# Visualizing and Simulating the Spatial Interaction of Internal Migration in Polycentric Mega-city Regions: An Application in the Great Bay Area, Guangdong Province, China

Candidate: Zunhao Zhang

Submission Date: September 14th 2020

Module Name: MRes Dissertation

Module ID: CASA0004

Supervisor: Prof. Michael Batty

Word Count: 11815

This dissertation is submitted in part requirement for the MRes in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.

# Abstract

Since the rapid development of meg-city regions around the world, the closely spatial interaction that takes place on urban networks has been a popular agenda in the studies on those urban agglomerations. 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, understanding and simulating the spatial interaction within the GBA will contribute to making tailored planning decisions to this area. Meanwhile, as one of the most commonly employed methods to understand and simulate spatial interaction, gravity model has a large number of empirical applications for different types of spatial interaction.

However, due to the limited availability of Origin-Destination (O-D) data, most studies on spatial interaction based on gravity models in China 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 visualizing and simulating the internal migration flows, one of the most common form of spatial interaction, within the GBA that includes the municipalities in Guangdong province. Unconstrained gravity models are used to simulate internal migration flows and measure the impacts of future inter-city railways on those flows within study areas. In addition, 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.

# Declaration

I, Zunhao Zhang, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 11815 words in length.

# Table of Contents

# List of Figures

# List of Tables

# List of Acronyms and Abbreviations

| | |
|---|---|
| **API** | Programming Interface |
| **GBA** | Great Bay Area |
| **Gd** | Guangdong |
| **GDP** | Gross Domestic Product |
| **O-D** | Origin-Destination Application |
| **PM-CR** | Polycentric Mega-city Region |

# Acknowledgements

This paper would not have been possible without the help of others, especially during this particularly difficult period of time in 2020 with the outbreak of Covid-19. First of all, I am sincerely grateful to my supervisor, Prof. Michael Batty, for his patient guidance and enthusiasm support throughout my dissertation. Prof. Michael Batty not only provide me with many inspired feedbacks, but also offer me the most understanding of the difficulties I encountered in writing my thesis, both personally and academically.

Secondly, I would like to thank the SIMETRI Project who gave me the valuable opportunity to join in this project.  Meanwhile, my thanks are also owed to all members in the SIMETRI Project, especially to Dr Adam Dennett, Minghang Hong, Bowen Zhang, Prof. Anthony Yeh and Dr Zifeng Chen, for their help and sharing data.

I also would like to express my thanks to Bin Chi, one of CASA's doctoral student, for her selfless advice and the time she spent discussing with me.

Finally, I owe deep gratitude to my mom, dad and fiancée who have accompanied and encouraged me throughout this year’s study.

# Chapter 1 Introduction

## 1.1 Context and Motivation

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 with high-close internal spatial interactions (He et al., 2017), and put forward many similar descriptive concepts, such as Mega-city Regions, Polycentric Mega-city Regions (Hall and Pain, 2006) and Super Mega-city Regions (Yeh and Chen, 2020). 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).

As an important bay area in China, the concept of the Greater Bay Area (GBA) in the Pearl River Delta was elevated to a national development strategy in 2017, covering nine cities in Guangdong (Gd) Province as well as Hong Kong and Macau. Since there are several cities normally considered to be the cores within the GBA, such as Guangzhou and Shenzhen, the GBA is also considered to be a polycentric mega-city region. Meanwhile, the scale of the GBA is on a par with the other three famous bay areas in New York, Tokyo and San Francisco in terms of different socio-economic aspects (Hall, 1999; Li et al., 2018). Furthermore, Chinese government is also committed to facilitate the coordinated development of the GBA through the implementation of the overall planning outline in 2019 and intercity rail transit planning in 2020. Predictably, the internal connections within the GBA will be closer and more frequent.

Therefore, it is of great significance to explore the current spatial interaction pattern 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.

## 1.2 Research Question and Objectives

The gravity model has been one of the most commonly used methods to understand and simulate the spatial interaction that takes place on urban networks (Batty, 1972; Haynes and Fotheringham, 1984; Poot et al., 2016; Wilson, 1971), and it has a large number of empirical applications for different types of spatial interaction (Borjas, 1989; Karemera et al., 2000; Reilly, 1931; Shen, 2017). However, due to the lack of the observed Origin-Destination (O-D) data, most studies on spatial interaction based on gravity models in China have remained at the theoretical level or have used assumed parameters values to estimate the flows (Gu and Pang, 2008). 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 and in the cities that exist within the Pearl River Delta region. Hence, the corresponding research questions are put forward as follows: the basic question is:

> What are the internal migration patterns within the GBA and Gd province and can we simulate these migration flows and the impacts of the planning policies on the future migration using gravity models?

Guided by this main question, there are several objectives for this paper which we list according to the analytical steps in this research. These are as follows:

1. Review the relevant literature and empirical studies on internal migration flows which have been simulated using gravity models;
2. Obtain and process the various datasets which can represent the long-term migration O-D data for the GBA and Gd province;
3. Visualise and understand the current spatial patterns of internal migration flows within the GBA and Gd province;
4. Simulate the internal migration patterns both for the GBA and Gd province by gravity model and make comparisons;
5. Interpret the influence of different factors on the internal migration flows within the GBA;
6. Quantitatively measure the potential impacts of future rail construction on the intercity migration flows within the GBA.

## 1.3 Research Scope

The scope of this research is to quantitatively explore the internal migration patterns within the GBA and Gd province at the intercity level and simulate these migration flows based on the gravity model famulated by Wilson (1970;1971). This paper accepts the theoretical basis of the gravity model that migration flows can be quantitatively measured by the attraction of the origins and destinations and the travel cost between them. The subjects of this study can be further defined as the aggregated migrants who have changed the cities of long-term residence by 2020 between cities within the study areas, and the historical migration processes are not considered, such as transit migration.

## 1.4 Report Structure

There are four more chapters after this introduction chapter involved in this report. Chapter 2 aims at reviewing the previous literature and empirical studies on the relevant concepts, modelling methods and research gap of human migration in the context of spatial interaction and urban network. The employed dataset, analytical framework and approaches and potential limitations will be presented in the Chapter 3. Chapter 4 will interpret and discuss the visualisation outcomes and simulation results of the internal migration flows within the GBA and Gd province. Finally, Chapter 5 will conclude some key arguments to answer the research questions and propose some suggestions about further studies and planning policies.

# Chapter 2 Literature Review

This chapter contextualizes relevant literature and empirical studies on migration to extract the research gap and applicable methods. The relationship between spatial interaction and polycentric mega-city region will be first introduced. Then the review turns to the development of gravity models employed in the analysis of the spatial interaction, and extends to some previous findings about human migration based on this method. The internal migration in the context of China will be discussed in the last part.

## 2.1 Spatial Interaction and Polycentric Mega-city Region

Under the rapid progression of globalization and urbanization, large regions with highly interconnected urban networks have gradually replaced the single city to become the basic pattern of spatial development with respect to global competition, which has also become the current popular agenda for urban research (Batty, 2013; Hanssens et al., 2013; He et al., 2017; Yeh and Chen, 2020). As Hall and Pain (2006) suggested, the Polycentric Mega-city Region (PM-CR) refers to a large urban region where there are strong spatial interactions between the internal cities or areas which have quite different urban functions, such as those involving their population, economy and industries. This kind of effective spatial interaction supported by well-connected transitions in traffic and information is contributing to the utilisation of different specializations of the nodes that comprise these networks (De Goei et al., 2010). In addition, the past 50 years have witnessed the rapid development of some PM-CRs in China, such as those in the Pearl River Delta, and planning policy for PM-CRs has also been strongly advocated, such as the recently proposed regional planning for the GBA (Yeh and Chen, 2020).

As Batty highlighted in his book (2013), the city is a carrier of the aggregated population and interactions, and it is necessary to better understand networks and spatial interaction in the modern urban system, instead of just focusing on the locational information. The definition of the spatial interactions in the network could be basically interpreted as the exchange flows between geographies in terms of different elements, such as the human movement, trade and business investment (Norris, 1972). As one of the most important types of flow, human migration within a spatial system could be normally described by a statistical matrix. Specifically, if we label the origins and destinations of migration as *i* and *j*, then the migration flows between them could be represented as $T_{ij}$. In practice, the numbers of origins are likely to be the same as destinations within a system, but the size of the migration flows are always unequal if we exchange the origins and destinations, $T_{ij} \neq T_{ji}$ . Hence, most quantitative studies of migration assume an asymmetric matrix to represent the statistical information of flows for further analyses and modelling.

## 2.2 Gravity Models for Spatial Interaction

Since the 1960s, the gravity model has been one of the most widely applied models in the understanding and simulation of different kinds of spatial interaction between different geographies, and it achieved good empirical results (Batty, 1972; Haynes and Fotheringham, 1984; Dennett, 2018; Poot et al., 2016; Wilson, 1971). Basically, the gravity model employed in spatial interaction study is derived from Newton's Law of Gravity (1687) as follows:

$$F_{ij} = G\frac{M_i M_j}{r^2} \qquad (2.1)$$

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 retailling 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 $O_i$, the total arrival $D_j$ and the total cost $C$, the most likely travel distribution should be when the entropy of this system is at its maximum (Batty, 1972; Wilson, 1967; 1970; 1971). The equations representing these three constraints are as followings:

$$\sum_j T_{ij} = O_i \quad (2.2); \qquad \sum_i T_{ij} = D_j \quad (2.3); \qquad \sum_i \sum_j c_{ij} = C \quad (2.4)$$

After a series of mathematical derivations, Wilson (1970) provided us with the most general of all such models that he termed '*double-constrained gravity' model.* From Wilson's paper (1970), this is formulated as:

$$T_{ij} = A_i B_j O_i D_j f(c_{ij}) \qquad (2.5)$$

where $A_i$ , $B_j$ are the constants used to ensure the satisfaction of the double-constrained total flows and $f(c_{ij})$ represents some function of travel cost, time or distance or something combination of these generic terms.

Based on this, Wilson extended gravity model into a family of related models according to the known systemic constraints (1971). Except for the double-constrained gravity model above, the unconstrained gravity model (as equation 2.6) will be generated when neither the constraint formula 2.2 nor formula 2.3 is known. When only one of these two equations hold, we get a '*production constrained model'* (as equation 2.7) or an '*attraction constrained model'* (as equation 2.8).

$$T_{ij} = K O_i D_j f(c_{ij}) \qquad (2.6)$$

$$T_{ij} = A_i O_i D_j f(c_{ij}) \qquad (2.7)$$

$$T_{ij} = B_j O_i D_j f(c_{ij}) \qquad (2.8)$$

In practice, the measure of spatial interactions $\boldsymbol{T_{ij}}$ is usually distributed as a power law with the distance or travel time between origins and destinations, namely the $\boldsymbol{f(c_{ij})}$ could be replaced by $\boldsymbol{c_{ij}^{\gamma}}$, where $\boldsymbol{\gamma}$ is usually negative (Stillwell, 1978). Meanwhile, there are commonly some '*parameters*' (like $\boldsymbol{\gamma}$ ) needed to give a certain elasticity to the importance of the mass of origins and destinations (Batty and Mackie, 1972). Therefore, the generalised form of the unconstrained model (equation 2.6) is also equivalent to the following equation:

$$T_{ij} = KO_i^{\alpha} D_j^{\beta} c_{ij}^{\gamma} \qquad (2.9)$$

Since the parameters in the model, especially the friction of the travel cost $\boldsymbol{\gamma}$, need to be estimated based on the observed panel data of the study system, Wilson's family of models required further calibration. Batty and Mackie's paper (1972) provides a review and compares on various early solutions for the calibration, such as in Hyman's (1969) and Evans' (1971) work.

However, only for the unconstrained gravity model in the family, there is a widely accepted calibration method by many scholars, namely transforming the Equation 2.9 into a linear formula (equation 2.10) by taking the logarithm of both sides and estimating the parameters by regression models (Flowerdew and Aitkin, 1982).

$$\ln T_{ij} = k + \alpha \ln O_i + \beta \ln D_j + \gamma \ln c_{ij} \qquad (2.10)$$

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 (Karemera et al., 2000; 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).

In terms of some theoretical debate about the gravity model, although Wilson has provided a theoretical basis for the gravity model in the field of statistical physics, the approach excludes interpretations of the micro choice-making mechanism of individuals within the system (Hua and Porell, 1979; Sheppard, 2010; Wilson, 1967; Yan, 2014; 2017). Therefore, in order to interpret the migration behaviours at a micro level, many other models have been generated, such as the Intervening Opportunity Model (Stouffer, 1940) and the Random Walk Model (Brockmann et al., 2006). However, as Poot et al. (2016) has pointed out, we cannot expect a unified model that provides a good understanding of behaviours at both the micro and macro levels. In conclusion, the gravity model offers effective and multiple applications in analysing the spatial interactions at the macro level, such as cross-regional or international migrations.

## 2.3 Previous Research and Visualisation of Human Migration

One of the most popular foci in the study of spatial interaction over decades involves a large amount of empirical analyses employed the gravity model and its calibration process to simulate different categories of population movement in various spatial systems (Borjas, 1989; Davis et al., 2013; Dennett and Wilson, 2013; Karemera et al., 2000; Poot et al., 2016; Shen, 2017). Most of the research on the aggregate population flow emphasizes two generic types of movement: one is daily commuting trips based on spatio-temporal data, and another one is international or internal migration flows with the change of residential address, which is also the core research scope of this paper.

Although the previous researches of human migration 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, such as distance. 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; Poot et al., 2016; Greenwood et al., 1991; Shen, 2017), 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, border effect and etc (Anderson and Wincoop, 2003; Karemera et al., 2000). 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, no matter where these are daily commutes or migration flows, 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, while the sizes or other attributes of different flows will be distinguished by their colour or thickness, which could be called a flow map or an O-D map. However, there are still existing challenges in generating and interpreting such flow maps, and as Batty highlighted in his book (2013), with the growth of the number of locations in the spatial system, it becomes difficult to identify meaningful information and patterns within the flow map. In addition, the asymmetrical flow matrices or the consideration of self-flows might further complicate the visualization work.

Over the past few decades, advances in computer technology have led to more and more attempts at the visualization of flows with facing the challenges mentioned above. In the flow visualisation work in Greater London (Batty, 2013; 2018), one of the attempts is to extract and visualise the biggest flows or average vectors to simplify the information. There are also some flow visualisations converting the flow information into the attributes of the location information by calculating the migration potential or the density of the migrating population of the origins and destinations (Wood et al., 2010). Furthermore, excepting the conventional flow map, some other visualisation methods are effective for some certain dataset or focus, such as the circular migration plot 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 (Batty, 2013; 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.

## 2.4 Internal Migration in China

Based on the unique *Household Registration System*, 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). Regarding the administrative hierarchy in China, the internal migration can be categorised as inter-provincial, inter-city and inter-district migrations from large spatial scale to small scale.

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. Comparing with the lack of long-term migration data, there are several spatio-temporary data sets about the daily movements in China, which recorded by the popular mobile applications that could gain the location changes of their users, such as WeChat, Weibo, Baidu Map and TenCent. Although most of these spatio-temporary data sets are not public, it still provides more opportunities for the studies on exploring spatial interaction in China.

Since there are no available O-D flows data at city level or more smaller geographical levels, the empirical studies on the long-term migration in China are still limited to a large spatial level (Fan, 2013; Li et al., 2017; Shen 1999; 2017), such as the inter-provincial level, and the situations with other types of spatial interaction relatively same. Due to rely on the observed O-D data to estimate its parameters, most studies using gravity models in China can only take assumed parameters to theoretically reflect the strength of connections between cities or at other smaller scales (Gu and Pang, 2008; Hu and Zheng, 2015; Mei et al., 2012).

Although there are some challenges with respect to the study of internal migration in China, some previous research statistically explores the relationships between some potential factors and migration patterns. For example, Xu and Yao (2018) proposed that the increase in the total number of immigrants to a city will significantly boost local housing prices. However, the influence of the housing price seems not to contribute to the migration flows, compared with the differences in income and distance between cities.

Moreover, in terms of the analysis of the spatial interactions and urban network between cities within the GBA and the Pearl River Delta, Sun et al. (2019) argued that the ‘dual-core’ (Guangzhou and Shenzhen) and ‘main Z-shape structure’ (Guangzhou, Shenzhen and Foshan, Dongguan) within the GBA is extremely significant both in terms of population movement and economic mobility. Meanwhile, Yang (2019) highlighted that in the symbiotic relationship between cities in the GBA, Guangzhou and Shenzhen have a positive radiation effect on the surrounding cities, while Hong Kong and Macao seem to have the opposite effect.

