

Solutions to cultural, organizational, and technical challenges in developing PECAS models for the cities of Shanghai, Wuhan, and Guangzhou

Ming Zhong

Wuhan University of Technology
mzhong@whut.edu.cn

John D. Hunt

University of Calgary
jdhunt@ucalgary.ca

Tao Chen

Tongji University
447807567@qq.com

Wei Yang

Wuhan Transportation Planning Institute
bdywzn@163.com

Wanle Wang (Corresponding Author)

Wuhan University of Technology
wlwang@whut.edu.cn

P. Haixiao Pan

Tongji University
hxpank@online.sh.cn

Jianzhong Li

Wuhan Transportation Planning Institute
ljzplan@163.com

Ke Zhang

Guangzhou Transport Planning Research
Institute
gzjtyjs@126.com

Abstract: Massive construction of transportation infrastructure and fast growth of private car ownership have brought unprecedented changes in land use and transportation systems to cities and regions in many developing countries. Traditional “four-step” travel demand models, which are not designed to assess transport policies under the case of rapid land-use change, cannot be used to achieve coordinated planning of transport and land use. Therefore, there is a pressing need to develop and use integrated land-use transport models (ILUTMs), which consider interactions among socioeconomic activities, urban land use, and transportation development, for policy analysis and for guiding the progressive urbanization process taking place in many parts of these countries. In light of this, efforts have been invested in developing production, exchange, and consumption allocation system (PECAS) models for the cities of Shanghai, Wuhan, and Guangzhou in mainland China. This paper presents the cultural, organizational, and technical challenges encountered in the development of PECAS models for the cities of Shanghai, Wuhan, and Guangzhou and the mitigating solutions from the development teams for taking up or working around them. The solutions and discussions presented in this paper should be interesting to researchers and practitioners for developing ILUTMs in the context of a developing country like China.

Article history:

Received: October 24, 2017

Received in revised form:

June 5, 2018

Accepted: July 20, 2018

Available online: December 4,
2018

Copyright 2018 Ming Zhong, Wanle Wang, John D. Hunt, P. Haixiao Pan, Tao Chen, Jianzhong Li, Wei Yang & Ke Zhang
<http://dx.doi.org/10.5198/jtl.u.2018.1329>

ISSN: 1938-7849 | Licensed under the [Creative Commons Attribution – Noncommercial License 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

1 Introduction

Development and use of integrated land-use/transport models (ILUTMs) for policy analysis in developing countries appear to be more pressingly needed than in developed countries, since they are experiencing unprecedented changes in their social/economic development, travel behavior, transportation infrastructure and land use. In particular, for the past few decades, the cities in China have been undergoing a rapid growth in their economy, built environment and transportation infrastructure. However, the problem of traffic congestion and air pollution has been becoming ever more serious with the expansion of the cities and skyrocketing car ownership. ILUTMs, with their capacity for considering the interactions between social/economic development, land use and transport, were developed and used to tackle the above issues in the context of developed and developing countries elsewhere (Wegener, 2004; Hunt, Kriger, & Miller, 2005; Iacono, Levinson, & El-Geneidy, 2008). They show their potential to study various policies related to economy, transportation, land use and social/natural environment in a holistic and realistic manner, which potentially eliminates the possibility of incomplete understanding of the impact of various policies and myopic solutions.

Although there are well known computable general equilibrium models and integrated land-use/transport models under relatively steady growth in developed countries, there are few models available in fast growing developing countries under rapid transformations in land use and transport. For the past two decades, great efforts have been made to study the interrelationship between transportation, land use and urban forms in China from one aspect to another (Mao & Yan, 2004, 2005; Pan, 2007; Pan & Zhang, 2008; Pan & Chen, 2009; Long & Shen, 2015). For example, Mao and Yan (2004) studied the corridor effects of urban transport arteries on land use based on a case study of Guangzhou Avenue in the City of Guangzhou. They combined RS (Remote Sensing) and GIS (Geographic Information System) to analyze the spatial attraction and spatial distribution effects. Mao and Yan (2005) also discussed the impacts of transport system on urban land use in a highly densely developed city — Guangzhou in terms of urban form and land value changes. Focusing on rail transit, Pan (2007), Pan and Zhang (2008) and Pan and Chen (2009) studied its impacts on land use, residential location choice and spatial structure. Long and Shen (2015) discussed in detail the application of geospatial analysis in the urban space and urban activity in one of the world's megacities — Beijing. They emphasized the role of planning support systems in urban planning, big data and big models. In these studies, however, only a few were conducted through the application or development of a land-use/transport model, which may be used to study such interrelationship from multiple perspectives. Zheng and Huo (2012) adopted UrbanSim to study the location and development of office buildings in Beijing. Shi, Tong, Zhang, and Tao. (2012) applied UrbanSim to explore the relationship between land-use change and energy consumption in Yizhuang, Beijing. Jin, Echenique, and Hargreaves (2013) built a recursive spatial equilibrium model based on urban general equilibrium, spatial interaction and non-equilibrium dynamic models for the planning of large-scale urban change. Jin et al. (2017) used a suite of dynamic land use, spatial equilibrium and strategic transport models for the medium to long term travel demand and associated emissions forecasting for Beijing municipality in China.

PECAS stands for Production, Exchange, and Consumption Allocation System (PECAS), which is one of the integrated land-use/transport modeling frameworks recently developed. Built upon the theories and experiences from its pioneers MEPLAN (Hunt & Simmonds, 1993) and TRANUS (de la Barra, 2012), the PECAS framework addressed several issues in the previous models and has been advanced with several distinct features, such as enhanced representation of socioeconomic system through social accounting matrix and microsimulation-based space development. Over the last 15 years, the PECAS framework has been implemented in more than a dozen locations worldwide, including Oregon

(statewide) (Weidner, Hunt, Abraham, Moeckel, Gregor & Knudson, 2008; Hunt, Weidner, Knudson, Bettinardi, & Wardell, 2010), California (statewide) (Gao, Lehmer, Wang, McCoy, Johnston, Abraham & Hunt, 2010), Atlanta (Abraham, Hunt, & Wang, 2014), San Diego, Los Angeles, San Francisco, USA; City of Edmonton, Province of Alberta, Canada (Hunt & Abraham, 2003); Caracas, Venezuela (Fuenmayor, 2016); Mumbai, India (Mir, Krishna Rao, & Hunt, 2010); Shanghai and Wuhan, China (Zhong et al., 2016).

This paper discusses key challenges encountered in the development of PECAS models for the cities of Shanghai, Wuhan and Guangzhou in China with respect to cultural, organizational and technical aspects, and corresponding solutions to solve or work around them. The discussions presented should be useful as a reference for researchers and practitioners who are or will be developing ILUTMs in the context of developing countries.

After a short introduction of PECAS modeling framework and the three ILUTM projects (for cities of Shanghai, Wuhan and Guangzhou), the challenges encountered and the mitigating solutions proposed are presented. The paper finishes with our conclusions and recommendations.

2 PECAS modeling framework and its implementations in China

2.1 Introduction to ILUTMs and the PECAS framework

The Lowry model developed in 1964 is seen as the classic ILUTM, which is the first operational framework that can explicitly consider the interactions between land use and transportation (Lowry, 1964). Then it was further enhanced to DRAM/EMPAL and then to ITLUP/METROPILUS by Putman (1983, 1991, 1994, 1996). Echenique and Lindsay (1969) and Echenique, Flowerdew, Hunt, Mayo, Skidmore & Simmonds (1990) introduced the MEPLAN model and used the model in Spain, West Germany and England, to predict the effects of local land use and transport policies. Boyce, Chon, Ferris, Lee, Lin & Eash (1985) firstly introduced a combined model of urban travel and location in the Chicago region, which provided transportation planners a decision-making tool based on computer methods and data. Miller and Salvini (2001) presented the architecture of the Integrated Land Use, Transportation, Environment (ILUTE) system, making it possible to simulate and analyze the environment impacts of different land use and transport plans. Waddell (2000) and Waddell, Peak, and Caballero (2004) proposed UrbanSim, which is featured with microsimulating urban land, real-estate, transport markets without requiring an equilibrium solution.

Based on the MEPLAN and TRANUS models, Hunt and Abraham (2003) proposed the PECAS framework to simulate the spatial economic systems under consideration, which is featured with higher spatial resolution and greater behavioral fidelity. PECAS is a generalization of the spatial I/O modeling approach used in the MEPLAN and TRANUS land-use/transport modeling systems. It incorporates the “integrated” structure of these models and also the “connected” structure used by UrbanSim. In an integrated model, the land-use component feeds the spatial distribution of flows of economic interactions between zones as origin and destination movements into the transport component, and the transport module performs the remaining steps. The transport component of a connected model performs the classical four steps, including trip distribution. Compared with a connected structure, the integrated option assures the best integration of the two components and avoids inconsistency in the use of O-D matrix. Both the “connected” and “integrated” structure can be set up within the PECAS framework, which is an important advantage to engage planning agencies when a full-fledge, four-step model is in place and a connected structure can be easily set up to run land-use and transport model concurrently. An integrated structure can be developed over a long-run when resources and opportunities permit. The unification of these two approaches into a single framework in PECAS is seen to be a step forward

which enables the benefits of both approaches to be realized. Compared to other ILUTM frameworks, this could be one advantage that makes PECAS more attractive. Besides, PECAS is very flexible in accommodating itself to a variety of data or the lack of data, which makes it applicable to situations where data shortage is the main obstacle of modeling. Depending on its design and purposes, a PECAS model usually consists of the following modules (Hunt & Abraham, 2009): Economy and Demographics Forecasting (ED module), Activity Allocation (AA module), Space Development (SD module) and Transport (TR module), as shown in Fig.1.

