haynes and Fotheringham, 1984; Poot et al., 2016), the empirical applications of gravity model in China have remained at the theoretical level or have used assumed parameters values to estimate the flows due to the lack of the observed Origin-Destination (O-D) data (Gu and Pang, 2008; Hu and Zheng, 2015; Mei et al., 2012). The studies of internal migration flows in China have the same problems, while most empirical research based on gravity models are at the provincial level rather than inter-city level (Fan, 2013; Li et al., 2017; Shen 1999; 2017). In order to contribute to the research in this field, the main aim of this paper is to better understand, visualise and simulate the internal migration flows within the GBA that includes the municipalities in Gd province by employing gravity models. The previous literature and empirical studies on modelling human migration by gravity model is firstly reviewed. The daily movement data during the return-home travel rush before the Spring Festival from Baidu Map is utilized to interpret migration flows, combining with the socio-economic indicators of each cities and O-D travel times. Finally, the results of visualisations and simulations are discussed.

Unconstrained Gravity Model for Spatial Interaction

The definition of the spatial interactions in spatial networks could be basically interpreted as the exchange flows between geographies in terms of different elements (Norris, 1972), and could be normally described by a statistical matrix though labelling the origins and destinations of flows as i and j. Basically, the gravity model employed in spatial interaction study is derived from Newton's Law of Gravity (1687) as follows:

$$F_{ij} = G\frac{M_i M_j}{r^2} \qquad (1)$$

The law of gravity has long been observed to be applicable in various fields of social science, namely the spatial interaction between two places is likely to be proportional to the mass or attraction of these two places and inversely proportional to the square of the distance between them (Davis et al., 2013). As early as 1931, Reilly suggested this kind of law in the context of retailing and creating his *Law of Retail Gravitation* (Batty, 1978). Later, Zipf's early research (1946) argued that the same principle exists in intercity movements, but he argued that flows were inversely proportional to distance, rather than raised to the power of 2 as in classical physics.

Although the applicability of the gravity model in physics to spatial interaction has been demonstrated in many early empirical studies, there was no fundamentally technical support until Wilson applied the principle of maximum entropy to travel distribution in 1967 (Williams, 2019). The core idea of Wilson's derivation process can be summarized as follows: In a closed system, with the constraints of the total departure, the total arrival and the total travel cost, the most likely travel distribution should be when the entropy of this system is at its maximum (Batty, 1972; Wilson, 1967; 1970; 1971). Based on this, Wilson (1971) extended gravity model into a family of related models according to the known systemic constraints. The unconstrained gravity model in Wilson's family of models is normally formulated as follows in practice:

$$T_{ij} = KO_i^{\alpha} D_j^{\beta} c_{ij}^{\gamma} \qquad (2)$$

where spatial interactions $\boldsymbol{T_{ij}}$ is the measure of flows, O_i and D_j represent the mass of origins and destinations, c_{ij} refers to the travel cost, such as travel time and distance, and the parameters of variables, α, β, γ, are used to give a certain elasticity to the importance of the mass of origins and destinations (Batty and Mackie, 1972). The parameter γ is also known as the friction of the travel cost

Since the parameters, especially the friction of the travel cost (Stillwell, 1978), $\boldsymbol{\gamma}$, need to be estimated based on the observed O-D data of the study system, this model is required further calibration. Transforming the equation 2 into a linear formula (equation 3) by taking the logarithm of both sides and estimating the parameters by regression models is a widely accepted approach for the calibration process (Flowerdew and Aitkin, 1982).

$$\ln T_{ij} = k + \alpha \ln O_i + \beta \ln D_j + \gamma \ln c_{ij} \qquad (3)$$

Meanwhile, in the choice of regression model regarding the analysis of the population movement, as some scholars argued, the Ordinary Least Squares (OLS) is one of the most common methods to estimate this logarithmic linear model (Martínez-Zarzoso, 2011; Ramos and Suriñach, 2017). However, Flowerdew and Aitkin (1982) also highlighted that, the Poisson and Negative Binomial Regression models might perform better than OLS, especially for predicting some extremely small flows. The reason can be summarized as follows: the predicted population migrations are normally non-negative integers, and it is highly likely to satisfy the Poisson distribution rather than the normal distribution (Dennett, 2018). Generally, regardless of the calibration methods used, there are many reasons to estimate the parameters in the model, such as comparing the difference of various indicators, predicting future interaction pattern, simulating the effects of the planned system change, and filling in missing historical data (Dennett, 2012; Fotheringham and O'Kelly, 1989).

Human Migration of Spatial Interaction

The human migration here is briefly defined as one form of population movement with the change of residential address. Based on the *Household Registration System*, internal long-term migrants in China could be clearly defined as people who do not live or work in the registered residence for any length of time, and this group is also known as the floating population (Zhang et al., 2013). In terms of the related data about the internal migration flows in China, there is annual O-D data only available for the inter-provincial migration recorded in the Statistical Yearbook. In addition, the total number of immigrants or emigrants from one city to the other cities within the same province is also recorded in the Statistical Yearbook of different provinces or cities. Meanwhile, this kind of total number of migrants at the district level can be obtained from decennial Census.

Although the previous researches of human migration using gravity models were generated within different spatial systems and at various geographic levels, there have been some common findings about the effects of different factors on migration. For example, most empirical studies across the world have revealed that the population size, income and travel cost always have the greatest influence on the migration flows both for international migrations and internal migrations (Borjas, 1989; Greenwood et al., 1991), which could be defined as '*gravitational demographic variables*' (Karemera et al., 2000).

On the other hand, the effects of some other indicators involved in the extended gravity models might be varied in different study areas or they might depend on different types of human migration. As Boyle et al. (1998) highlighted in their study of the inter-ward migration within the non-metropolitan counties in UK, housing requirements and types have a crucial role on the flows. The study proposed by Shen (2017) revealed that the temperature severity and rural development at the destination might be significant to inter-provincial migration in China. In terms of international migration flows, the indicators which should be considered in the extended gravity models are much complicated, including migration policy, differences in culture and border effect. (Anderson and Wincoop, 2003). Therefore, the variables considered in the gravity model and their significance to the migration flows might vary in different geographic scales or countries.

Furthermore, the visualisation of the aggregate movements could not only help us better understand their characteristics and patterns, but also offer more significant reference for urban planning and policy decision-making (Batty, 2013; 2018). A common way to visualize flows is to draw the vector lines between origins and destinations, which could be called a flow map or an O-D map. Excepting the conventional flow map, some other visualisation methods are effective for some certain dataset or focus, such as the flow density maps which converts flows into the attributes of the location information (Wood et al., 2010), circular migration plots which easily demonstrate the self-flows (Charles-Edwards et al., 2015; Sander et al., 2018), compressing and abstracting the flow information into the commuting path like underground and street framework (Dorling, 2012; Wood et al., 2011). All in all, how to balance simplifications of flow information and the readability of visual map must be considered during the process of visualisation.

Study Area and Data

Study Area and Geographical Scales

The Version of the GBA studied here only contains 9 cities in Gd province and excludes Hong Kong and Macau (Figure 3.1), due to the limitations of the available migration data and a consideration of the political influences on internal migration. The inter-city migration between the 9 cities within the GBA contain total 72 migration flows based on 9*9 O-D matrix regardless of the intrazonal flows.

Data

All original datasets employed in this research are public. There are mainly three categories of data sets employed and processed in this paper.

Long-term Migration Flows

Firstly, the scaled index of daily movements at inter-city level provided by *Baidu Map Migration Platform* is processed to represent the actual volumes of long-term migration flows. *Baidu Map* collects aggregated O-D daily trips between cities in China mainly by making use of its 'Location Based Service' function, namely to record the changing locations of the users who have installed or called the Baidu Map application. As demonstrated in Figure 1, the around two weeks leading up to the Spring Festival are often defined as the peak period for long-term migrants returning to their home city in the context of the Chinese cultural activities (Zhang and Yang, 2013). This return-home travel rush migration is from Jan. 10th to Jan. 25th in 2020, while it was Jan. 21st to Feb. 5th in 2019, since the Chinese 'New Year' varies depending on the Lunar Calendar. Based on the discussion and verification of the feasibility in the *Flow Data Processing Section of the Supplementary Materials*, this paper argues that the sum of return-home movements between Jan. 10th 2020 and Jan. 25th 2020 (16 days) can be used to represent the actual long-term migration flows in China.

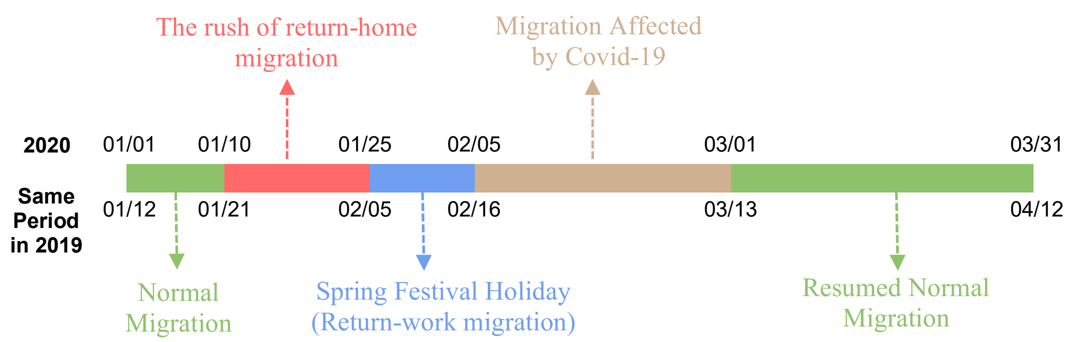


Figure 1. *The Explanatory Timeline of Spring Festival Migration in 2020 and the Same Period in 2019. (Source: Author)*

Socio-economic Variables

The data set of the socio-economic condition of each city within the GBA applied in the further simulation, is obtained from the *2019 Statistical Yearbooks of Gd province*. Most of these variables are panel data through to the end of 2018, but also include one GDP growth rate between 2017 and 2018, and this enable us to check the potential impacts of the economic growth speeds on long-term migration.

Travel Cost

In terms of the travel cost c_{ij} between origins and destinations, this research adapts three types of travel time data based on the different transportation modes: minimum driving time, minimum travel time by any public rail and minimum travel time only by the ordinary trains. Instead of using the physical distance as in many previous empirical studies (Champion et al., 1998; Dennett and Wilson, 2013), travel time more realistically represents the deterrence between places due to the widespread use of roads and rail systems today. Firstly, the minimum driving time excluding the influence of traffic conditions between places are captured from *Baidu Map Route Matrix API* by using Python scripts. Secondly, the minimum travel time by any kinds of public transportation and only by ordinary train are both collected from all train schedules provided by *12306 China Railway*, which is the only direct and official channel for the sales of all kinds of railway tickets in China. Not surprisingly, most of the minimum travel times by any public transportation are by high-speed trains or bullet trains.

Methodology

Building Gravity Models

Since one of the main objectives of this research is to quantitatively explore the relationships between different indictors and the long-term migration flows, the unconstrained gravity model (equation 2) in the family of related models is employed.