## 2.5 Conclusion

As one of the most common and crucial types of spatial interaction, migration flows have always been a pivot in the urban studies to provide an understanding of urban networks, especially in the context of PM-CRs. After a period of evolution and practice, the effect of the gravity model in simulating aggregated migration flows is worthy of affirmation, but there are also some challenges, such as in the visualization process.

However, in China, due to limited availability of the observed O-D data, studies on spatial interaction based on gravity models mainly remain at the theoretical level and the empirical research on internal migration flows at an intercity level or smaller scale are extremely lacking. This situation largely hinders our understanding of urban networks and migration patterns, making it difficult to provide more quantitative measurements or predictions for the planning decision makers in the context of the dynamic development of PM-CRs in China. Therefore, this research tries to contribute to this gap by developing new models of migration based on spatial interaction for this region.

# Chapter 3 Methodology

This chapter introduces the study area, different geographical levels focused on in this research, and then demonstrates the workflow of the data analysis. The detailed explanations of data processing and three analytical stages with relevant methods will follow from the framework. Finally, the limitations and ethical statement of this methodology will be interpreted.

## 3.1 Study Area

The investigation of the internal migration flows covered in this research is mainly concentrated on the Great Bay Area (GBA) of the Pearl River Delta in Guangdong Province (Gd), China. 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. This study will first explore the aggregated migration flows between the 9 cities within GBA, based on the total 72 migration flows based on 9*9 O-D matrix regardless of the intrazonal flows. Meanwhile, the migration flows between 21 cities within the whole Gd province (420 flows=21*20) are also considered in this study and this provides a comprehensive understanding and further comparisons with GBA.

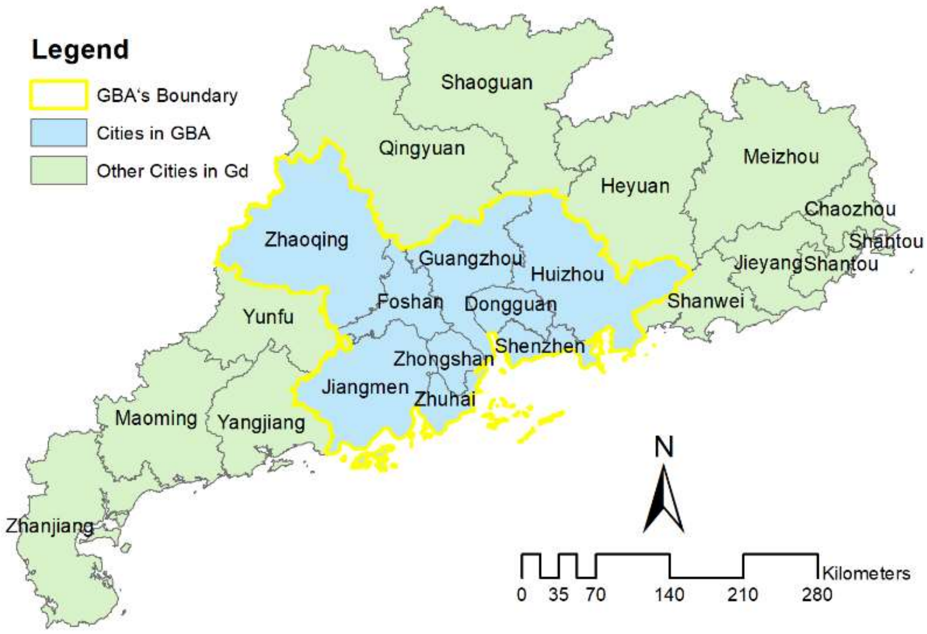



*Figure 3.1: The Map of Cities within Guangdong Province* (Source: Author)

## 3.2 Research Framework

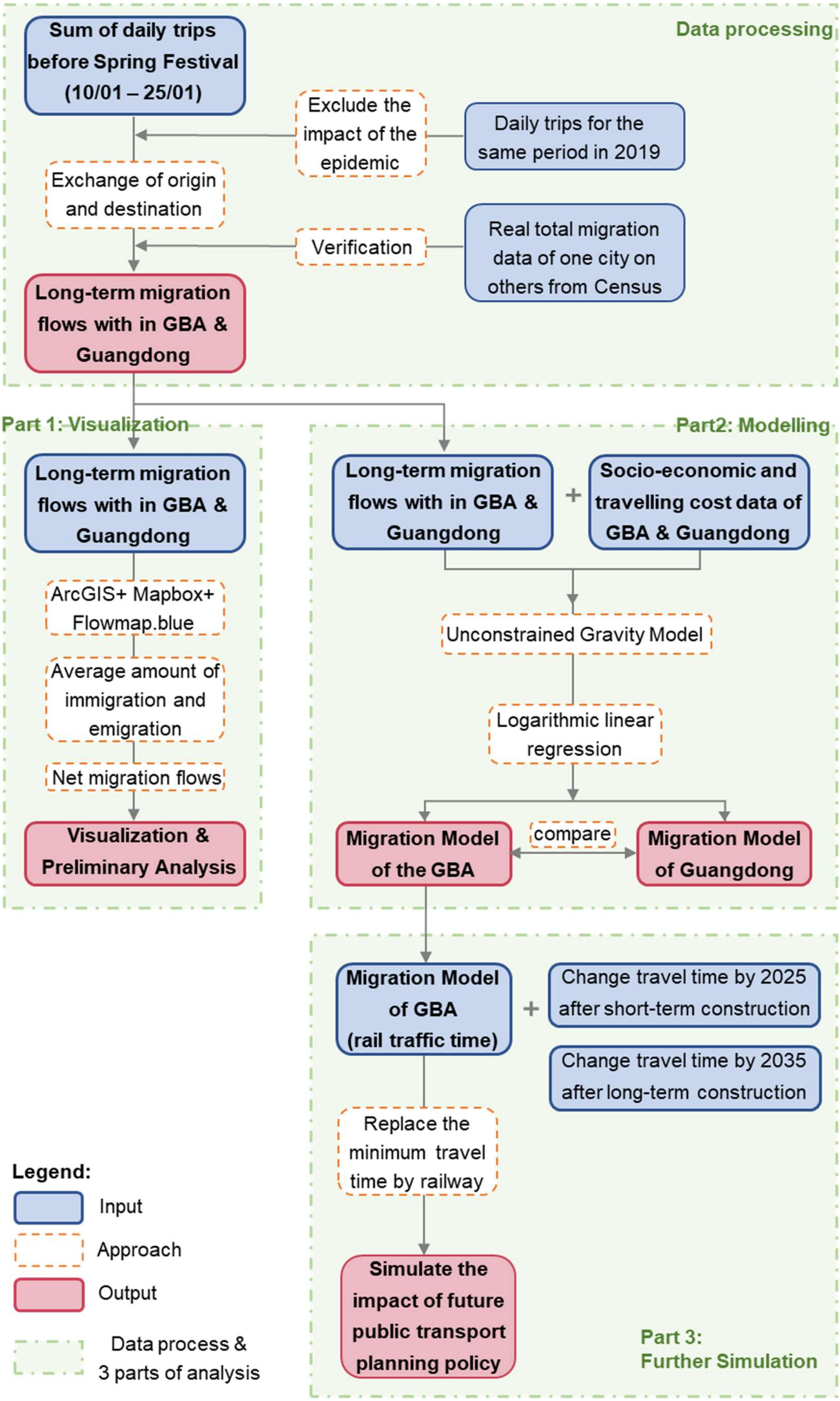



*Figure 3.2: The Research Framework based on Key Workflows* (Source: Author)

As shown in Figure 3.2, the research framework contains one data processing flow and three analytical stages. First of all, in data processing, the aggregated migration flows between cities within the study area are derived from the sum of daily trips before the 2020 Spring Festival. The feasibility of this kind of processing will also be discussed and verified in depth in this part through the comparisons with a few samples of historical migration data.

In terms of the three parts of the analyses, the main objective of the first part is to briefly visualize the internal migrations obtained from data processing, making some preliminary analyses, to reveal characteristics of the spatial interactions and mega-city network of the GBA. In the next stage, by introducing some socio-economic and travel-cost indicators, we attempt to build two calibrated gravity models to simulate internal migration within the GBA, as well as Gd province. In the last simulation stage, we are committed to modelling and visualizing the impact of the completion of future rail transit planning policies on migration within the GBA.

## 3.3 Data Source and Processing

All original datasets employed in this research are public and the summary list of all data sources can be found in Appendix A. In addition, all the processed data and the codes employed to collect the original data have been published on the author's Github site: https://github.com/ZZH-H/Migration-Dataset-for-the-GBA-Dissertation.git.

### 3.3.1 Inter-city Migration Flows

From the review of the literature and data sources, the unavailability of the actual O-D data between cities is one of the biggest obstacles to the empirical study of the complex spatial interaction patterns in China (Hu and Zheng, 2015; Gu and Pang, 2008). However, regarding historically observed migration data, we can still obtain 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. Taking Guangzhou as an example, by the end of 2018, we know that Guangzhou had a total of 111658 immigrants from 20 other cities in the Guangdong province, while 23124 people had moved out of Guangzhou to the 20 other cities (*Statistical Yearbook of Guangdong*, 2020). 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 3.3 (Stewart, 1950).

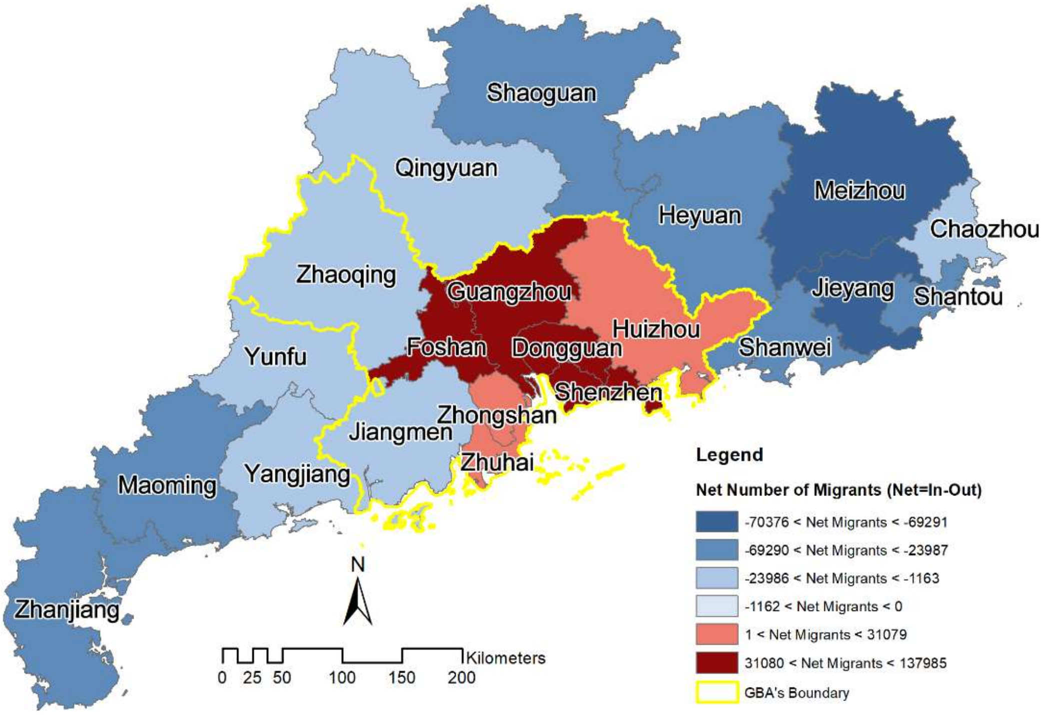



*Figure 3.3: The Map of the Net Migrants of Cities with in Gd Province from 2019 Statistical Yearbook of Gd* *(Source: Author)*

Therefore, in order to develop the subsequent calibration of the simulation models, this research attempts to indirectly reflect the internal migration flows by employing and processing the daily trips data from *Baidu Map Migration Platform*, which are available at: http://qianxi.baidu.com/. The python scripts used to capture these data could be found in Appendix B. This interactive platform from Baidu Map records the matrix of aggregated O-D daily trips at the city level across the whole of China, covering the period time between Jan. 1st to May 3rd, 2020. Baidu Map collects these movements 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. Meanwhile, the Baidu Map application is widely used across the whole of china, similar in frequency to Google Maps in the West.

It is also worth noting that, for privacy reasons in the original data, Baidu's short-term migration data is not the real number of movements between two cities, but a scaled index of movements that have been functionally processed by Baidu Map company. Although we cannot determine the exactly functional relationship between the obtained scale index and the actual number of movements, its numerical value can directly reflect the number of trips and can be used for horizontal comparisons between different cities.

As demonstrated in Figure 3.4, the 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), which are similar to the increases in travel before the Charismas Holliday in the West. Since Chinese 'New Year' varies in time as it depends on the Chinese Lunar Calendar, this travel rush based on return-home migration is from Jan. 10$^{th}$ to Jan. 25$^{th}$ in 2020, while it was Jan. 21$^{st}$ to Feb. 5$^{th}$ in 2019.

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 3.5 and Figure 3.6.

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.

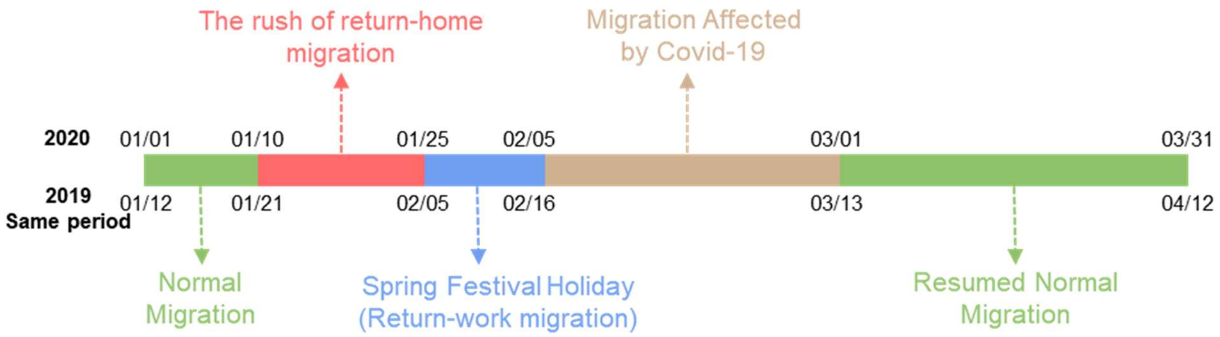



*Figure 3.4: The Explanatory Timeline of Spring Festival Migration in 2020 and the Same Period in 2019* *(Source: Author)*

