



Delineating urban functional zones using mobile phone data: A case study of cross-boundary integration in Shenzhen-Dongguan-Huizhou area

Bowen Zhang^a, Chen Zhong^{b,*}, Qili Gao^{b,c}, Zahratu Shabrina^a, Wei Tu^c

^a Department of Geography, King's College London, London, United Kingdom

^b Centre for Advanced Spatial Analysis, University College London, London, United Kingdom

^c Department of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China

ARTICLE INFO

Keywords:

Urban functional zones (UFZs)
Human mobility
Spatial interaction model
Mega-city region
Mobile phone data

ABSTRACT

As cities continuously expand and with the emergence of mega-city regions, the urban functional zones (UFZs) have spread beyond their original administrative boundaries. An accurate and updated delineation of the UFZs is crucial for assessing the functional integration between cities within a mega-city region. Mobility data provides a chance to depict the UFZs from actual human activities at a finer spatial scale. Existing studies mostly adopted network-based approaches relying on the topological relationship but ignoring spatial factors, causing the lack of sensitivity in detecting the cross-cities integration of the functional region. This research proposed a novel regionalisation algorithm that redraws non-overlap boundaries of urban functional zones based on the commuting origin-destination matrix, representing the spatial interactions within cities and cross-cities. In particular, functional zones are drawn by searching for the best partition with the best goodness of fitting in the hierarchical spatial interaction model. The algorithm was applied to a case study of a mega-city region, Shenzhen-Dongguan-Huizhou (SDH) area in China using mobile phone signalling data. By adopting two different settings, this model evaluated the current status and predict the future trend of urban integration respectively. The results show the current boundary of UFZs in the SDH area almost coincides with administrative boundaries. Meanwhile, the results of long-term predictions might be utilised by policymakers to give more attention to the areas near the Dongguan-Huizhou boundary to promote industry cooperation and avoid mismatches between the functional and administrative regions.

1. Introduction

City regions have been formed as a consequence of urban growth and a vast improvement in inter-city connectivity since the second half of the 20th century (Hall & Pain, 2006). Governments worldwide encourage regional cooperation to acquire efficiency premiums from agglomeration economies (Brenner, 2002). For example, the Chinese government has issued a series of policies since 2004 to encourage cities within one city-region to integrate as one city, which attempts to promote industrial cooperation and sharing urban function (Li, Wu, & Hay, 2015; Wu, 2016). As cities continuously expand and interact with other cities, human daily activities related to urban function (e.g., work, residence, recreation) have expanded beyond their original administrative boundaries and occur in different cities, forming urban functional zones (UFZs) (Gao, Janowicz, & Couclelis, 2017; Yeh & Chen, 2019; Zhai et al., 2019; Zhong, Arisona, Huang, Batty, & Schmitt, 2014). These increasing

cross-boundary trips and the ambiguity of UFZs raised new challenges for regional planning and management in response to the rapid development of mega-city regions.

Despite the widespread existence of this phenomenon, there are limitations in assessing how UFZs have been integrated across cities due to the lack of large-scale data and the corresponding analytical methods. The emerging mobility data provides an opportunity for a breakthrough to delineate UFZs from intra-city and intercity trips. The current mobility data covers the daily movement flows of a huge population. Moreover, unlike the traditional survey data conducted separately by individual local authorities, the new form of data (e.g., mobile signalling data and social media data) enables us to analyse a finer-grained network beyond city/county boundaries with a unified dataset. Previous research on detecting spatial structure and community detections using mobility data mainly applied network analysis (Jin et al., 2021; Shen & Batty, 2019; Wu, Smith, & Wang, 2021; Zhong et al., 2014).

* Corresponding author.

E-mail address: c.zhong@ucl.ac.uk (C. Zhong).

<https://doi.org/10.1016/j.compenvurbsys.2022.101872>

Received 7 February 2022; Received in revised form 17 May 2022; Accepted 17 August 2022

Available online 14 September 2022

0198-9715/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

However, the network-based analysis shows limitations as the distance decay effect often has not been appropriately reflected in the topology relationship (Liu, Sui, Kang, & Gao, 2014; Yin, Soliman, Yin, & Wang, 2017), which may result in insensitivity to changes in cross-boundary flows (Liu et al., 2014). Thus, a novel method for detecting UFZs with more sensitivity for cross-boundary travel flow and distance decay effect is needed.

For this study, a critical hypothesis is that the boundary of a functional zone is highly associated with local travel behaviour, that is, distance decay effects in our case. Zipf (1946) proposed that human mobility follows a spatial distribution with a distance decay from centres to the periphery. This concept has been accepted and applied in previous trip estimation models (Batty & Milton, 2021; Masucci, Serras, Johansson, & Batty, 2013; Wilson, 1971). Thus, the heterogeneity of trip distribution can be seen as an indicator to reveal the discontinuity of urban functional space or the “border effect” of urban functional zones (Brown, Dar-Brodeur, & Tweedle, 2020; Jin et al., 2021). When crossing different urban functional zones, the border effect can be observed in human travel activity. Such border effects could be used as indicators for delineating urban functional zones (Jin et al., 2021; Rinzivillo et al., 2012; Shen & Batty, 2019). As one of the most widely used methods for predicting interaction flows, spatial interaction models predict the strength of spatial interaction based on the distance decay effect. Previous studies confirmed the border effect can be represented by spatial heterogeneity in the spatial interaction model (Jin et al., 2021; McCallum, 1995), which provides a new method for delineating the UFZs and overcoming the limitations of network-based methods. Therefore, this study would answer a research question on *how to delineate the UFZs of cities within a mega-city region by spatial interaction model using human mobility data?*

This study first applies a two-level hierarchical spatial interaction model (HSIM) to generate the flow of spatial interaction between zones, then redraws non-overlap boundaries of urban functional zones by searching for the best partition with the best goodness of fitting in HSIM. The algorithm is applied to delineate the cities’ functional regions within a specific mega-city region, Shenzhen-Dongguan-Huizhou (SDH) area in two different settings. The results prove that the goodness of fitting in HSIM can represent reasonable cities’ boundaries. The empirical study of SDH area shows current UFZs almost coincide with administrative boundaries. Meanwhile, our results of long-term predictions remind policymakers to give more attention to the areas near the Dongguan-Huizhou boundary to promote industry cooperation and avoid mismatches between functional and administrative regions providing implications for related regional planning policies.

2. Literature review

2.1. Urban spatial structure and functional region

The urban spatial structure is topical research in urban geography. Various factors, such as politics, economic activities, topography, history, infrastructures, and policies, interact with the urban spatial structure and eventually form how city elements are geographically located (Dadashpoor & Yousefi, 2018; Engelfriet & Koomen, 2018). The current prevailing interpretations of urban spatial structure can be categorised into morphological structures and functional structures, distinguished based on the data sources and how urban structures are interpreted (Green, 2007). The urban spatial structures are traditionally described using morphological properties based on traditional geographical data, demographical data and built-up areas. Previous research about the urban spatial structure or boundary was dedicated to identifying built environment areas from remote sensing data by utilising classification algorithms (Henderson, Yeh, Gong, Elvidge, & Baugh, 2003; Lu, Li, Kuang, & Moran, 2014). Other studies also attempted to define or detect the urban boundaries from morphological observation (Tannier, Thomas, Vuidel, & Frankhauser, 2011), using data

analytic frameworks such as the scaling law (Alvioli, 2020; Arcaute et al., 2015; Cottineau, Finance, Hatna, Arcaute, & Batty, 2019) and the transport network density (Long, 2016; Long, Zhai, Shen, & Ye, 2018).

In contrast to the concept of morphology, functional structure emphasises the socio-economic links between urban areas. Compared with the morphological structure, the functional structure is more temporal and dynamic (Wu et al., 2021), which would better correspond to the rapid changes in the urban environment. Various urban flows like commuting and logistic flows within the city are being used to describe the urban spatial structure by spatial interactions (Burger & Meijers, 2012; Sohn, 2005; Zhong et al., 2014). Therefore, two distant areas can be integrated into a community because of the strong links of functional elements (Zhang, Marshall, Cao, Manley, & Chen, 2021). This feature would benefit the understanding of urban functional integration across cities.

Researchers have conducted studies on the urban spatial structure at two levels due to spatial scale differences: the intra-city level and the city-region level. For intra-city polycentric spatial structures, scholars have focused on the location, and morphological attributes of the newly emerged centres in the evolution of cities from monocentric to polycentric and then analysed the systemic characteristics and interrelationships between the internal centres. Meanwhile, in regional-level or country-level polycentric studies of spatial structure, studies usually take administrative cities as the centre of regional spatial structure rather than searching the urban centres by detecting method (Huang, Liu, & Zhao, 2015; Gao et al., 2017). Therefore, delineating the urban functional zones between cities is essential for discussing the functional spatial structure in a city region or larger scope. This can be done by using some well-known existing regionalisation algorithms for delineating regions based on indicators or objective functions. Some examples of such methods include P-regions and max-p which are based on a defined objective function, while REDCAP and SKATER are methods based on hierarchical structure reflecting neighbourhood relationship (Duque, Church, & Middleton, 2011; Guo, 2008; Helbich, Brunauer, Hagenauer, & Leitner, 2013). These methods mainly use social-economic indicators (e.g., house price, income) to find the spatial clustering or non-spatial similarity rather than using flow data to assess the connection between areas.