Table 1. *The Definitions of Variables Employed in Gravity Models （Source: Author）*

The Definitions of Variables in the Gravity Models			
Dimensions	**Indicators (Variables)**	**Labels**	**Unit**
Population Size	Permanent Population	P_i^{a1} P_j^{a2}	10000 pers
Income & Economy	Per Capita Disposable Income of Urban Permanent Households	I_i^{a3} I_j^{a4}	¥
	The Proportion of Tertiary Industry in GDP	TI_i^{a5} TI_j^{a6}	%
	The Growth Rate of GDP (preceding year =100%)	$GDP\_G_i^{a7}$ $GDP\_G_j^{a8}$	%
Housing Price	Average Housing Price of Destination	H_j^{a9}	10000 ¥ / ㎡
Education	The Density of Secondary Schools of Destination	S_j^{a10}	Unit /10000 pers
Travelling Time (3 Types of c_{ij}^{γ})	Minimum travelling time by driving	cd_{ij}^{b1}	hours
	Minimum travelling time by any public transportation (mainly High-speed trains)	cp_{ij}^{b2}	hours
	Minimum travelling time only by ordinary trains	ct_{ij}^{b3}	hours

Furthermore, all socio-economic indicators and three types of travel times used to represent the mass of origins and destinations are listed in the Table 1. Except for the traditionally gravitational demographic variables, other potential indicators are also included based on considering the particular context of spatial interactions in China and several empirical research (Fan, 2013; Karemera et al., 2000; Shen, 2017; Xu and Yao, 2019). Finally, we build an unconstrained gravity model as follows:

$$\begin{aligned}\ln M_{ij} &= a_0 + a_1 \ln P_i + a_2 \ln P_j + a_3 \ln I_i + a_4 \ln I_j + a_5 \ln TI_i \\ &+ a_6 \ln TI_j + a_7 \ln GDP\_G_i + a_8 \ln GDP\_G_j + a_9 \ln H_j \\ &+ a_{10} \ln S_j + b \ln c_{ij} \qquad i, j = 1 \dots N, \quad i \neq j \qquad (4)\end{aligned}$$

Selection of Travel Cost

The travel cost c_{ij} is always defined as the combination of the expense and time of travelling. However, the physical distance between i and j is always used to represent the travel cost, the relationship between distance and commuting cost and time is often nonlinear in real life (Zhao et al., 2016). Therefore, the minimum travel time is more appropriate to measure the travel cost. Since there are three measures of travel time employed in this research, their different sensitivity to the migration flows need to be first discussed.

Figure 2 display the scatter plots of $\ln M_{ij}$ against three cost types of $\ln c_{ij}$ for the GBA, and these combine the corresponding results of linear regression models. Based on the comparisons of the values of the adjusted R squares, the travel time on only ordinary trains has a relatively lower explanation on the variations of long-term migration than the other two travel time measures. The reason behind this finding might be interpreted by the current travel-behaviour patterns in China, namely people prefer high-speed trains rather than ordinary trains due to the relative affordability and high coverage of the high-speed rail system. Therefore, this study introduces the minimum driving time and minimum travel time by any public transportations into the subsequent calibration of gravity models.

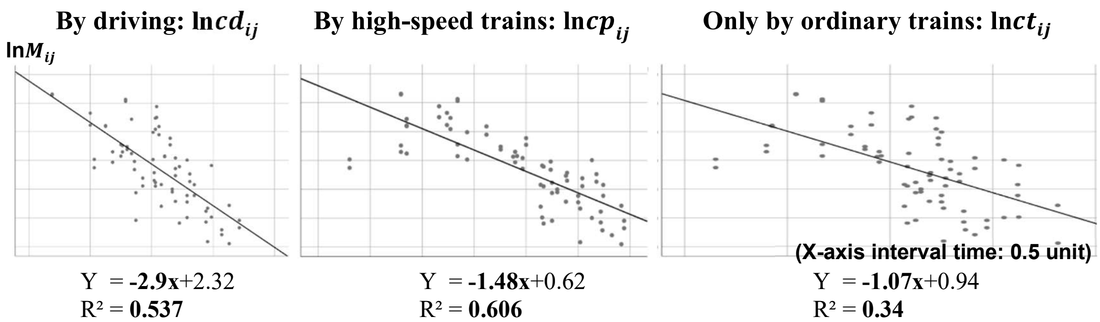


Figure 2. *Scatter Plots of 9 Cities within GBA: $\ln M_{ij}$ against $\ln c_{ij}$*

Calibration

Since the flow data used in the study is not the actual number of migrants but the functionalized migration indices from *Baidu Map* which are fractional, this research adopts the logarithmic linear regression for the calibration of models (equation 4) rather than Poisson regression. Moreover, the Variance Inflation Factor (VIF) and corresponding tolerances will be calculated to identify the potential multicollinearity in the regression model, namely the ratio of the overall model variance to the variance of only putting one single independent variable in the model (Mansfield and Helms, 1982). The acceptable significance level of the coefficients is 5% in this study.

Simulating the Impacts of Inter-city Rail Planning

The further simulation about the impacts of future railway construction on the long-term migration flows is mainly based on the calibrated gravity model of the GBA using the minimum travel time by public transportation. Two estimated sets of travel times after the accomplishment of the new railway construction – short-term planning by 2025 and long-term planning by 2035 – will separately replace the original travel time, c_{ij}, to generate the estimated flows both for the short-term and long-term constructions. The process of estimating travel time changes based on these two phases of planning can be checked in the *Estimation of Travel-time Changes section in the Supplementary Materials*. It is worth mentioning in advance that this further simulation does not involve the change of socio-economic indicators of cities in the GBA. Therefore, the final results can only represent the influence of intercity rail transit construction, rather than the prediction of the real long-term migration flows in the future.

In addition, the change rates of the newly estimated flows to the original flows will be utilised to quantitatively access the influence of these two planning phases as below:

$$R_{ij} = \frac{|\widehat{M'_{ij}} - \widehat{M_{ij}}|}{\widehat{M_{ij}}} \qquad (5)$$

where $\widehat{M'_{ij}}$ is the estimated flow after the future railway construction and $\widehat{M_{ij}}$ is on behalf of the original flow between i and j.

Visualisation Approaches

In order to visualize the migration patterns for the GBA as well as the whole Gd province, three tools will be employed successively: *ArcGIS Map*, *Mapbox* and *flowmap.blue*. Firstly, the centroids of each cities within the whole Gd province are extracted from their administrative boundaries by *ArcGIS Map*, and these thus represent the nodes in the migration network. Furthermore, the shape file of the administrative boundaries is uploaded to the Mapbox in order to customize the base map of the visualisation outcomes in *flowmap.blue*.

The main process of visualisation is to create the online interactive flow maps by using the *flowmap.blue*, which is an open source script published on GitHub by Ilya Boyandin (2019). With this online tool, we just need to upload our dataset in the required form and our Mapbox's access token to generate the interactive flow maps. In addition to the convenience, there are two other advantages to using the *flowmap.blue* instead of the traditional visualization methods, such as the "*xy to line*" methods in *ArcGIS Map.* Firstly, the interactive visualisation maps provide readers more opportunities to explore and customize the detailed information of their own interests, especially for those complicated flow maps. Secondly, compared with visualizing maps in ArcGIS or R, it can better visualize the asymmetric migration directions of flows and intelligently cluster some origins or destinations as required. In the following selection of this results presentation, this study will provide the links to the visualized outcomes and present screen captures for further analysis.

Results and Discussion

Visualisations of Current Migration Patterns

Based on the *Flowmap.blue* visualisation tool, the interactive flow maps of the long-term migration between the nine cities within the GBA can be accessed at:

https://flowmap.blue/1BSGRUmwvalzlBuYceECx3eMyzmb7vu5etezZ88Rf7hI?v=23.081405,112.620695,7.50,0,0&a=0&as=1&b=1&bo=73&c=1&ca=0&cz=7&d=1&fe=1&lt=1&lfm=ALL&col=Reds&f=37

In order to offer a comprehensive understanding of the status of the GBA in the whole Gd province. The interactive flow maps of the internal migration flows for the whole Gd province can be accessed at:

https://flowmap.blue/1HbTVERFGswZMwWPI5F8K4d7XmFlgO36Rqn08gT7ocls?v=23.128972,112.924107,6.79,0,0&a=0&as=1&b=1&bo=56&c=1&ca=1&d=1&fe=1&lt=1&lfm=ALL&col=Reds&f=49

According to Figure 3, there is no doubt that the interaction between Guangzhou and Foshan is the highest one in the GBA. The long-term migration among Foshan, Guangzhou, Dongguan, Shenzhen and Huizhou are also apparently higher than others, which form a Z-shaped geographic structure (Sun et al., 2019). Meanwhile, these five cities are also the most important nodes in the regional structure, considering their total amount of migrations. Except for this visually striking Z-shaped structure, the connection between Zhuhai and Zhongshan is also significant. On the other hand, Jiangmen is the least connected with the other eight cities regarding the long-term migration, followed by Zhaoqing and Zhuhai. In addition, the nine cities of the GBA are closely connected regarding long-term migration than the others within Gd province, and the Z-shaped structure mentioned above is extremely prominent even in the context of the whole Gd province (see the Figure S9).

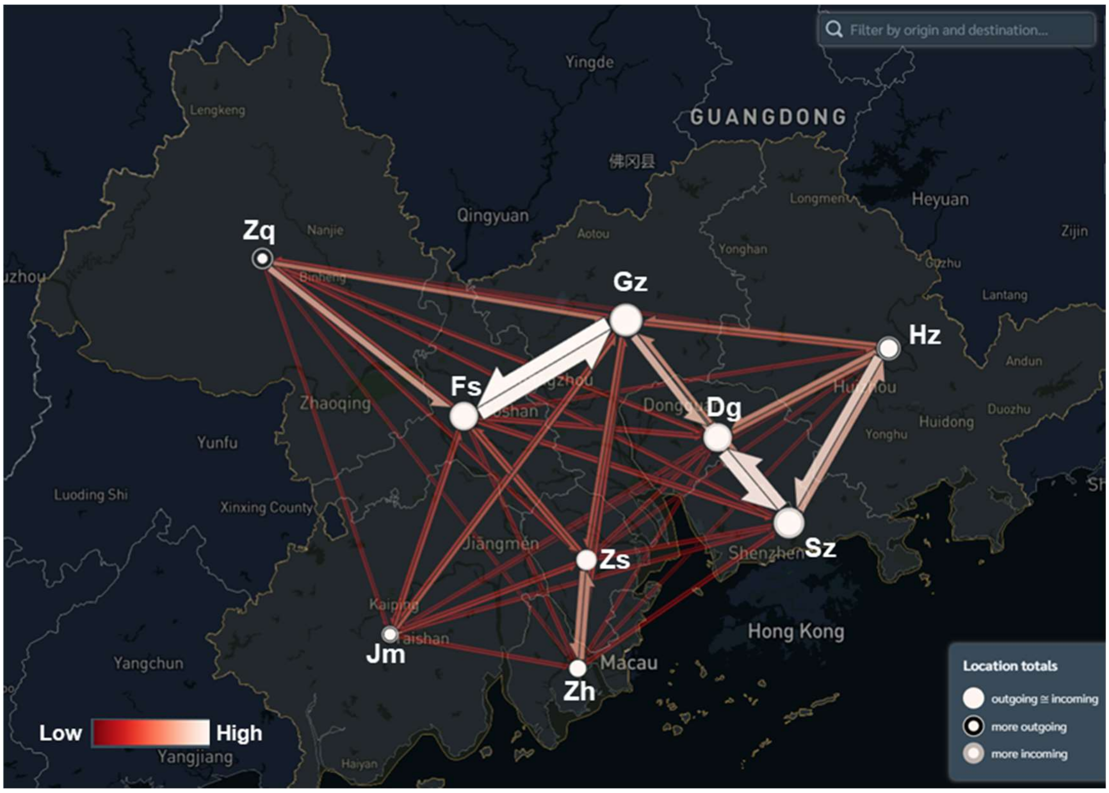


Figure 3. *The Flow Map of the Long-term Migration Flows between the nine cities within the GBA（Source: Author）*

Secondly, Figure 4 identifies the net migration flows of the GBA. What should be made clear in advance is that the thickness and colour of the lines cannot be compared between different figures, but only horizontally within the same figure. Zhaoqing has the largest total amount of net outflows within the GBA with all net outflows to the rest of the cities, and the two main directions of the outflows are Foshan and Guangzhou. In an opposite way, Shenzhen is the only one city with all net immigration flows from other cities, followed by Dongguan and Foshan which also have high levels of net immigration. However, as another developed city in the GBA, Guangzhou has relatively significant net outflows to Dongguan and Shenzhen.