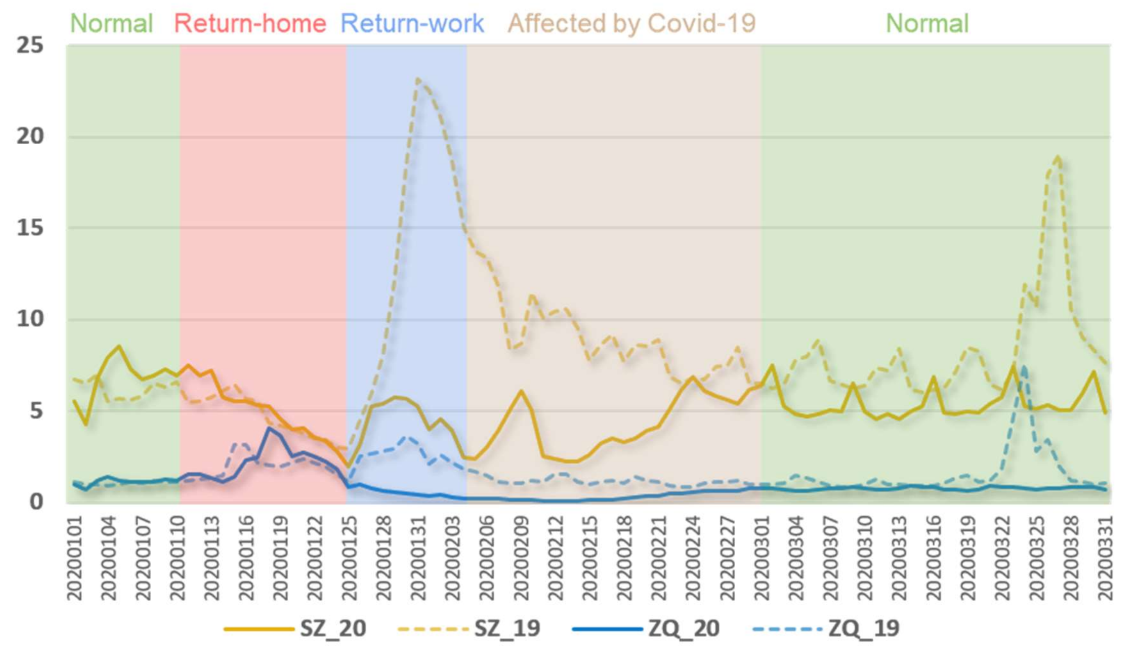
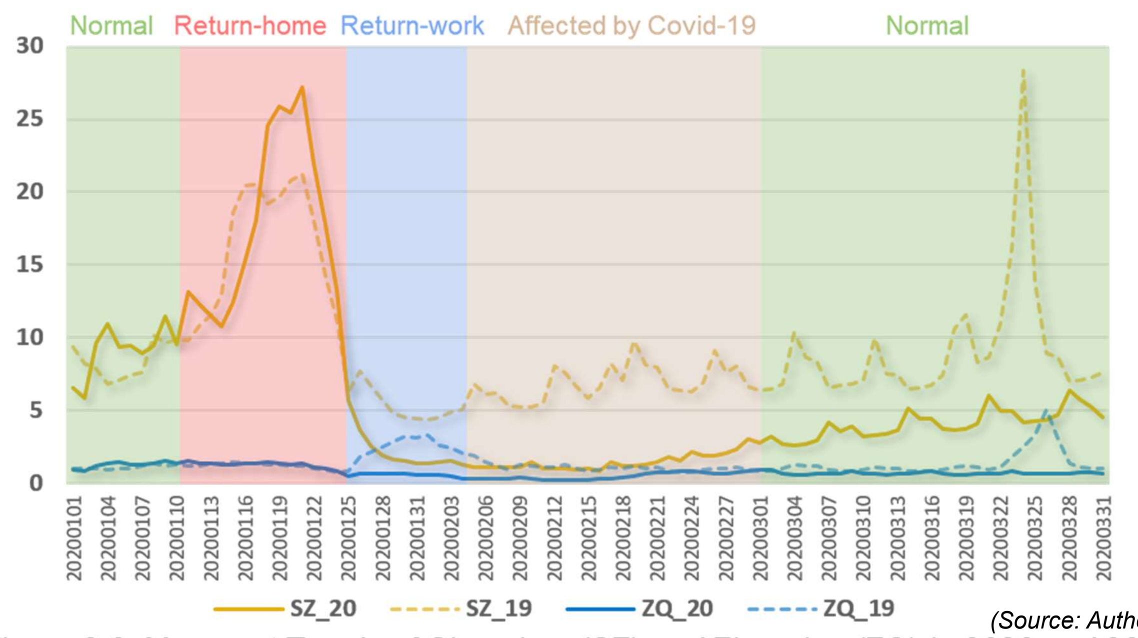



*Figure 3.5: Move-in Trends of Shenzhen (SZ) and Zhaoqing (ZQ) in 2020 and 2019* *(Source: Author)*




*(Source: Author)*

*Figure 3.6: Move-out Trends of Shenzhen (SZ) and Zhaoqing (ZQ) in 2020 and 2019*

In addition, by comparing the movement index of the same period in 2019 (dotted lines in Figure 3.5 and 3.6), 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.

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 3.7, which is also relatively consistent with the actual long-term migration pattern recorded in the *2019 Statistical yearbooks* mentioned above (Figure 3.3). Furthermore, the scatter plot of these two types of migration data which we show in Figure 3.8 has a fitting line which demonstrates that there is a strong linear relationship between them.

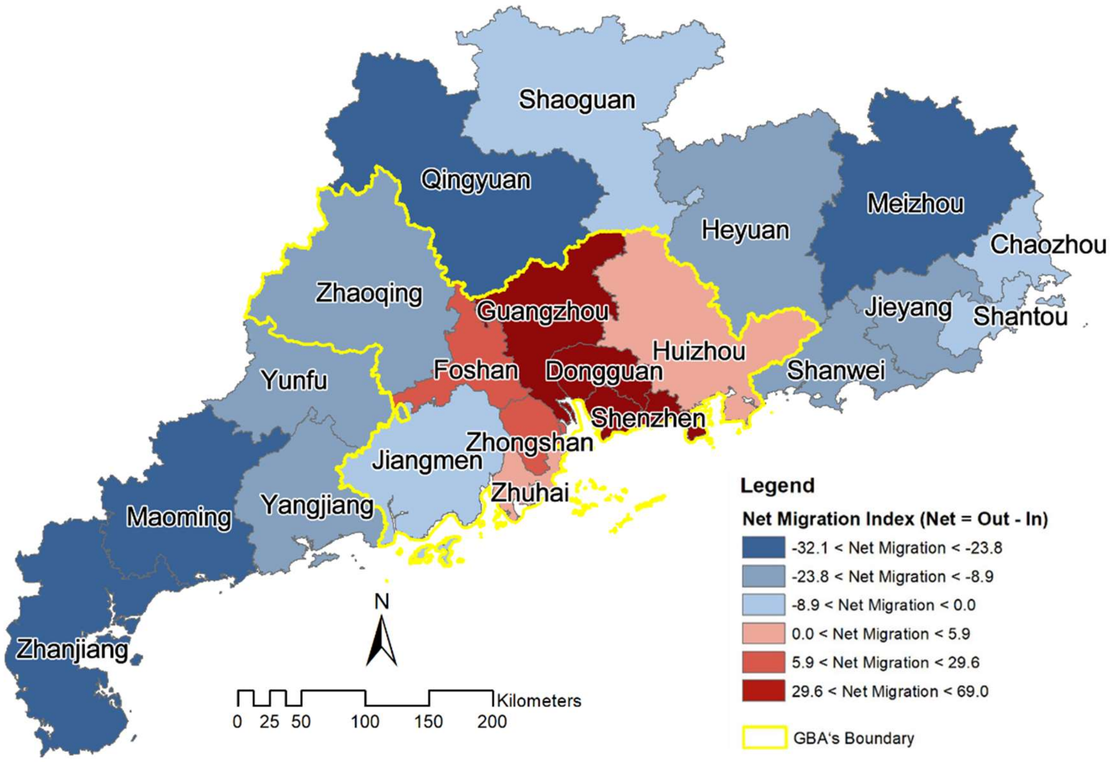



*Figure 3.7: The Map of the Sum of Net Movement Indices of Cities with in Gd Province between Jan.10th to Jan.25th* (Source: Author)

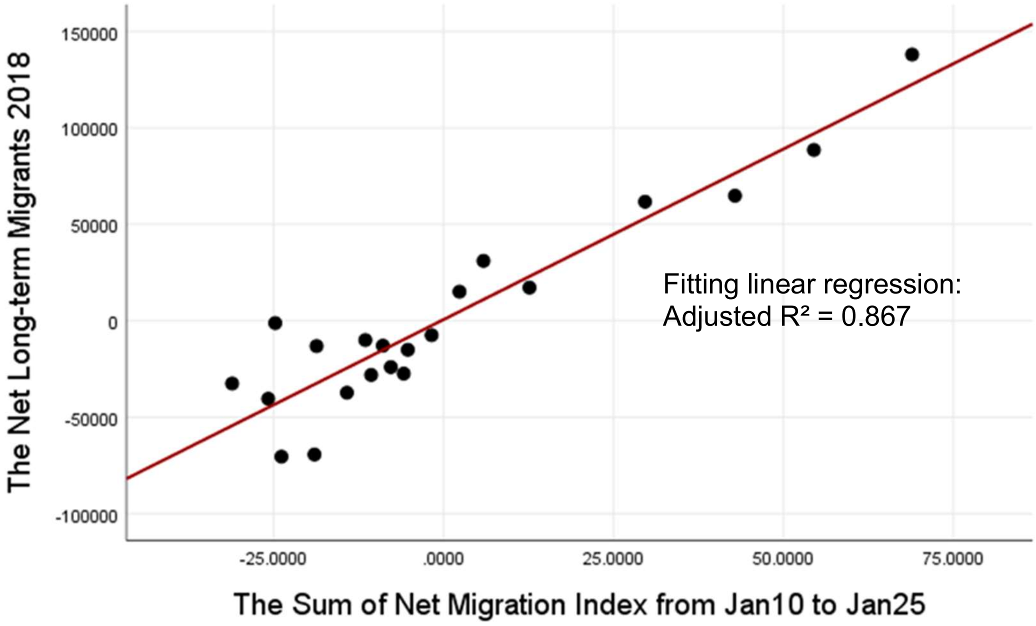
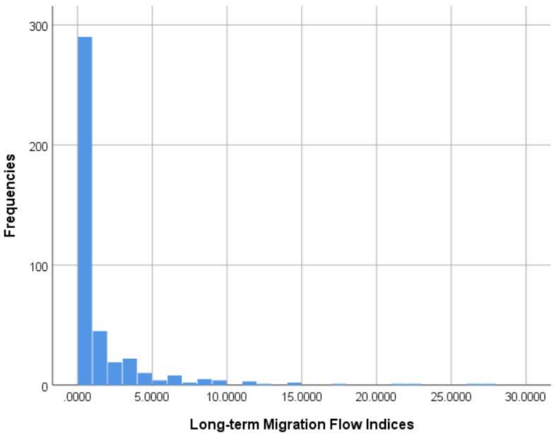



*Figure 3.8: The Scatter Plot of Actually Long-term Net Migrants (2019 Statistical Yearbook) with the Sum of Net migration Indices from Jan.10$^{th}$ to Jan.25$^{th}$of 21 Cities within Gd Province* *(Source: Author)*

In conclusion, the asymmetric matrix of O-D long-term migration flows of 21 cities within Gd province (Appendix C) 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 3.9), 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 3.9: The Frequencies and Descriptive Statistics of the Final Long-term Migration Dataset for the whole Gd province* *(Source: Author)*

### 3.3.2 Socio-Economic & Travel Time Data

According the previous literature, there are two types of data employed to simulate the internal migration flows obtained above: one is based on several socio-economic indicators which reflect the masses of origins or destinations, and another dataset is about the travel cost or distance between these origins and destinations. In the context of China, the indicators of these two types will be introduced into further gravity models (Table 3.1).

Firstly, there are totally 10 socio-economic indicators at the city level within Gd province applied in the further simulation, which could all be accessed from the *2019 Statistical Yearbooks* of corresponding cities. Most of these indicators are panel data through to the end of 2018, but also include one GDP growth rate variable between 2017 and 2018, and this enable us to check the potential impacts of the economic growth speeds on long-term migration. These socio-economic indicators only need to be sorted according to the origins and destinations to satisfy the conditions for subsequent analyses.

*Table 3.1: The Definitions of Variables Employed in the Gravity Models*
*（Source: Author）*

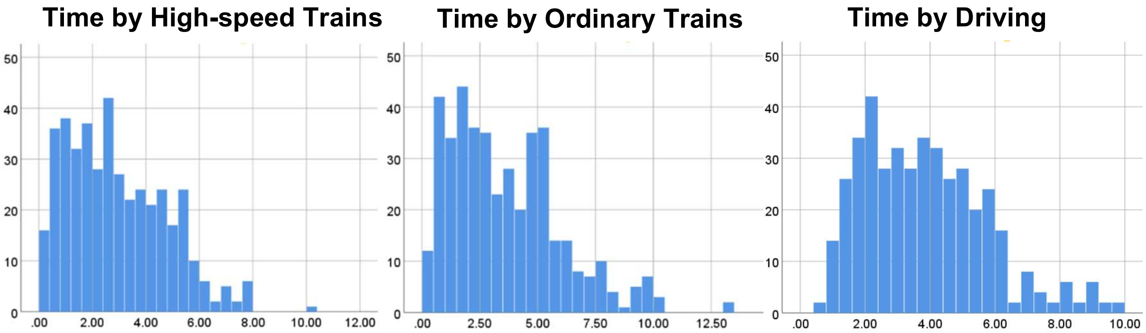
| The Definitions of Variables in the Gravity Models | | | |
|---|---|---|---|
| **Dimensions** | **Indicators (Variables)** | **Labels** | **Unit** |
| Population Size | Permanent Population | $P_i^{a1}$ $P_j^{a2}$ | 10000 pers |
| Income & Economy | Per Capita Disposable Income of Urban Permanent Households | $I_i^{a3}$ $I_j^{a4}$ | ¥ |
| | 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 $c_{ij}^{\gamma}$) | 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 |

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 driving time data $cd_{ij}$ between places are captured from *Baidu Map RouteMatrix API* by using Python. Through this service, users can calculate the distance and travel time by driving the suggested routes according to the start and end coordinates. The suggested route provided by Baidu Map is based on the choice of most users and is basically equivalent to the shortest driving route. More explanation of this and its returned variables can be found at: http://lbsyun.baidu.com/index.php?title=webapi/route-matrix-api-v2. In order to exclude the influence of traffic conditions, all driving time data was collected between 2 am and 5 am Chinese time. Meanwhile, the geographic coordinates of origins and destinations are selected as the local government office. In this way, we finally get the driving time matrix between 21 cities in Gd province.

Secondly, the minimum travel time by any kinds of public transportation ( $cp_{ij}$) and only by ordinary train ( $ct_{ij}$) are all obtained from *12306 China Railway*, which is the only direct and official channel for the sales of all kinds of railway tickets in China. Through the ticket inquiry function of the website of *12306 China Railway* (available at: https://www.12306.cn/index/), we can get all the train schedules between any two cities over the next 30 days, which includes all available railway schedules between all the railway stations both in the origin and destination cities. Since there is no available API from *12306 China Railway*, this study obtained the matrix of the minimum travel times by any inter-city rail transits between 21 cities within Gd province by manually searching the train schedules through inputting every origin and destination. Similarly, the matrix of the minimum travel times only by the ordinary trains was collected by only considering the ordinary train schedules between cities from the search results. Not surprisingly, most of the minimum travel times by any public transportation are by high-speed trains or bullet trains. Furthermore, if there is no direct rail schedule between the origins and destinations, the minimum travel times by transfer will be adopted, including the waiting time between the two schedules.

In conclusion, the frequencies and descriptive statistics of these 3 types of travel time data (Unit: hours) are shown in Figure 3.10. The processed travel time matrices between 21 cities within Gd province based on the minimum driving times can be checked in Appendix D as an example.




| N | 420 |
|---|---|
| Mean | 2.89 |
| Min. | 0.18 |
| Max. | 10.03 |
| Std. Deviation | 1.84 |
| Variance | 3.40 |

| N | 420 |
|---|---|
| Mean | 3.59 |
| Min. | 0.18 |
| Max. | 13.10 |
| Std. Deviation | 2.39 |
| Variance | 5.71 |

| N | 420 |
|---|---|
| Mean | 3.83 |
| Min. | 0.73 |
| Max. | 9.93 |
| Std. Deviation | 1.90 |
| Variance | 3.60 |

*Figure 3.10: The frequencies and descriptive statistics of Three Types of Travel Time Data (unit: hours)* (Source: Author)

### 3.3.3 The Travel time after Future Rail Transit Construction of the GBA

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 3.11) 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 3.11, 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 3.2 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 3.1, since there is extremely limited information about the long-term rail planning up until now.