2.2. Trip distribution laws and models

Most widely used transport and urban models assume that flows of people follow a spatial distribution with a distance decay from urban centres to the periphery (Anderson, 2011). Zipf’s law states that the volume of movement flow would be directly proportionate to the product of the population sizes between any two communities $P_i \times P_j$ and inversely proportional to the transport distance d_{ij} (Eq. (1)), which construed the base of modern analytics of human mobility patterns (Zipf, 1946).

$$T_{ij} \propto \frac{P_i \times P_j}{f(d_{ij})} \quad (1)$$

This law has been extensively studied since then and has been developed as the gravity model because its form is in analogy with Newton’s law of gravitation (Lenormand, Bassolas, & Ramasco, 2016; Schneider, 1959; Wilson, 1970; Wilson, 1971). The distance decay function $f(d_{ij})$ may vary depending on research topics and dataset, but the negative power and the exponential forms are the most common forms used in previous research (Lenormand et al., 2016). The gravity model and its variant has been widely used in previous research to estimate flows of people, trade and information (Krings, Calabrese, Ratti, & Blondel, 2009; Kwan, 1998; Lewer & Van den Berg, 2008; Liang, Zhao, Dong, & Xu, 2013; Liu et al., 2014). There are many variant forms of the gravity model and other spatial interaction models such as the intervening opportunities model. They provide various options of

formulation, including pre-set parameters or parameter-free ones, for estimating the flows of spatial interaction (Kang, Liu, Guo, & Qin, 2015; Masucci et al., 2013; Wilson, 1971).

Most of the previous spatial interaction models assume that the inner space of the modelling area has spatial isogeneity, which means the distribution of trips only follows one general law related to $f(d_{ij})$. But considering the real-world situation, spatial heterogeneity exists in this distribution due to zoning system, administrative boundaries, transport linkage, jobs and housing balance, as well as other complicated urban contexts. Therefore, some previous research discussed this spatial heterogeneity effect and referring as the Modifiable Areal Unit Problem (MAUP) caused by the setting of the zoning system. Since spatial heterogeneity may cause the inconsistent results of spatial interaction models, previous research has regarded spatial heterogeneity as a “problem”, and has attempted to find an optimal zoning system or technical solution to mitigate its effect (Arbia & Petrarca, 2011; Marceau, 1999; Openshaw, 1977). Besides, a few researchers attempted to adopt hierarchical structures for eliminating the MAUP issues during estimating interactions (Masser & Brown, 1975). Following the conception of MAUP, some researchers proposed that applying a hierarchical structure in the spatial interaction model may eliminate the spatial heterogeneity between each sub-system, improving the overall performance of prediction (Fotheringham, Nakaya, Yano, Openshaw, & Ishikawa, 2001; Nazara, Hewings, & Sonis, 2006; Qian et al., 2020). The hierarchical spatial interaction model distinguished the trips between inner and inter sub-systems for estimating the flows respectively. By applying this framework, spatial heterogeneity on borders between sub-systems can be largely eliminated (Qian et al., 2020).

On the other hand, some scholars also argued that MAUP issues have the ‘bright side’ for detecting the agglomeration effect (Menon, 2012). Considering the cities’ functional space depends on how citizens perceive their activity space and interact with their urban environments (Lynch, 1960), some researchers were aware of the linkage between the border effect and spatial heterogeneity in the spatial interaction model and attempted to quantify the border effect between zones by spatial interaction model (Engel & Rogers, 1994; McCallum, 1995; Yin et al., 2017). However, how to use the variability of hierarchical boundary in the spatial interaction model as the indicator for delineating the urban functional zone or other types of communities has not been further discussed. Thus, it would be an interesting perspective for observing the boundary effect by trip distribution, providing a solid reference for delineating the boundaries of the urban functional zones.

2.3. Using emerging human mobility data to understand urban functional zones

Origin-destination (OD) flow matrix generated from human mobility data (e.g., taxi and bus swipe cards, mobile phone signalling data) can be used as a proxy of the interaction between regions (González, Hidalgo, & Barabási, 2008). Based on that, some studies have been developed to identify urban functional zones and urban spatial structures. Network-based methods are a commonly used approach based on the intensity of human interactions between different spatial units (Jiang & Miao, 2015; Louail et al., 2015; Zhang, Fang, Zhou, & Zhu, 2020; Zhong, Manley, Müller Arisona, Batty, & Schmitt, 2015). Each spatial unit is seen as a node, and human interactions are represented as edges between the two nodes.

Zhong et al. (2014) detected and depicted urban structures in Singapore using a graph-based community detection algorithm, and it is one of the representative studies for urban functional zones detection. The network method may explain the composition via structural shifts of transient sub-centres. For example, it can describe the increasing interaction between certain developing sub-centres (Zhang et al., 2021). Shen and Batty (2019) detected community structures in the London Metropolitan area based on disaggregated flow data, suggesting that the functional structure may vary for people with different occupations.

Zhang et al. (2021) analysed several years of transport smart card data in London and the results of network community detection show that Greater London can be clustered into five communities based on the travel pattern, but London moved towards a more polycentric and compact urban structure. However, the traditional network analysis and most community detection algorithms usually only consider the absolute value of flow volume (edges) for dividing the partitions regardless of the spatial factors such as distance decay or time consumption (Adam, Delvenne, & Thomas, 2018; Hong & Yao, 2019; Jin et al., 2021).

Some researchers have been aware of this shortcoming and tried to apply spatial interaction to improve their method. Jin et al. (2021) identified the activity broad within Shenzhen city and discussed the boundary effect by using a modified spatial interaction model. Yin et al. (2017) proposed a method to delineate urban boundaries for Great Britain based on the physical space inferred from human activities on social media and then verified the results by a gravity model. However, both still applied a network-based algorithm to identify the functional urban regions. Currently, due to the limitation of the data and computation, most of the previous studies on spatial structure from movement flow investigated one city only (Jin et al., 2021; Wu et al., 2021; Zhang et al., 2021; Zhong et al., 2014). This study would explore functional urban spatial structure at a larger scale by delineating the urban functional zones based on human mobility and movement flows. In addition, Because of the importance of distance factors and the absence of a method that can detect urban boundaries by distance-based trip distribution, we believe it is worth establishing a new method for depicting the form of urban boundaries.

3. Data collection

The case study area of this study is Shenzhen-Dongguan-Huizhou (SDH) area. SDH area covers a total of 15,800 km², with a resident population of 26.25 million and a total GDP of RMB 3.7 trillion in 2019 (according to *Guangdong Statistical Yearbook 2020*). This area has been experiencing rapid urban growth and changes in urban spatial structure since the 1980s and became one of the most open and economically vibrant regions in China. SDH persistent attracts national and global focuses, especially after the Guangdong-Hong Kong-Macao Greater Bay Area was proposed in 2015.

This research uses the mobile phone data provided by one of the main mobile phone operators called China Unicom. The data contains Origin sub-district ID, Destination sub-district ID, the volume of travel flow and travel time. The spatial resolution of the original data is collected as 500 m*500 m but is provided as aggregated form into 172 sub-district level zones (“jiedao level” or “街道级” in China). The specific study units are shown in Fig. 1 below. The mobile phone data detected 13,588,846 commuters, including both intra-city and inter-city. The observed period of the data is February 2019. To identify commuting trips, home and workplaces are first determined from one-month sequent locations of mobile phones.

Specifically, the site with the most prolonged stay during the observation period (09:00 pm-08:00 am) in a day is considered as the candidate place of residence. When a candidate residence lasts for more than 15 days in a month, it is deemed to be valid. Similarly, the location with the most prolonged stay between 09:00 am and 05:00 pm is determined to be the workplace. Commuting is defined as a journey from one’s home to the workplace. Individual commuting trips of mobile phone users are aggregated at the street scale, generating links between streets across the study area (SDH region).

The data used for our study is at the sub-district level. In total, there are 8921 pairs of Origin-Destinations (OD) summarised from commuting trips. For each pair of ODs, the data records the original street ID, the destination street ID, the number of commuters, average commuting time and distance. Fig. 1 shows the distribution of inter-city flows within the SDH area.