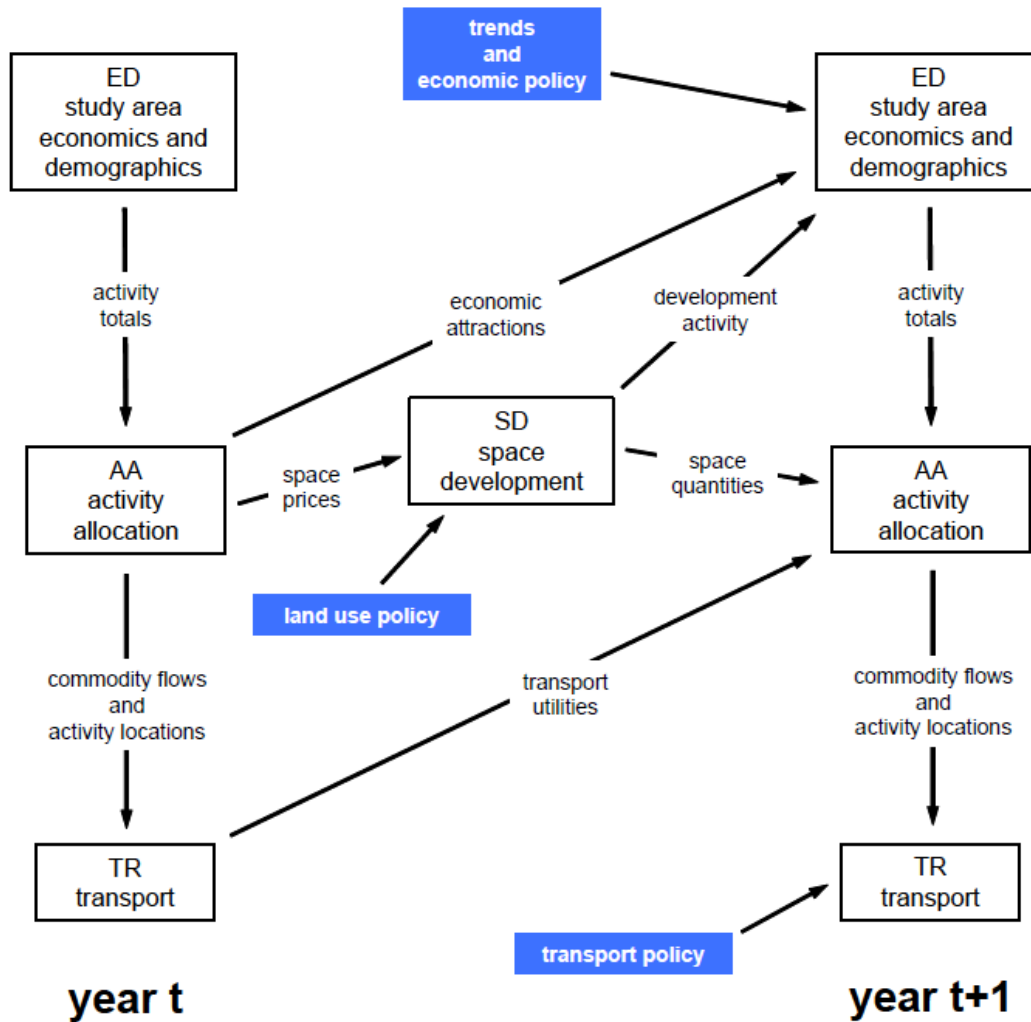


Figure 1: Modules and information flows simulating temporal dynamics in the PECAS framework

ED module usually uses structure equations, regression analysis or time series analysis to forecast the size of activities within the studied area, which provides the activity totals, normally in terms of monetary units, to the AA module where the activities are allocated to different zones based on random utility theory. Flows of commodities from production zones to exchanges zones and from exchanges zones to consumption zones are determined by nested logit models that take into account exchange prices and transport (dis)utilities. SD module simulates developer actions according to the exchange prices for space. The PECAS framework provides a static equilibrium framework, but it can be run annually. It integrates interregional trade with and supports modeling of freight due to the relationship between

trade and the movement of goods. Different policies related to economy, land use and transport can be investigated with the framework.

Abraham, Gordon, and Hunt (2005) presented the development of the PECAS Model for the region of Sacramento, which was developed based on a series of pioneering land-use/transport interaction models (including TRANUS and MEPLAN) used by the Sacramento Area Council of Governments (Hunt, Johnston, Abraham, Rodier, Garry, Putman & de la Barra, 2001). Zhong, Hunt, and Abraham (2007) developed a province-wide PECAS Demo model for Alberta, Canada for the provincial government, which was intended for policy analysis, in order to make more informed economic, land use and transportation infrastructure decisions. Hunt, Abraham, De Silva, Zhong, Bridges & Mysko (2008) developed a PECAS model to simulate space development in Baltimore. PECAS models have also been developed or being developed in many other locations, such as Atlanta, USA and Mumbai, India.

Regarding the three PECAS models under development for the cities of Shanghai, Wuhan and Guangzhou, an integrated structure has been set up, where the land-use model will carry out “trip generation and distribution” and the transport model takes care of “mode split” and “network assignment.” The land-use model will take the congested travel/transport time from the transport model as one of the inputs, as shown in Fig.1. The three PECAS models will be briefly introduced below.

2.2 Shanghai PECAS demo model

The City of Shanghai, located in the southeast region of China and the Delta region of the Yangtze River, is one of the most developed parts in China. Like other top cities in China, the City of Shanghai is also undergoing a progressive economic and social development, accompanied by a dramatic increase of population and urbanization, and a large scale of investment in space and transportation infrastructure development. However, the planning of land use and the planning of transport have been traditionally carried out by independent departments of the city government and problems arising from uncoordinated plans become apparent and costly to mitigate. Therefore, an effective tool to address this issue is particularly useful. In this regard, an integrated land-use/transport model (ILUTM) is clearly the choice.

Due to the shortage of time, budget and data, a PECAS Demo model is first developed. Fig.2 shows the design diagram of Shanghai PECAS demo model, which only considers 16 broad categories of activities (including 6 types of households classified by their income level (low, medium and high) and rural/urban designation), 13 types of goods, 5 types of labors and 7 types of space. The rows show producing and consuming activities, while the columns are commodities (including goods, services, space and labors). The commodity production is represented in a “make table format” at the top of the figure, and the commodity consumption is represented in a corresponding “use table format” at the bottom of the figure. Commodities flow down the columns from where they are produced to where they are consumed (Hunt and Abraham, 2009). A large number of data items have been assembled and then used to design and develop the model during this stage. Currently, the Activity Allocation (AA) module

	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	L01	L02	L03	L04	L05	S01	S02	S03	S04	S05	S07
produce activities																								
agriculture operations		4952796																						
agriculture office		230211																						
manufacturing operations		69648066																						
manufacturing office				5791475																				
retail hospitality catering (H0)						4505802																		
retail office							46937																	
other the third industry									1806313															
III high income urban											9074503.721			345053.7465	2597331.591	245146.7889		0	127011.4541					
III medium income urban											2948879.50			642295.697	1064667.29	516445.656		0	96756.6955					
III low income urban											264006.91			9145.87730	10042.6705	272325.501		0	36498.3814					
III high income rural											1277611.136			2902.928156	102063.1193	69996.33025		0	23.52063965					
III medium income rural											1306527.468			559.3989943	18067.01325	22698.77675		0	11.76031983					
III low income rural																								
government account																								
capital account																								
total demand																								
Input (fact total)																								
Total supply (Production + Import)																								
Export																								
consume activities																								
agriculture operations																								
agriculture office																								
retail hospitality catering (H0)																								
retail office																								
other the third industry																								
III high income urban											98246			772	143978	84218		0	599892					
III medium income urban											1278551			5059713	324249	2532929		0	1633					
III low income urban											412027			24179095	254943	1702490		0	4590					
III high income rural											1910272			7246314	2311592	602234		0	4113					
III medium income rural											3006796			14252866	7777599	4157800		0	517					
III low income rural																		0	19918.60941					
government account																		0	29877.91412					
capital account																		0	3522.91039					
total demand																		0	7132.2					
Input (fact total)																		0	2032					

Figure 2: The aggregate economic flows table for Shanghai PECAS demo model

of Shanghai PECAS Demo model has been built and its Space Development (SD) module is still under development.

AA module is run under different scenarios in Shanghai which include congestion tolling and polycentric development strategy. The results are shown in Fig.3 and Fig.4 respectively. The coded colors from blue to red represent quantities of activities from low to high. As can be seen from Fig.3, under the congestion tolling scenario, more agricultural activities will be allocated to the southwest area of the city, more industrial activities to the northwest of the city and more Retail, Hospitality and Catering (RHC) activities to the northeast area where the Pudong District is located. In Fig.4, under polycentric development strategy, both the agricultural and the industrial activities will concentrate in the southwest area of the city, while RHC activities will generally move to the north of the city with few concentrations in different parts of the north.

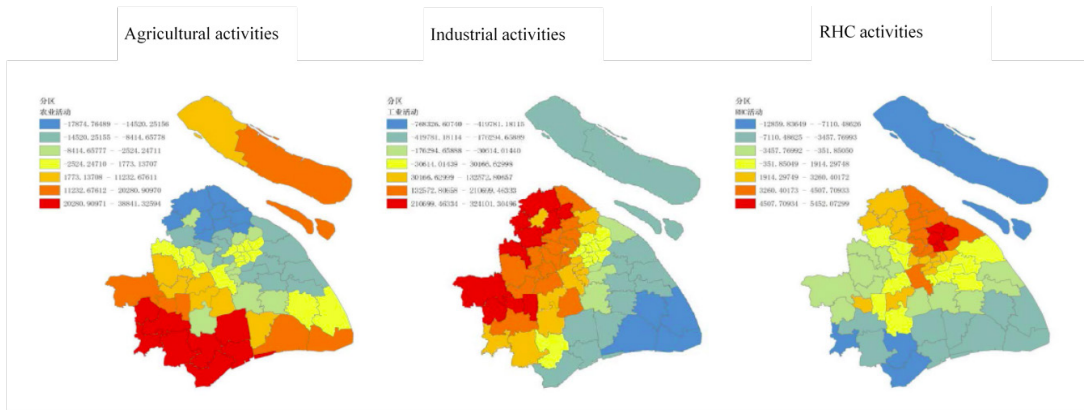


Figure 3: Spatial allocation pattern of sample activities under congestion tolling

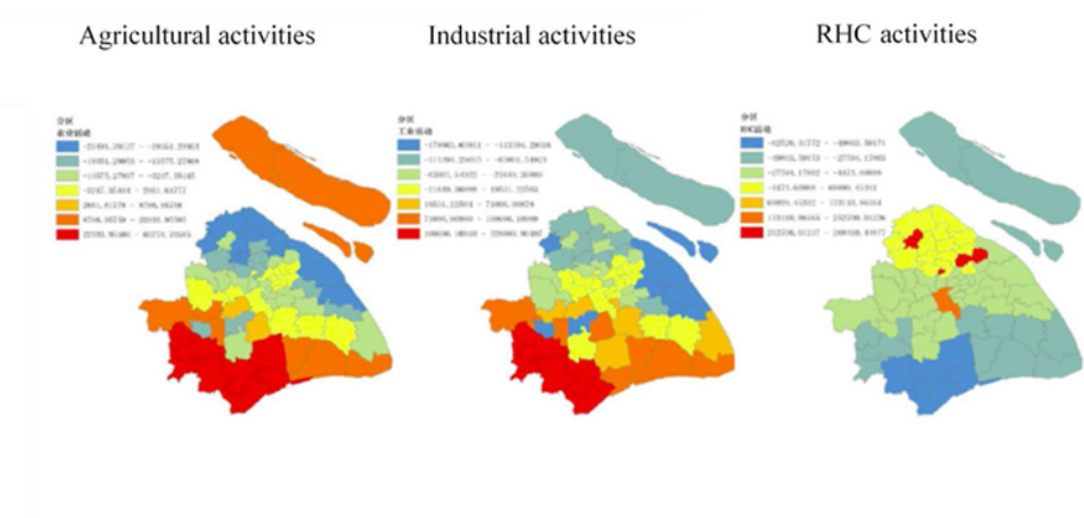


Figure 4: Spatial allocation pattern of sample activities under polycentric development strategy

2.3 Wuhan PECAS demo model

The City of Wuhan, located in the central part of China, is the capital of Hubei province, the largest city in Central China and the largest transportation hub in China. The City of Wuhan is undergoing a rapid

economic and social development, and therefore a careful plan is required for a sustainable development. Similar to the case of the City of Shanghai, the planning of land use and the planning of transport of the city have also been carried out by independent government branches, resulting in uncoordinated plans which could potentially result in many side effects on the city development. In this regard, an effective integrated land-use/transport model can play a significant role for supporting the development of a coordinated plan.