Figure 4. *The Net Flow Map of the Long-term Migration Flows between the nine cities within GBA（Source: Author）*

Comparing Figure 3 and Figure 4, some potential findings about the functional transfer between core cities and surrounding cities can be identified. As the most developed cores within the GBA, Guangzhou and Shenzhen both have significant emigration flows to the surrounding satellite cities: Foshan, Dongguan and Huizhou, although there are large number of immigrants from other cities to these three cities too. To some extent, this situation can be attributed to the transfer of some urban functions from the central cities to surrounding satellite cities, such as the transfer of some particular industries and residential functions from the core cities (Sun et al., 2019). Taking the huge migration flow from Foshan to Guangzhou as an example, due to the improved transportation network between these two cities in recent years, such as the direct connection of subway network between them, some manufacturing industries and private enterprises in Guangzhou have begun to move to Foshan, and more and more people working in Guangzhou choose to settle in Foshan where housing prices are much lower (Yang, 2019) .

Estimates of the Gravity Models for the GBA

The estimates of the two gravity models for the GBA - one using the minimum driving time and the other one based on the minimum travel time by any public transportation (Estimates within brackets) - are presented in Table 2.

Table 2. *The Estimations of Two Gravity Models for the GBA （Source: Author）*

N = 72	**Coefficients: using $c_{ij} = cd_{ij}^{b1}$** (Coefficients: using $c_{ij} = cp_{ij}^{b2}$)				**Collinearity using $c_{ij} = cd_{ij}^{b1}$** (Collinearity: using $c_{ij} = cp_{ij}^{b2}$)	
	Unstandardized	Standardized	t	Sig.	Tolerance	VIF
Constant	**-14.89** (-22.85)		**-1.29** (-2.34)	**0.201** (0.022)		
P_i	**0.622** (0.586)	**0.296** (0.279)	**3.25** (3.50)	**0.002** (0.001)	**0.385** (0.385)	**2.60** (2.60)
P_j	**0.437** (0.420)	**0.208** (0.200)	**2.22** (2.44)	**0.030** (0.018)	**0.364** (0.364)	**3.00** (2.75)
I_i	**-1.150** (-0.421)	**-0.191** (-0.070)	**-1.95** (-0.84)	**0.056** (0.404)	**0.333** (0.353)	**0.20** (2.83)
I_j	**2.384** (2.948)	**0.385** (0.476)	**2.71** (3.87)	**0.009** (0.000)	**0.159** (0.162)	**6.30** (6.18)
TI_i	**0.677** (-0.398)	**0.081** (-0.048)	**1.03** (-0.68)	**0.309** (0.502)	**0.513** (0.493)	**1.95** (2.03)
TI_j	**-1.066** (-1.931)	**-0.127** (-0.231)	**-1.47** (-2.97)	**0.148** (0.004)	**0.423** (0.404)	**2.36** (2.47)
GDP_G_i	**-0.533** (-0.463)	**-0.183** (-0.159)	**-2.18** (-2.18)	**0.033** (0.033)	**0.456** (0.462)	**2.20** (2.17)
GDP_G_j	**0.160** (0.162)	**0.055** (0.056)	**0.574** (0.67)	**0.568** (0.508)	**0.352** (0.352)	**2.84** (2.84)
H_j	**0.439** (0.289)	**0.197** (0.130)	**1.77** (1.37)	**0.081** (0.176)	**0.260** (0.273)	**3.84** (3.67)
S_j	**1.086** (0.644)	**0.194** (0.115)	**2.22** (1.54)	**0.030** (0.128)	**0.421** (0.442)	**2.38** (2.26)
cd_{ij} (cp_{ij})	**-2.557** (-1.298)	**-0.646** (-0.671)	**-9.26** (-11.4)	**0.000** (0.000)	**0.657** (0.709)	**1.52** (1.41)
Adjusted R^2	**0.773** (0.826)					

Note: *The estimations of the gravity model which is employed the minimum travel time by any public transportations are shown in the brackets.*

In general, the independent variables employed in these two gravity models explain the variations of dependent variables to a large extent with respect to the adjusted R squares: 0.773 and 0.826. Meanwhile, since the results of VIF test are all below 10 and the tolerances are all above 0.1, there is no serious multicollinearity between the independent variables (Hair et al., 1995). However, the VIF value of income of destination I_j is the only one of all variables to exceed 3.0, which means that it may have a certain correlation with other socio-economic indicators (Hair et al., 2019), but it is considered acceptable in this experiment.

In terms of the significance of all 11 independent variables, these two models only present the differences at 5% level in 3 variables: the proportion of the tertiary industry at destinations, the growth rate of GDP at origins and the education situation in the destinations. The estimates are relatively consistent with the previous researches in migration (Borjas, 1989; Poot et al., 2016; Greenwood et al., 1991; Shen, 2017): the populations of origins, income of destinations and travel time are all apparently significant to variations of long-term migration flows (which are all below 0.01 significance level). Considering the similarities between these two models, this paper will just use the estimates based on the minimum driving time to identify the further relationships between each significant variable and migration flows below.

Population

The coefficients of the population in the origins and destinations are both positive with a high level of significance 0.2% and 3%, which indicates that an increase in population regardless of origin or destination might encourage long-term migration between them. Meanwhile, by comparing their standardized coefficients (0.296 and 0.208), the origin population has a greater influence on increasing the long-term migration flows than the destination population. In addition, the cities with large populations might not mean more saturated employment but more attractive opportunities for immigrants.

Economy

In terms of the effects of economic indicators, the income of destinations has a significantly positive correlation with the migration flows at 1% level. Moreover, the magnitude of its coefficient is greater than 2 (2.384), and this reveals that the cities with high average salaries are sensitively attractive to the migrants. Although the high-income origins seem to effectively reduce the opportunities of long-term migration between i and j (Coeff. = -1.15), the significance of I_i is 5.6% slightly exceeding the 5% level. According to the results of TI_i and TI_j, the influence of the proportion of tertiary industry on migration are not significant both for the origins and destinations. The impact of the growth speed of the urban economy on long-term migration is partly in line with expectations, namely the rapid economic development of origins might discourage the emigration, since the coefficient of GDP_G_i is estimated about -0.533 at a 3.3% significance level. Therefore, in addition to the traditional use of static panel data to represent the mass of the cities, some dynamic changes in the cities might also be worth considering, such as the urban development potential.

House Price and Education

Although the high housing prices have long been the focus of social discussion in China, the negative effect of high housing prices on the destinations of long-term migration is not identified as being significant in this empirical study of the GBA, and this is similar to the findings argued by Xu and Yao (2018). On the contrary, the coefficient of S_j is estimated as 1.086 at 3% significance level, which indicates that the better educational resources in destinations might attract more immigrants. This kind of attraction could be relatively explained as the fact that the available choices of middle and high schools for the next generations are directly linked to household's location of the domicile in the context of China's household registration system.

Travel Cost

The minimum driving time with the highest absolute value of the standardized coefficient (-0646), has the largest impact on the long-term migration of all variables. Specifically, based on the equation (2) and equation (4), if the minimum driving time between two places within the GBA doubles and other variables remain the same, the long-term migration flows between these two places will be about only 17% of the original. Generally, compared to the absolute values of the standardized coefficients of all significant variables (< 5%), the effects of travel cost, income and population on long-term migration rank in the top three.

Furthermore, based on the visualisation map for the residuals in **Figure S1**, we could argue that the lines with the higher error percentage are those with the lower original migration, such as the migrations between Zhaoqing and Jiangmen. This might be largely because the method of logarithmic linear regression for model calibration is difficult in making an acceptable prediction of the migration flows with quite small values (Flowerdew et al., 1982).

Impacts of Inter-city Railway Planning of the GBA on Migration Patterns

In order to simulate the long-term migration flows M_{ij} within the GBA after the completion of the inter-city railway planning, the model based on the minimum public transportation above is employed by excluding insignificant variables (5% significance level) and then stepwise regression processing again. The results are shown in the equation (6) with the adjusted R square: 0.825.

$$M_{ij} = \exp\left(-28.928 + 0.463 * \ln P_i + 0.432 * \ln P_j + 2.760 * \ln I_j - 1.371 \ln TI_j - 0.298 * \ln GDP\_G_i - 1.208 * \ln cp_{ij}\right) \quad (6)$$

Furthermore, since the future rail infrastructure will only lead to shorter travel times between some cities, the affected flows will theoretically increase. The growth rates of the two scenarios for the short-term planning (2025) and long-term planning (2035) are calculated as below, to quantitatively measure the influence of these two phases, where the $\widehat{M_{ij}}$ is the original estimates, $\widehat{M_{ij2025}}$ is the estimated flows for the short-term planning and $\widehat{M_{ij2035}}$ represents the flows affected by the long-term rail planning.

$$R_{ij2025} = \frac{\widehat{M_{ij2025}} - \widehat{M_{ij}}}{\widehat{M_{ij}}} \quad (7)$$

$$R_{ij2035} = \frac{\widehat{M_{ij2035}} - \widehat{M_{ij2025}}}{\widehat{M_{ij2025}}} \quad (8)$$

Impacts of the Short-term Rail Construction

The estimated migration flows affected by the short-term railway constructions are visualized in Figure 5. By comparing this short-term simulation with the current flows shown in Figure 3, the recent railway construction will highly contribute to pulling previously marginalized cities, such as Zhongshan, into this close urban network of the long-term migration. In general, by 2025, the originally Z-shaped core mentioned at the beginning will be enriched by the increases in migration flows between Foshan and Dongguan and Shenzhen, while Zhongshan might be successfully connected to this core structure through its remarkable migration to Guangzhou. In addition, the number of the emigrants from Jiangmen to Zhongshan is expected to rise to a significant level.