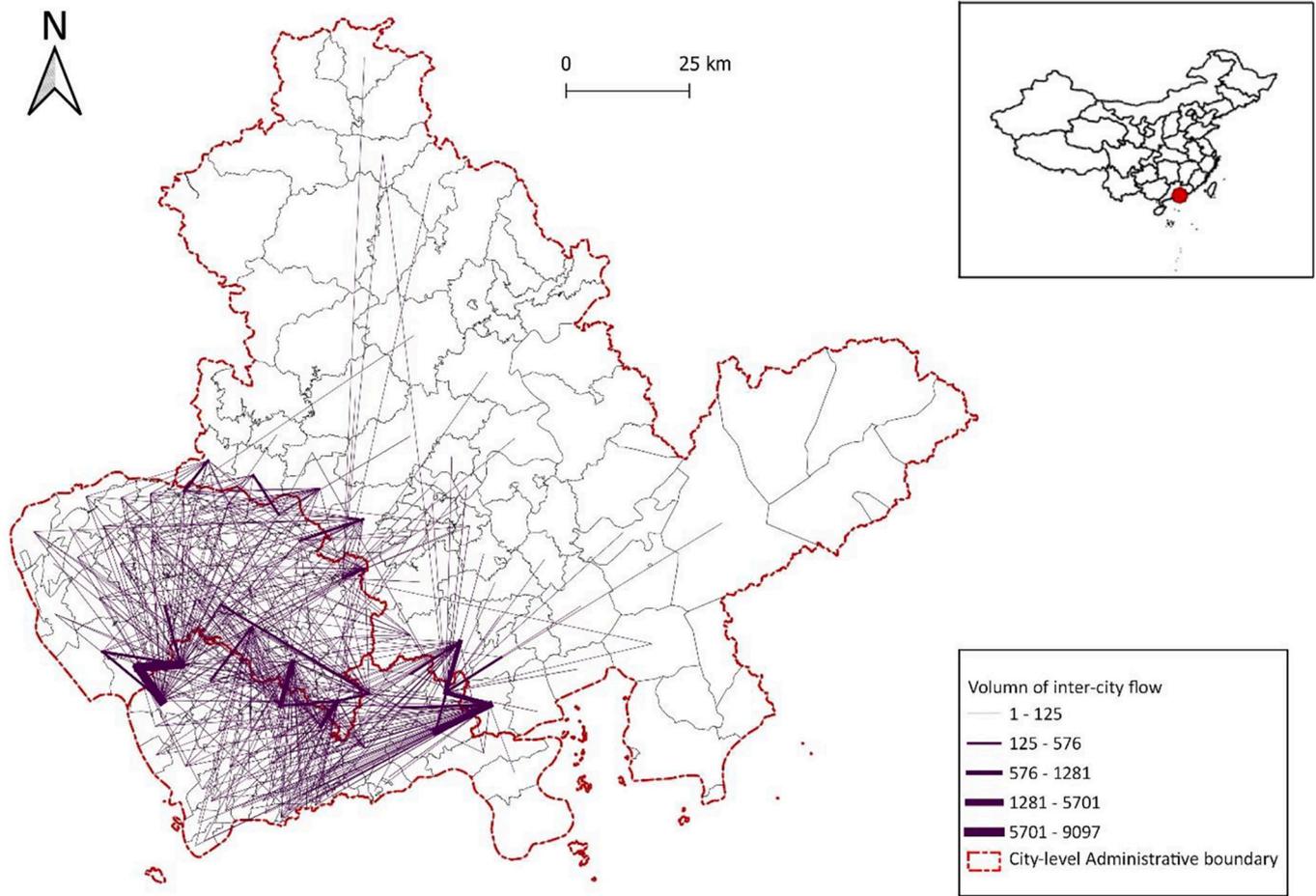


Fig. 1. Distribution of cross-city flows within the SDH area.

4. Methodology

This study adopted disaggregated spatial interaction model for simulating the flow of spatial interaction between zones. Furthermore, a Hierarchical Spatial Interaction Model (HSIM) was applied to reflect the boundary effect between cities. For detecting the urban functional zones, this research proposed a novel regionalisation algorithm that redraws non-overlap boundaries of urban functional zones by searching for the best partition with the best goodness of fitting in HSIM.

4.1. Basic spatial interaction model

In this study, we established a set of spatial interaction models using the singly constrained gravity model, which assumes the distribution of trips roughly follows the format of the negative-power function for predicting the flow between zones. The core spatial interaction model can be represented as the following Eq. (2):

$$T_{ij} = O_i^{obs} \frac{D_j^{obs} c_{ij}^{-\beta}}{\sum_k D_k^{obs} c_{ik}^{-\beta}} \quad (2)$$

where O_i^{obs} is observed origins totals from zone i and D_j^{obs} refers observed destinations totals to zone j , c_{ij} is the main travel time between origins and destinations, β is a parameter related to the travel cost.

The basic framework of this model is a form of classic gravity models (Wilson, 1971). The calibrating processing is parameter-free since the model picks the distance decay parameters β by continually executing

the iterations of standard non-linear optimised (Batty, 1976; Batty & Milton, 2021) until the difference between the predicted mean trip cost C and the observed mean trip cost C^{obs} is less than the pre-set threshold ϵ (the default is 5% for balancing the calculation time and accuracy) (Eqs. (3)–(4)).

$$|C^{pre} - C^{obs}| < \epsilon \quad (3)$$

$$C = \frac{\sum_i \sum_j T_{ij} c_{ij}}{\sum_i \sum_j T_{ij}} \quad (4)$$

4.2. Hierarchical spatial interaction model (HSIM)

Although spatial heterogeneity exists within a mega-city region, most traditional (or “global”) spatial interaction models assume the inner space of the modelling area is spatial homogeneity, which means all trips follow one general law. Thus, we further adopted a two-level hierarchical spatial interaction model for estimating the travel flow between zones (Fig. 2). By applying this framework, it can eliminate spatial heterogeneity because of the boundary between sub-systems. It divides the global spatial interaction model into some intra-city interaction models. For each sub-model, the form is the same as the basic spatial interaction model introduced in Section 4.1. Since the distance decay parameter β is an auto-fitted value that is different in each sub-system, each sub-system describes a distinguished travel pattern. The model can be written as Eqs. (5)–(7).

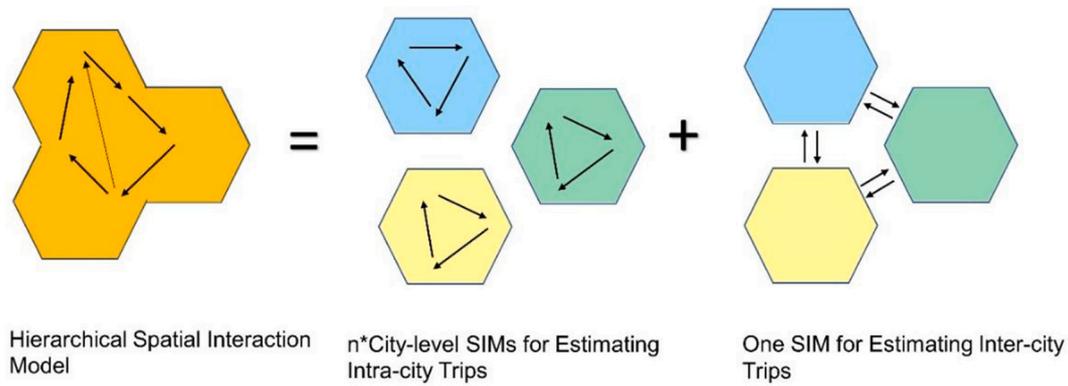


Fig. 2. The Hierarchical Spatial Interaction Model: The total predicting trips equals city-level intra-city models plus one inter-city model, Eq. (5) = Eq. (6) + Eq. (7).

$$T_{total} = \sum T_{intra(n)} + T_{inter} \tag{5}$$

$$T_{intra(n)} = \sum_{i \in n} \sum_{j \in n} O_i^{obs} \frac{D_j^{obs} c_{ij}^{-\beta_n}}{\sum_k D_k^{obs} c_{ik}^{-\beta_n}} \tag{6}$$

$$T_{inter} = \sum_{i \in n} \sum_{j \in m} O_i^{obs} \frac{D_j^{obs} c_{ij}^{-\beta}}{\sum_k D_k^{obs} c_{ik}^{-\beta}} \quad (n \neq m) \tag{7}$$

As a key condition, the difference in the boundary of cities can affect the performance of this model because of the boundary effect and spatial heterogeneity. If the cities' boundaries in this model coincide with the boundary effect, the overall performance of this model will improve since the processing of splitting has eliminated the spatial heterogeneity between sub-models. Therefore, we believe the goodness of fitting of this model can be an indicator for assessing the reasonableness when drawing the functional boundary of cities. The detailed proof of this hypothesis will be described in Section 5.1. Since all trip flows can be allocated to one of the sub-models, the sum of the total trip and the constrained factor (volume origin trips in our model) would keep constant by applying the HSIM framework without any loss of information.

4.3. Regionalisation algorithm for delineating urban functional zones

Based on the Hierarchical Spatial Interaction Model, we propose a novel regionalisation algorithm for delineating urban functional zones by searching for the best partition with the best goodness of fitting in the HSIM. After determining a predefined number of regions (in this case study is three because there are three cities-level governments within this area), our iteration-based algorithm will run several times until the best partition which has the highest R². This algorithm design takes the conception of the tabu search algorithm and referenced to previous works by Openshaw and Rao (1995). As an evolutionary method of local search, it inherits the basic concepts of greedy algorithms that continually choose the optimal choice at each step to find the optimal solution to reduce complexity and time consumption. At the same time, it can avoid being trapped in local optima by adopting the "tabu list", where each reassigned zone will be recorded for not being considered in the following R times iterations.