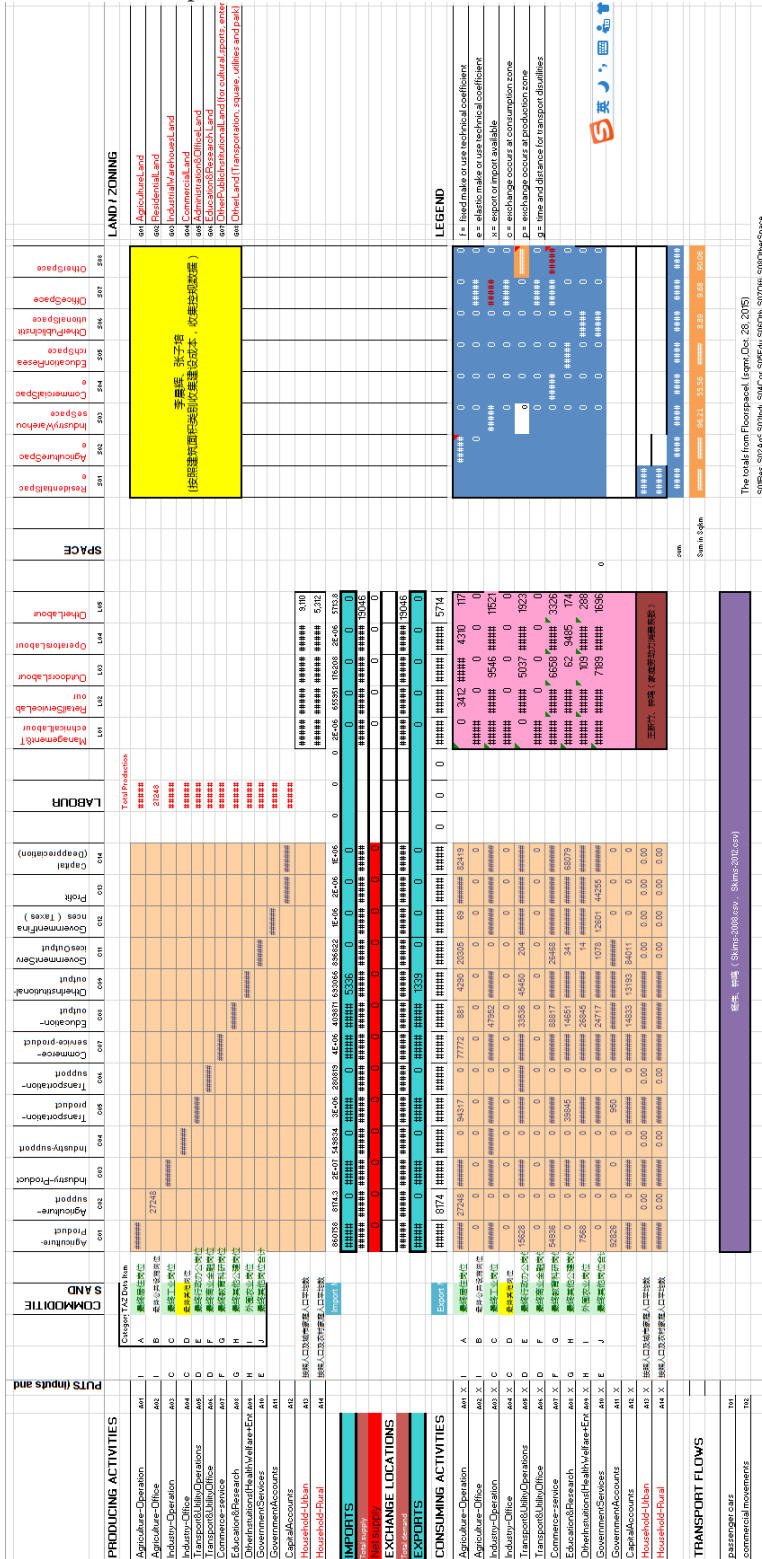


Figure 5: The aggregate economic flows table for Wuhan PECAS demo model

Like in the case of Shanghai, due to the shortage of time, budget and data, a PECAS Demo model is first developed. Fig.5 shows the design diagram of Wuhan PECAS demo model. It is featured with 14 types of socioeconomic activities, 13 types of goods, 5 types of labor and 8 types of space, which were determined based on the resources available to the project and consideration of integration with the existing travel model owned by the government agency. The model will cover the entire City of Wuhan (including its urban and rural area) and have 690 transportation analysis zones (TAZs) and 147 land-use zones (LUZs). A large number of data items about the city have been assembled. At present, the joint AA-SD version of Wuhan PECAS Demo model is running and subject to further calibration and validation. Some preliminary results from the Demo model are presented as follows.

1) ED module

To provide the activity totals for the PECAS framework, initially, separate ordinary least square (OLS) regression model was developed for each of the sectors considered (Industry sectors and Households shown in Table 1) based on historical data, but it was found that the models developed did not correctly represent the expected trend using historical gross output totals of each specific sector. The forecast results are inconsistent with the expected development trend, where a positive but declining growth rate would take place as the time moves along. Therefore, several other techniques, including regression models with polynomial variables and ARIMA, were tried, but they were all found to be unable to represent the expected trend described above. Finally, a set of exponential smoothing functions were chosen, as they were found to be able to represent the trend identified based on the data from the last few years and they were used to forecast the totals for the 12 Industry sectors and the two Household categories.

The 12 Industry sectors are categorized into the three industries in the city and the GDP growth rate of each industry and the overall GDP growth rate are calculated based on the economic data from years 2002–2016. Assuming a declining and steady economic growth in the planning horizon (see Fig.6), the forecasted results of 12 Industry sectors from years 2017 to 2038 are illustrated in Fig.7.

The rural and urban population data are collected from the Yearbook of 2002–2016. It should be noted that the urban population is the urban resident population plus floating population. Floating population is defined from the census point of view, as the population that have different residential places as their places of household registration and their stay in the current location is less than 6 months. Since the floating population has a great fluctuation before the year of 2015, it is impossible to forecast the floating population directly. Therefore, we forecasted the total population (Fig.8 (a)), rural resident population (Fig.8 (b)) and urban resident population (Fig.8 (c)) separately. The floating population is obtained by subtracting both the resident population from the total population, as shown in Fig.8 (d). The urban population is calculated and shown in Fig.8 (e).

Table 1: Industry sectors and households considered in Wuhan PECAS demo model

Industry sector	Household
Agriculture-Operation	Household-Urban
Agriculture-Office Support	Household-Rural
Industry-Operation	
Industry-Office Support	
Transport&UtilityOperations	
Transport&UtilityOffice	
Commerce-service	
Education&Research	
OtherInstitutions (HealthWelfare+EntertainSportCulture)	
GovernmentServices	
GovernmentAccounts	