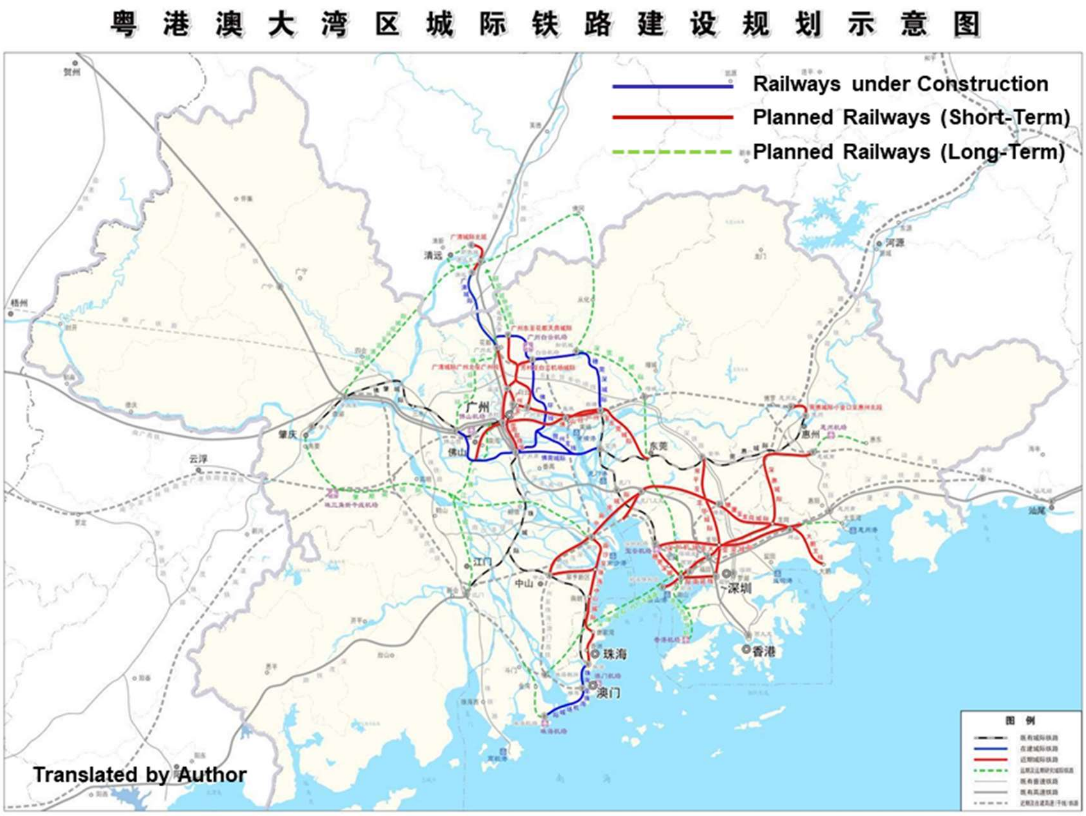



*Figure 3.11 The Concept Map of the Intercity Railway Construction Plan for the GBA (Source: National Development and Reform Commission, July 31 2020, avaiable at:* https://www.ndrc.gov.cn/xxgk/zcfb/tz/202008/t20200804_1235517.html*)*

*Table 3.2 The Table of the Estimated Travel time by Railway after the Completion of the Short-term Rail Construction* *(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.*

## 3.4 Analytical Stages and Methods

### 3.4.1 Preliminary Analysis and Visualisation

In order to visualize the internal migration flows both for the GBA and 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 this part 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 and can be accessed directly at: https://flowmap.blue/. 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 with the accessible link. 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.

### 3.4.2 Build and Calibrate the Gravity Models

In order to simulate the spatial interactions between cities within the study area, the gravity model is employed in this research, which has been theoretically demonstrated (Wilson, 1970;1971) and tested in many previous empirical studies (Borjas,1989; Dennett and Wilson, 2013; Karemera et al.,2000; Shen, 2017). Since the main objective of this research is to quantitatively explore the relationships between different indictors and the long-term migration flows, the unconstrained gravity model in the family of related models is built to simulate the intercity migration flows as Equation 3.1 (Batty and Mackie, 1972; Wilson, 1971).

$$M_{ij} = K\frac{O_i^{\alpha} D_j^{\beta}}{{c_{ij}}^{\gamma}} \quad i,j = 1 \dots N, \quad i \neq j \qquad (3.1)$$

As reviewed in the literature, the long-term migration flows between origin *i* and destination *j* could be explained by the scaling constant *K*, the supply factor $O_i^{\alpha}$, the attractor factor $D_j^{\beta}$ and the travel cost between i and j, ${c_{ij}}^{\gamma}$. Meanwhile, the parameters $\alpha$ and $\beta$ represent the scaling importance of the mass of origins and destinations, and the parameter $\gamma$ is defined as the friction of the travel cost (Batty and Mackie, 1972; Wilson, 1970;1971).

Furthermore, the total 10 socio-economic indicators used to represent the mass of origins and destinations are listed in the Table 3.1 in the data process part above (p:31). Except the traditional indicators, population and income, other potential indicators are also included based on considering the particular context of spatial interactions in China and the empirical research (Fan, 2013; Karemera et al., 2000; Shen, 2017; Xu and Yao, 2019). Through taking the logs of both sides of the above equation and extending the variables, the equivalent model is as follows:

$$
\begin{aligned}
\ln M_{ij} = a_0 &+ a_1 \ln P_i + a_2 \ln P_j + a_3 \ln I_i + a_4 \ln I_j + a_5 \ln TI_i \\
&+ a_6 \ln TI_j + a_7 \ln GDP\_G_i + a_8 \ln GDP\_G_j + a_9 \ln H_j \\
&+ a_{10} \ln S_j + b \ln c_{ij} \qquad i, j = 1 \ldots N, \quad i \neq j \qquad (3.2)
\end{aligned}
$$

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. In addition, three types of minimum travel time mentioned in the data process part will be introduced into the model separately to compare their sensitivity to the internal migration flows.

In terms of the calibration of the model built above, although the logarithmic linear regression might have some limitations compared to Poisson regression (Flowerdew et al., 1982), it can still be applied to the calibration in this research because the O-D data used in the study is not the number of migrants but the functionalized migration indices which are fractional. The acceptable significance level of the coefficients is 5% in this study. The calibration is all processed using Python scripts.

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). This paper mainly adopts the standards proposed by Hair in 1995: The independent variable can be considered seriously collinear with others if its VIF is greater than 10 or tolerance is less than 0.1.

Finally, the results of the models for the GBA and the whole Gd province will be visualized based on flowmap.blue, and then interpreted and further compared. The residuals of the gravity model for the GBA will also be discussed.

### 3.4.3 Further Simulation

The further simulation about the impacts of future railway planning 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 construction. 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. Similarly, the visualisation outcomes of this further simulation are all achieved by *flowmap.blue.*

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:

$$R_{ij} = \frac{|\widehat{M'_{ij}} - \widehat{M_{ij}}|}{\widehat{M_{ij}}} \qquad (3.3)$$

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

## 3.5 Limitations

There are two main limitations of the methodology in this research. Firstly, although the linear relationship between the total amount of movements during the return-home rush before the Spring Festival and real long-term migration flows has been discussed, there is no denying that this alternative flow dataset still involves a small number of short-term trips. However, as there is no historical O-D data of long-term migration in China, this limitation seems inevitable.

Secondly, the minimum travel time by public transportation between cities is much more complicated in the real world. It might involve 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. However, these complications might have more limited influence on the long-term migration than short-term trips.

## 3.6 Ethical Statement

There is no personal or private data involved in this study and all data employed is public and freely accessible. Besides, all data sources have been illustrated and the original dataset has been published online by author for anyone to check to avoid any potential ethical risks.

# Chapter 4 Results and Discussion

This chapter presents the results of visualisations and simulations and provides analyses based on these outcomes, organising in the following three parts. Firstly, the visualisation outcomes of the internal migration flows and corresponding preliminary analysis will be demonstrated. The next part focuses on the results and analysis of the gravity models for the GBA and Gd province. Finally, there is a further simulation using the calibrated models in the last part.

## 4.1 Visualisation and Preliminary Analysis of Migration Flows

### 4.1.1 Visualisation Outcomes of the GBA

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

Through this link, we can customize the visual map to our own interests and view the original dataset published through a Google Sheet, which lew us select all flows or net flows and one or some particular cities. Generally, as shown in the Figure 4.1, the thickness and colour of the lines are proportional to the size of the migration between cities, while the size of the nodes represents the total amount of migrations.

First of all, the average amount of migration between two places could relatively reflects the closeness of the population mobility, where the total thickness of the lines with two directions visualises these volumes in the figure. There is no doubt that the interaction between Guangzhou and Foshan is the highest one in the GBA. In addition, 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 four cities are also the most important nodes in the regional structure, considering their total amount of migrations. Except for the 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.

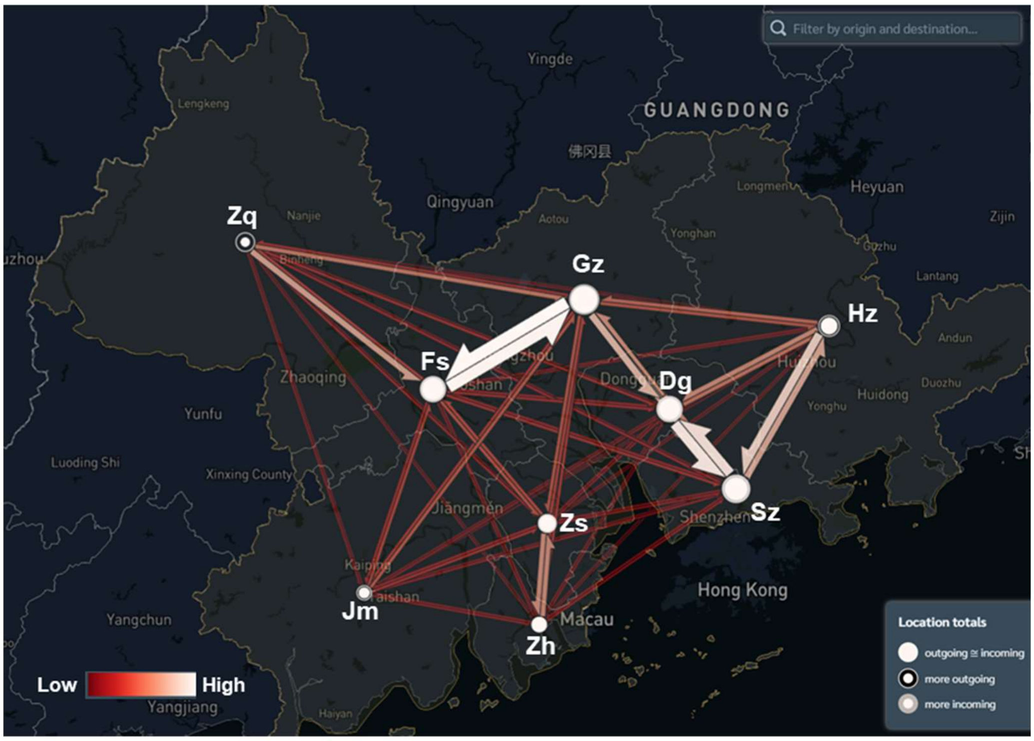



*Figure 4.1 The Flow Map of the Long-term Migration Flows between the nine cities within the GBA – The Screenshot of Interactive Outcomes* (Source: Author)

Secondly, the Figure 4.2 identifies the net migration flows of the GBA and some selected cities can be checked in Figure 4.3. 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. Similarly, Huizhou and Jiangmen are also dominated by significant net emigrations. 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. This means that Shenzhen is more attractive to other cities than other cities to Shenzhen. However, as another developed city in the GBA, Guangzhou has relatively significant net outflows to Dongguan and Shenzhen.


*Figure 4.2 The Net Flow Map of the Long-term Migration Flows between the nine cities within GBA – The Screenshot of Interactive Outcomes* (Source: Author)

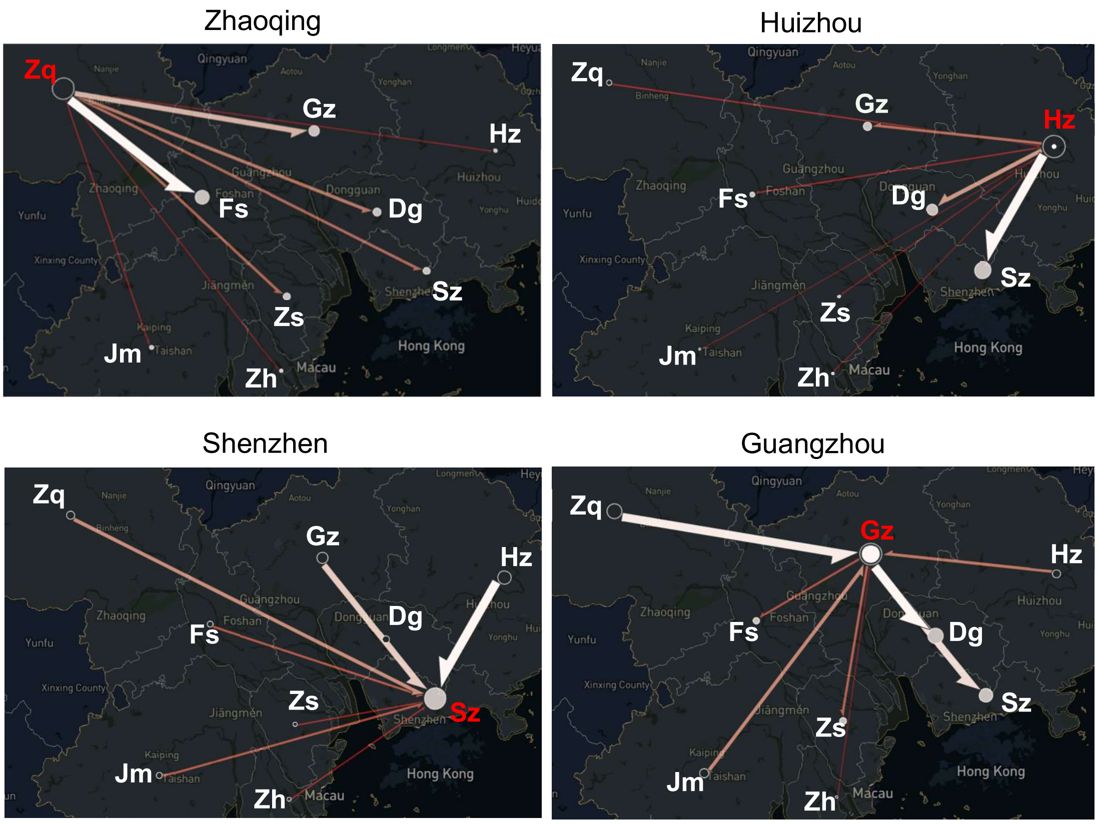



*Figure 4.3 The Net Flow Maps of the Long-term Migration Flows for Zhaoqing, Huzhou, Shenzhen and Guangzhou* (Source: Author)

In addition, by comparing Figure 4.1 and Figure 4.2, 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). For example, the migration-scale index between Guangzhou and Foshan is very close respectively at 27.48 and 26.74. Based on the previous studies, the huge migration flow from Foshan to Guangzhou is in line with expectations that more developed, higher income regions have a great attraction for long-term migrants On the other hand, Due to the improved transportation network between Guangzhou and Foshan in recent years, such as the direct connection of subway 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) .

### 4.1.2 Visualisation Outcomes of Guangdong Province

Since the main focus of this study is the GBA, the visualisation outcomes of the whole Gd province will also be demonstrated briefly in the Figure 4.4 below, 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

Generally, the nine cities of the GBA are closely connected regarding long-term migration than the others within Gd province, and the Z-shape mentioned above is extremely prominent even in the context of the whole Gd province. Figure 4.5

is derived from clustering some cities based on Flowmap.blue to concisely explore the relationships between other cities and the GBA. After clustering the 4 cities in the Z shape as a core, we find that the large number of immigrants to this core are mainly from five origins except other than the origins within the GBA, including 3 single cities and 2 clusters.