When the algorithm starts, it will initialise an empty tabu list and then each iteration repeats the steps below:

Step one: The algorithm would first attempt to reassign candidate zones (not in the tabu list) to another city respectively and calculate the goodness of fitting in HSIM, in order to identify the zone improve the result most. If any improvement identified, execute it, then jump to *Step four*.

Step two: When the algorithm finds that reassigning zones (not in the tabu list) cannot improve the result anymore, the algorithm will re-assess if reassigning zones currently in the tabu list can further improve the result, called the "aspiration move". If any aspiration move can be found, execute it, then jump to *Step four*.

Step three: If no further improvement or aspiration move can be made and the stopping criterion has not been satisfied yet, the algorithm would reassign the zones with the best result even if this assigning would worsen the result.

Step four: Record the reassigned zones in the tabu list. If the current results are better than the 'the best result ever', replace the best result with the current one and record the current boundaries. Update the travel flow according to the settings, then back to step one for starting a new iteration.

Before executing *step three*, here is a stopping criterion to avoid endless iterations. The iteration will terminate if the best partition is not updated after N (N = 20 in this study) iterations then output the "the best result ever" boundary as the result.

We have conducted a sensitivity analysis for these two parameters, which determines R = 11, which can maximise the goodness of fitting. The results keep the same when R is within the range 1– 10. Then the result slightly improved and then kept the same when R equals 11 or continually increased until reaching the length of the candidate zone list. In addition, the sensitivity analysis finds that the algorithm is not sensitive to the value of N. Whether the N increase to 50 or 100, the result would not change.

Although theoretically, the candidate zones could be any zone within the region, we could customise a set of prioritised zones to improve the algorithm's efficiency. For instance, in our experience, we set the scope of search space to all zones whose intercity commuters are more than 1%, which matches the average ratio of inter-city commuters in the case study area. A flow chart of this algorithm for reassigning the boundaries can be found in Fig. 3.

To provide appropriate decision support, this algorithm should not only assess the current functional zones but also predict the long-term situation. Thus, we designed two different settings that have a minor difference when we execute the algorithm. The first setting is based on the situation that current cities' core functional regions can only spill over to zones close to the administrative boundary due to the local authority's current land-use planning and management scope. Therefore, the proportion of inter-city trips in each zone would not change further by updating functional zones during each step of iterations. In other words, the inter-city flows for each zone are static according to the administrative boundary.

The second setting is the inter-city flows for each zone are dynamically updated according to the current boundary in iteration processing. That is, inter-city flows may be re-classified as an intra-city flow after

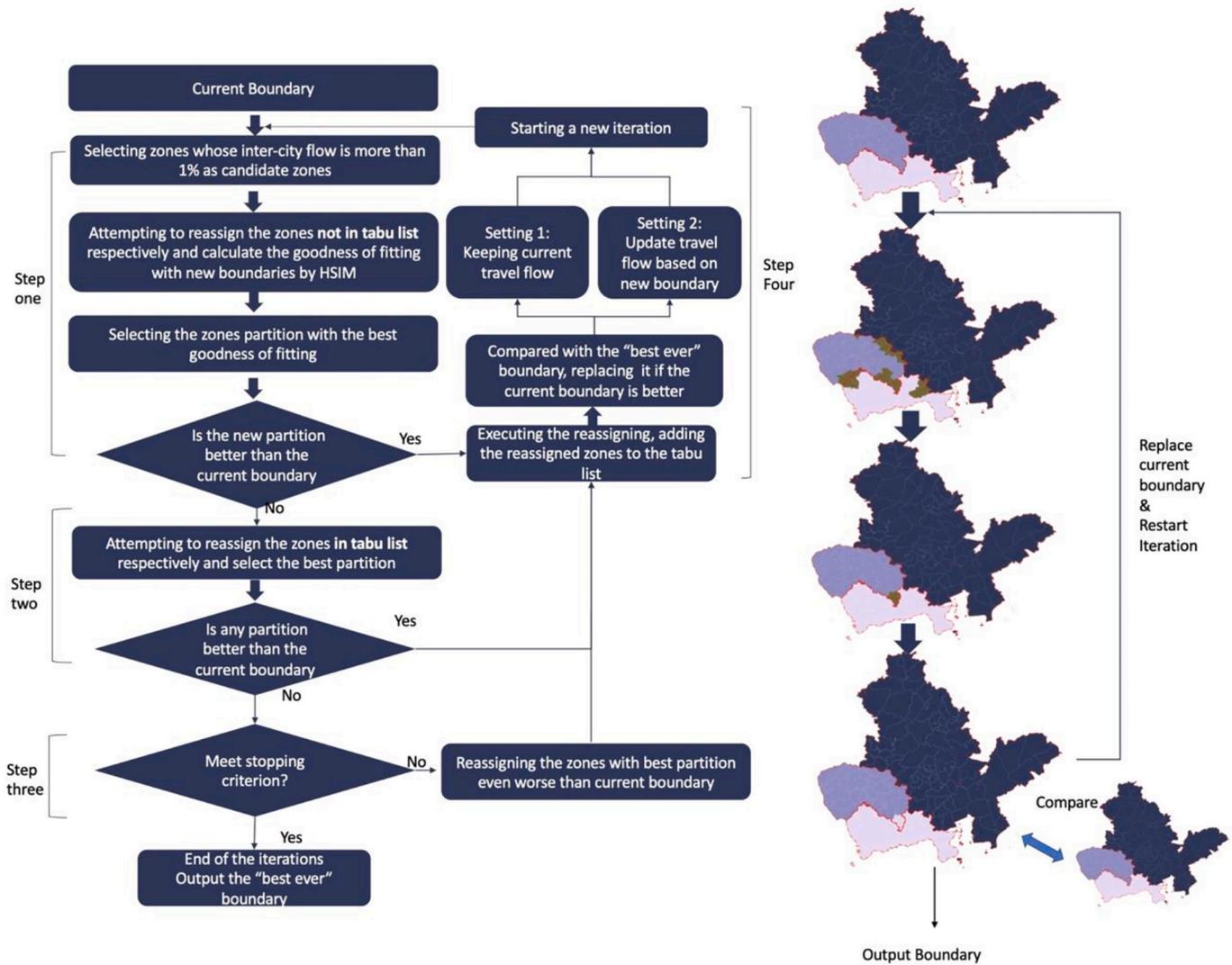


Fig. 3. Flowchart of the redrawing boundaries tabu search algorithm.

iterations of a boundary. This setting could be used for predicting long-term scenarios in which the cities' core functional zones can spill over freely without restriction by the current administrative boundary, forecasting the potential functional boundary in the long term.

5. Results

5.1. Goodness of fitting for hierarchical spatial interaction model (HSIM)

For this case study, we split the global spatial interaction model of the whole SDH area into four sub-SIM models: three intra-city trips models for Shenzhen, Dongguan, Huizhou respectively based on its original administrative boundaries, plus one model for only predicting the inter-city trip.

To verify the hypothesis, we raised before that the goodness of fitting can be an indicator for reflecting the boundary effect of cities, and we introduced a controlled group since the spatial heterogeneity and the boundary effect are often more significant around the boundaries between cities. This controlled group is set as it still has the same trips and zoning system (172 sub-district level zones) but with random urban boundaries. The boundaries applied in models are shown in Fig. 4.

There are some flow trips produced by the spatial interaction model and hierarchical spatial interaction model (HSIM) introduced in Sections 4.1 and 4.2 according to the different boundaries. The goodness of fitting can be an indicator for assessing the reasonableness when drawing the functional boundary of cities. The estimated distance decay parameters in sub-models have been attached in the Appendix 1, Table 3.

For assessing the goodness of fitting, we calculated the Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and R-square (R^2) compared with the observed flow, the results are represented in Table 1 below.

As the statistical measures are shown in Table 1, compared with the traditional GSIM model, the R^2 for the HSIM model with random boundaries and administrative boundaries sharply rise to 0.6564 and 0.8165 from 0.45310. Meanwhile, MSE, MAE, and RMSE decreased significantly, which shows that the HSIM model largely shortened the difference between the estimated and actual values. Thus, all statistical measure indicators prove that the hierarchical spatial interaction can significantly improve the goodness of fitting from the traditional Global methods in regional-scale scenarios. This result indicates the broader effect can be partly represented by the random boundaries.