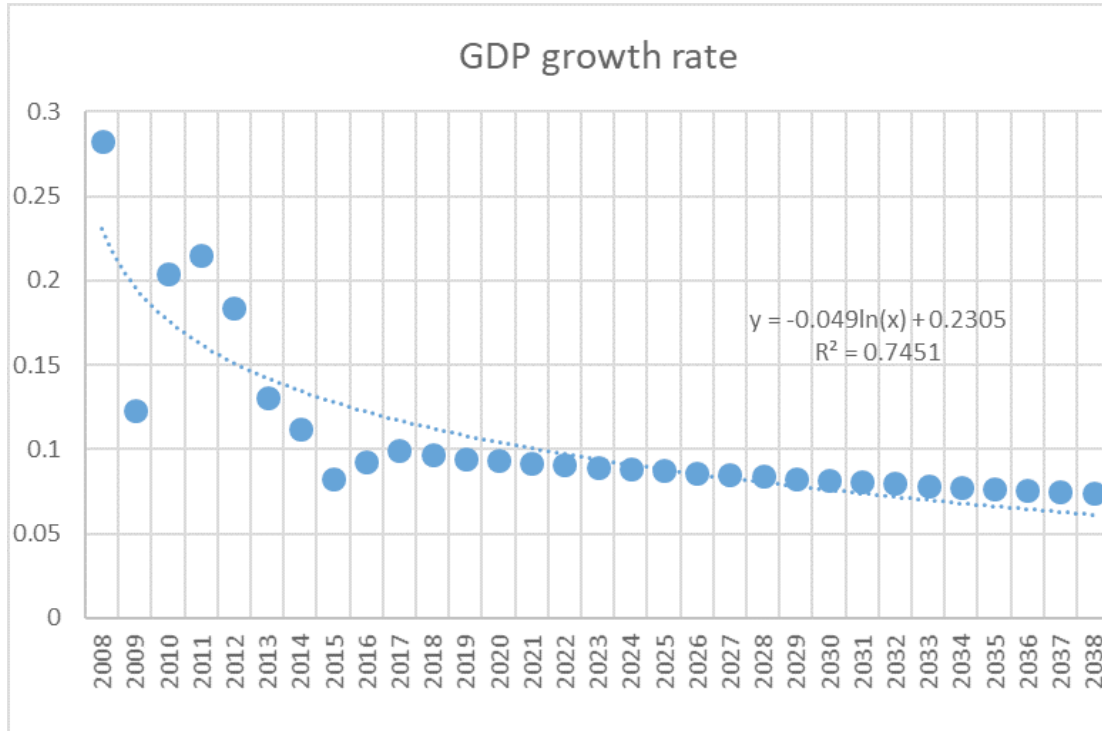


Figure 6: Forecasted results of GDP growth rate for Wuhan city

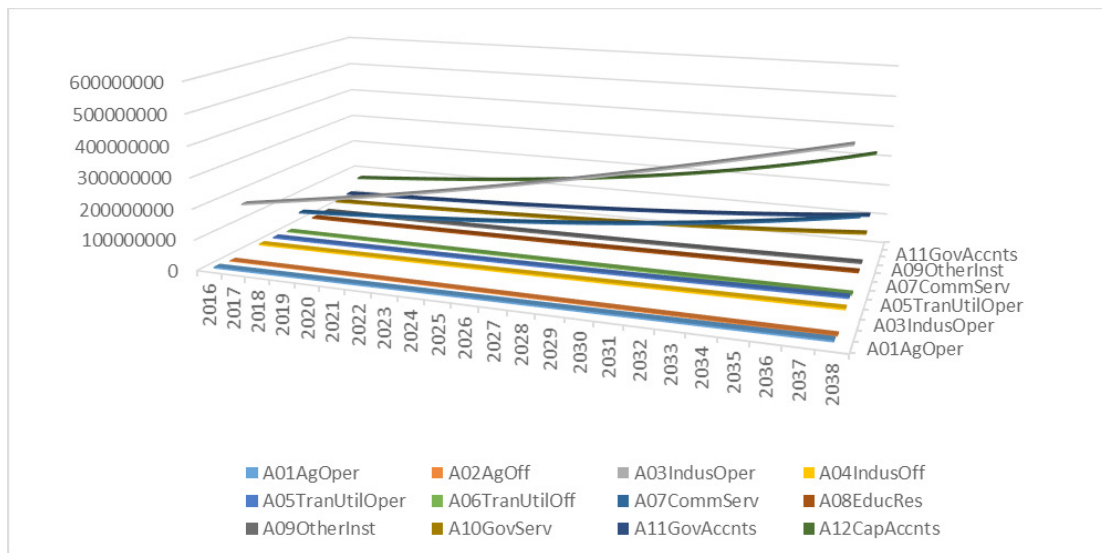
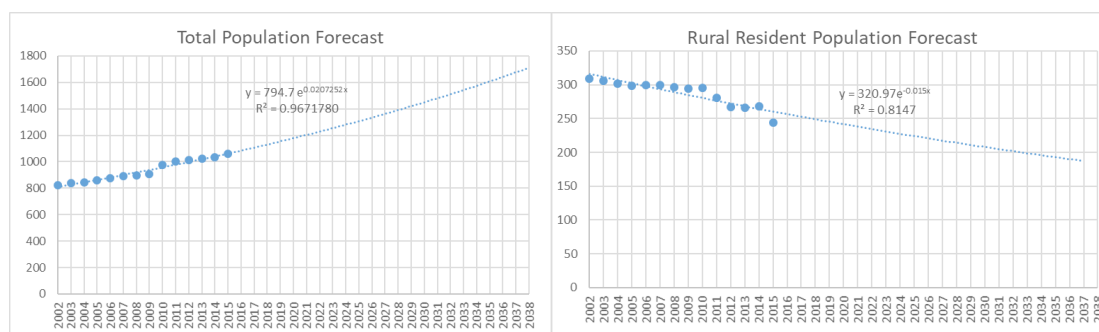


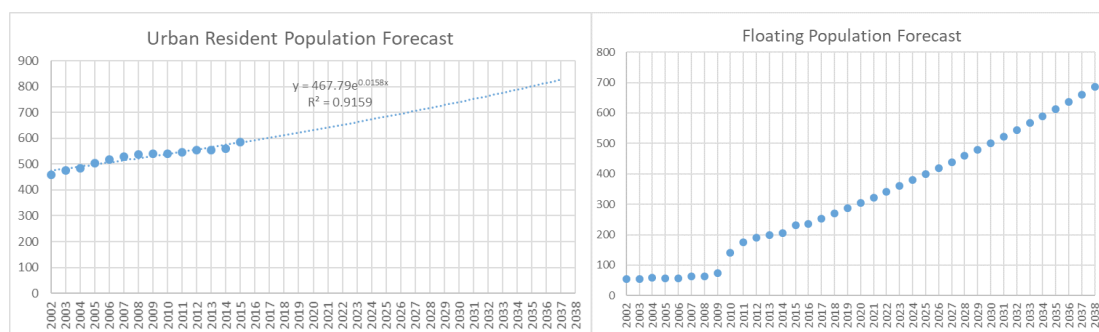
Figure 7: Forecasted results of activity totals by sectors through time from 2016 to 2038¹

¹Note: All activity totals are in 10,000 yuan.



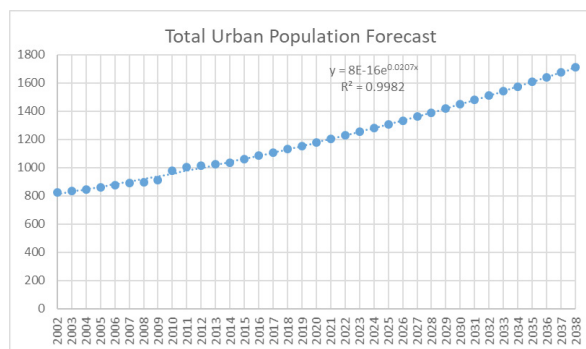
(a) Total population forecast

(b) Rural resident population forecast



(c) Urban resident population forecast

(d) Floating population forecast



(e) Total urban population forecast

Figure 8: Forecasted results of population for the city of Wuhan²

2) AA module

Activity Allocation Module in the Wuhan PECAS Demo model determines how total activity forecasted (by sector) by the ED module will be allocated in each land-use zone (LUZ) or traffic analysis zone (TAZ) and how these activities interact with each other from years 2008 to 2038. Fig.9 shows how Commercial Service activity will be allocated into zones from years 2008 to year 2038. Fig.10 shows trip volumes of L01ManTechLabour (management & technical labor) between selected zones in 2012.

²Note: All population figures are in ten-thousand.

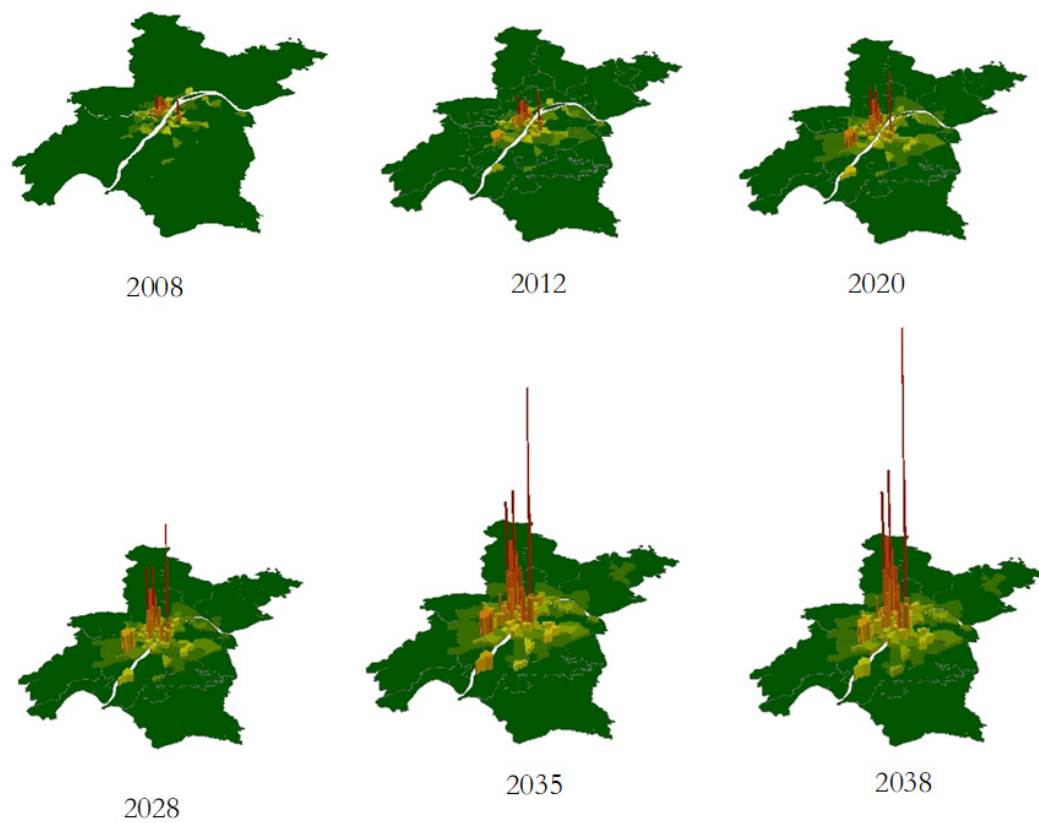


Figure 9: Allocated commercial service activity over the zones for sample cross-section years

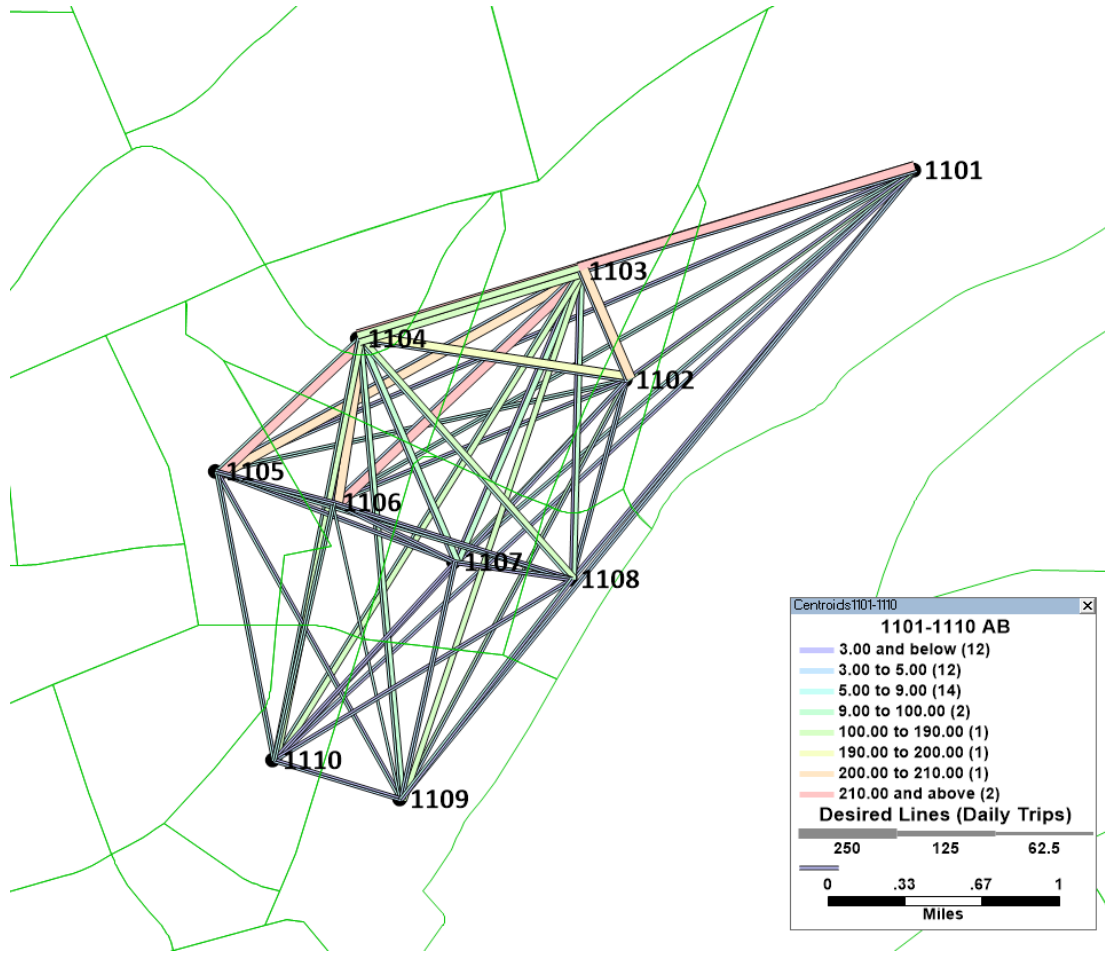


Figure 10: Trip volumes of L01ManTechLabour (management and technical labor) between selected zones in 2012