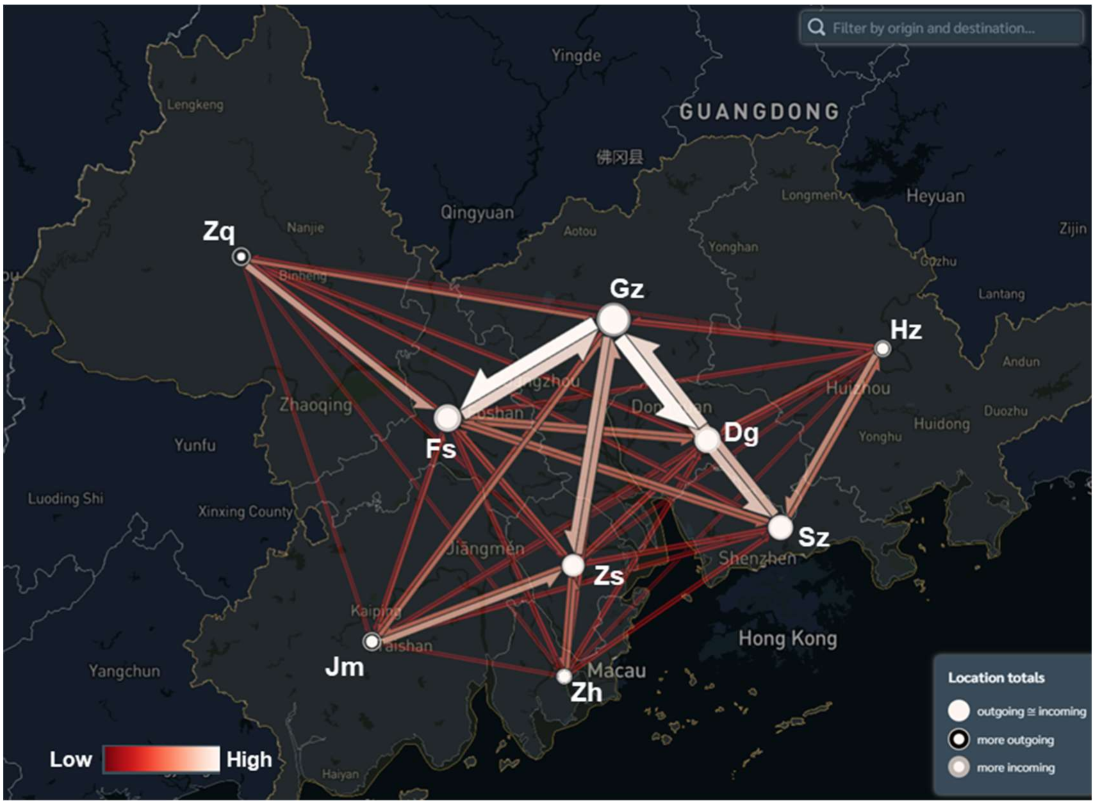


Figure 5. *The Estimated Flows based on the Impacts of the Short-term Railway Planning by 2025 of the GBA （Source: Author）*

Furthermore, the growth rates of all 16 affected flows by the short-term rail constructions are visualized in Figure 6. Overall, in the short term, the long-term migration flows that will increase greatly by 2025 are all around Dongguan, among which the number of migrants in Foshan and Dongguan might more than double the original estimates based on the current railway system, as well as Guangzhou. Therefore, we could conclude that Dongguan is likely to become the most important focus in the long-term migration network of the GBA in the next five years, along with large numbers of migrants from or to the surrounding cities.

Meanwhile, under the assumption that other socio-economic indicators remain unchanged, the specifically quantitative changes of the total amount of migration for each city by 2025 are displayed in the Figure 7. Since the short-term track construction does not involve Jiangmen and Zhaoqing, two relatively marginal cities, their growth rate is 0. In the other seven cities, the growth rates of total emigration are all slightly higher than that of total emigration. The number of immigrants or emigrants of

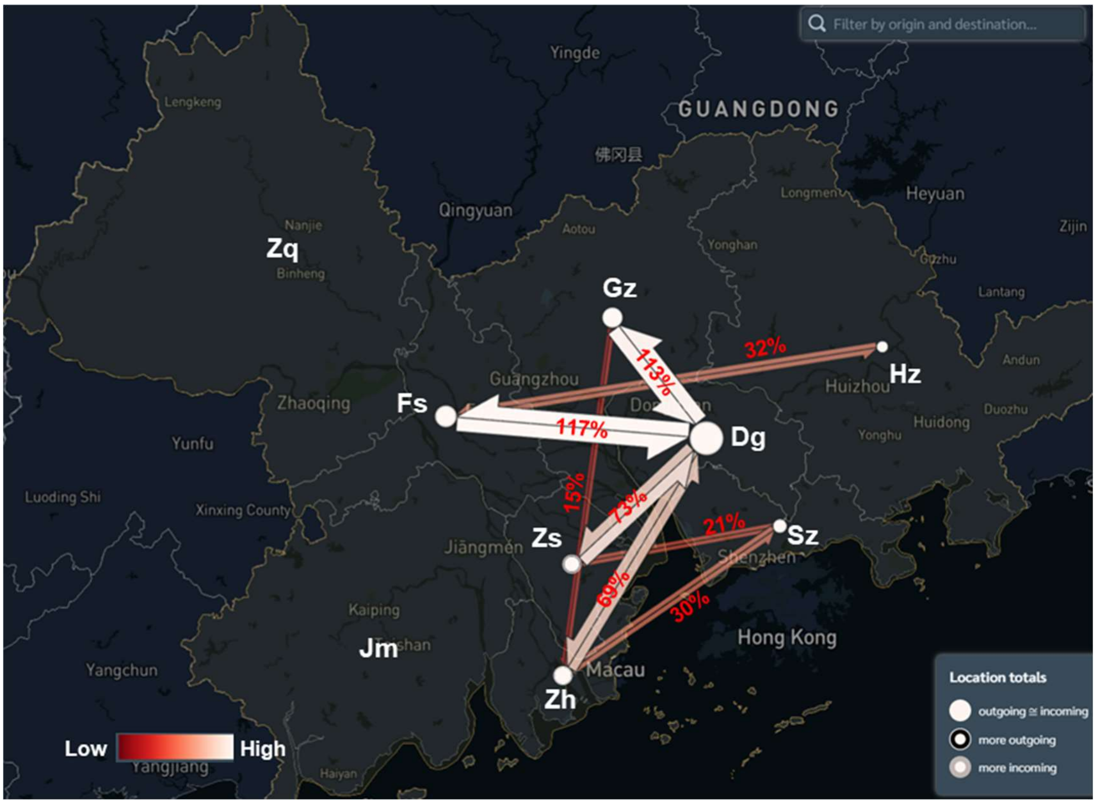


Figure 6. The Growth Rates of the Estimated Flows Affected by the Short-term Railway Planning by 2025 to the Original Estimates （Source: Author）

Dongguan within the GBA in 2025 will be far more than 50% higher than the predicts of current situations, because we do not take into account the growth of other socio-economic indicators. However, the completion of the short-term railway planning seems to have limited contribution to the increase of total migration for Huizhou and Zhongshan. It means that with the implementation of new intercity rail transit in the next five years, Dongguan is likely to face tougher demand for housing, competition for jobs and potential social segregation problems.

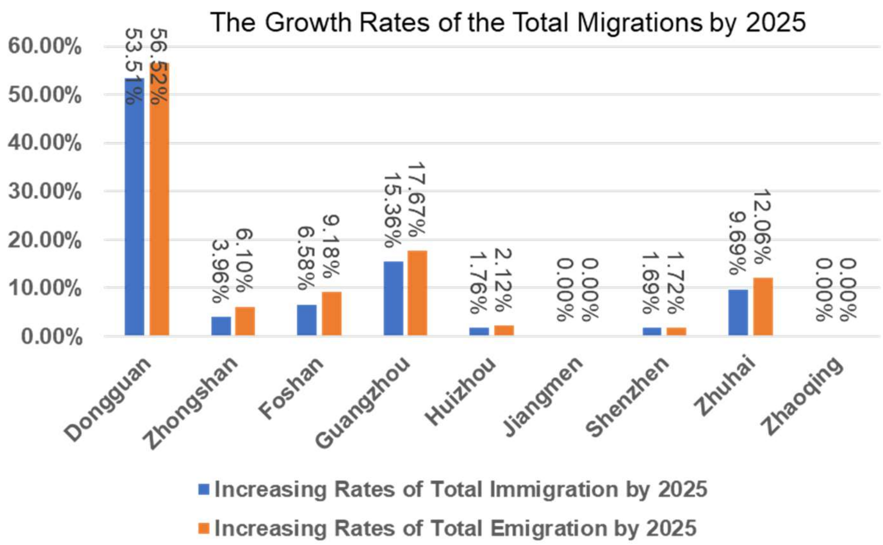


Figure 7. The Histogram of the Growth Rates of the Total Immigration or Emigration by 2025 in the GBA （Source: Author）

Impacts of the Long-term Rail Construction

The estimated migration pattern in 2035 (Figure 8) after the achievement of the 'One-hour Metropolitan Area' of the GBA looks quite similar to the estimated flows in 2025 (Figure 5). The reason behind this limited difference could be interpreted by the growth rates of the affected migration flows, as shown in Figure 9. Although the accomplishment of the one-hour commuting circle within the GBA will facilitate many long-term migration flows, especially the flows from or to Huizhou, these flows are all originally quite small in the estimated picture in 2025. Therefore, these increases did not result in significantly visual changes across the whole migration network. However, it is worth to noting that the long-term planning may greatly enhance the possibility of population mobility between the four cities on the geographical edge of the GBA, including: Huizhou, Zhaoqing, Jiangmen and Zhuhai, where the commutes now takes an hour or even two.

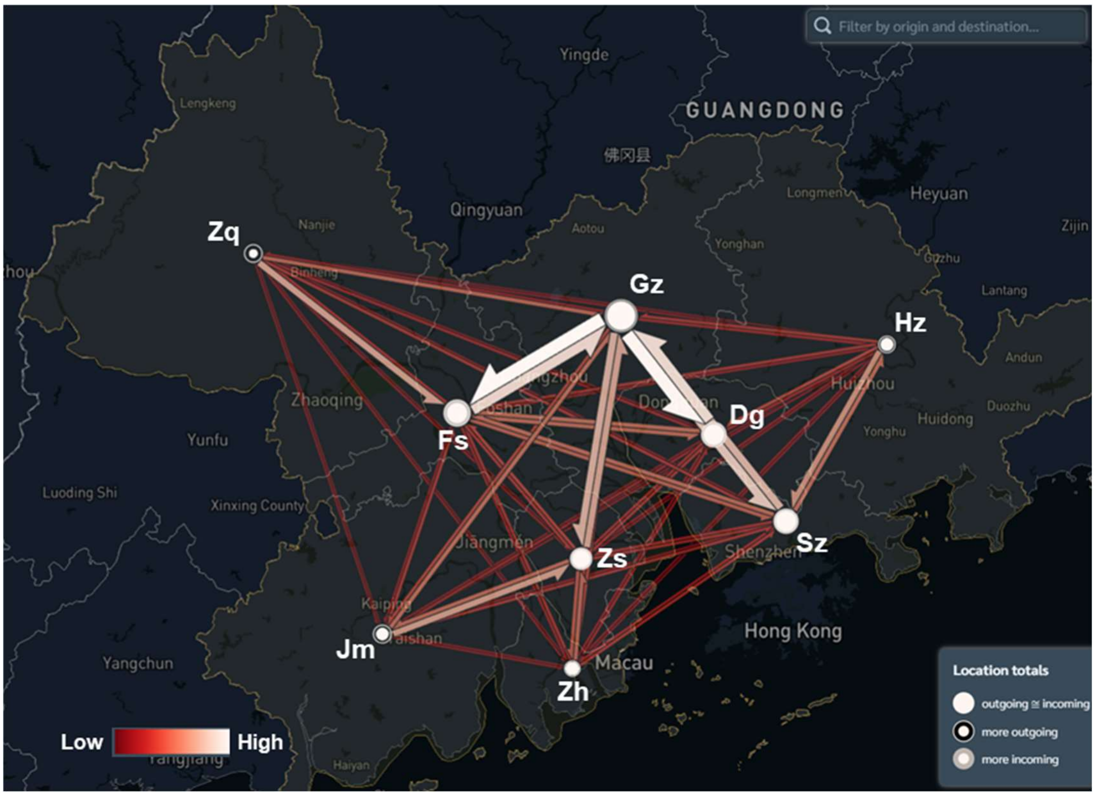


Figure 8. *The Estimated Flows based on the Impacts of the Long-term Railway Planning by 2035 of the GBA （Source: Author）*

The growth rates of the total migration by 2035 are shown below in the Figure 10. Since the minimum travel times by railway between Guangzhou and all other cities are currently lower than 1 hour, the total number of migrants from or to Guangzhou will stay the same. In contrast to the growth rate in 2025, the growth rate of total migration in all cities will be greater than that of total migration out in 2035. From 2025 to 2035, Huizhou and Zhuhai will witness the biggest increases in the long-term migration, without considering the changes in other socio-economic indicators.