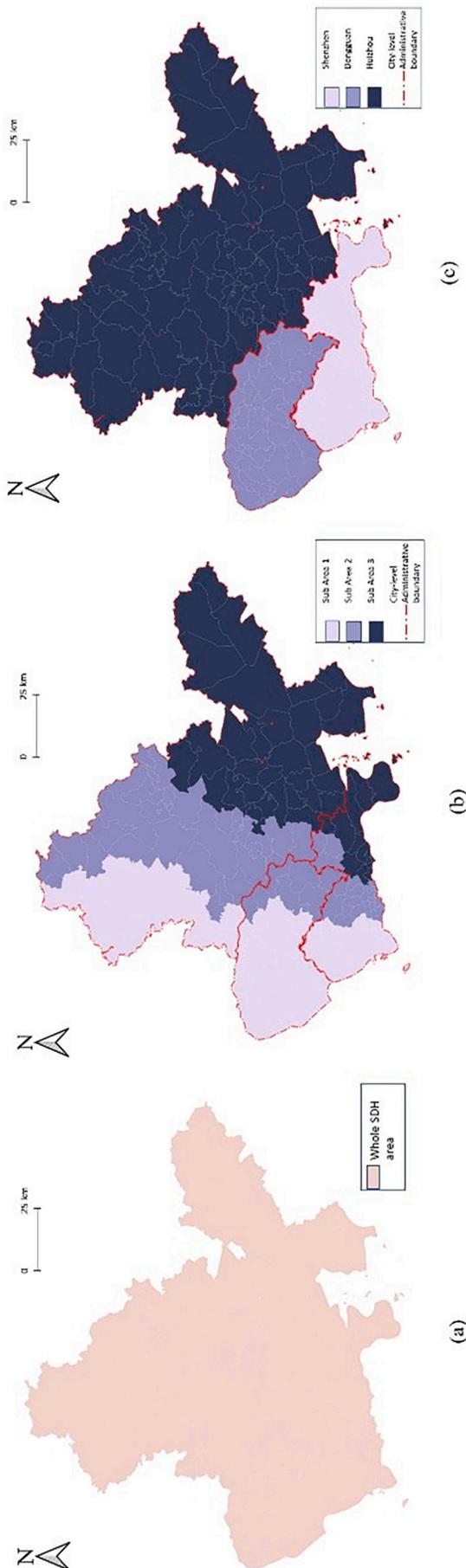


Fig. 4. (a) the GSIM model with one whole modelling area (left); (b) the HSIM model with random boundaries (middle); (c) the HSIM model with administrative boundaries (right).

Table 1

Goodness of fitting for Global-SIM and HSIMs (Note: AB means administrative boundaries; CB means current boundaries; and LB means long-term boundaries.)

	Global-SIM	HSIM with RB	HSIM with AB
MSE	14,945,798	9,082,981	5,013,151
MAE	335.73	278.188	239.90
RMSE	3865.80	3013.79	2239.01
R ²	0.4531	0.6564	0.8165

By comparing the result of HSIM with arbitrary boundaries and the HSIM with administrative boundaries, all statistical measure indicators reveal that applying appropriate boundary that reflects the spatial heterogeneity in HSIM would significantly improve the model's performance. This finding suggests that the travel behaviours of people who belong to the same functional city may yield better performance in fitting the specific distribution of trips. In other words, in the case of the spatial resolution and number of sub-models keeping constant, the goodness of fitting by HSIM can be an indicator for assessing the reasonability of functional boundaries of cities. This finding provides a solid theoretical reference for the algorithm that will be introduced in the next section.

5.2. Result of urban functional zones in SDH area

Table 2 reports the statistical result for the models with different settings. Although the performance of the basic scenario with the administrative boundary is already good enough, the models of both settings with new boundaries still slightly outperform the basic model. Comparing the minor improvement of statistical measurements, the new boundary itself is more meaningful for assessing the urban functional integration. The estimated distance decay parameters in sub-models have been attached in Appendix 1, Table 4.

5.2.1. Setting 1-current functional boundary

Fig. 5 shows the result of setting 1 (statistic inter-city flow), indicating the current functional boundary within the SDH area. This result suggests that the current administrative boundary explains the boundary effect of trip distribution well. The statistical results in Table 2 support it. Compared with the current administrative boundary in the base scenario, the R² and other statistical indicators improved very slightly as only a few zones changed their origin belonging. For example, the functional core of Dongguan city is in the west of its administrative boundary because of its good transport connection with Shenzhen and Guangzhou. Therefore, the only zone in Dongguan that should be re-assigned to Shenzhen is Fenggang, as it has been known as the 'sleep city' for workers in Shenzhen, which is a typical example of cross-city functional integration. Moreover, a few zones near the Dongguan-Huizhou boundary will be re-allocated to Huizhou from Dongguan because these zones are away from the city centre and lack commuting connections with the city centre. It might be the main reason why trips in these zones would better fit the trip distribution in Huizhou rather than Dongguan. Similarly, a few zones in east Shenzhen have been re-assigned to Huizhou.

Table 2

Goodness of fitting for HSIM in different scenarios. (Note: AB means administrative boundaries; CB means current boundaries; and LB means long-term boundaries.)

	Base Scenario (AB)	Setting 1 (CB)	Setting 2 (LB)
MSE	5,013,151	4,842,008	4,377,221
MAE	239.90	242.71	231.41
RMSE	2239.01	2200.456	2092.18
R ²	0.8165	0.8228	0.8398

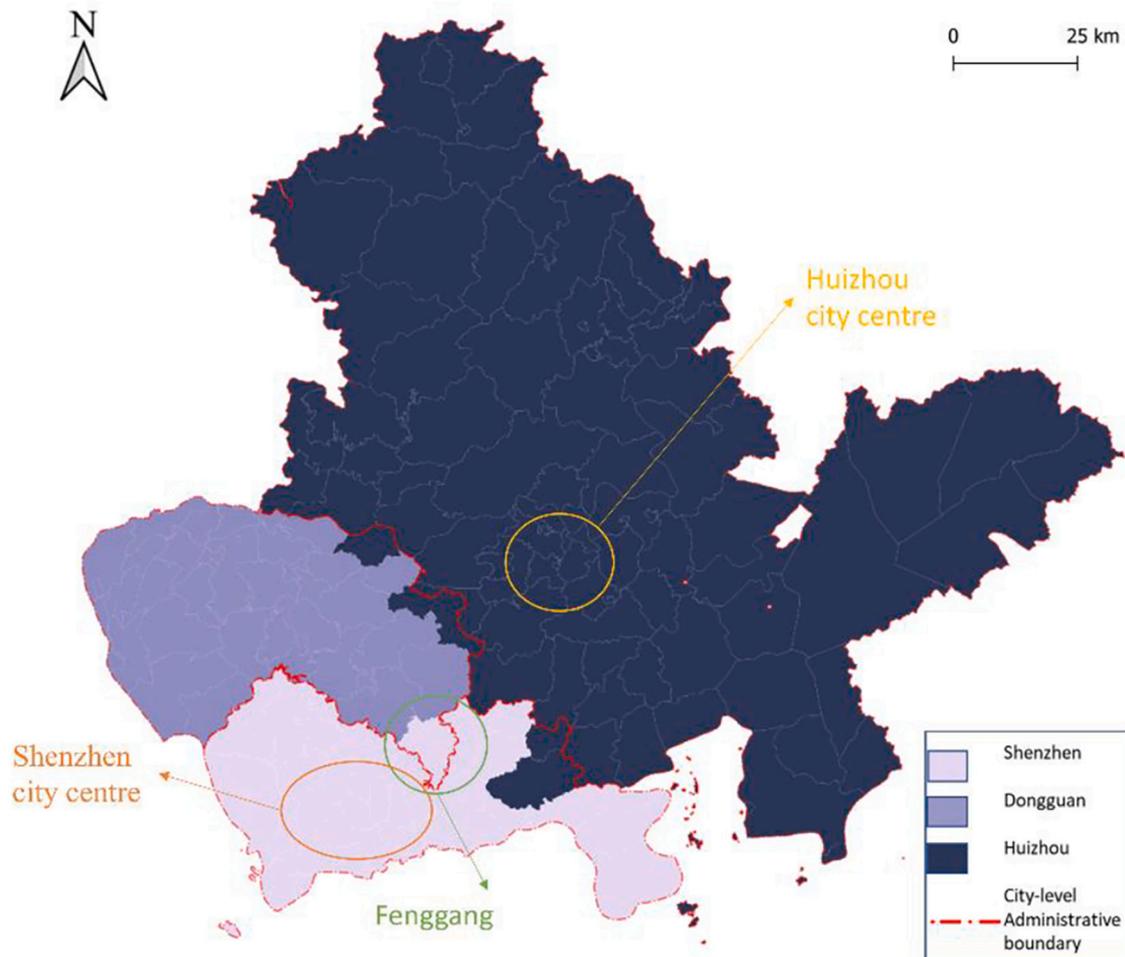


Fig. 5. Setting 1- Current functional boundary within SDH area.