The forecasted trip length distribution of each commodity reported by the AA module is analyzed (only the agricultural products and industrial products are shown in Fig.11 as the examples). Both the buying and selling activity of agricultural products require a distance mostly longer than 2 km, but the agricultural products for selling are transported by shorter distance than the products for buying. On the other hand, the trip length distribution of industrial products shows a similar preference for buying and selling, where the selling of the industrial products has much shorter trip lengths than the buying. The results seem reasonable, since the selling of the agricultural products and industrial products are mostly driven by the markets and therefore the suppliers/sellers are required to transport them to the exchange zones to be close to the customers to meet the demand of the local markets and therefore a relatively shorter trip length. On the other hand, the buying of the two products, especially with a large amount of quantity, would encourage the buyers (usually wholesalers and enterprise customers) to consider the regional even national markets in order to receive a competing price and therefore the buying often requires much longer trip lengths.

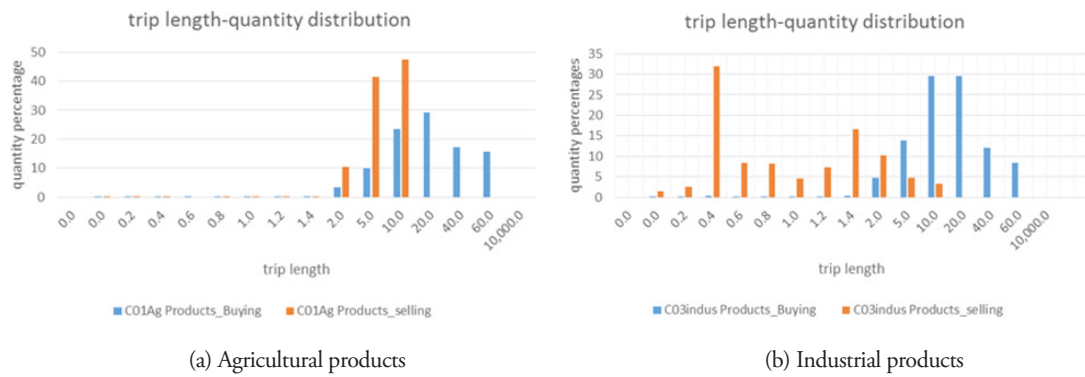


Figure 11: Trip length distributions of agricultural and industrial products in 2008

3) SD module

The forecasted quantities of the industrial space developed for several cross-section years from years 2008 to 2038 are presented in Fig.12. Fig.13 shows the cumulative residential space developed by each year in several sample LUZs selected. It can be seen that generally the residential spaces in these zones experience a quick increase from 2008 to 2020, especially zone 78, and then become very stable after that. A possible reason is that the quantity has reached the maximum allowed by the zoning regulation or the space quantity is enough for supporting the forecasted residential activity in these zones (and therefore stabilized space pattern).

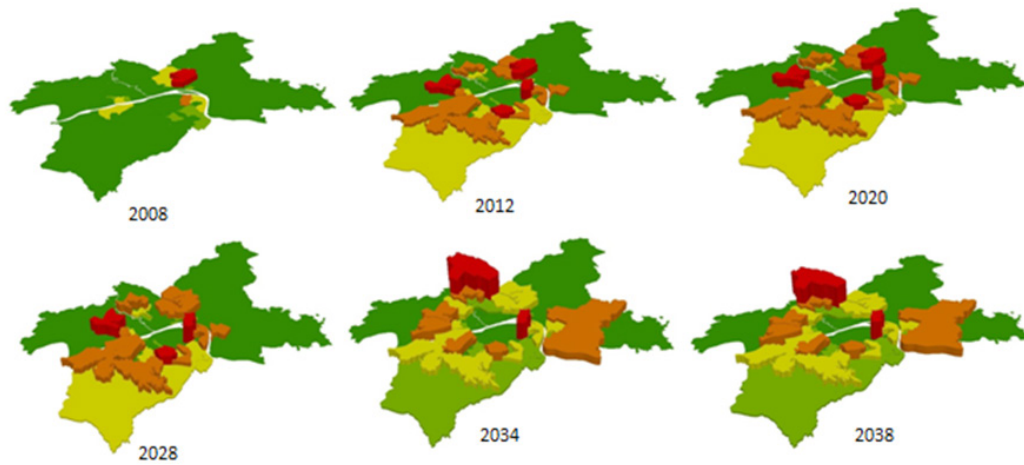


Figure 12: Forecasted quantities of the industrial space developed for sample cross-section years

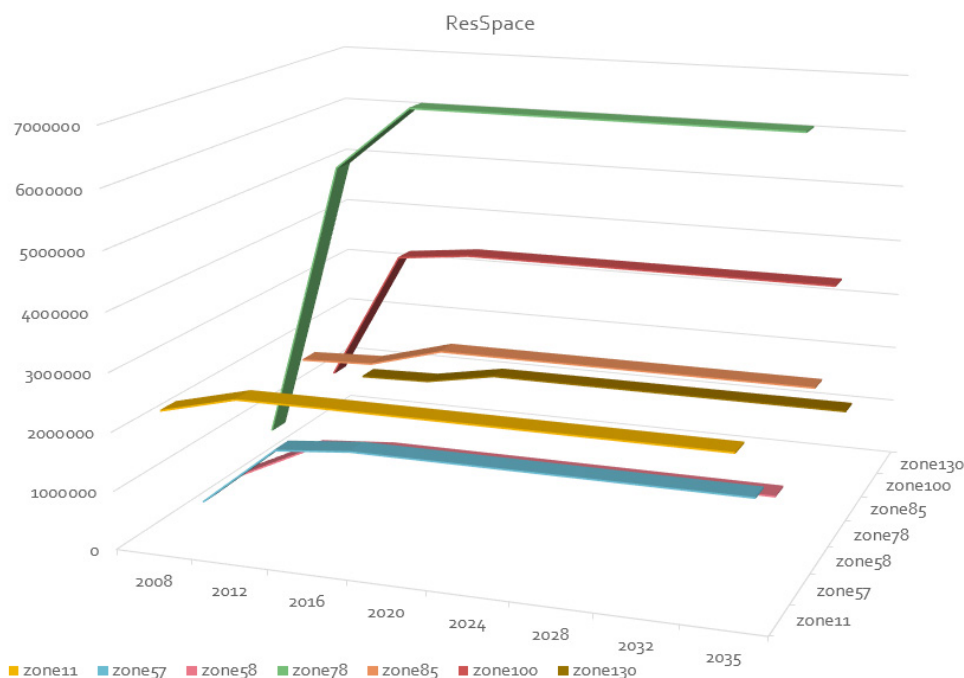
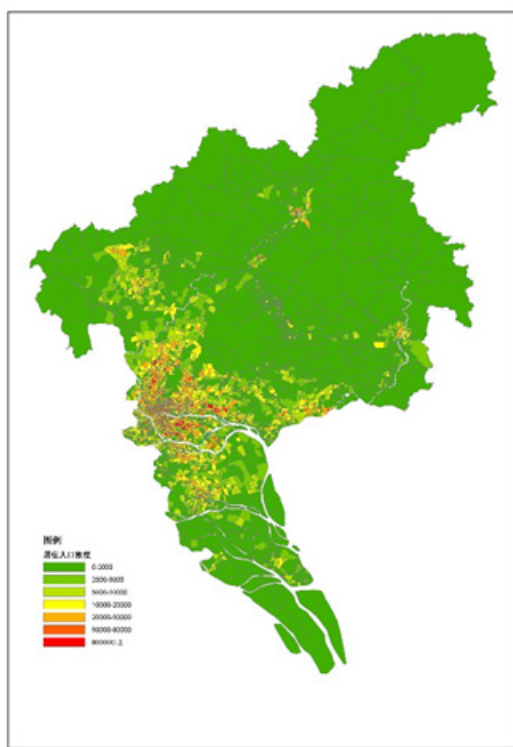


Figure 13: Cumulative residential space in sample land-use zones

The model can be applied to investigate real policy questions that cannot be addressed by simple travel models. For example, a new river-side business district is going to be developed based on an “old area” of the city in Hankou District (right on the north side of Yangtze River), which features with limited number of small-size streets and low-density, informal development. The developers wish to have a high-density development planned, but the capacity of the streets within the area would not be able to accommodate all the traffic associated with such a development. The proposed ILUTM for the city will not only correctly forecast the phases and quantities of developed space (by category by zone) and therefore be able to support the study of changes of traffic patterns within the area, but also provide estimates of highest development density allowed that is restrained by the capacity of surrounding streets. As another example, the government of the City of Wuhan is examining the plan of building a 2.5 ring-road, which will be placed between the 2nd and the 3rd ring road of the city. However, the exact location of this ring road and its impact on the urban form (sprawled vs. concentrated) has been under intense debate. With an ILUTM, the desirable location of this ring road can be better located rather than with a travel demand model alone.