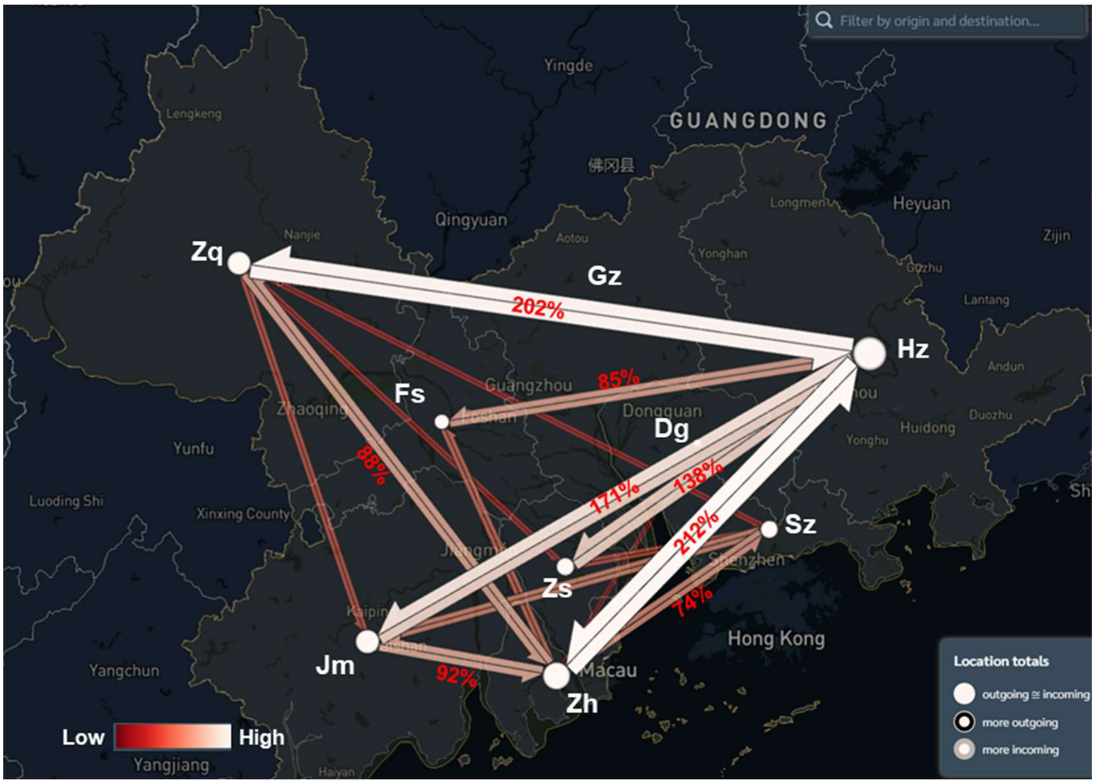


Figure 9. *The Growth Rates of the Estimated Flows Affected by the Long-term Railway Planning by 2035 to the Original Estimates* （Source: Author）

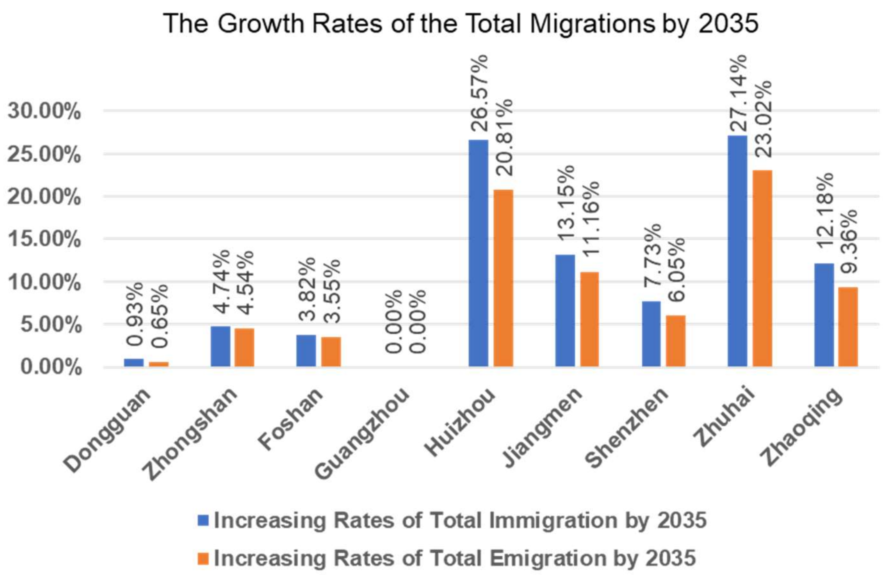


Figure 10. *The Histogram of the Growth Rates of the Total Immigration or Emigration by 2035 in the GBA* （Source: Author）

Conclusion

This paper has visualised and simulated the internal migration flows and the potential impacts of short-term and long-term inter-city railway planning on these migration flows within the GBA mainly by using the unconstrained gravity models. In addition, the relationships between these migration flows and different indicators, including socio-economic and travel cost factors, have also been quantitatively revealed based on the results of models. These findings can serve the planning decision makers to optimize urban functions and the allocations of public resources to respond to the changes in future migration patterns within the GBA.

Although the gravity model employed in this study shows significant applicability in demonstrating aggregate spatial interaction in China, this modelling approach excludes interpretations of the micro choice-making mechanism of individuals within the system (Hua and Porell, 1979; Sheppard, 2010; Yan, 2014; 2017). Therefore, future studies could employ many other models that attempt to simulate the migration behaviours at a micro level, such as the Intervening Opportunity Model (Stouffer, 1940) and the Random Walk Model (Brockmann et al., 2006), thus to verify or compare with the results of this experiment.

Furthermore, since there are more and more spatio-temporary movement data provided by mobile-phone network operators or mobile applications that record users' location changes, the method that using the total daily movement before the Spring Festival can be employed to smaller geographical levels in future studies to overcome the limited availability of long-term migration O-D data. Except for more granular analyses and visualisations, future studies could also be devoted to enriching and optimizing the indicators that introduced into the model employed in this study, especially the indicator of travel cost that might be much complicated in real world with the considerations about the number of train schedules with different commuting times, the possibility of transfers in intermediate cities, and the relative local accessibility of the train stations.

References

Anderson J and Wincoop E (2003) Gravity with Gravitas: A Solution to the Border Puzzle. *The American economic review, 93(1),* pp.170–192.

Baidu Map Migration Platform (2020) *Migration data of cities in china during the travel Rush of Spring Festival,* (http://qianxi.baidu.com/) – accessed March 2020.

Baidu Map Route Matrix API (2020) *Travel time by driving between different locations,* (http://lbsyun.baidu.com/index.php?title=webapi/route-matrix-api-v2) – accessed June 2020.

Batty M and Mackie S (1972) The calibration of gravity, entropy, and related models of spatial interaction. *Environment and Planning A*, *4*(2), pp.205-233.

Batty M (1972) Recent developments in land-use modelling: a review of British research. *Urban Studies*, *9*(2), pp.151-177.

Batty M (1978) Reilly's challenge: new laws of retail gravitation which define systems of central places. *Environment and planning A*, *10*(2), pp.185-219.

Batty M (2013) *The new science of cities.* MIT press.

Batty M (2018) Visualizing aggregate movement in cities. *Philosophical transactions. Biological sciences,* 373(1753), p.20170236.

Borjas GJ (1989) Economic theory and international migration. *International migration review*, *23*(3), pp.457-485.

Boyle PJ, Flowerdew R and Shen J (1998) Modelling inter-ward migration in Hereford and Worcester: the importance of housing growth and tenure. *Regional studies*, *32*(2), pp.113-132.

Brockmann D, Hufnagel L and Geisel T (2006) The scaling laws of human travel. *Nature*, *439*(7075), pp.462-465.

Champion AG, Fotheringham S, Rees P, Boyle P and Stillwell J (1998) *The determinants of migration flows in England: a review of existing data and evidence: a report prepared for the Department of the Environment, Transport and the Regions*. Newcastle upon Tyne: Department of Geography, University of Newcastle upon Tyne.

Charles-Edwards E, Wilson T and Sander N (2015) Visualizing Australian internal and international migration flows. *Regional Studies, Regional Science*, *2*(1), pp.432-434.

Chen@CSDN (2020) *Python scripts to capture data from Baidu Migration Platform,* (https://blog.csdn.net/Leaze932822995/article/details/104731505?ops_request_misc=%7B%22request%5Fid%22%3A%22159979427319724839205033%22%2C%22scm%22%3A%2220140713.130102334..%22%7D) – accessed March 2020.

Davis KF, D'Odorico P, Laio F and Ridolfi L (2013) Global spatio-temporal patterns in human migration: a complex network perspective. *PloS one*, *8*(1), p.e53723.

Dennett A (2012) Estimating flows between geographical locations: 'get me started in' spatial interaction modelling. *Technical report*.

Dennett A (2018) Modelling population flows using spatial interaction models. *Australian Population Studies,* 2(2), pp.33–58.

Dennett A and Wilson A (2013) A multilevel spatial interaction modelling framework for estimating interregional migration in Europe. *Environment and Planning A*, *45*(6), pp.1491-1507.

Dorling D (2012) *The visualization of spatial social structure.* John Wiley & Sons.

Fan CC (2013) Modeling Interprovincial Migration in China, 1985-2000. *Eurasian geography and economics,* 46(3), pp.165–184.

Flowerdew R and Aitkin M (1982) A method of fitting the gravity model based on the Poisson distribution. *Journal of regional science*, *22*(2), pp.191-202.

Fotheringham AS and O'Kelly ME (1989) *Spatial interaction models: formulations and applications* (Vol. 1, p. 989). Dordrecht: Kluwer Academic Publishers.

Li F, Feng Z, Li P. and You Z (2017) Measuring directional urban spatial interaction in China: A migration perspective. *PloS one*, *12*(1), p.e0171107.

Li X, Li X and Du Z (2019) Spatio-temporal Characteristics and Factors of Migration in the Rapid Urbanization Area: Based on the Points-based Participants in 2010-2015. *Scientia Geographica Sinica, 39(5*), pp.742-749.

Li X, Zhou J, Huang Y and Huang M (2018) The spatial structure of Guangdong-Hong Kong-Macao Greater Bay Area examined from the perspective of mega-city region. *Progress in Geography, 37(12*), pp.1609-1622.

De Goei B, Burger MJ, Van Oort FG and Kitson M (2010) Functional polycentrism and urban network development in the Greater South East, United Kingdom: Evidence from commuting patterns, 1981–2001. *Regional Studies*, 44(9), pp.1149-1170.

Greenwood MJ, Hunt GL, Rickman DS and Treyz GI (1991) Migration, regional equilibrium, and the estimation of compensating differentials *The American Economic Review*, *81*(5), pp.1382-1390.

Greenwood MJ and McDowell JM (1991) Differential economic opportunity, transferability of skills, and immigration to the United States and Canada. *The Review of Economics and Statistics*, pp.612-623.

Gu C and Pang H (2008) The spatial interaction and hierarchical structure of urban systems in China based on gravity model. *Geographical Research, 1(1)*.

Hair JF Jr, Anderson RE, Tatham RL and Black WC (1995) *Multivariate Data Analysis (3rd ed,*. New York: Macmillan.

Hair JF, Risher JJ, Sarsted M and Ringle CM (2019) When to use and how to report the results of PLS-SEM. *European Business Review*.

Hall PG and Pain K eds (2006) *The polycentric metropolis: learning from mega-city regions in Europe*. Routledge.