5.2.2. Setting 2- predicted functional boundary within SDH area in long-term

As for the dynamic inter-city flow setting, the result (shown in Fig. 6) predicts that the functional areas will have more reassigning between Huizhou and Dongguan. The re-assigned zones in Dongguan are mainly from the 'East industrial park', Songshan Lake, and Dalang. Historically, East industrial park areas are the cluster of manufactory industries but lack commuting connection with the city centre. The algorithm also finds the Songshan Lake area in the middle of Dongguan is reassigned. This region has been assigned because of its strong linkage with the 'East industrial park' area. Local governments have recently emphasised such connections in their planning report. Besides, the Dalang area may also be reassigned for long-term prediction. Unlike the Fenggang area, this area lacks road linkage directly with Shenzhen although it is physically close to Shenzhen. Thus, this area has been re-assigned to Huizhou following the reassigning of the Songshan Lake area, which is one of the main workplaces for residents in Dalang. These results show that there will be more potential interaction and functional integration opportunities between zones between Dongguan and Huizhou because of the chain reaction in long-term prediction.

5.3. Policy implications for city integration in Shenzhen-Dongguan-Huizhou area

These empirical-based results can help the governments and planners to understand the spatial structure in mega city-region and support

their urban integration policy. Previous studies have always focused more on the north-western part of Shenzhen and the south-western part of Dongguan since it has the most volume of the cross-boundary trip statistically. Our study argues that in the case of balanced bidirectional flows present, the functional boundary effect between the cities would not change obviously. Because trips in this area fit their original intra-city trip distributions, the high inter-city flow might be a natural consequence of the high population density and spatially relatively close to their original urban centres. In contrast, this study reveals that Fenggang and Shenzhen have a very high degree of functional integration, indicating that the urban function (e.g., housing or employment) are shared within these areas. When considering such an integration between the two regions, the policymakers should pay more attention to amenities and public services for inter-city commuters.

For the long-term prediction, zones in mid-Dongguan should be given more attention. These areas are very 'sensitive' to any change of trips since fits in these areas do not fit the intra-city trip distribution of their original cities and are far away from city centres. Thus, our algorithm predicts that a severe mismatch between functional zones and administrative boundaries could occur in these areas, even with tiny inter-city interactions. This result proves that transport linkages are vital for reshaping the urban functional zones in the long term because of the chain reaction of the previous reassigned zones. An example is the Dalang area. Though it is physically close to Shenzhen, it has been re-assigned to Huizhou because the road linkage with Songshan Lake is better than those with Shenzhen.

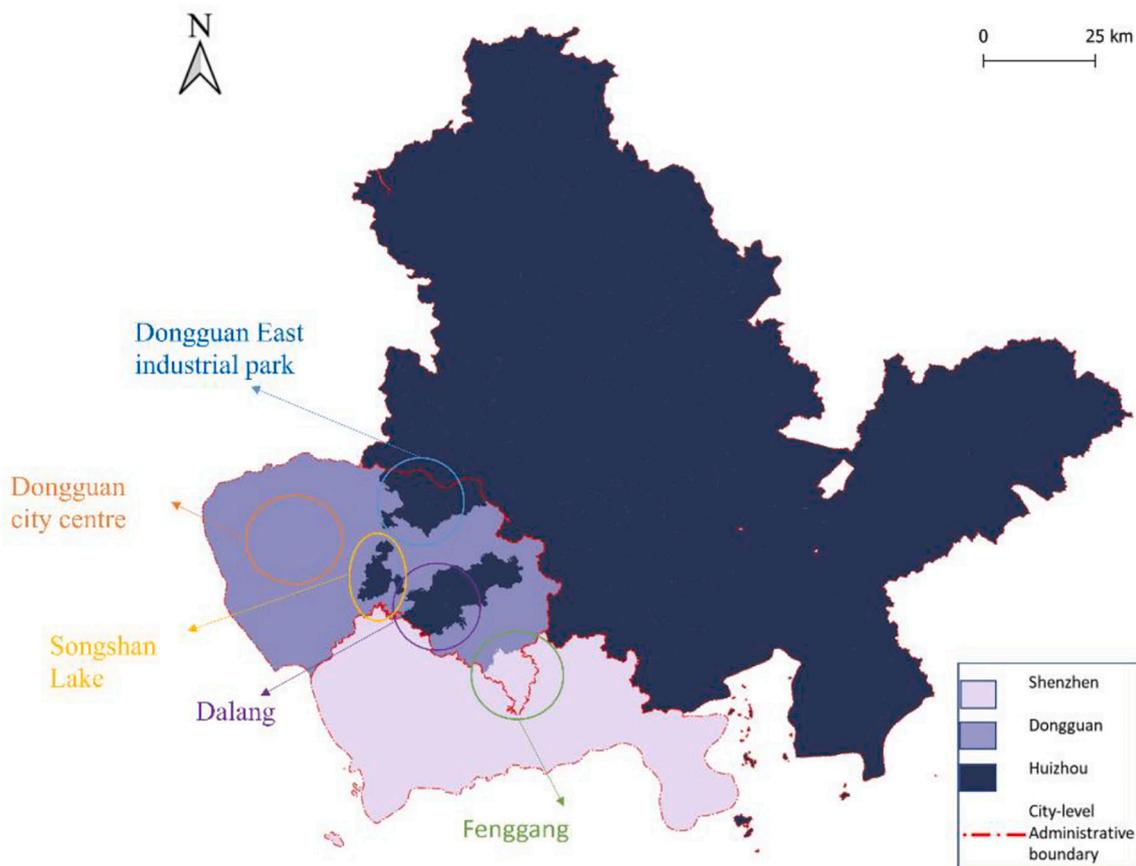


Fig. 6. Setting 2- Predicted functional boundary within SDH area in long-term.

Overall, these empirical results imply that there will be more potential interaction and functional integration opportunities between zones between Dongguan and Huizhou in the future. Besides, policymakers should consider improving transport connectivity between the reassigned areas and Dongguan city centres to eliminate the boundary effect of city centres in trip distribution. Such measures would also avoid severe mismatches between functional zones and administrative regions, which may cause extra difficulty for management.

6. Conclusion and discussion

This research has contributed to the existing literature in several ways. First, this research confirms that the hierarchical spatial interaction model (HSIM) results can assess if the boundary of subsystems appropriately represents the inter-city boundary effect in trip distribution. Furthermore, this study proposes a novel method to delineate UFZs by searching for the best partitions in HSIM. By adopting the proposed model to a specific mega-city region, China, Shenzhen-Dongguan-Huizhou (SDH) area, this research confirmed the model's effectiveness in delineating UFZs based on spatial interaction from the perspective of human activity behaviour.

6.1. Spatial interaction methods vs network-based methods

The network-based community detection method is the mainstream method employed for detecting the boundary of communities and functional spatial structure in the cities-level in previous studies. However, there is some limitation as well. The traditional network analysis and most community detection algorithms usually only consider the absolute value of flow volume for dividing the partitions but overlook

the spatial factors like travel distance/cost (Liu et al., 2014; Yin et al., 2017).

Typical community detection algorithms (e.g., the Louvain algorithm) are always trying to search a partition for maximising the ratio of intra-city flows to overall flows. However, because the percentage of inter-city trips is usually tiny (3% or less compared to intra trips) among all trips, thus when ignoring the spatial factors, the traditional community detection prefers to split space within the origin of administrative boundaries rather than break it. Similar to the phenomenon observed by previous research (Liu et al., 2014), the detected communities are precisely the same as the original administrative boundary when we use the Louvain algorithm and adjust the minimum resolution point to let the number of communities equal to the number of cities. According to the definition by OECD, the city or town whose 10% of the population exhibits cross-boundary commuting behaviour can be regarded as the satellite city of the mega-city. Thus, the traditional network analysis is not sensitive enough for cross-boundary commuting trips, which may fail to support planners and policymakers appropriately when discussing the cross-boundary integration of the functional region.

In contrast with the network analysis-based method, our proposed spatial interaction-based algorithm will consider the distance decay effect when detecting the boundary effect reflecting spatial heterogeneity. Because zones close to the cities' boundary are usually spatially far from the city centre, our algorithm would be more sensitive to cross-boundary trips even with a relatively small volume.

Besides, another limitation of network-based methods is the difficulty of predicting the future situation. Almost all research applied community detection methods must base on the existing data of travel flow. If the data is unavailable, the estimation of flows would still rely on

spatial interaction models (Wu et al., 2021). It will cause more deviation when switching between the multi-methods. Because of the strength for estimating the travel flow, the spatial interaction-based methods would have a special advantage for the prediction and simulation of future urban regions dynamically.

6.2. Limitation and further research

There are some limitations, and several directions can be further explored. First, different forms of spatial interaction models can be adopted for predicting the trip distribution. This paper only employed the most widely used gravity model with a negative power functional form. Therefore, more conditions of the spatial interaction model, including the intervening opportunity and radiation models, can be discussed and employed in future work.

The second point is the “scaling issue”. Additional experiments have been conducted to validate the model with more communities and different boundaries. One of the experiments attempted to extend the case study area to a border area, the Great Bay Area (GBA) in Pearl River Delta China, for nine cities with the same zoning system (sub-district level units). It confirms algorithm still works appropriately for this extended area, but the algorithm would yield different results for local results in the SDH areas (the results of the GBA area have been attached as Appendix 2). The reason could be that the added areas and the additional trips will affect the existing results when applying the inter-city trip estimation models. The difference would be extended in long-term prediction due to the chain reaction. Thus, choosing the spatial extent needs to be associated with the specific research question and focus study area.