2.4 Guangzhou PECAS demo model

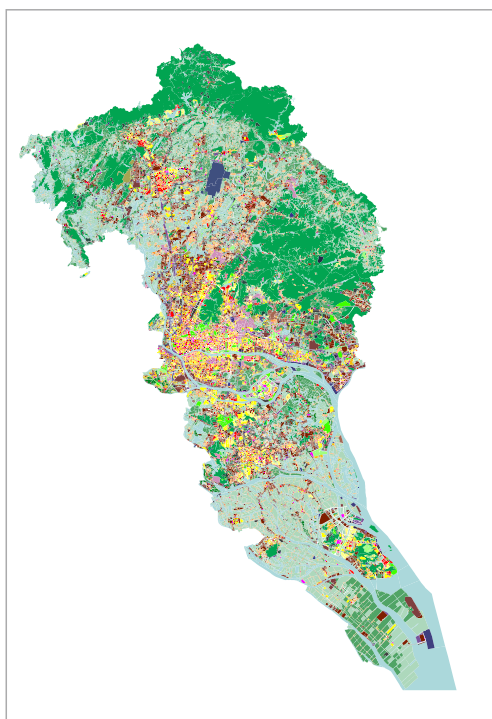
The City of Guangzhou, located on the Pearl River, is China’s third-largest city and a thriving commercial center. Its proximity to Hong Kong has made it a strategic port for centuries. The residents’ density and employment density of the city for the year of 2017, and the land-use classification under detailed regulatory planning for the year of 2030 are shown in Fig.14.



(a) Residents' density for 2017



(b) Employment density for 2017



(c) Land-use classification under detailed regulatory planning for 2030

Figure 14: Residents' density, employment density and land-use classification of Guangzhou city

For the past decades, the traditional four-step model was criticized for its incompetency in dealing with the relationship between urban master plan, land-use plan and transportation plan. For example, when a land-use plan is made, the distribution of the production factors for each industry should be taken into consideration. The plan is feasible only if the factors required for production are distributed within a certain range (measured by production cost, which includes the value of commodity and corresponding transport cost). To coordinate different planning practices and facilitate the sustainable development of the city, Guangzhou PECAS Demo model is being developed. The design diagram of Guangzhou PECAS Demo model is shown in Fig.15.

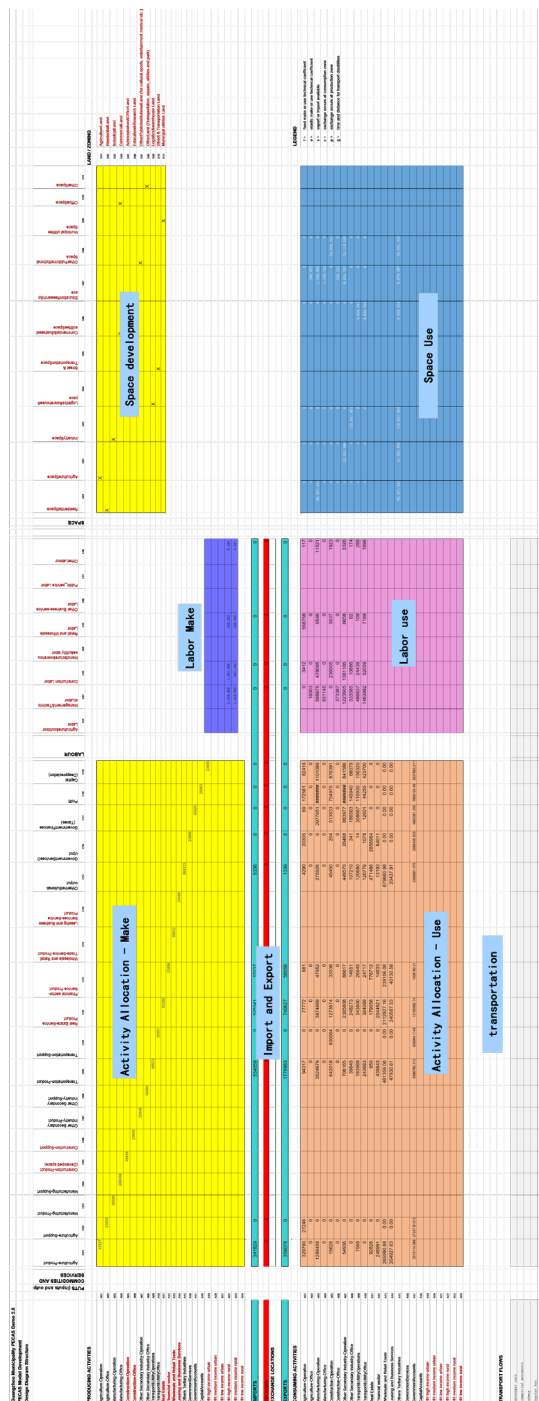


Figure 15: The aggregate economic flows table for Guangzhou PECAS demo model

Compared with the Wuhan PECAS Demo model, the Guangzhou PECAS Demo model has the following characteristics:

1) Activities

Besides the activities considered in the Wuhan PECAS Demo model, construction, real estate, and wholesale and retail industries are added. The inclusion of construction industry is to determine the impact of the industry on the land development and space supply of the city, and the inclusion of real estate is to analyze the impact of the land and space trade and transfer of ownership on other relevant industries (e.g., Construction and Finance) and eventually urban form. As to the wholesale and retail industry, it is one of the pillar industries of Guangzhou city. Canton Fair (or the China Export Commodities Fair) held annually in Guangzhou city is the largest trade fair in China. With such a design, it is then possible to explore the impact of the development of the above industries on the extension of transportation infrastructure, the land and space development and resulting urban form with the model. In total, 6 more activities are added in the Guangzhou PECAS Demo model, namely, Construction-operation, Construction-office, Real-estate, Wholesale and retail, Financial service, Leasing and business service.

2) Commodities

For the above 6 additional activities added into the Guangzhou PECAS Demo model, the following commodities are taken into consideration accordingly: Construction-operation product - developed space, Construction support, Wholesale and retail service, Financial service, Real-estate service and Leasing and business service.

3) Land use and space

In the model, 11 land use and space types are considered to conform to the next round of detailed zoning or regulatory planning, and predict the spatial distribution of different socioeconomic activities considered in the current 4-step travel model. This is particularly important to the agencies (Guangzhou Land-Use Planning Bureau and their subsidies) in charge of the travel model, as this will empower them to systematically consider the following plans: Master Development Plan, Land-Use Plan, Transportation Plan, Environment Protection Plan, and Water and Resource Plan within a consistent framework.

4) Labor types

Eight types of labor are included to cover the newly added activities. New types are Construction workers, Wholesale and retail trade service workers, Leasing and business service workers, and Public service workers, which are consistent to the activities specified for the model.

5) Household classification

In the Wuhan PECAS Demo model, only urban and rural households are classified, while for the Guangzhou PECAS Demo model, households are divided into 6 according to rural/urban and household income. The more household categories increase the complexity in the household production and consumption blocks and the data demand, but this will provide a way to more accurately model the behavior of rural/urban households, especially the impact of their income onto some critical behaviors relevant to urban modeling, such as housing and car ownership.

6) Transport model

In the transport model, car, truck, subway, bus and non-motorized vehicles (bike and electro-mobility) are considered in order to be consistent with their current and next generation of transport demand model in the future. Consideration of the above modes within a large city like Guangzhou seems to be a minimum of any travel demand model developed for similar cities.

The Guangzhou PECAS Demo model is designed considering the city characteristics, however, the design is too complicated in terms of a demo model. It requires a large amount of resources to complete and a great amount of challenges have to be overcome, which will be discussed in the following sections.

3 Challenges and mitigating solutions

With the three applications of PECAS model to the major cities in China, a number of experiences and lessons can be drawn. During the development process, different types of problems have emerged with respect to cultural, organizational and technical issues. Specifically, they have to do with:

- 1) Modeling culture;
- 2) Institutional barrier;
- 3) Distrust from engineers and politicians regarding ILUTMs;
- 4) Lack of human resources required for ILUTM work;
- 5) Limited resources and urgency of decision making faced by modelers;
- 6) No I/O table for administrative units lower than province and I/O tables do not provide detailed commodities produced or consumed;
- 7) Scissor differences between rural and urban areas;
- 8) Lack of detailed land use/space information at LUZ level;
- 9) Lack of population/employment/land-use data at disaggregate zonal level;
- 10) Lack of freight transport data;
- 11) Data inconsistency; and
- 12) Complexity of developing, calibrating, validating and using ILUTMs.

We discuss these issues in more detail in the following sections of the paper.

3.1 Modeling culture

Generally, the modeling culture is weak, especially for integrated modeling, in a developing country like China. The history of land use and transportation modeling dates back to the later 1950s (Batty, 1979). Afterwards, various kinds of experiments with integrated modeling were carried out. As to the modeling work in China, it was not until the early 1980s that quantitative methods were implemented in transportation planning. After that, transportation modeling started to flourish with the Reform and rapid growth of cities. For example, the first travel survey of the City of Wuhan was conducted in 1987 and the very first travel demand model for the city was developed based on the data collected from the survey. Later two travel surveys for the city were carried out respectively in 1998 and 2008, and travel models were redeveloped accordingly, which all followed the line of traditional “four-step” travel demand model. The Wuhan PECAS Demo model is the first of its kind (Zhong, Hunt, Abraham, Li, Li, Yang & Wang, 2016). A weaker culture about data and modeling has forced the team to use aggregate commodity, occupation, and space categories in the Wuhan PECAS model and a demo model had to be developed in order to demonstrate it does, in fact, functionally show the interactions between land use and transport systems. Considering enormous transformation of the city is taking place, it is no longer a good planning practice to assume that land-uses remain fixed when making transportation investment decisions and evaluating the impacts of major transport projects. In this regard, relevant scholars should take actions for advocating ILUTMs which should be eventually enforced by legislation, e.g., Clean Air Act Amendment in the USA. Only in this way, people will take modeling seriously and make informed decisions based on modeling results.

3.2 Institutional barrier

Institutional rather than technical barrier dominates. Relevant city models often become monopoly and black boxes in cities. Land-use modelers usually cannot access the information from transport models and similarly for transport modelers to access land-use models, which makes it difficult to integrate these

two models or entails work for replicating models again. For the sake of a better modeling community, we suggest that models be transparent and open to public. People should have access to information and data. It should also be made clear to all stakeholders that staying behind is dangerous.

3.3 Distrust from engineers and politicians regarding ILUTMs

The distrust from engineers and politicians regarding ILUTMs makes it hard to reach a consensus for building an integrated model in the first place. Due to weak modeling culture, politicians, engineers and even transport modelers doubted the usefulness of integrated models and successfulness of developing such a “complicated” model based on limited data and other resources (e.g., funding and staff). In such a case, the PECAS team had to develop a Demo model to demonstrate that it is feasible to develop such a model within a reasonable time and budget. It should be clear that modeling is not about to provide an accurate forecast of the future, but a tool for understanding relative impacts of a range of policies and corresponding actions. A set of assumptions exist behind every type of models.