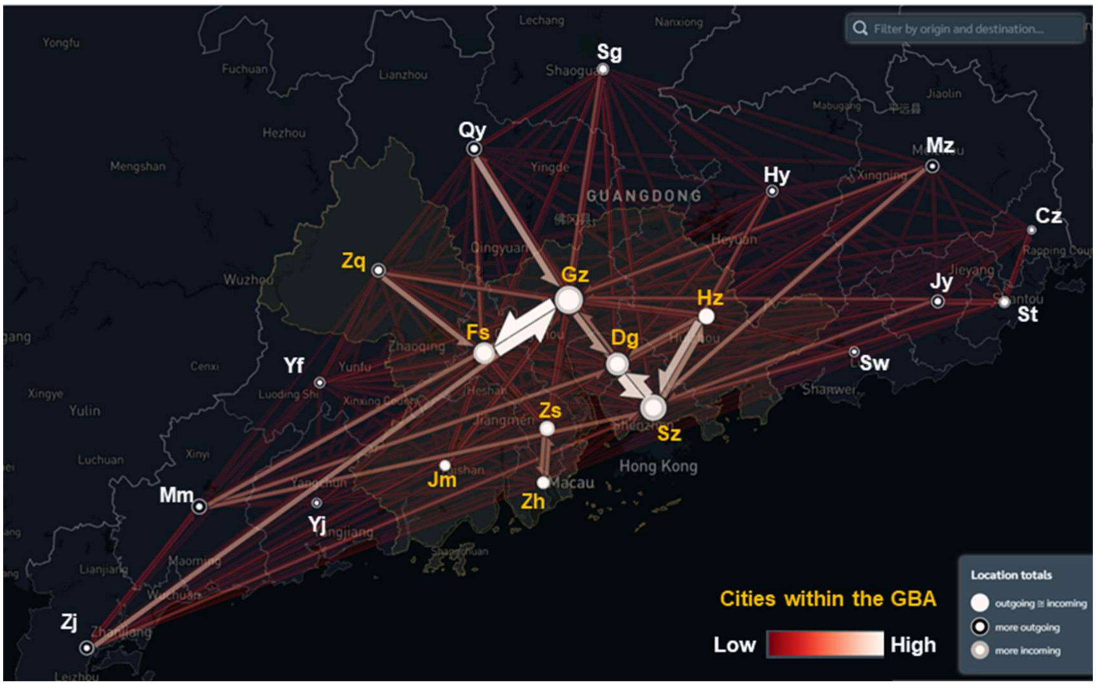



*Figure 4.4 The Flow Map of the Long-term Migration Flows between the 21 cities within Guangdong Province* (Source: Author)

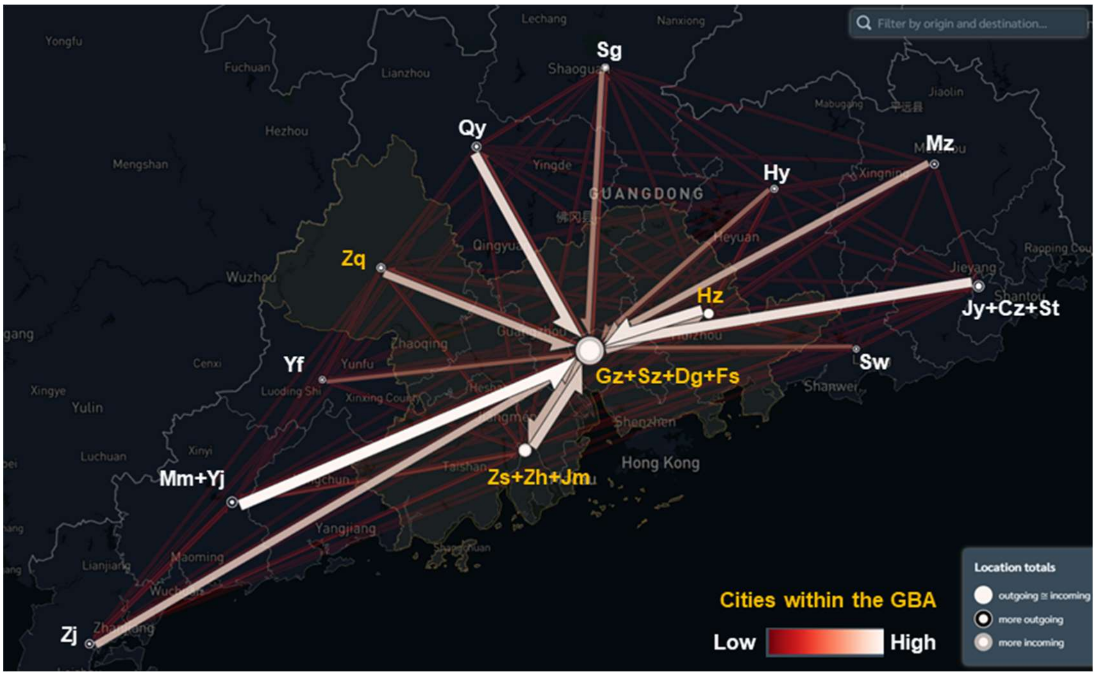



*Figure 4.5 The Flow Map of the Long-term Migration Flows between the Clustered cities within Guangdong Province* (Source: Author)

## 4.2 The Results of Gravity Models

As mentioned in Chapter 3, an unconstrained gravity model is applied to simulate the long-term migration flows with some extended variables (can be checked in the Table 3.1, p: 31) and the logarithmic version of extended gravity model is further employed to estimate the parameters. For the ease of reading, the two equations for the gravity model are repeated as follows:

$$M_{ij} = K\frac{O_i^{\alpha} D_j^{\beta}}{c_{ij}{}^{\gamma}} \quad i,j = 1 \ldots N, \quad i \neq j \qquad (3.1)$$

*Logarithmic Version of Extended Gravity Model:*

$$\ln M_{ij} = a_0 + a_1 \ln P_i + a_2 \ln P_j + a_3 \ln I_i + a_4 \ln I_j + a_5 \ln TI_i$$

$$+a_6 \ln TI_j + a_7 \ln GDP\_G_i + a_8 \ln GDP\_G_j + a_9 \ln H_j$$

$$+a_{10} \ln S_j + b \ln c_{ij} \qquad i,j = 1 \ldots N, \quad i \neq j \qquad (3.2)$$

### 4.2.1 Migration Flows & Three Types of Travel time

Since we have already got three measures of travel time, the relationships between the migration flows and these different measures of deterrence need to be first discussed. Figure 4.6 and Figure 4.7 display the scatter plots of $\ln M_{ij}$ against three cost types of $\ln c_{ij}$ for the GBA and Gd province, and these combine the corresponding results of linear regression models.

First of all, all the coefficients of $\ln c_{ij}$ are between -1 to -3, which is consistent with the previous findings concerning typical values of the friction of distance (Boyle et al., 1988; Stillwell, 1978). Based on the comparisons of the values of the adjusted R squares, whether for the GBA or Gd province, 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, except for the considerations of further simulation requirements, this is another reason to only introduce the minimum driving time and minimum travel time by any public transportations into the subsequent calibration of gravity models.

*Figure 4.6 Scatter Plots of 9 Cities within GBA: $\ln M_{ij}$ against $\ln c_{ij}$*

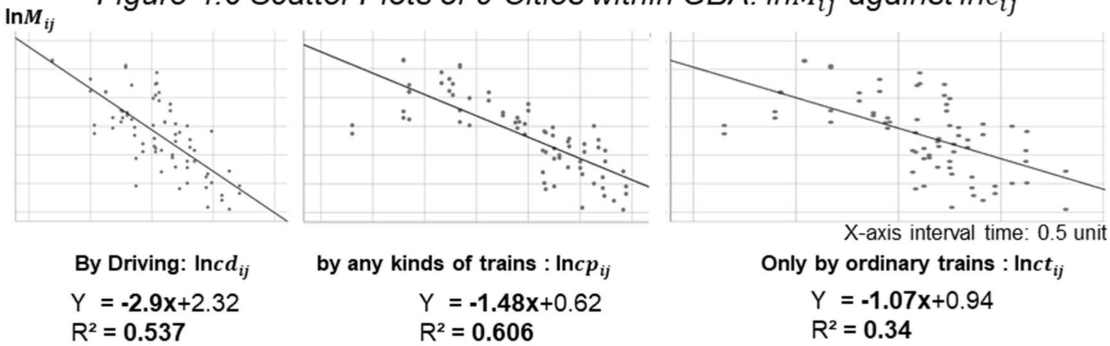
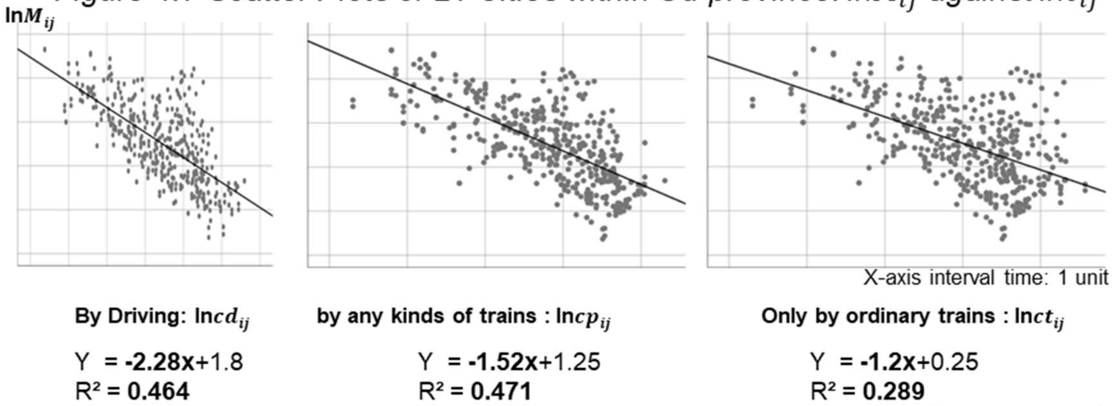



*Figure 4.7 Scatter Plots of 21 Cities within Gd province: $\ln M_{ij}$ against $\ln c_{ij}$*




*(Source: Author)*

### 4.2.2 The Results of the Calibration 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 the Table 4.1.

In general, the independent variables employed in these two logarithmic regression 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.

*Table 4.1 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 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 and the possible limitations of the shortest public traffic data (mentioned in Chapter 3), 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.

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.

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.

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.

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 (3.1) and equation (3.2), 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.

### 4.2.3 The Visualization of the Estimated Flows and Residuals

The estimated flows generated from the gravity model based on the minimum driving time are shown in Figure 4.8 (the same access link mentioned above, p:41). Although the thickness of the lines between the different graphs generated by *Flowmap.blue* cannot be directly compared vertically, we can still see the differences in the overall structure by compared the original flow map mentioned at the beginning (Figure 4.1). As the maximum index value of the GBA, the migration between Guangdong and Foshan is estimated to be 37.88 and 43.04 respectively, higher than the original 27.48 and 26.74. This makes it difficult to distinguish between small and medium migration flows because the thickness of the lines varies according to the size of the values.

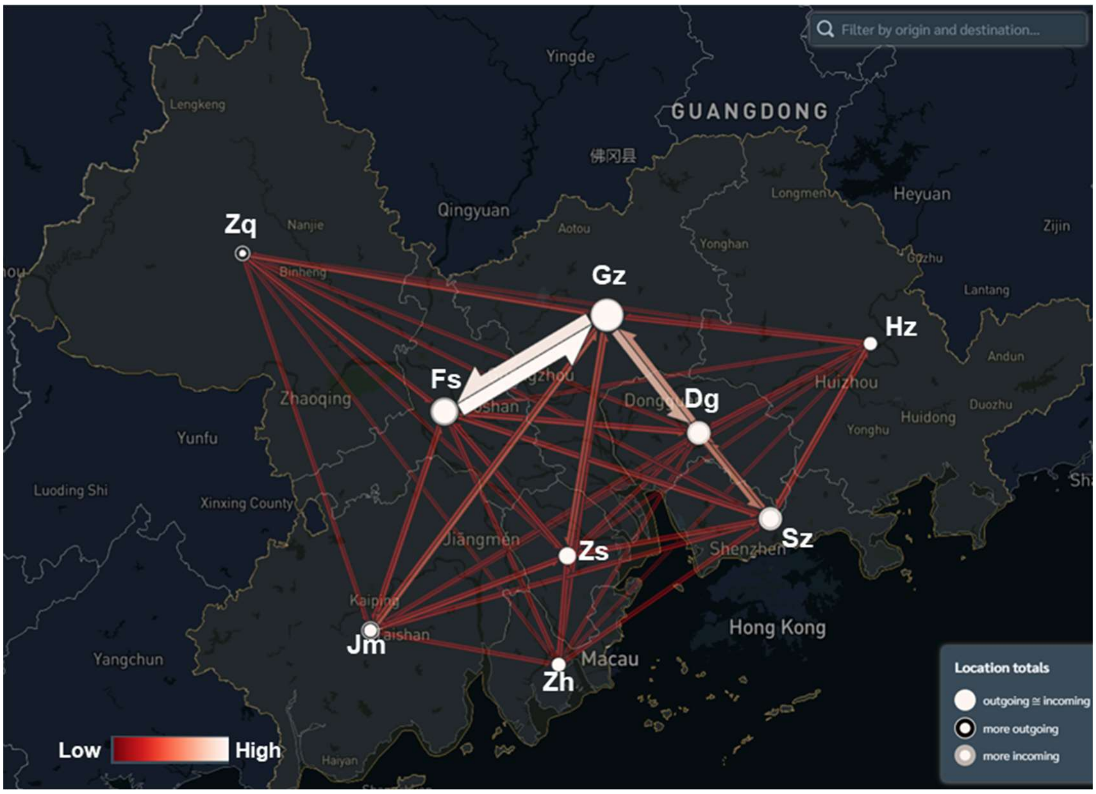



*Figure 4.8 The Estimated Flows based on the Driving Time between the 9 cities within the GBA* (Source: Author)

In terms of the residuals, the frequencies of the residuals are shown in the Figure 4.9 which reveals a near normal distribution. Figure 4.10 displays the scatterplot of standardized residuals with standardized estimates and all the points are floating about two standard deviations above or below the zero axis. Furthermore, the rates of the residuals to the original migration value - $\frac{|\widehat{M_{ij}}-M_{ij}|}{M_{ij}}$ - are employed to measure the estimated errors on the interactive map (Figure 4.11). Comparing the Figure 4.11 with Figure 4.1, 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).

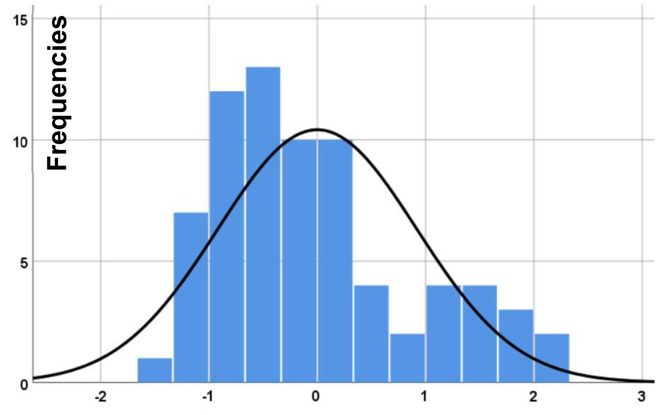



*Figure 4.9: The Frequency Histogram of the Residuals* (Source: Author)

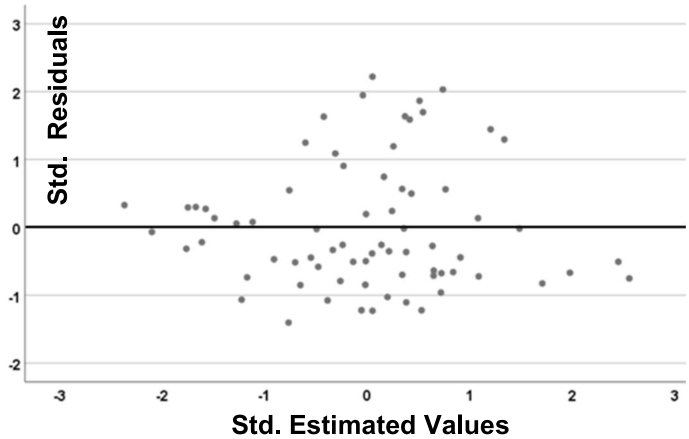



*Figure 4.10: The ScatterPlot of the Std. Residuals against Std. Estimates* (Source: Author)

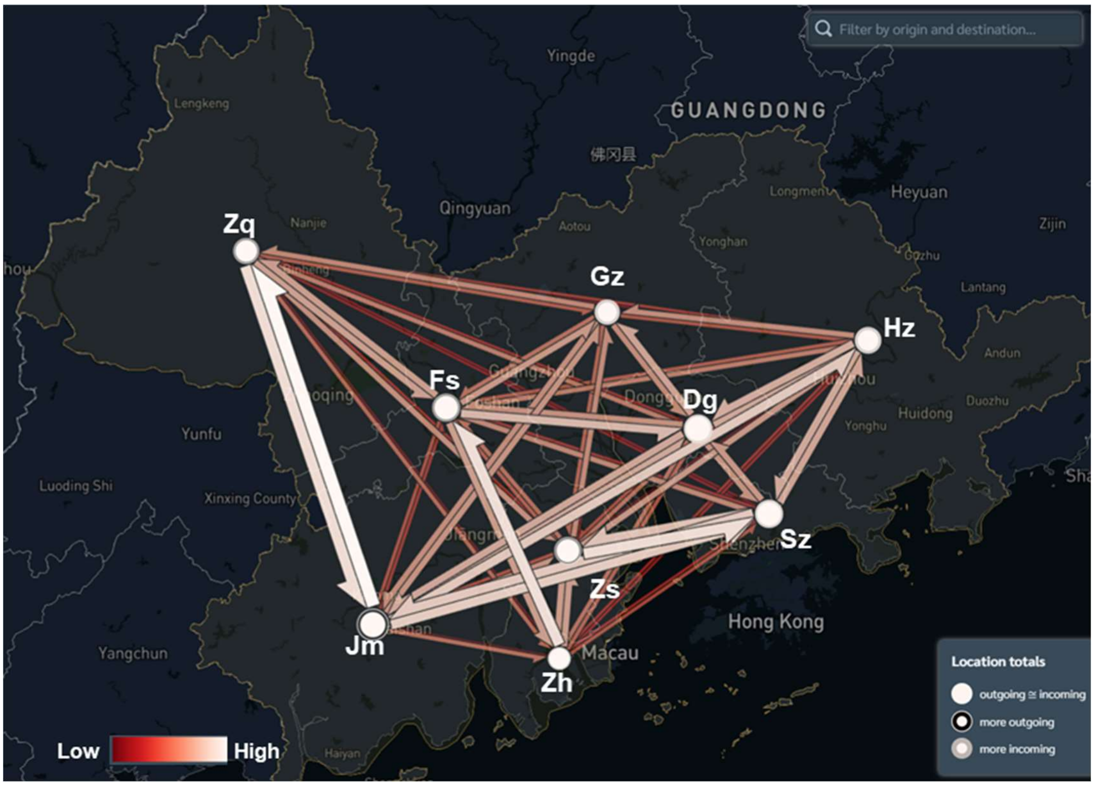



*Figure 4.11 The Absolute Rates of the Residuals to the Original Flows based on the Driving Time between the 9 cities within the GBA (unit: %)* (Source: Author)

### 4.2.4 Comparisons with the Estimates for Gd Province

Similarly, the estimates of the gravity models for the 21 cities within Gd province are presented in the Table 4.2, and theses provide further comparisons with the results for the GBA (Table 4.1) on the variables' impacts on long-term migration.