Hall P (1999) Planning for the mega-city: A new Eastern Asian urban form? [M]. Hunt, UK: Ashgate

Hanssens H, Derudder B, Van Aelst S and Witlox F (2013) Assessing the functional polycentricity of the mega-city-region of Central Belgium based on advanced producer service transaction links. *Regional Studies*, *48*(12), pp.1939-1953.

He J, Li C, Yu Y, Liu Y and Huang J (2017) Measuring urban spatial interaction in Wuhan Urban Agglomeration, Central China: A spatially explicit approach. *Sustainable cities and society*, *32*, pp.569-583.

Hua CI and Porell F (1979) A Critical Review of the Development of the Gravity Model *International Regional Science Review,* 4(2), pp.97–126.

Hu K and Zheng X (2015) Modelling the inter-provincial migration in China. *Science of Surveying and Mapping, 40(6),* pp.61-66.

Hyman GM (1969) The calibration of trip distribution models. *Environment and Planning A*, *1*(1), pp.105-112.

Ilya Boyandin (2019) *Visualisation Tool of Flowmap.blue,* (https://flowmap.blue/) – accessed August 2020.

Karemera D, Oguledo VI and Davis B (2000) A gravity model analysis of international migration to North America. *Applied economics*, *32*(13), pp.1745-1755.

Mansfield ER and Helms BP (1982) Detecting multicollinearity. *The American Statistician*, *36*(3a), pp.158-160.

Martínez-Zarzoso I (2011) The log of gravity revisited. *Applied Economics*, *45*(3), pp.311-327.

Mei Z, Xu S, Ou Y and Shi C (2012) The spatio-temporal evolution of the spatial interaction in the urban agglomeration within the Pearl River Delta in recent 20 years. *Scientia Geographica Sinica, 32(6),* pp.694-701.

National Development and Reform Commission (China) (2020) *Guangdong-Hong Kong-Macao Greater Bay Area Intercity Railway Construction plan,* (https://www.ndrc.gov.cn/xxgk/zcfb/tz/202008/t20200804_1235517.html) – accessed August 2020.

Norris RE (1972) Migration as Spatial Interaction. *Journal of Geography,* 71(5), pp.294–301.

Poot J, Alimi O, Cameron MP and Maré DC (2016) The gravity model of migration: the successful comeback of an ageing superstar in regional science.

Ramos R and Suriñach J (2017) A gravity model of migration between the ENC and the EU. *Tijdschrift voor economische en sociale geografie*, *108*(1), pp.21-35.

Reilly WJ (1931) The Law of Retail Gravitation (Pilsbury, New York), republished in 1953.

Samelltiger@GitHub *Python scripts to capture data from Baidu Migration Platform,* (https://github.com/samelltiger/baidu_qx) – accessed March 2020.

Sander N, Stillwell J and Lomax N (2018) Circular migration plots. In *The Routledge Handbook of Census Resources, Methods and Applications.* Routledge, pp. 203–209.
Shen J (1999) Modelling Regional Migration in China: Estimation and Decomposition. *Environment and Planning A,* 31(7), pp.1223–1238.
Shen J (2017) Modelling interregional migration in China in 2005–2010: The roles of regional attributes and spatial interaction effects in modelling error. *Population, Space and Place*, *23*(3), p.e2014.
Sheppard. ES (2010) Theoretical Underpinnings of the Gravity Hypothesis. *Geographical analysis,* 10(4), pp.386–402.
Stewart JQ (1950) Potential of Population and Its Relationship to Marketing. *Operational research quarterly, 1(3), p.47.*
State Council of the People's Republic of China (2019*) Outline Development Plan for the Guangdong-Hong Kong-Macao Greater Bay Area,* (http://www.gov.cn/gongbao/content/2019/content_5370836.htm) – accessed April 2020.
Statistics Bureau of Guangdong Province (China) (2020) *Statistical Yearbook of Guangdong in 2019,* (http://stats.gd.gov.cn/) – accessed May 2020.
Stillwell JCH (1978) Interzonal migration: some historical tests of spatial-interaction models. *Environment and Planning A*, *10*(10), pp.1187-1200.
Stouffer SA (1940) *Intervening Opportunities: A Theory Relating Mobility and Distance. American sociological review,* 5(6), pp.845–867.
Sun WY, Qiu KF, He B, Wang H (2019) Illustration of the Guangdong-Hong Kong-Macao Greater Bay Area based on big data. *Cheng Plus,* (https://mp.weixin.qq.com/s/xCapdkapz0eY79YbitzVWA) – accessed May 2020.
Xu T and Yao Y (2018) Urban Population Migration and Housing Price Fluctuation: An Empirical Research Based on the Census Data and Baidu Migration Data. *Journal of Jiangxi University of Finance and Economics*, (1), p.2.
Williams H (2019) The contribution of Sir Alan Wilson to spatial interaction and transport modelling. *Interdisciplinary Science Reviews*, *44*(3-4), pp.232-248.
Wilson AG (1967) A Statistical Theory of Spatial Distribution Models. *Transportation Research* 1 (3): 253–269.
Wilson A (1970) *Entropy in urban and regional modelling*. London: Pion.
Wilson A (1971) A family of spatial interaction models, and associated developments. *Environment and Planning A*, *3*(1), pp.1-32.
Wood J, Dykes J and Slingsby A (2010) Visualisation of origins, destinations and flows with OD maps. *The Cartographic Journal*, *47*(2), pp.117-129.
Wood J, Slingsby A and Dykes J (2011) Visualizing the dynamics of London's bicycle-hire scheme. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *46*(4), pp.239-251.
Xu T and Yao Y (2018) The relationship between inter-city migration and housing price – the empirical study based on Census and migration data from Baidu. *Journal of Jiangxi University of Finance and Economics, (1),* pp.11-19.
Yan XY, Zhao C, Fan Y, Di Z and Wang WX (2014) Universal predictability of mobility patterns in cities. *Journal of The Royal Society Interface*, *11*(100), p.20140834.
Yan X (2017) Research progress of spatial interaction and networks. *Science & Technology Review, 35(14),* pp.15-22.

Yang P (2019) Verification of urban symbiosis model in Guangdong-Hong Kong-Macao Greater Bay Area based on WeChat data. *2019 Annual National Planning Conference in China (16 Regional Planning and Urban Economy).*

Yeh AGO and Chen Z (2020) From cities to super mega city regions in China in a new wave of urbanisation and economic transition: Issues and challenges. *Urban Studies*, *57*(3), pp.636-654.

Zhang Z and Yang S (2013) The review of concepts, data and issues in the study of floating population. *Chinese Journal of Population Science, (6),* pp.102-112.

Zhao YM, Zeng A, Yan XY, Wang WX and Lai YC (2016) Unified underpinning of human mobility in the real world and cyberspace. *New Journal of Physics*, *18*(5), p.053025.

Zipf GK (1946) The P 1 P 2/D hypothesis: on the intercity movement of persons. *American sociological review*, *11*(6), pp.677-686.

12306 China Railway (2020) *Travel time by trains and train schedule,* (https://www.12306.cn/index/) – accessed June 2020.

Supplementary Materials

Flow Data Processing

Since there are no available O-D migration data at inter-city level, we obtained the total number of immigrants or emigrants from one certain city to all other cities within the same province from the annual *Statistical Yearbook* released by local government. This kind of the total migration from one certain city could be used to relatively measure the potential attractiveness of one city to other cities in the same province as shown in Figure S1 (Stewart, 1950).

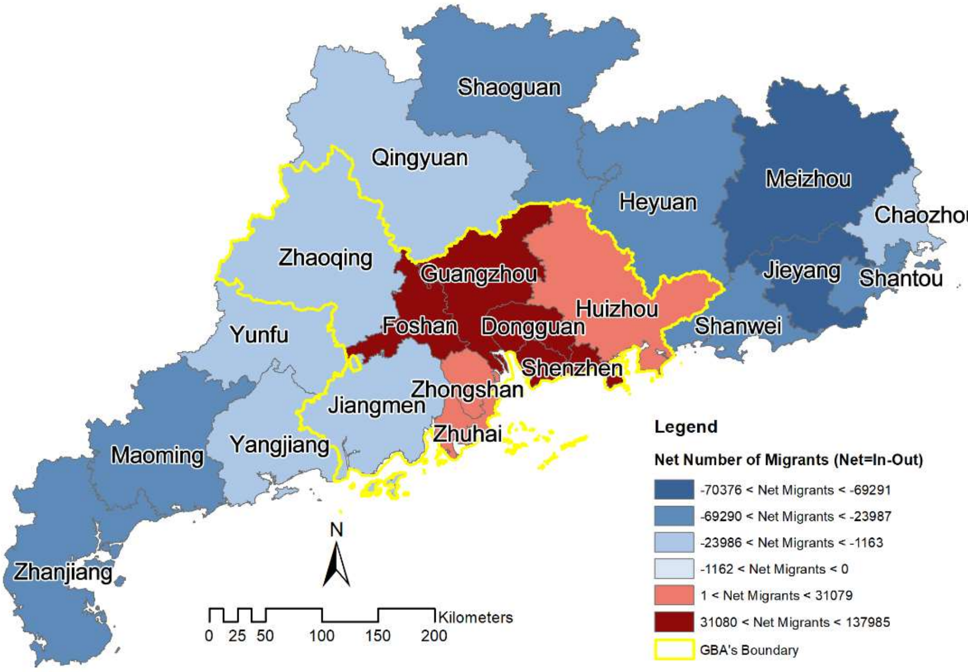


Figure S1*. The Map of the Net Migrants of Cities with in Gd Province from 2019 Statistical Yearbook of Gd* （*Source: Author*）

In order to better understand the relationship between short-term trips during this period and the long-term migration, this study takes the two cities – Shenzhen and Zhaoqing - in the GBA - with their aggregated move-in and move-out index within the Gd province for further illustration as we show in Figure S3 and Figure S4. As one of the most developed cities in Gd province, it can be clearly found that during return-home period (10/1-25/01), a large number of move-out trips happened from Shenzhen to other cities with a sharp drop in arrivals. In terms of Zhaoqing where the urban scale is only at the lower-middle level among the 21 cities in Gd Province, there has been a huge increase in the aggregate move-in trips. The different movement patterns of these two cities during the return-home rush are relatively consistent with the previously theoretical findings of long-term migrations, namely the bigger and more developed a city is, the more attractive it is to migration from other cities (Zipf, 1946). Meanwhile, similar findings on the migration patterns during this special period also apply to other cities in Gd province.

Comparing the movement index of the same period in 2019 (dotted lines in Figure S3 and S4), we can find that around Jan. 25th 2020, there is like a watershed. It seems that the number of trips before Jan. 25th has not been affected by the epidemic (Covid-19), which is basically consistent with the number in 2019. However, after Jan. 25, the outbreak of *Covid-19* has largely hindered the movement of people.