3.4 Lack of human resources required for ILUTM work

Lack of human resources required for ILUTM work at both universities and planning agencies is a hard fact the modeling community has to face. The development of an ILUTM requires expertise in various subjects, such as transportation planning and modeling, land-use planning, geography, computer science, statistics and economics. Dedicated education and training programs are needed. Relevant universities and institutions should take the responsibility of hosting ILUTMs.

3.5 Limited resources and urgency of decision making

The development of ILUTMs requires a long-term commitment and investment, but limited resources and urgency of decision-making make the efforts of modeling suspicious. Due to the shortage of time, budget and data, it is often impossible to develop a full-fledged, operational model at once. Therefore, we suggest a long-term incremental and short-term agile approach be taken for the development of ILUTMs. For example, a three-stage development strategy is proposed to the City of Wuhan, with which a PECAS Demo model, a Preliminary model and an Operational model will be developed over an incremental process through time. Through the designed incremental process, more disaggregated activities, goods, labors, land and space categories and higher spatial and temporal resolution will be introduced and represented inside the model. The short-term agile approach splits a large project into several smaller modules that could be completed in a short period of time like two to three weeks, which allows for flexible planning of the development of different modules according to available time, fund, data and staff. The design diagrams of Demo models for Shanghai PECAS and Wuhan PECAS are shown in Fig.2 and Fig.5 respectively. Demo models consider limited categories of activities, commodities, labors, and land and space categories. The design diagram of a Preliminary model (Fig.16) and an Operational model (Fig.17) for Wuhan PECAS are also developed, which will be developed over an incremental process through time. It can be seen that through the designed incremental process, more disaggregated activities, goods, labors, land and space categories and higher spatial and temporal resolution are adapted into the models. Compared to the Demo model, the Preliminary model will include finer level of sectors in the first, secondary and tertiary industry. Specific improvements are: 1) Expanding 14 socioeconomic activities into 26 sectors plus six types of households (in total, 32 activities). Most of the additional activities are from the tertiary industry–service sectors and by splitting urban/rural households into 6 finer categories (small, medium and large) based on their size; 2) Adding the “slow traffic mode,” which

includes walk, bike and electronic bike, into the transport model; 3) Replacing the existing 13 products and services with 27; 4) Using a more elaborate 7 types of labor instead of the original 5 categories, basically by splitting “Office employment” into “Management,” “Professional/Technical” and “Office Support” Labor. The eight land and space categories are maintained within the Preliminary PECAS model. In terms of spatial and temporal resolution, the initial design of the Wuhan PECAS Demo and Preliminary model is to divide the whole City of Wuhan into 690 traffic analysis zones (TAZs) and 147 land-use zones (LUZs). With one year as the update step size for the land-use module, and five-year as an update step size for the transport module. The TAZs used are consistent with the current travel model in Wuhan and the LUZs used are developed largely based on the street blocks used in the census.

The representation of the socioeconomic activities, products, services, labor, floorspace, and land categories of the above Wuhan PECAS Preliminary model is largely in the similar order with several PECAS operational models. However, the modeling team suggest that the following improvements be made to the Preliminary model, in order to provide a “powerful” Operational model: 1) Adding the consideration of consumption of the natural resources (water, minerals, fuel) of the socioeconomic activities and their environmental impacts (such as water, carbon dioxide emissions), into the model; 2) Expanding 6 types of families into 18 categories according to urban/rural status, household income (high, medium, low) and household size (large, medium, small); 3) Explicitly consider the following economic flows: the central and local financial, investment and rent, and taxes and financing. Considering the spatial and temporal resolution, the Operational model may use 1100 TAZs covering the entire city with most of the newly added TAZs in the rural areas for better representation of these parts of the city. The number of LUZs will also be increased from 147 to about 200 to 300. The land use and the transport module will be run and updated each year over the next 30 years or so, in order to support the planning analyses under the reality that unprecedented infrastructure development is taking place in the City of Wuhan (Completing 11 subway lines over the next 6 years and the overall mileage of the Wuhan subway network will be 400 km by 2021).

be carried out and relative results can be borrowed. Moreover, detailed production/consumption data by sectors can be compiled through production data from various sources. However, due to lack of appropriate data, the choice of commodity, space occupation categories had to be set at a much aggregated level within the Wuhan PECAS Demo model, with 13 types of commodities, 5 types of labor and 8 types of space. Efforts are underway to develop “better I/O tables” using more advanced non-survey methods, such as “Best Practice Firms” proposed by Miernyk (1965) and “RAS” procedure (also known as a “biproportional” matrix balancing technique) (see Stone, 1961: and Bacharach 1970), which will be used in our operational models.

3.7 Scissor differences between rural and urban areas

Significant “scissor differences” of socioeconomics and corresponding behaviors between rural and urban households in typical municipalities in China are found to exist and have an important implication to urban land use and transportation planning policies, which are usually targeted to mitigate such differences. In the current 690 TAZs of the City of Wuhan within which 399 urban-zones and 291 rural-zones are identified. A household travel survey for collecting socioeconomics and travel behavior data of urban and rural households of the city was carried out in 2008. This data set is used to study the “scissor differences” between rural and urban households in terms of their income, size, space and vehicle consumption and trip frequency/distance. Our study results show that “scissor differences” are statistically significant in nearly all of the above aspects. Table 2 provides the mean, minimum, maximum, range, Standard Deviation (SDV) and Covariance (COV) of household size (HS) (unit: number of family members), building area (BA) (unit: m²), household income (HI) (unit: 10,000 yuan), highest household mobility (HHM), and travel distance (TD) (unit: km) of the urban and rural households. The highest household mobility refers to 7 grades which represent nil vehicle (0), bicycle (1), electric bicycle (2), motorcycle (3), other vehicles (4), private car (5) and company car (6). Other vehicles refer to truck, carriage, tractor and so on. In Table 2, the mean of HS, BA and HHM of rural households are higher than that of the urban households, while the mean of HI and TD of rural households are lower than those of the urban households. The rural households have a higher COV in HI, HS, and TD, but lower COV in BA and HHM, which indicates that the rural households show a larger variation in the first three aspects, but a smaller variation in the last two aspects than the urban households. The p-values for comparing household size, household income, building area consumed, highest household mobility and travel distance between urban and rural households through a set of Mann-Whitney U tests (a nonparametric paired test) are provided in the last column of the table. It is clear that the two household groups show significant differences for all variables with a p-value of 0.000 in each case, except for travel distance. A p-value of 0.157 was reported for comparing the travel distance between the rural and urban households, which indicates that the hypothesis that the mean of travel distance of the two types of residents is the same cannot be statistically rejected. However, the outliers and higher variation (COV is 67.5 for the urban residents vs. 69.9 for the rural) could result in the same conclusion. The histogram of travel distance by rural and urban residents is shown in Fig.18. As can be seen from the figure, there is an obvious difference of travel distance between rural and urban households.

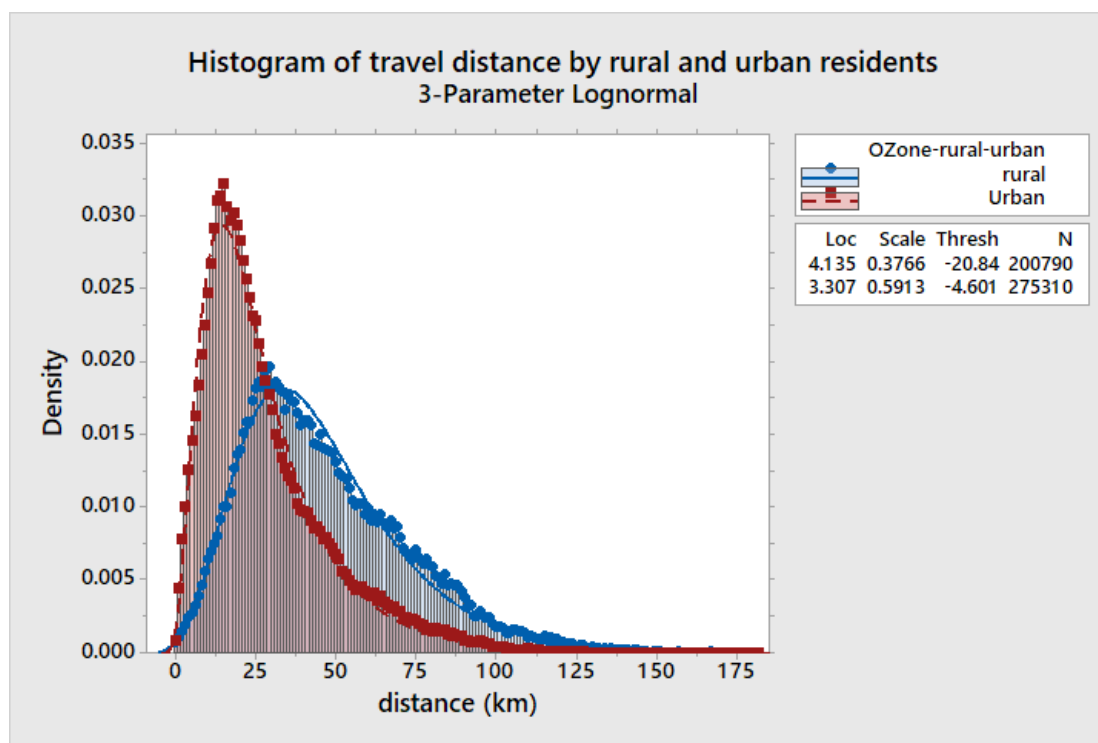


Figure 18: Distribution of travel distance by rural and urban residents

A further examination was also carried out by further dividing rural/urban households into 9 categories (categorized by three household income levels and three household sizes) and it is found that many subgroups began to show significant differences in their travel distance (Tang, 2017).

Table 2: Descriptive statistics and p-values of corresponding Whitney U test for urban and rural household socioeconomics, space consumption, mobility choice and travel distance

	Mean	Min.	Max.	Range	SDV	COV(%)	P-value
Urban household size	2.9	0	11	11	0.84	28.9	.000
Rural household size	3.2	1	10	9	1.1	34.4	
Urban household income (in 10,000 yuan)	5.1	0	18	18	3.8	74.5	.000
Rural household income (in 10,000 yuan)	4.2	0	18	18	3.5	83.3	
Urban building area (in m ²)	71.8	6	255	249	36.8	51.3	.000
Rural building area (in m ²)	108.8	10	250	240	44.7	41.1	
Urban highest household mobility	1.2	0	6	6	1.4	116.7	.000
Rural highest household mobility	1.8	0	6	6	1.5	83.3	
Urban travel distance (in km)	27.7	0.01	90.2	90.2	18.7	67.5	.157

As there are significant scissor differences of socioeconomics, travel and space consumption behavior between the rural and urban households in the City of Wuhan, such differences should be explicitly considered in integrated modeling exercises by separately setting up urban and rural household groups. Indeed, as the number of household categories considered within the model increases, the amount of

data required and parameters to be estimated for developing a similar PECAS model would increase in an exponential manner. For example, currently, only urban and rural households are considered in the model, but it can be extended into a cross classification considering household urban/rural designation (2 classes), income (3-classes: low, medium and high), size (3 classes: small, medium and large), and car ownership (2 classes), with a total number of household classes of 36. In many cases, under such a household classification plan, there would not be enough data to estimate corresponding household parameters for some underreported households.