*Table 4.2 The Estimations of Two Gravity Models for Gd Province* (Source: Author)

| N = 420 | **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 | **-38.821** (-40.017) | | **-10.47** (-10.4) | **0.000** (0.000) | | |
| $P_i$ | **1.091** (1.024) | **0.327** (0.307) | **13.82** (11.81) | **0.000** (0.000) | **0.608** (0.610) | **1.65** (1.64) |
| $P_j$ | **0.750** (0.701) | **0.225** (0.210) | **8.64** (7.34) | **0.000** (0.000) | **0.502** (0.503) | **1.99** (1.99) |
| $I_i$ | **-0.520** (-0.487) | **-0.087** (-0.082) | **-3.44** (-2.91) | **0.001** (0.004) | **0.525** (0.521) | **1.90** (1.92) |
| $I_j$ | **2.987** ( 3.303) | **0.502** (0.555) | **9.95** (10.05) | **0.000** (0.000) | **0.133** (0.135) | **7.50** (7.42) |
| $TI_i$ | **0.445** (0.326) | **0.032** (0.023) | **1.37** (0.91) | **0.171** (0.362) | **0.625** (0.624) | **1.60** (1.60) |
| $TI_j$ | **0.141** (0.040) | **0.010** (0.003) | **0.41** (0.10) | **0.682** (0.917) | **0.555** (0.554) | **1.80** (1.80) |
| $GDP\_G_i$ | **-0.578** ( -0.592) | **-0.129** (-0.132) | **-5.99** (-5.57) | **0.000** (0.000) | **0.734** (0.731) | **1.36** (1.37) |
| $GDP\_G_j$ | **-0.124** ( -0.141) | **-0.028** ( -0.032) | **-1.23** (-1.27) | **0.220** (0.205) | **0.670** (0.668) | **1.49** (1.50) |
| $H_j$ | **-0.085** (-0.243) | **-0.027** (-0.078) | **-0.66** (-1.73) | **0.509** (0.085) | **0.201** (0.202) | **4.98** (4.96) |
| $S_j$ | **0.475** (0.523) | **0.062** (0.069) | **1.95** (1.95) | **0.053** (0.052) | **0.331** (0.331) | **3.02** (3.02) |
| $cd_{ij}$ ($cp_{ij}$) | **-1.694** (-1.062) | **-0.506** (-0.479) | **-24.50** (-20.6) | **0.000** (0.000) | **0.797** (0.761) | **1.26** (1.31) |
| **Adjusted $R^2$** | **0.858** (0.828) | | | | | |

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

Comparing the results of the GBA and Gd province, there was slight difference between the variables which are identified as significant to the long-term migration. The average income of destinations for Gd province has higher significance level than the estimation for the GBA (1% vs. 5.6%), while the significance level of educational resources becomes lower (5.3% vs 3%).

Since the sample size was expanded to the 420 migration flows between 21 cities in Guangdong province, including 72 flows within the GBA, we could find some changes in the degree of the influence of variables on migration. Firstly, the negative impacts of travel time on long-term migration for Gd province is lower than the GBA according to their coefficients' value: -1.694 for Gd province and -2.557 for the GBA. It indicates that distance is more sensitive to the migration in GBA than Gd province. On the other hand, the positive influences of population size and income on long-term migration for Gd province become higher than the GBA.

In addition, by comparing their standardized coefficients' absolute values, income and travel time had almost the same effective abilities on the migration in Gd province (std. coefficients.: 0.502 and -0.506), while there is still a gap in their influence in the GBA (std. coefficients.: 0.385 and -0.646). Based on this kind of difference, it could be indirectly speculated that the migrants from or to other 12 cities in Guangdong province except the GBA might be slightly more concerned with the income of destinations than the distance.

Furthermore, the reasons resulting in this difference on the magnitudes of coefficients between the GBA and Gd province could be identified as following: Firstly, because the GBA is a collection of several most developed cities in Gd province, the average income gap within the GBA is smaller than the gap in the context of the whole Gd province. Secondly, since the migrants of other cities in Gd province have a greater consideration for the income of destinations than the migrants of the GBA, the negative consideration for distance is correspondingly lower than the GBA. Thirdly, since the cities within the GBA are more closely linked by roads or rails, and the geographic scale itself is smaller than that of the entire Gd province, the impact of the difference in commuting time on people's willingness to migrate may therefore be magnified.

## 4.3 Further Simulation

In order to simulate the long-term migration flows $M_{ij}$ within the GBA after the completion of the intercity railway planning, the model based on the minimum public transportation above is employed by excluding insignificant variables (5% level) and then stepwise regression processing again. The results are shown in the equation (4.1) with the adjusted R square: 0.825.

$$M_{ij} = \exp(-28.928 + 0.463 * \ln P_i + 0.432 * \ln P_j + 2.760 * \ln I_j - 1.371 \ln TI_j - 0.298 * \ln GDP\_G_i - 1.208 * \ln cp_{ij}) \qquad (4.1)$$

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. All visualisation outcomes in this part have also been published on the same link of *Flowmap.blue* (p:40).

$$R_{ij2025} = \frac{\widehat{M_{ij2025}} - \widehat{M_{ij}}}{\widehat{M_{ij}}} \qquad (4.2)$$

$$R_{ij2035} = \frac{\widehat{M_{ij2035}} - \widehat{M_{ij2025}}}{\widehat{M_{ij2025}}} \qquad (4.3)$$

### 4.3.1 The Impacts of the Short-term Rail Construction

The estimated migration flows affected by the short-term railway constructions are visualized in the Figure 4.12. By comparing this short-term simulation with the current flows shown in the Figure 4.1 (p:41), 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-shape 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.

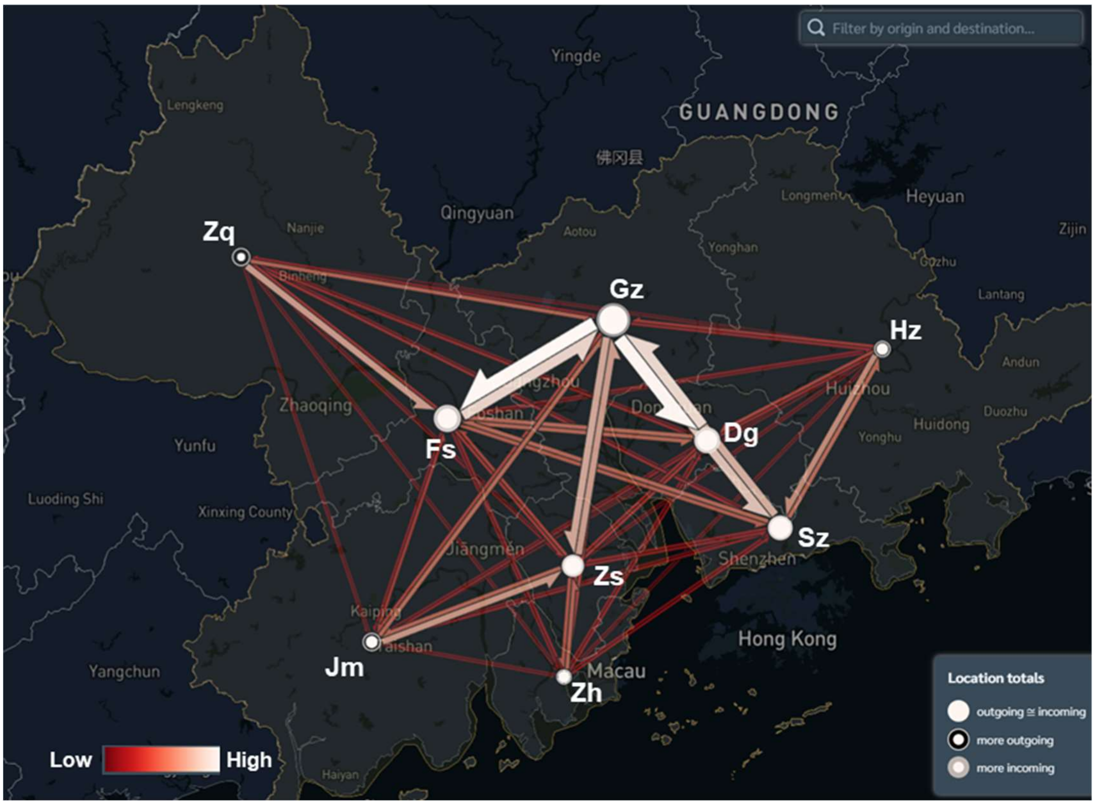



*Figure 4.12 The Estimated Flows based on the Impacts of the Short-term Railway Planning by 2025 of the GBA* *(Source: Author)*

Furthermore, the growth rates of all 16 affected flows by the short-term rail constructions are visualized in the Figure 4.13. 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.

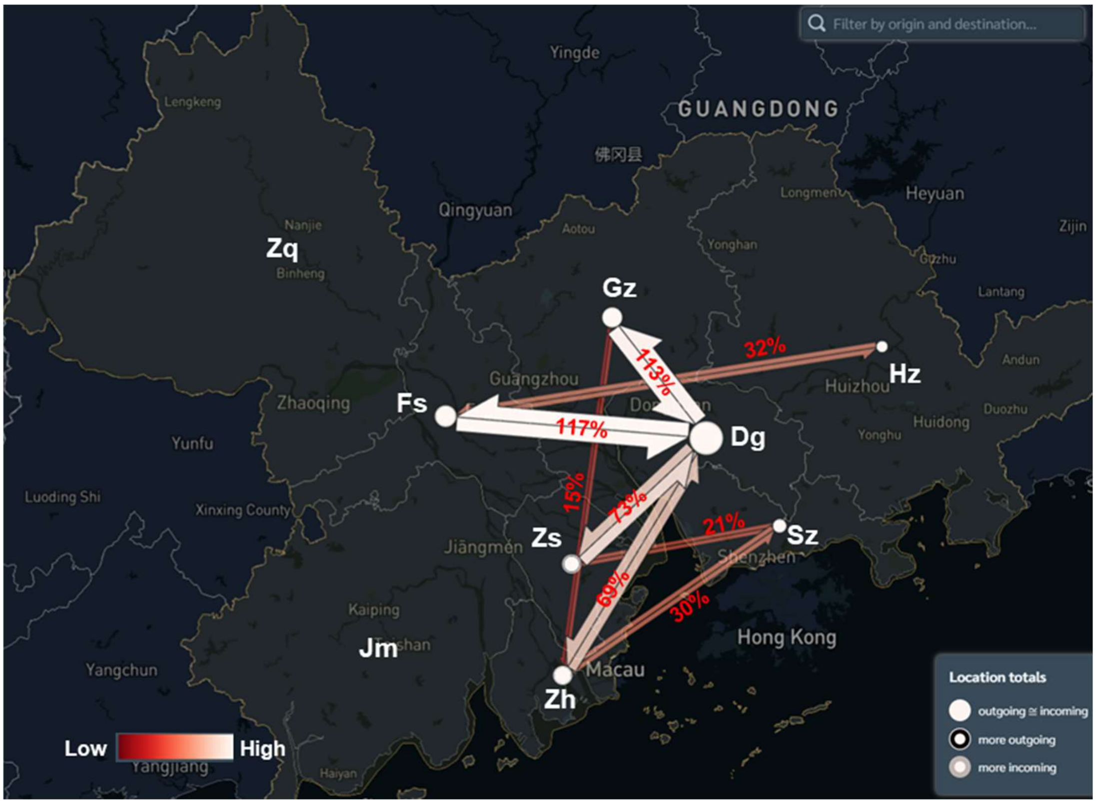



*Figure 4.13 The Growth Rates of the Estimated Flows Affected by the Short-term Railway Planning by 2025 to the Original Estimates* (Source: Author)

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 4.14. 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 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.

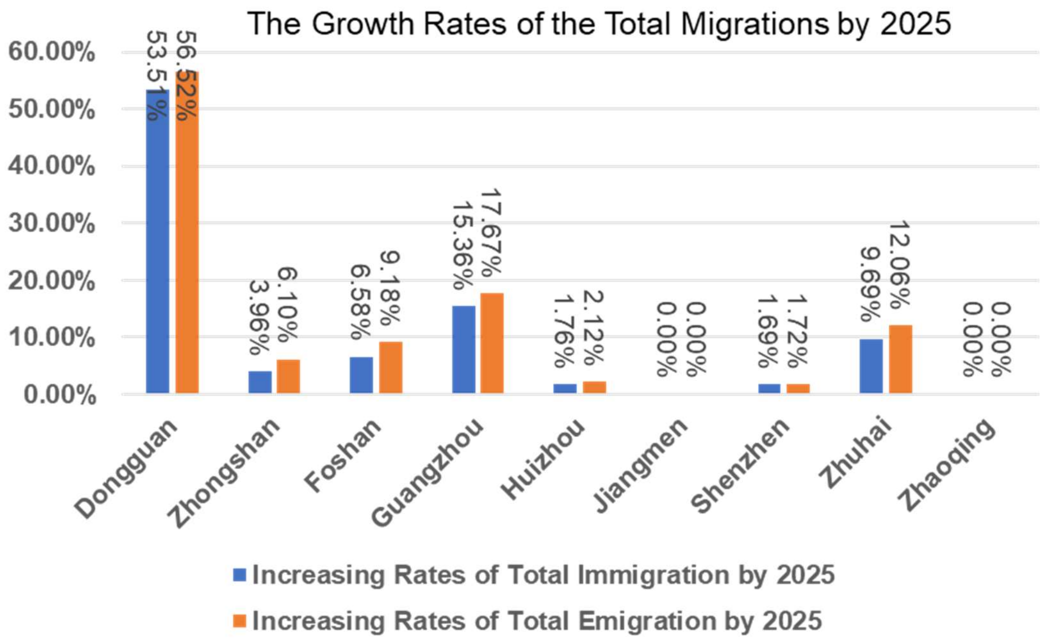



*Figure 4.14 The Histogram of the Growth Rates of the Total Immigration or Emigration by 2025 in the GBA* *(Source: Author)*

### 4.3.2 The Impacts of the Long-term Rail Construction

The estimated migration pattern in 2035 (Figure 4.15) after the achievement of the 'One-hour Metropolitan Area' of the GBA looks quite similar to the estimated flows in 2025 (Figure 4.12). The reason behind this limited difference could be interpreted by the growth rates of the affected migration flows, as shown in the Figure 4.16. 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.