Lastly, our regionalised algorithm considers connections and flows between any pairs of units, not just neighbours. The spatial factors have mainly been reflected by travel time in this study. On the one hand, this

is one of the advantages of emphasising mobility flows compared with other regionalisation algorithms. However, on the other hand, spatial adjacency is crucial in some cases (land-use planning, air pollution, etc.). Thus, spatial constraints on physical distance may need to be added to this algorithm to handle more situations.

CRediT authorship contribution statement

Bowen Zhang: Methodology, Investigation, Writing – original draft. **Chen Zhong:** Conceptualization, Supervision, Writing – review & editing. **Qili Gao:** Validation, Writing – review & editing. **Zahratu Shabrina:** Writing – review & editing. **Wei Tu:** Data curation.

Declaration of Competing Interest

The authors have no conflict of interest.

Data availability

The authors do not have permission to share data.

Acknowledgement

This research was supported by the Postdoctoral Science Foundation of China (Grant No. 2021M692163), Natural Science Foundation of China (Grant No. 42071360), the Tencent-Rhino bird Fund (JR-WXG-2021131), the China Scholarship Council (CSC) from the Ministry of Education of P.R. China (CSC201908060169), the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (Grant agreement No. 949670), and from ESRC under JPI Urban Europe/NSFC (Grant No. ES/T000287/1).

Appendix A. The distance decay parameters in sub-models

Table 3

The distance decay parameters in sub-models in Section 5.1. (Note: RB means random boundaries, and AB means administrative boundaries)

	Global-SIM	HSIM with RB	HSIM with AB
Area 1	2.5425	3.2006	2.4759
Area 2	N/A	2.3841	3.4837
Area 3	N/A	2.6334	2.0937
Inter-city	N/A	1.7279	1.6927

Table 4

The distance decay parameters in sub-models in Section 5.2. (Note: AB means administrative boundaries; CB means current boundaries; and LB means long-term boundaries)

	Base scenario (AB)	Setting 1 (CB)	Setting 2 (LB)
Shenzhen	2.4759	2.4594	2.4594
Dongguan	3.4837	3.5455	3.7646
Huizhou	2.0937	2.0552	2.3514
Inter City	1.6927	1.7429	1.7746

Appendix B. The delineated functional boundary for the GBA area

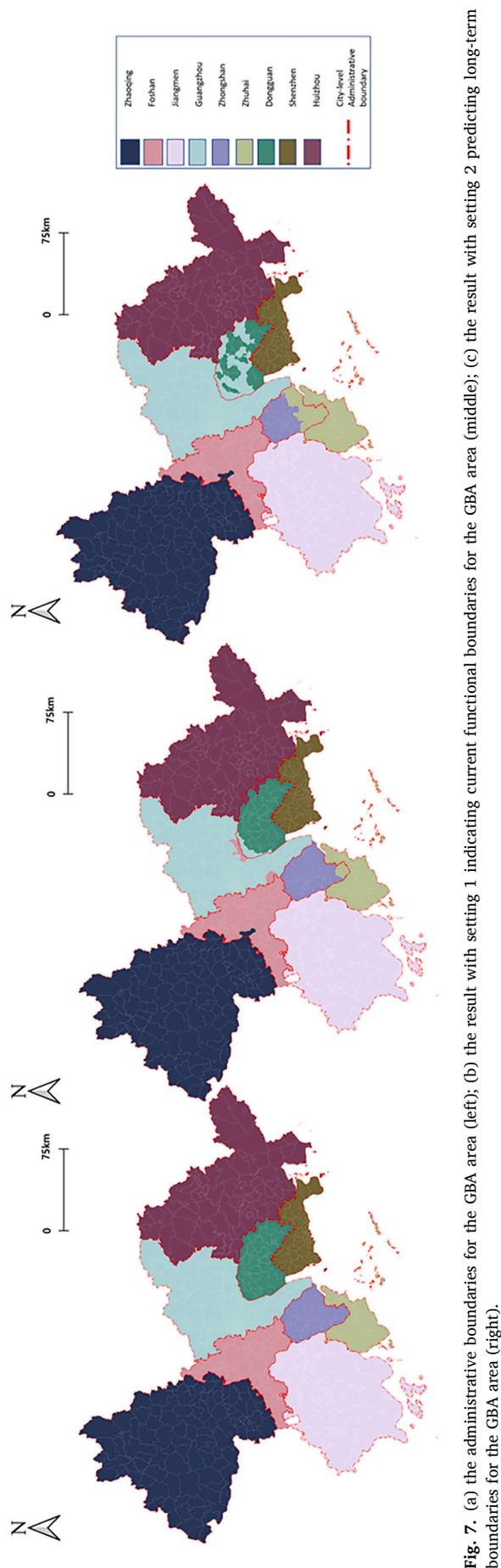


Fig. 7. (a) the administrative boundaries for the GBA area (left); (b) the result with setting 1 indicating current functional boundaries for the GBA area (middle); (c) the result with setting 2 predicting long-term boundaries for the GBA area (right).

References

Adam, A., Delvenne, J.-C., & Thomas, I. (2018). Detecting communities with the multi-scale Louvain method: Robustness test on the metropolitan area of Brussels. *Journal of Geographical Systems*, 20(4), 363–386.

Alvioli, M. (2020). Administrative boundaries and urban areas in Italy: A perspective from scaling laws. *Landscape and Urban Planning*, 204, Article 103906. <https://doi.org/10.1016/j.landurbplan.2020.103906>

Anderson, J. E. (2011). The gravity model. *Annual Review of Economics*, 3(1), 133–160. <https://doi.org/10.1146/annurev-economics-111809-125114>

Arbia, G., & Petrarca, F. (2011). Effects of MAUP on spatial econometric models. *Letters in Spatial and Resource Sciences*, 4(3), 173.

Arcaute, E., Hatna, E., Ferguson, P., Youn, H., Johansson, A., & Batty, M. (2015). Constructing cities, deconstructing scaling laws. *Journal of the Royal Society Interface*, 12(102), 20140745.

Batty, M. (1976). *Urban Modelling. Algorithms, Calibrations, Predictions.*

Batty, M., & Milton, R. (2021). A new framework for very large-scale urban modelling. *Urban Studies*, 58(15), 3071–3094. <https://doi.org/10.1177/0042098020982252>

Brenner, N. (2002). Decoding the newest “metropolitan regionalism” in the USA: A critical overview. *Cities*, 19(1), 3–21. [https://doi.org/10.1016/S0264-2751\(01\)00042-7](https://doi.org/10.1016/S0264-2751(01)00042-7)

Brown, W. M., Dar-Brodeur, A., & Tweedle, J. (2020). Firm networks, borders, and regional economic integration. *Journal of Regional Science*, 60(2), 374–395.

Burger, M., & Meijers, E. (2012). Form follows function? Linking morphological and functional Polycentricity. *Urban Studies*, 49(5), 1127–1149. <https://doi.org/10.1177/0042098011407095>

Cottineau, C., Finance, O., Hatna, E., Arcaute, E., & Batty, M. (2019). Defining urban clusters to detect agglomeration economies. *Environment and Planning B: Urban Analytics and City Science*, 46(9), 1611–1626.

Dadashpoor, H., & Yousefi, Z. (2018). Centralization or decentralization? A review on the effects of information and communication technology on urban spatial structure. *Cities*, 78, 194–205. <https://doi.org/10.1016/j.cities.2018.02.013>

Duque, J. C., Church, R. L., & Middleton, R. S. (2011). The p-regions problem [<https://doi.org/10.1111/j.1538-4632.2010.00810.x>]. *Geographical Analysis*, 43(1), 104–126. <https://doi.org/10.1111/j.1538-4632.2010.00810.x>

Engel, C., & Rogers, J. (1994). How wide is the border?. In *National Bureau of Economic Research Cambridge, Mass., USA.*

Engelfriet, L., & Koomen, E. (2018). The impact of urban form on commuting in large Chinese cities. *Transportation*, 45(5), 1269–1295. <https://doi.org/10.1007/s11116-017-9762-6>

Fotheringham, A. S., Nakaya, T., Yano, K., Openshaw, S., & Ishikawa, Y. (2001). Hierarchical destination choice and spatial interaction modelling: A simulation experiment. *Environment and Planning A: Economy and Space*, 33(5), 901–920. <https://doi.org/10.1068/a33136>

Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, 21(3), 446–467.

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779–782. <https://doi.org/10.1038/nature06958>

Green, N. (2007). Functional Polycentricity: A formal definition in terms of social network analysis. *Urban Studies*, 44(11), 2077–2103. <https://doi.org/10.1080/00420980701518941>

Guo, D. (2008). Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). *International Journal of Geographical Information Science*, 22(7), 801–823.

Hall, P. G., & Pain, K. (2006). *The polycentric metropolis: Learning from mega-city regions in Europe.* Routledge.

Helbich, M., Brunauer, W., Hagenauer, J., & Leitner, M. (2013). Data-driven regionalization of housing markets. *Annals of the Association of American Geographers*, 103(4), 871–889.