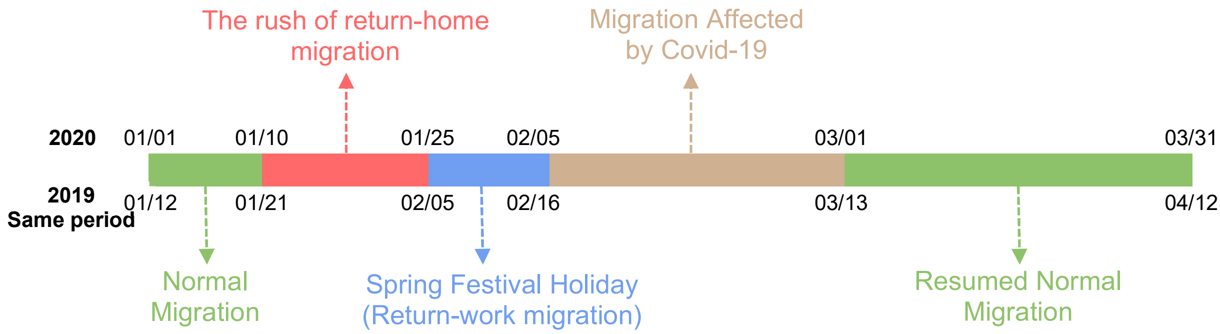


Figure S2. *The Explanatory Timeline of Spring Festival Migration in 2020 and the Same Period in 2019*

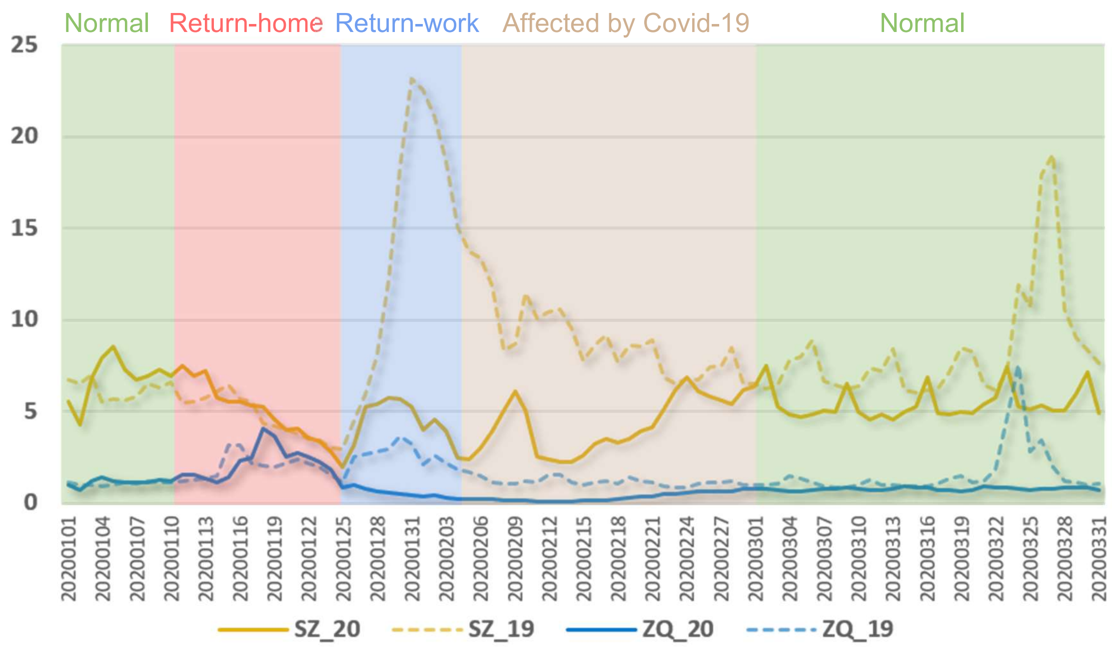


Figure S3. *Move-in Trends of Shenzhen (SZ) and Zhaoqing (ZQ) in 2020 and 2019*

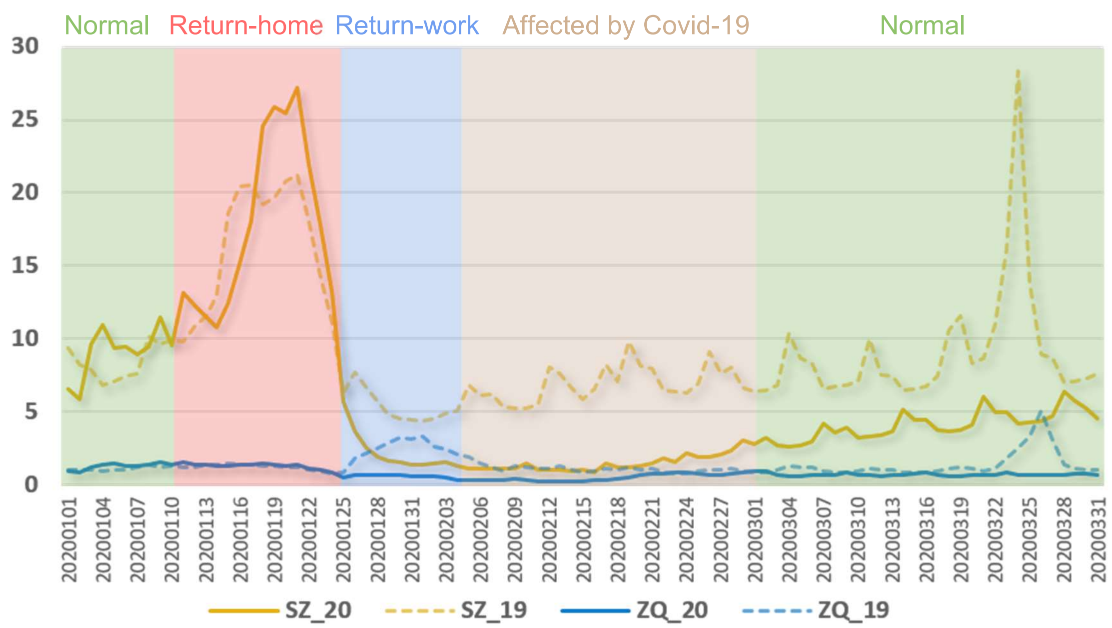


Figure S4. *Move-out Trends of Shenzhen (SZ) and Zhaoqing (ZQ) in 2020 and 2019*

Therefore, based on the above findings, this study suggests that the sum of return-home movements between Jan. 10th and Jan. 25th (16 days) can be used to represent the actual long-term migration in China. The totally net short-term trips during the return-home period for all 21 cities within Gd province are demonstrated in the Figure S5, which is also relatively consistent with the actual long-term migration pattern recorded in the *2019 Statistical yearbooks* mentioned above (Figure S1). Furthermore, the scatter plot of these two types of migration data which we show in Figure S6 has a fitting line which demonstrates that there is a strong linear relationship between them.

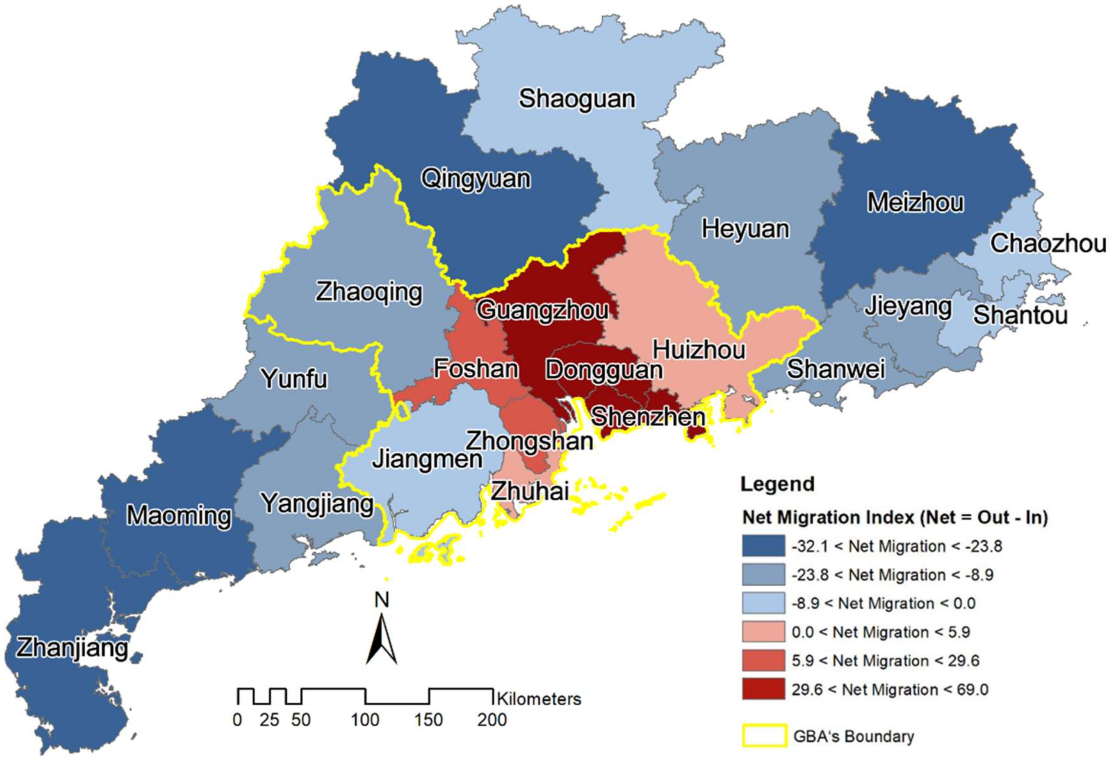


Figure S5. *The Map of the Sum of Net Movement Indices of Cities with in Gd Province between Jan.10th to Jan.25th* (Source: Author)

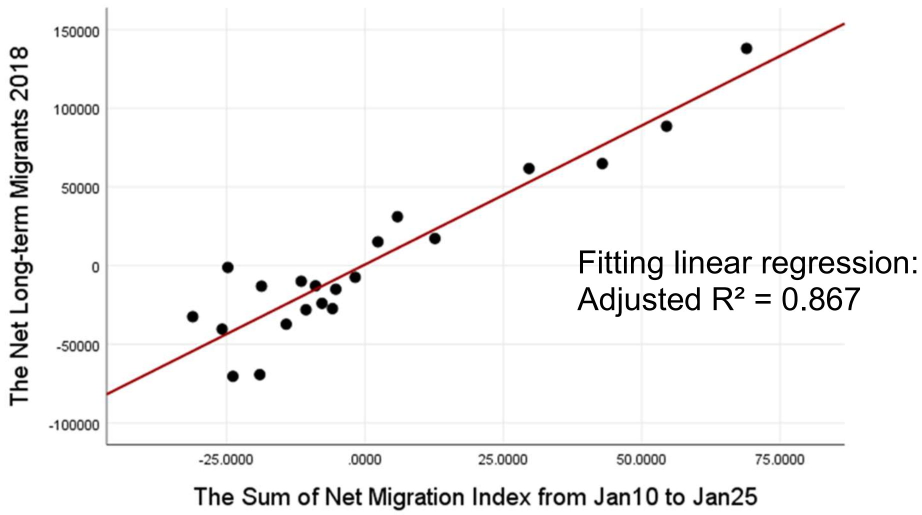
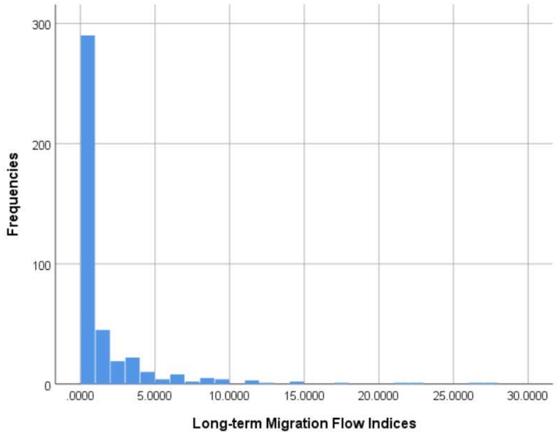


Figure S6. *The Scatter Plot of Actually Long-term Net Migrants (2019 Statistical Yearbook) with the Sum of Net migration Indices from Jan.10th to Jan.25th of 21 Cities within Gd Province* (Source: Author)

In conclusion, the asymmetric matrix of O-D long-term migration flows of 21 cities within Gd province were now obtained from the sum of short-term movement indices between Jan.10th to Jan.25th. In the data processing, after calculating the sum of the return-home trip indices, we also need to exchange the destinations and origins, because the direction of the return-home rush is opposite to the direction of long-term migration. The frequencies and descriptive statistics of the final long-term migration dataset of the 21 cities are presented below (Figure 7), containing a total of 420 flows.