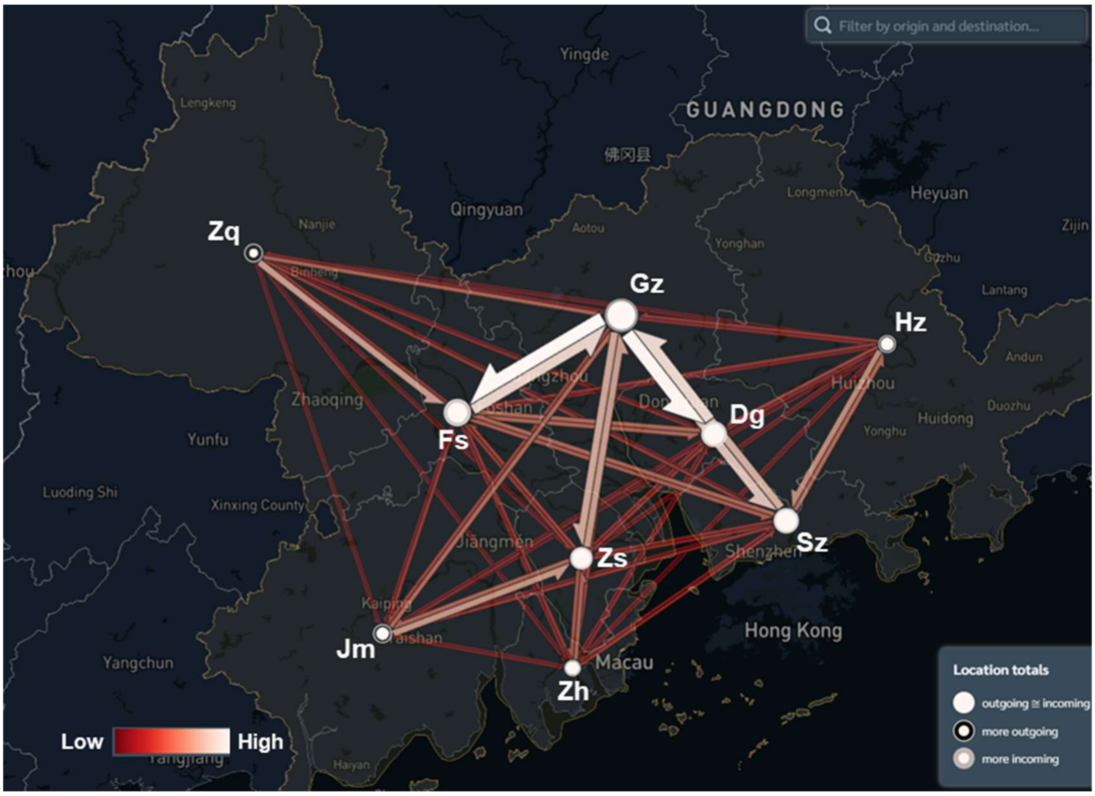



*Figure 4.15 The Estimated Flows based on the Impacts of the Long-term Railway Planning by 2035 of the GBA* (Source: Author)

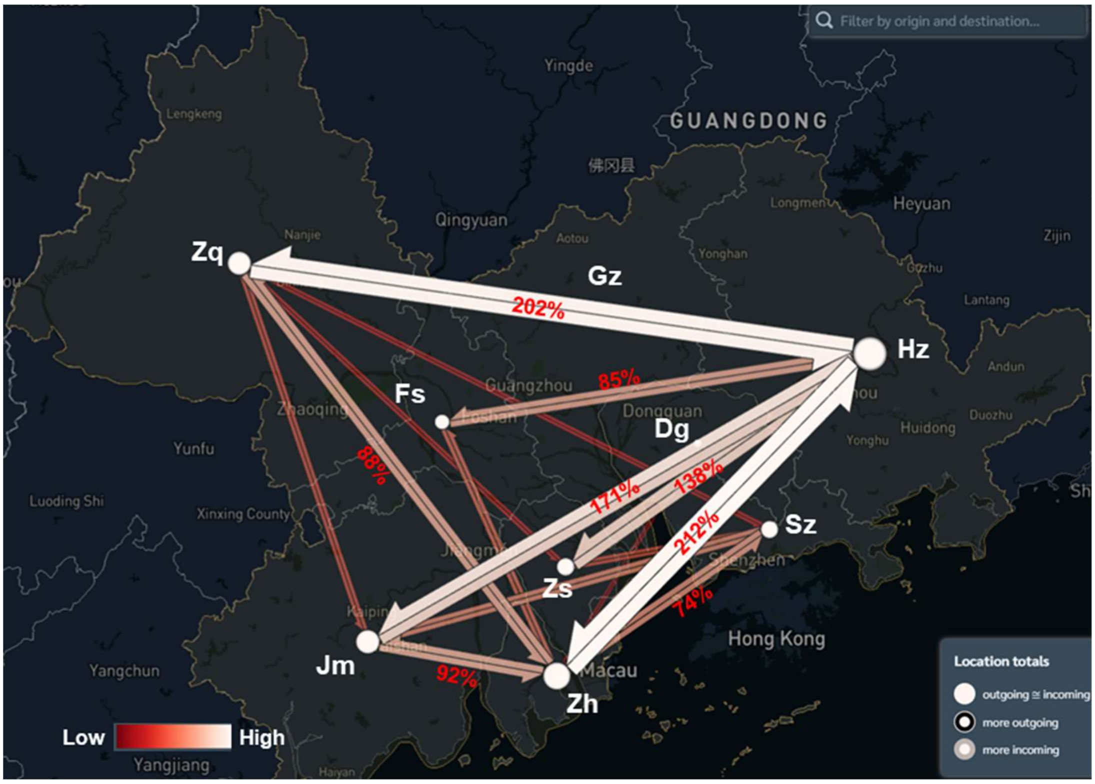



*Figure 4.16 The Growth Rates of the Estimated Flows Affected by the Long-term Railway Planning by 2035 to the Original Estimates* (Source: Author)

The growth rates of the total migration by 2035 are shown below in the Figure 4.17. 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.

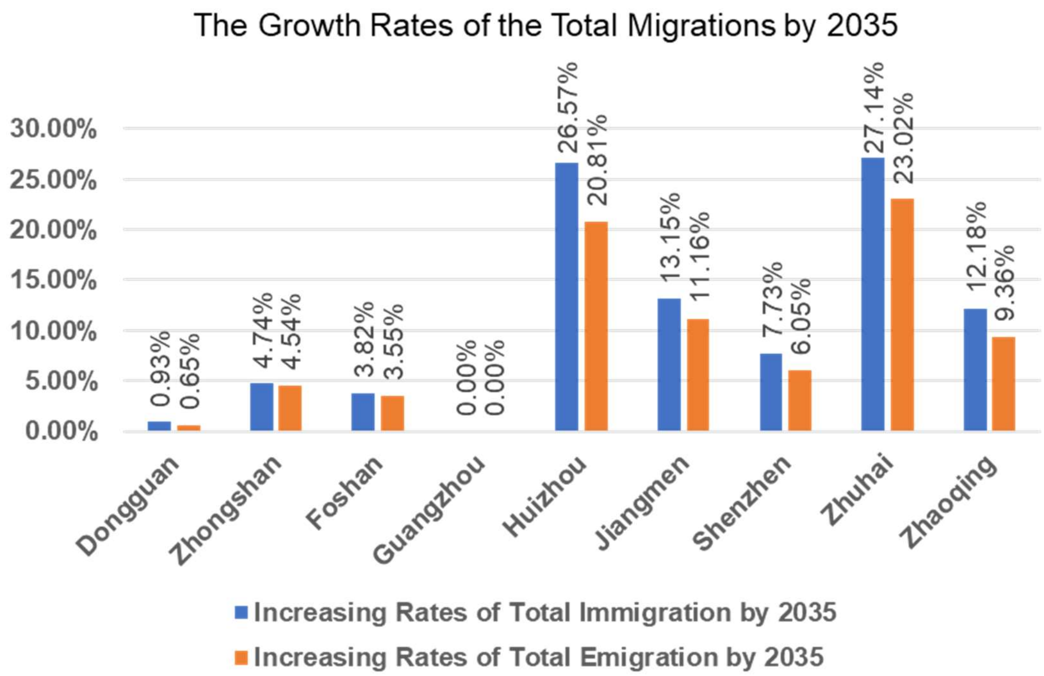



*Figure 4.17 The Histogram of the Growth Rates of the Total Immigration or Emigration by 2035 in the GBA* *(Source: Author)*

# Chapter 5 Conclusion

In order to contribute to the empirical studies that address spatial interaction in terms of long-term migration within the PM-CRs in China, this paper has visualised and simulated the internal migration flows and the potential impacts of inter-city railway planning on those migration flows within the GBA mainly by using the unconstrained gravity models. In addition, the relationships between those migration flows and different indicators, including socio-economic and travel cost factors, have also been quantitatively revealed based on the results of models.

Briefly, the strong connections within the Z-shaped geographic structure including 5 cities are significantly prominent in the current long-term migration pattern for whether the GBA or Gd province, and the migration flows from or to these cities will further increase in the next five years with the implementation of short-term railway planning, especially the flows from or to Dongguan. However, from 2025 to 2035 will witness the closer spatial interaction between the other four cities on the geographical edge of the GBA. 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 there are some limitations mentioned in the Chapter 3, the relatively good fitness of the models demonstrates the significant applicability of gravity model in interpreting the spatial interaction between Chinese cities. 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 can 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.

# 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, G J (1989) Economic theory and international migration, *International migration review*, *23*(3), pp.457-485.

Boyle, P J, 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, A G, 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, K F, 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.

Evans, A W (1971) The calibration of trip distribution models with exponential or similar cost functions, *Transportation research*, *5*(1), pp.15-38.

Fan, C C (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, A S and O'Kelly, M E (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, M J, Van Oort, F G 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, M J, Hunt, G L, Rickman, D S and Treyz, G I (1991) Migration, regional equilibrium, and the estimation of compensating differentials, *The American Economic Review*, *81*(5), pp.1382-1390.

Greenwood, M J and McDowell, J M, (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, J F Jr, Anderson, R E, Tatham, R L and Black, W C (1995) *Multivariate Data Analysis (3rd ed.,* New York: Macmillan.

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

Hall, P G 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.

Haynes, K E and Fotheringham, A S (1984) *Gravity and spatial interaction models* (Vol. 2). Beverly Hills, CA: Sage.

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, C I 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, G M (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, V I and Davis, B (2000) A gravity model analysis of international migration to North America, *Applied economics*, *32*(13), pp.1745-1755.

Mansfield, E R and Helms, B P (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, R E (1972) Migration as Spatial Interaction, *Journal of Geography,* 71(5), pp.294–301.

Poot, J, Alimi, O, Cameron, M P and Maré, D C (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, W J (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, E S (2010) Theoretical Underpinnings of the Gravity Hypothesis, *Geographical analysis,* 10(4), pp.386–402.

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.

Stewart, J Q (1950) Potential of Population and Its Relationship to Marketing, *Operational research quarterly, 1(3), p.47.*

Stillwell, J C H (1978) Interzonal migration: some historical tests of spatial-interaction models, *Environment and Planning A*, *10*(10), pp.1187-1200.

Stouffer, S A (1940) *Intervening Opportunities: A Theory Relating Mobility and Distance, American sociological review,* 5(6), pp.845–867.

Sun, W Y, Qiu, K F, 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, A G (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, X Y, Zhao, C, Fan, Y, Di, Z and Wang, W X (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, A G O 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, Y M, Zeng, A, Yan, X Y, Wang, W X and Lai, Y C (2016) Unified underpinning of human mobility in the real world and cyberspace, *New Journal of Physics*, *18*(5), p.053025.

Zipf, G K (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.

# Appendix

## Appendix A:

The Summary List of All Data Source Employed in this Research

| Data Name | Source Name | Source Link |
|---|---|---|
| Observed Number of Migrants within Gd Province by the end of 2018 | 2019 Statistical Year books of Gd Province and Cities within Gd | http://stats.gd.gov.cn/ |
| Socio-economic Data of Cities within Gd Province, including Population, Income, Economy, Housing Price and Education | | |
| Daily Movements during the Spring Festival Travel Rush | Baidu Map Migration Platform | http://qianxi.baidu.com/ |
| Minimum travel time by driving | Baidu Map Route Matrix API | http://lbsyun.baidu.com/index.php?title=webapi/route-matrix-api-v2 |
| Minimum Travel Time by any Kinds of Public Transportation | 12306 China Railway | https://www.12306.cn/index/ |
| Minimum Travel Time by Ordinary Trains | | |
| The Administrative Boundaries of the Cities within Gd Province | - | https://mp.weixin.qq.com/s/ilTvdgQscwL89DCvFjjYOw |

## Appendix B:

The Scripts of Python to Capture the Baidu Migration Index

(Note: The process of crawling the data took place in March, 2020, and *Baidu Map Migration Platform* has been closed since May 8, 2020. Therefore, this research will not be responsible for the applicability of the following code after the platform is closed.)

The scripts used in this research are mainly referenced from Xiuyi Chen@CSDN, accessed at:
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

And Samelltiger@GitHub, accessed at:

https://github.com/samelltiger/baidu_qx

### B1: The Scripts Used to Capture the Total Migration Index of Cities within Gd:

```
import requests

import json

import time

import xlsxwriter


CitiesCode = {'广州市':440100,'韶关市':440200,'深圳市':440300,

'珠海市':440400,'汕头市':440500,'佛山市':440600,'江门市':440700,

'湛江市':440800,'茂名市':440900,'肇庆市':441200,'惠州市':441300,

'梅州市':441400,'汕尾市':441500,'河源市':441600,'阳江市':441700,

'清远市':441800,'东莞市':441900,'中山市':442000,'潮州市':445100,

'揭阳市':445200,'云浮市':445300,}


def migration_index(FileTittle,classname,direction,CodeDict):

    if direction == 'in' :

        nameofdire = '迁入'

    if direction == 'out':

        nameofdire = '迁出'
```

```
workbook = xlsxwriter.Workbook(f'{FileTittle} {nameofdire}规模指数.xlsx')
    worksheet = workbook.add_worksheet('Sheet')
if direction == 'in' :
        nameofdire = '迁入'
    if direction == 'out':
        nameofdire = '迁出'
    CitiesOrder = {}
    worksheet.write(0 , 0 , '城市代码')
    worksheet.write(0 , 1 , '城市')
    times = 1
    for key , value in CodeDict.items():
        worksheet.write(times , 0 , str(value))
        worksheet.write(times , 1 , str(key))
        CitiesOrder[str(key)] = times
        times += 1
for key , value in CodeDict.items():
url=f'http://huiyan.baidu.com/migration/historycurve.jsonp?dt={classname}&id={value}&type=move_{dire
ction}'
        print(f'{key}:{url}')
        response=requests.get(url, timeout=2) # #json
        time.sleep(3)
        r=response.text[3:-1]
        data_dict=json.loads(r)
        if data_dict['errmsg']=='SUCCESS':
            data_list=data_dict['data']['list']
            counter_date = 2        #date counter
            datelist = []
            for date , index in data_list.items():      #date order
                datelist.append(date)
            datelist.sort()
            for date in datelist:
                index = data_list[date]
                # print(f'{date} : {index}')
                worksheet.write(0 , counter_date , float(date))
                worksheet.write(CitiesOrder[str(Area)] , counter_date , float(index))
                counter_date += 1   # date + 1
        else:
            print('Error')
```

```
    workbook.close()
if __name__=="__main__":
    for key,value in CitiesCode.items():
        migration_index(key,'city','in',CitiesCode)
        migration_index(key,'city','out',CitiesCode)
    print('All done')
```