Henderson, M., Yeh, E. T., Gong, P., Elvidge, C., & Baugh, K. (2003). Validation of urban boundaries derived from global night-time satellite imagery. *International Journal of Remote Sensing*, 24(3), 595–609. <https://doi.org/10.1080/01431160304982>

Hong, Y., & Yao, Y. (2019). Hierarchical community detection and functional area identification with OSM roads and complex graph theory. *International Journal of Geographical Information Science*, 33(8), 1569–1587.

Huang, D., Liu, Z., & Zhao, X. (2015). Monocentric or polycentric? The urban spatial structure of employment in Beijing. *Sustainability*, 7(9), 11632–11656.

Jiang, B., & Miao, Y. (2015). The evolution of natural cities from the perspective of location-based social media. *The Professional Geographer*, 67(2), 295–306. <https://doi.org/10.1080/00330124.2014.968886>

Jin, M., Gong, L., Cao, Y., Zhang, P., Gong, Y., & Liu, Y. (2021). Identifying borders of activity spaces and quantifying border effects on intra-urban travel through spatial interaction network. *Computers, Environment and Urban Systems*, 87, Article 101625. <https://doi.org/10.1016/j.compenvurbsys.2021.101625>

Kang, C., Liu, Y., Guo, D., & Qin, K. (2015). A generalized radiation model for human mobility: Spatial scale, searching direction and trip constraint. *PLoS One*, 10(11), Article e0143500. <https://doi.org/10.1371/journal.pone.0143500>

Krings, G., Calabrese, F., Ratti, C., & Blondel, V. D. (2009). Urban gravity: A model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07), L07003.

Kwan, M. P. (1998). Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis*, 30(3), 191–216.

- Lenormand, M., Bassolas, A., & Ramasco, J. J. (2016). Systematic comparison of trip distribution laws and models. *Journal of Transport Geography*, 51, 158–169.
- Lewer, J. J., & Van den Berg, H. (2008). A gravity model of immigration. *Economics Letters*, 99(1), 164–167. <https://doi.org/10.1016/j.econlet.2007.06.019>
- Li, Y., Wu, F., & Hay, I. (2015). City-region integration policies and their incongruous outcomes: The case of Shantou-Chaozhou-Jieyang city-region in East Guangdong Province, China. *Habitat International*, 46, 214–222. <https://doi.org/10.1016/j.habitatint.2014.12.006>
- Liang, X., Zhao, J., Dong, L., & Xu, K. (2013). Unraveling the origin of exponential law in intra-urban human mobility. *Scientific Reports*, 3(1), 1–7.
- Liu, Y., Sui, Z., Kang, C., & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS One*, 9(1), Article e86026. <https://doi.org/10.1371/journal.pone.0086026>
- Long, Y. (2016). Redefining Chinese city system with emerging new data. *Applied Geography*, 75, 36–48. <https://doi.org/10.1016/j.apgeog.2016.08.002>
- Long, Y., Zhai, W., Shen, Y., & Ye, X. (2018). Understanding uneven urban expansion with natural cities using open data. *Landscape and Urban Planning*, 177, 281–293. <https://doi.org/10.1016/j.landurbplan.2017.05.008>
- Louail, T., Lenormand, M., Picornell, M., García Cantú, O., Herranz, R., Frias-Martinez, E., ... Barthelemy, M. (2015). Uncovering the spatial structure of mobility networks. *Nature Communications*, 6(1), 6007. <https://doi.org/10.1038/ncomms7007>
- Lu, D., Li, G., Kuang, W., & Moran, E. (2014). Methods to extract impervious surface areas from satellite images. *International Journal of Digital Earth*, 7(2), 93–112. <https://doi.org/10.1080/17538947.2013.866173>
- Lynch, K. (1960). *The image of the city* (Vol. 11). MIT Press.
- Marceau, D. J. (1999). The scale issue in the social and natural sciences. *Canadian Journal of Remote Sensing*, 25(4), 347–356.
- Masser, I., & Brown, P. J. (1975). Hierarchical aggregation procedures for interaction data. *Environment and Planning A*, 7(5), 509–523.
- Masucci, A. P., Serras, J., Johansson, A., & Batty, M. (2013). Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E*, 88(2). <https://doi.org/10.1103/physreve.88.022812>
- McCallum, J. (1995). National Borders Matter: Canada-U.S. regional trade patterns. *The American Economic Review*, 85(3), 615–623. <http://www.jstor.org/stable/2118191>.
- Menon, C. (2012). The bright side of MAUP: Defining new measures of industrial agglomeration. *Papers in Regional Science*, 91(1), 3–28.
- Nazara, S., Hewings, G. J., & Sonis, M. (2006). An exploratory analysis of hierarchical spatial interaction: The case of regional income shares in Indonesia. *Journal of Geographical Systems*, 8(3), 253–268.
- Openshaw, S. (1977). Optimal zoning systems for spatial interaction models. *Environment and Planning A*, 9(2), 169–184.
- Openshaw, S., & Rao, L. (1995). Algorithms for reengineering 1991 census geography. *Environment and Planning A: Economy and Space*, 27(3), 425–446. <https://doi.org/10.1068/a270425>
- Qian, Y., Zhou, W., Pickett, S. T. A., Yu, W., Xiong, D., Wang, W., & Jing, C. (2020). Integrating structure and function: Mapping the hierarchical spatial heterogeneity of urban landscapes. *Ecological Processes*, 9(1), 59. <https://doi.org/10.1186/s13717-020-00266-1>
- Rinzivillo, S., Mainardi, S., Pezzoni, F., Coscia, M., Pedreschi, D., & Giannotti, F. (2012). Discovering the geographical borders of human mobility. *KI-Künstliche Intelligenz*, 26(3), 253–260.
- Schneider, M. (1959). Gravity models and trip distribution theory. *Papers in Regional Science*, 5(1), 51–56.
- Shen, Y., & Batty, M. (2019). Delineating the perceived functional regions of London from commuting flows. *Environment and Planning A: Economy and Space*, 51(3), 547–550. <https://doi.org/10.1177/0308518x18786253>
- Sohn, J. (2005). Are commuting patterns a good indicator of urban spatial structure? *Journal of Transport Geography*, 13(4), 306–317.
- Tannier, C., Thomas, I., Vuidel, G., & Frankhauser, P. (2011). A fractal approach to identifying urban boundaries. *Geographical Analysis*, 43(2), 211–227. <https://doi.org/10.1111/j.1538-4632.2011.00814.x>
- Wilson, A. (1970). *Entropy in urban and regional modelling* (Vol. 1, p. Routledge).
- Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A: Economy and Space*, 3(1), 1–32. <https://doi.org/10.1068/a030001>
- Wu, C., Smith, D., & Wang, M. (2021). Simulating the urban spatial structure with spatial interaction: A case study of urban polycentricity under different scenarios. *Computers, Environment and Urban Systems*, 89, Article 101677. <https://doi.org/10.1016/j.compenvurbsys.2021.101677>
- Wu, F. (2016). China's Emergent City-region governance: A new form of state spatial selectivity through state-orchestrated rescaling. *International Journal of Urban and Regional Research*, 40(6), 1134–1151. <https://doi.org/10.1111/1468-2427.12437>
- Yeh, A. G.-O., & Chen, Z. (2019). From cities to super mega city regions in China in a new wave of urbanisation and economic transition: Issues and challenges. *Urban Studies*. <https://doi.org/10.1177/0042098019879566>, 0042098019879566.
- Yin, J., Soliman, A., Yin, D., & Wang, S. (2017). Depicting urban boundaries from a mobility network of spatial interactions: A case study of Great Britain with geo-located twitter data. *International Journal of Geographical Information Science*, 31(7), 1293–1313. <https://doi.org/10.1080/13658816.2017.1282615>
- Zhai, W., Bai, X., Shi, Y., Han, Y., Peng, Z.-R., & Gu, C. (2019). Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Computers, Environment and Urban Systems*, 74, 1–12. <https://doi.org/10.1016/j.compenvurbsys.2018.11.008>
- Zhang, W., Fang, C., Zhou, L., & Zhu, J. (2020). Measuring megaregional structure in the Pearl River Delta by mobile phone signaling data: A complex network approach. *Cities*, 104, Article 102809. <https://doi.org/10.1016/j.cities.2020.102809>
- Zhang, Y., Marshall, S., Cao, M., Manley, E., & Chen, H. (2021). Discovering the evolution of urban structure using smart card data: The case of London. *Cities*, 112, Article 103157. <https://doi.org/10.1016/j.cities.2021.103157>
- Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, G. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, 28(11), 2178–2199. <https://doi.org/10.1080/13658816.2014.914521>
- Zhong, C., Manley, E., Müller Arisona, S., Batty, M., & Schmitt, G. (2015). Measuring variability of mobility patterns from multiday smart-card data. *Journal of Computational Science*, 9, 125–130. <https://doi.org/10.1016/j.jocs.2015.04.021>
- Zipf, G. K. (1946). The P 1 P 2/D hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6), 677–686.