Descriptive Statistics

N	420
Mean	1.592
Min.	0.0052
Max.	27.478
Std. Deviation	3.331
Variance	11.097

Figure S7. *The Frequencies and Descriptive Statistics of the Final Long-term Migration Dataset for the whole Gd province* (*Source: Author*)

Estimation of Travel-time Changes

In order to simulate the impacts of the future Rail Transit Planning on the long-term migration flows of the GBA, the changing travel time by the railway between cities are required to collect. Since the Chinese government just approved the concept map of the future railway planning for the GBA (Figure S8) on July 31st 2020, most of the changes in travel time could only be roughly estimated from the available length and speed for some railway lines. This railway plan is divided into the short-term phase (2020-2025) and the long-term phase (2025-2035), and the ultimate goal of this plan is to keep the commuting time between the cities within the GBA no more than 1 hour by 2035, implementing namely "One-hour GBA".

In terms of the short-term planning, as shown in the Figure S8, Jiangmen and Zhaoqing are the only cities without any railways under construction or approved for construction by 2025. The estimated travel time by intercity railways from the short-term planning are summarized in the Table S1 below and the related railway information can be found in Appendix E. In addition, the travel times for the long-term planning by 2035 will be estimated by setting all travel time over 1 hour to 1 hour based on the results in the Table S1, since there is extremely limited information about the long-term rail planning up until now.

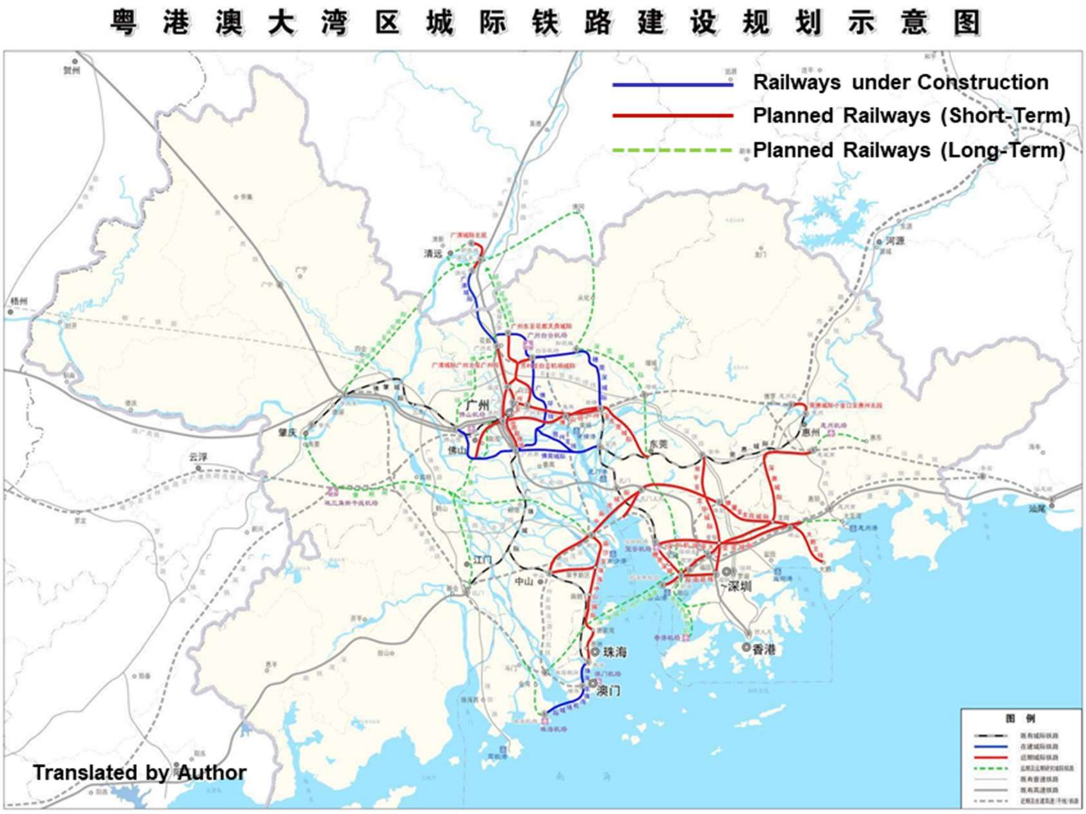


***Figure S8**. The Concept Map of the Intercity Railway Construction Plan for the GBA (Source: National Development and Reform Commission, July 31 2020, avaiable at:* https://www.ndrc.gov.cn/xxgk/zcfb/tz/202008/t20200804_1235517.html*)*

***Table S1.** The Table of the Estimated Travel time by Railway after the Completion of the Short-term Rail Constructions（Source: Author）*

	Dg	Fs	Gz	Hz	Jm	Sz	Zh	Zq	Zs
Dg		1.267 (0.667)	0.467 (0.25)	0.683	1.15	0.517	1.65 (1.067)	0.933	1.183 (0.75)
Fs	1.267 (0.667)		0.3	2.1 (1.667)	0.9	0.967	1.35	0.317	0.783
Gz	0.467 (0.25)	0.3		1	0.517	0.483	0.883 (0.817)	0.567	0.3
Hz	0.683	2.1 (1.667)	1		2.283	0.433	2.567	2.5	2.05
Jm	1.15	0.9	0.517	2.083		1.467	1.717	1.283	0.183
Sz	0.517	0.967	0.483	0.433	1.467		1.967 (1.584)	1.167	1.483 (1.267)
Zh	1.65 (1.067)	1.35	0.917 (0.817)	2.567	1.717	1.967 (1.584)		1.683	0.317
Zq	0.933	0.317	0.567	2.5	1.283	1.167	1.683		1.183
Zs	1.283 (0.75)	0.783	0.3	2.05	0.183	1.483 (1.267)	0.317	1.183	

***Note:** The values in red color are the estimated travel time. First column is the origins.*

Visualisation Map of Current Migration Pattern for the Gd Province

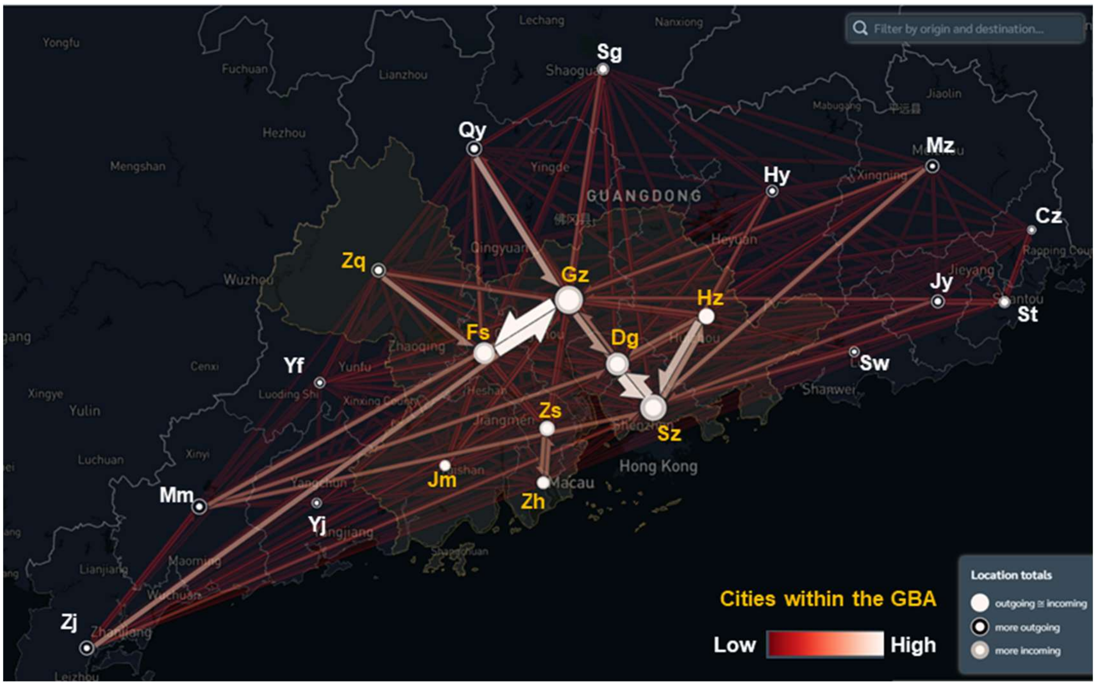


Figure S9. *The Flow Map of the Long-term Migration Flows between the 21 cities within Guangdong Province* （*Source: Author*）

Visualisation Map of Change Rates of the Residuals

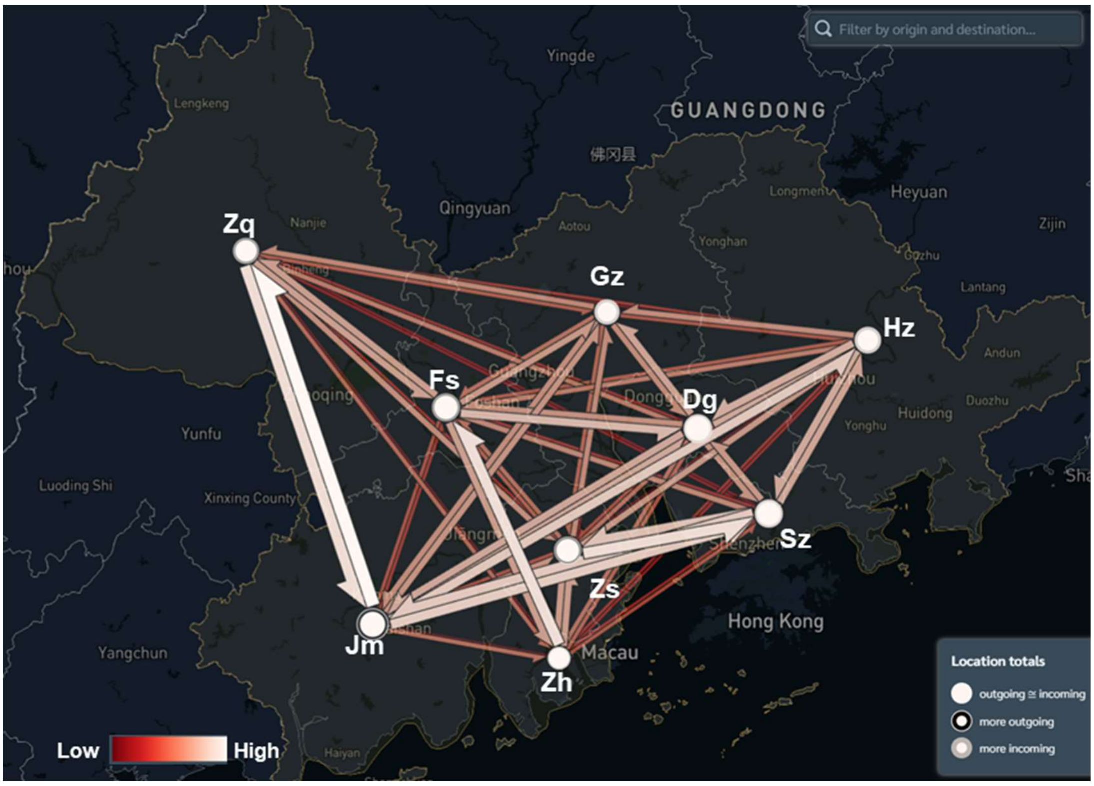


Figure S10. *The Absolute Rates of the Residuals to the Original Flows based on the Driving Time between the 9 cities within the GBA (unit: %)* （*Source: Author*）