## B2: The Scripts Used to Capture the Particular Percentage of One certain Origin or Destination to the Total Index:

```
import requests
import json
import time
import xlrd
import xlwt

CitiesCode = {'广州市':440100,'韶关市':440200,'深圳市':440300,'珠海市':440400,
'汕头市':440500,'佛山市':440600,'江门市':440700,'湛江市':440800,'茂名市':440900,
'肇庆市':441200,'惠州市':441300,'梅州市':441400,'汕尾市':441500,'河源市':441600,
'阳江市':441700,'清远市':441800,'东莞市':441900,'中山市':442000,'潮州市':445100,
'揭阳市':445200,'云浮市':445300,}

def migration_all_date(areaname,classname,no,direction):
    if no == -1 :
        no = CitiesCode[str(areaname)]

    workbook = xlwt.Workbook(encoding = 'utf-8')
    worksheet = workbook.add_sheet('Sheet', cell_overwrite_ok=True)

    if direction == 'in' :
        nameofdire = '迁入来源地'
    if direction == 'out':
        nameofdire = '迁出目的地'
    CitiesOrder = {}
    worksheet.write(0 , 0 , label='城市代码')
```

```
    worksheet.write(0 , 1 , label=str(nameofdire))
    times = 1
    for key , value in CitiesCode.items():
        worksheet.write(times , 0 , label=str(value))
        worksheet.write(times , 1 , label=str(key))
        CitiesOrder[str(key)] = times
        times += 1

    datelist = []
    counter_data = 2                    # Date Counter
    for date1 in range(20200101,20200132):
        datelist.append(date1)
    for date2 in range(20200201,20200230):
        datelist.append(date2)
    for date3 in range(20200301,20200302):
        datelist.append(date3)
    for date in datelist:
        datename = date
        time.sleep(1)
url=f'http://huiyan.baidu.com/migration/cityrank.jsonp?dt={classname}&id={no}&type=move_{direction}&
date={date}'
        print(url)
        response=requests.get(url, timeout=2) #json
        time.sleep(1)
        r=response.text[3:-1]
        data_dict=json.loads(r)
        if data_dict['errmsg']=='SUCCESS':
            data_list=data_dict['data']['list']
            time.sleep(1)

            worksheet.write(0 , counter_data , label=datename)
            for a in range(len(CitiesCode)):
                worksheet.write(a+1 , counter_data , label=0)

            for i in range (len(data_list)):
                city_name=data_list[i]['city_name']
                value=data_list[i]['value']
```

```
            worksheet.write(CitiesOrder[str(city_name)] , counter_data , label=value)
          counter_data += 1
    workbook.save(f"{areaname}-{nameofdire}.xls")
def circu_exe_direction(areaname,classname,no):
    mukous = ['in','out']
    for mukou in mukous:
        migration_all_date(areaname,classname,no,mukou)
    print(str(areaname)+'---','Success')
if __name__=="__main__":
    for key,value in CitiesCode.items():
        circu_exe_direction(key,'city',-1)
    print('All Done')
```

## Appendix C:

The Final Matrix of O-D Long-term Migration Flows between 21 Cities within Gd Province

| | Gz | Dg | Fs | Hz | Jm | Sz | Zh | Zq | Zs | Cz | Hy | Jy | Mm | Mz | Qy | Sg | St | Sw | Yf | Yj | Zj |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Gz** | | 14.14 | 27.48 | 4.68 | 3.16 | 11.86 | 3.10 | 2.81 | 4.56 | 0.36 | 1.08 | 0.91 | 1.26 | 1.02 | 5.63 | 4.30 | 0.98 | 0.74 | 1.01 | 0.92 | 1.48 |
| **Dg** | 9.29 | | 1.97 | 6.65 | 0.55 | 22.82 | 0.64 | 0.37 | 1.11 | 0.09 | 0.64 | 0.40 | 0.47 | 0.37 | 0.54 | 0.79 | 0.24 | 0.31 | 0.19 | 0.20 | 0.31 |
| **Fs** | 26.74 | 2.62 | | 0.80 | 2.55 | 2.96 | 0.91 | 3.46 | 4.39 | 0.07 | 0.23 | 0.18 | 0.46 | 0.22 | 1.65 | 1.20 | 0.17 | 0.15 | 0.80 | 0.38 | 0.52 |
| **Hz** | 5.90 | 8.85 | 1.17 | | 0.20 | 17.80 | 0.34 | 0.15 | 0.46 | 0.08 | 1.92 | 0.39 | 0.13 | 0.47 | 0.22 | 0.45 | 0.25 | 0.90 | 0.05 | 0.06 | 0.13 |
| **Jm** | 4.73 | 1.17 | 3.67 | 0.24 | | 1.60 | 1.47 | 0.34 | 2.82 | 0.01 | 0.04 | 0.03 | 0.23 | 0.04 | 0.17 | 0.19 | 0.04 | 0.02 | 0.35 | 0.61 | 0.23 |
| **Sz** | 8.08 | 21.39 | 2.05 | 12.21 | 0.65 | | 1.30 | 0.38 | 1.35 | 0.22 | 1.26 | 1.00 | 0.54 | 0.93 | 0.47 | 0.96 | 0.75 | 1.41 | 0.21 | 0.23 | 0.56 |
| **Zh** | 3.18 | 0.80 | 1.00 | 0.26 | 1.12 | 1.70 | | 0.16 | 8.93 | 0.02 | 0.05 | 0.07 | 0.22 | 0.07 | 0.13 | 0.22 | 0.07 | 0.05 | 0.09 | 0.23 | 0.24 |
| **Zq** | 7.25 | 2.75 | 11.44 | 0.57 | 0.88 | 2.28 | 0.51 | | 2.19 | 0.04 | 0.07 | 0.07 | 0.16 | 0.06 | 0.60 | 0.20 | 0.10 | 0.05 | 1.65 | 0.13 | 0.13 |
| **Zs** | 3.66 | 1.26 | 3.98 | 0.32 | 2.11 | 1.77 | 9.15 | 0.31 | | 0.03 | 0.09 | 0.08 | 0.22 | 0.10 | 0.21 | 0.32 | 0.07 | 0.06 | 0.19 | 0.27 | 0.24 |
| **Cz** | 2.13 | 0.43 | 0.36 | 0.23 | 0.03 | 1.33 | 0.12 | 0.04 | 0.17 | | 0.03 | 0.99 | 0.02 | 0.29 | 0.03 | 0.05 | 4.23 | 0.06 | 0.01 | 0.01 | 0.02 |
| **Hy** | 3.05 | 3.44 | 0.98 | 4.91 | 0.13 | 6.72 | 0.24 | 0.08 | 0.44 | 0.04 | | 0.11 | 0.04 | 0.57 | 0.12 | 0.68 | 0.10 | 0.08 | 0.03 | 0.02 | 0.04 |
| **Jy** | 6.16 | 3.01 | 1.12 | 1.59 | 0.15 | 7.30 | 0.41 | 0.11 | 0.54 | 1.00 | 0.16 | | 0.04 | 0.69 | 0.10 | 0.23 | 5.31 | 0.79 | 0.02 | 0.02 | 0.06 |
| **Mm** | 9.42 | 8.16 | 3.03 | 1.02 | 0.90 | 8.36 | 1.38 | 0.43 | 1.74 | 0.02 | 0.07 | 0.04 | | 0.06 | 0.20 | 0.24 | 0.05 | 0.03 | 0.41 | 0.94 | 3.33 |
| **Mz** | 6.63 | 3.68 | 1.74 | 2.54 | 0.24 | 9.63 | 0.55 | 0.18 | 0.87 | 0.51 | 0.82 | 0.76 | 0.05 | | 0.20 | 0.41 | 0.56 | 0.15 | 0.03 | 0.03 | 0.06 |
| **Qy** | 14.29 | 5.60 | 6.68 | 1.05 | 0.45 | 4.33 | 0.53 | 0.65 | 1.40 | 0.04 | 0.11 | 0.08 | 0.08 | 0.07 | | 2.75 | 0.10 | 0.08 | 0.11 | 0.06 | 0.10 |
| **Sg** | 6.97 | 3.25 | 2.89 | 0.75 | 0.32 | 3.50 | 0.43 | 0.15 | 1.00 | 0.03 | 0.30 | 0.07 | 0.06 | 0.09 | 1.28 | | 0.07 | 0.04 | 0.04 | 0.04 | 0.08 |
| **St** | 4.68 | 0.91 | 0.67 | 0.70 | 0.09 | 4.20 | 0.37 | 0.09 | 0.28 | 2.69 | 0.10 | 3.74 | 0.05 | 0.28 | 0.09 | 0.17 | | 0.22 | 0.02 | 0.02 | 0.07 |
| **Sw** | 2.97 | 1.51 | 0.65 | 2.20 | 0.08 | 6.24 | 0.27 | 0.06 | 0.36 | 0.06 | 0.10 | 0.73 | 0.02 | 0.11 | 0.06 | 0.06 | 0.34 | | 0.02 | 0.01 | 0.03 |
| **Yf** | 3.85 | 2.03 | 3.51 | 0.30 | 0.72 | 1.98 | 0.35 | 2.12 | 1.16 | 0.01 | 0.05 | 0.02 | 0.34 | 0.03 | 0.17 | 0.12 | 0.03 | 0.02 | | 0.34 | 0.11 |
| **Yj** | 3.11 | 1.76 | 1.53 | 0.30 | 1.14 | 1.80 | 0.76 | 0.21 | 1.36 | 0.01 | 0.03 | 0.02 | 0.87 | 0.02 | 0.09 | 0.09 | 0.02 | 0.03 | 0.37 | | 0.33 |
| **Zj** | 11.19 | 3.59 | 3.46 | 0.71 | 0.74 | 6.11 | 1.57 | 0.27 | 1.76 | 0.02 | 0.06 | 0.04 | 3.33 | 0.06 | 0.19 | 0.20 | 0.07 | 0.04 | 0.11 | 0.39 | |

**Column: Origins** **Row: Destinations**

GBA Cities

## Appendix D:

The Sample Matrix of the Minimum Driving Time between the 21 Cities within the Whole Guangdong Province (Unit: hours)

| | Gz | Dg | Fs | Hz | Jm | Sz | Zh | Zq | Zs | Cz | Hy | Jy | Mm | Mz | Qy | Sg | St | Sw | Yf | Yj | Zj |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Gz** | | 1 | 0.733 | 1.917 | 1.283 | 1.7 | 1.833 | 1.667 | 1.3 | 4.667 | 2.334 | 4.333 | 4.333 | 4.4 | 1.117 | 2.617 | 4.833 | 3.217 | 1.967 | 3 | 5.133 |
| **Dg** | 1 | | 1.433 | 1.75 | 1.783 | 1.333 | 2.2 | 2.2 | 1.7 | 4.55 | 2.383 | 4.3 | 4.533 | 4.183 | 2.067 | 3.267 | 5.05 | 3.167 | 3.133 | 3.633 | 5.283 |
| **Fs** | 0.733 | 1.433 | | 2.333 | 1.2 | 2 | 1.95 | 1.4 | 1.35 | 5.05 | 2.783 | 4.733 | 3.767 | 4.8 | 1.583 | 3.183 | 5.367 | 3.75 | 1.833 | 2.667 | 4.55 |
| **Hz** | 1.917 | 1.75 | 2.333 | | 2.917 | 1.717 | 3.367 | 3.1 | 2.567 | 3.417 | 1.383 | 3.117 | 5.733 | 3.283 | 2.517 | 3.767 | 3.767 | 1.783 | 3.733 | 4.767 | 6.5 |
| **Jm** | 1.283 | 1.783 | 1.2 | 2.917 | | 2 | 1.533 | 1.45 | 1.033 | 5.5 | 3.467 | 5.183 | 3.3 | 5.4 | 2.217 | 3.617 | 5.85 | 3.95 | 1.8 | 2.25 | 4.133 |
| **Sz** | 1.7 | 1.333 | 2 | 1.717 | 2 | | 2.25 | 2.683 | 1.683 | 4.2 | 2.467 | 3.883 | 5 | 4.383 | 3.65 | 4.2 | 4.117 | 2.35 | 3.133 | 3.867 | 5.8 |
| **Zh** | 1.833 | 2.2 | 1.95 | 3.367 | 1.533 | 2.25 | | 2.6 | 1.133 | 6.1 | 3.85 | 5.8 | 4.033 | 6.183 | 3.4 | 4.4 | 6.2 | 4.4 | 3.117 | 2.717 | 4.867 |
| **Zq** | 1.667 | 2.2 | 1.4 | 3.1 | 1.45 | 2.683 | 2.6 | | 2.067 | 5.617 | 3.383 | 5.3 | 3.2 | 5.267 | 1.933 | 3.417 | 5.733 | 4.317 | 0.95 | 2.633 | 4 |
| **Zs** | 1.3 | 1.7 | 1.35 | 2.567 | 1.033 | 1.683 | 1.133 | 2.067 | | 5.733 | 3.517 | 5.517 | 4.133 | 5.633 | 2.7 | 3.817 | 5.667 | 4.067 | 2.7 | 3.067 | 4.817 |
| **Cz** | 4.667 | 4.55 | 5.05 | 3.417 | 5.5 | 4.2 | 6.1 | 5.617 | 5.733 | | 3.283 | 0.967 | 8.433 | 2.133 | 5 | 5.367 | 1.167 | 2.433 | 6.233 | 7.55 | 9.2 |
| **Hy** | 2.334 | 2.383 | 2.783 | 1.383 | 3.467 | 2.467 | 3.85 | 3.383 | 3.517 | 3.283 | | 3 | 6.2 | 2.383 | 2.833 | 3.167 | 3.367 | 2.817 | 4.05 | 5.417 | 6.917 |
| **Jy** | 4.333 | 4.3 | 4.733 | 3.117 | 5.183 | 3.883 | 5.8 | 5.3 | 5.517 | 0.967 | 3 | | 8.317 | 1.85 | 4.75 | 5.067 | 1.2 | 2.183 | 5.917 | 7.283 | 8.883 |
| **Mm** | 4.333 | 4.533 | 3.767 | 5.733 | 3.3 | 5 | 4.033 | 3.2 | 4.133 | 8.433 | 6.2 | 8.317 | | 8.083 | 4.467 | 6.067 | 9.083 | 6.817 | 2.85 | 1.733 | 1.367 |
| **Mz** | 4.4 | 4.183 | 4.8 | 3.283 | 5.4 | 4.383 | 6.183 | 5.267 | 5.633 | 2.133 | 2.383 | 1.85 | 8.083 | | 4.683 | 4.35 | 2.267 | 3.133 | 5.85 | 7.167 | 8.833 |
| **Qy** | 1.117 | 2.067 | 1.583 | 2.517 | 2.217 | 3.65 | 3.4 | 1.933 | 2.7 | 5 | 2.833 | 4.75 | 4.467 | 4.683 | | 2.283 | 5.15 | 3.867 | 2.183 | 3.667 | 5.133 |
| **Sg** | 2.617 | 3.267 | 3.183 | 3.767 | 3.617 | 4.2 | 4.4 | 3.417 | 3.817 | 5.367 | 3.167 | 5.067 | 6.067 | 4.35 | 2.283 | | 5.667 | 5 | 3.833 | 5.333 | 6.833 |
| **St** | 4.833 | 5.05 | 5.367 | 3.767 | 5.85 | 4.117 | 6.2 | 5.733 | 5.667 | 1.167 | 3.367 | 1.2 | 9.083 | 2.267 | 5.15 | 5.667 | | 2.367 | 6.333 | 8.117 | 9.933 |
| **Sw** | 3.217 | 3.167 | 3.75 | 1.783 | 3.95 | 2.35 | 4.4 | 4.317 | 4.067 | 2.433 | 2.817 | 2.183 | 6.817 | 3.133 | 3.867 | 5 | 2.367 | | 0.133 | 6.383 | 7.667 |
| **Yf** | 1.967 | 3.133 | 1.833 | 3.733 | 1.8 | 3.133 | 3.117 | 0.95 | 2.7 | 6.233 | 4.05 | 5.917 | 2.85 | 5.85 | 2.183 | 3.833 | 6.333 | 0.133 | | 2.217 | 3.7 |
| **Yj** | 3 | 3.633 | 2.667 | 4.767 | 2.25 | 3.867 | 2.717 | 2.633 | 3.067 | 7.55 | 5.417 | 7.283 | 1.733 | 7.167 | 3.667 | 5.333 | 8.117 | 6.383 | 2.217 | | 2.567 |
| **Zj** | 5.133 | 5.283 | 4.55 | 6.5 | 4.133 | 5.8 | 4.867 | 4 | 4.817 | 9.2 | 6.917 | 8.883 | 1.367 | 8.833 | 5.133 | 6.833 | 9.933 | 7.667 | 3.7 | 2.567 | |

**Column: Origins** **Row: Destinations**

**GBA Cities**

**Appendix E:**

The Information about Some Railways in the Short-term Rail Planning for the GBA Effected the Minimum Travelling Time

| Railway Name | Kilometres | Main Access Places | Time Node |
|---|---|---|---|
| Foshan-Dongguan Railway | 36.7 | Foshan, Guangzhou, Dongguan | Completed by the end of 2020 |
| Nansha-Zhuhai Railway | 79 | Guangzhou, Zhongshan, Zhuhai | Start before 2025 |
| Zhongshan-Tangxia Railway | 63 | Zhongshan, Guangzhou, Dongguan | Start before 2025 |
| Hu-Long Railway | 12.5 | Dongguan, Shenzhen | Start before 2025 |
| Zhong-Hu-Long Railway | 54+12.5 | Zhuhai, Zhongshan, Dongguan, Shenzhen | Start before 2025 |