To further discuss the implications of different household classification schemes to integrated modeling, different household classification schemes are designed for different models according to urban/rural and different categories of HS and HI combinations (as shown in Table 3) and compared about the model accuracy (Mean Square Error, MSE) and complexity (data demand, DD).

The following within-group mean square error (WG-MSE) is used to indicate model accuracy:

$$S^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2 + \dots + \sum_{i=1}^{n_n} (y_{in} - \bar{y}_n)^2}{n_1 + n_2 + \dots + n_n - n} = \frac{SSE}{n_1 + n_2 + \dots + n_n - n} \quad (1)$$

Where

n_n ---- The number of samples in group n;

y_{in} ---- The sample value in group n;

\bar{y}_n ---- The average of the samples in group n;

SSE---- Sum of Squares for Error.

The input data required for developing a full land-use transport model based on the PECAS framework, under different household classification schemes, is calculated as the following:

$$D = n_g \times (n_s + n_t) \quad (2)$$

Where

n_g ----Number of household groups after classification;

n_s ---- Number of groups of household socioeconomics data;

n_t ---- Number of groups of household travel data.

The MSE represents how much variation is left out when using an average as a modeling parameter for describing the behavior of a given class of household. In general, the lower it is, the better the modeling accuracy. On the other hand, the more the household categories are, the more data are required for household outputs and consumption variables, increasing the complexity of the model. We define the DD (data demand) indicator (Eq.(2)) to represent the complexity of the model.

Table 3: Six household classification schemes for PECAS modeling

Model	Number of household categories	Description
Model-1	2	Urban/Rural
Model-2	4	Urban/Rural with household size (small/large)
Model-3	6	Urban/Rural with household size (small, medium and large)
Model-4	8	Urban/Rural, household size (small and large) and household income (low and high)
Model-5	12	Urban/Rural, household size (small, medium and large) and household income HI (low and high)
Model-6	18	Urban/Rural, household size (small, medium and large) and household income (low, medium and high)

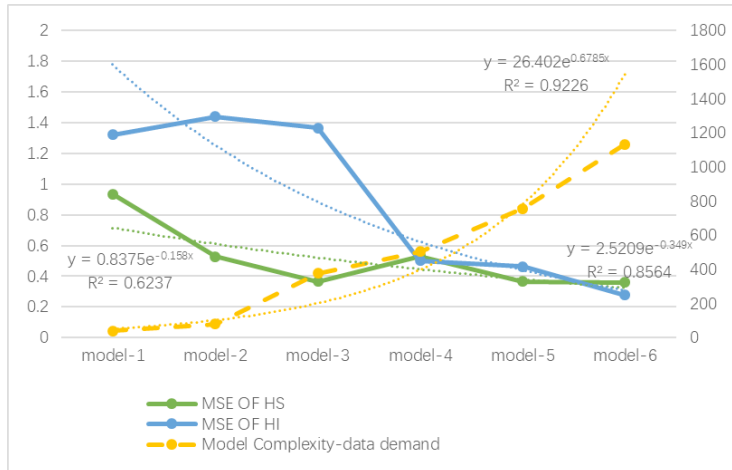
The results of MSE and DD calculations by Eq. (1) and (2) for different household classification schemes are shown in Fig.19 (a) and Fig.19 (b). The x-axis in Fig.19 (a) represents the 6 models with 6 schemes of household classifications we defined. From model 1 to model 6, the household categories increase from 2 to 18. The left y-axis shows the MSE of grouping variables like household size and household income; the right y-axis shows the data required for the construction of the model.

In Fig.19 (a), the blue solid line represents the MSE variation curve of household income from model 1 to model 6, the green solid line stands for the MSE variation curve of household size and the yellow dashed line shows the variation curve of data demand. It can be observed that from model 1 to model 6, with the increase of household categories, the MSE curves for household size and household income generally decline, whereas data demand curve rises. For example, in model 1, data required is 40 and the MSE of household size is 0.93; in model 2, data required is 80 and the MSE of household size is 0.53; in model 3, data required is 378 and the MSE of household size is 0.36; in model 4, data required is 504 and the MSE of household size is 0.53; in model 5, data required is 756 and the MSE of household size is 0.36; in model 6, data required is 1134 and the MSE of household size is 0.35. It is quite understandable the coarse classification of household into urban and rural in model 1 generates the highest MSE. With more categories, the MSE is lower. Model 3 has one more category of household size (medium size) than model 2. In model 2, the small-size household has less than 5 family members while the large-size household has 5 or more family members; while in model 3, the small-size household has less than 3 family members, the medium-size household has 3 to less than 7 while the large-size household has 7 or more family members. The MSE of household size reduces from 0.53 for model 2 to 0.36 for model 3.

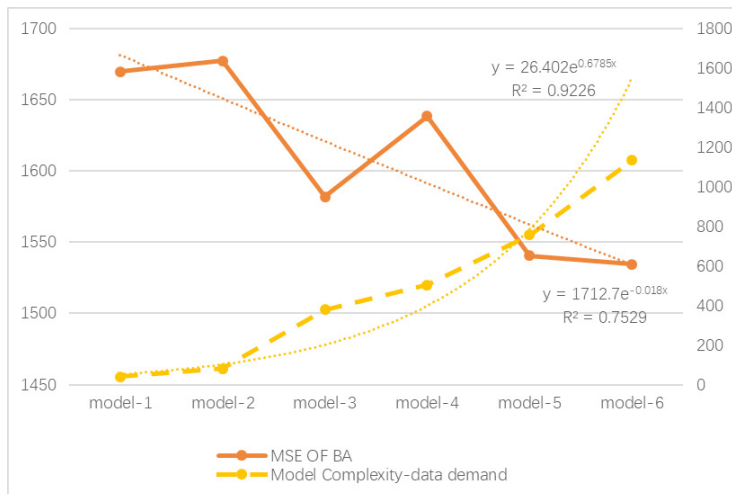
The x-axis of Fig.19 (b) also represents the 6 models with 6 different household classification schemes. The left y-axis shows the MSE of building area and the right y-axis represents the complexity of the model (data demand, DD). The MSE of building area and data demand is compared in Fig.19 (b), which demonstrates a similar general declining trend as in Fig.19 (a). However, the MSE of building area displays an obvious declining and fluctuation trend.

The fitting curves and corresponding formula for each variable are also shown in Fig.19 (a) and Fig.19 (b). In Fig.19 (a), it can be seen that the MSE of household income has a steeper declining trend than that of household size, which suggests that household income thresholds used for classifying households into different income groups may play a more important role in determining whether the members from different household groups would show a similar or different behavior.

It appears that the accuracy of the PECAS model improves with more divided household categories. Meanwhile, the data demand of the PECAS model increases, especially sharply from model 4, which adds huge amount of workload in modeling practice. This suggests that model 3 is the best option with the tradeoff between accuracy and data demand in our case.



(a) MSE of household income and household size and data demand trend curves



(b) MSE of building area and data demand trend curves

Figure 19: Comparison of MSE and data demand for different models

The above results show that the scissor differences between rural and urban households in cities of China would affect model results in a very significant manner if such differences were not well considered and structured within the model. The study results above clearly show that such differences are statistically significant and this fact, in turn, calls for a refined treatment of household categorization inside an ILUTM, with at least rural/urban households classified by some socioeconomic variables (such as income and size) being included. This is going to be a salient difference from those applications in North America and other developed countries.

The PECAS framework or any other ILUTM framework should be able to represent such differences easily, but structural challenges may arise, as the amount of data required will increase in an exponential manner when the number of the household categories is doubled or tripled.

3.8 Lack of detailed land-use/space data

The problem of lack of detailed land-use/space information at LUZ level, is common even in major cities in China. In most ILUTM applications, floorspace data are obtained simply by transforming household and employment census data into floorspace via conversion factors. To tackle this problem, we suggest that various technologies, such as remote sensing (LiDAR + Optical, as shown in Fig.20 and Fig.21), online mapping (land-use function) and GIS data (Points of Interest (POI); Mobile Phone Location data) be applied to extract land use and population/employment information. LiDAR is a relatively new, active (in contrast to optical which is passive in nature) remote sensing technology, which can be used to produce digital elevation models and digital surface models in a timely fashion. The preprocessed LiDAR data require filtering and classification prior to building and floorspace data extraction. A set of tools (LAStools) is used to classify LiDAR point clouds into three main land-cover groups, namely buildings, ground and vegetation. When the extraction software identifies the building points, boundaries are drawn using polygons to represent building footprints. The next step is to calculate the amount of floorspace from the LiDAR-extracted building layer by category and by zone. The estimated floorspace in each zone is obtained by multiplying the number of floors (based on the building height and a floor height estimate) by the building footprint. The land-use category of each building is identified by overlaying a parcel layer. According to our study results (Shiravi, Zhong, BeyKaei, Hunt, & Abraham, 2015), in general, the LiDAR-based method tends to slightly underestimate floorspace, due to problem with data density and overhanging trees; nevertheless, such a method can produce fairly accurate building representations and floorspace estimates compared to the current census data-based method. Furthermore, a land use extraction expert system as shown in Fig.22 has been proposed to extract land-use information at a very disaggregate spatial scale (Beykaei, 2013). As can be seen from the figure, different data sources will be used in the land use extraction process. Results of land cover classification including parking lots, vegetation area, and bare soil are essential to the morphological analysis of urban features. Geographic data and land cover extracted are used to examine and recognize different land-use patterns at Dissemination Block (DB) and also parcel level. Different LU classification tools, including Fuzzy-Decision Tree (FDT), Logistic-Decision Tree (LDT), and Artificial Neural Network (ANN), are used in the morphological analysis in order to extract LU information at the DB-level. At this level, only morphological analysis is applied, as the spatial arrangement analysis to the relative locations of DBs with different LUs will not help extract any more detailed LU information. The LU extraction process at the parcel-level continues to the spatial arrangement analysis to develop a more sophisticated LU extraction expert system. In this regard, misclassifications occurring in the morphological analysis at the parcel-level are further corrected through spatial arrangement analysis. The accuracies of the final results from the LU extraction system at each analysis level (DB or parcel) and at each step (Morphological or spatial arrangement analysis) are assessed to evaluate the utility of the LU extraction expert system developed.

Classification results

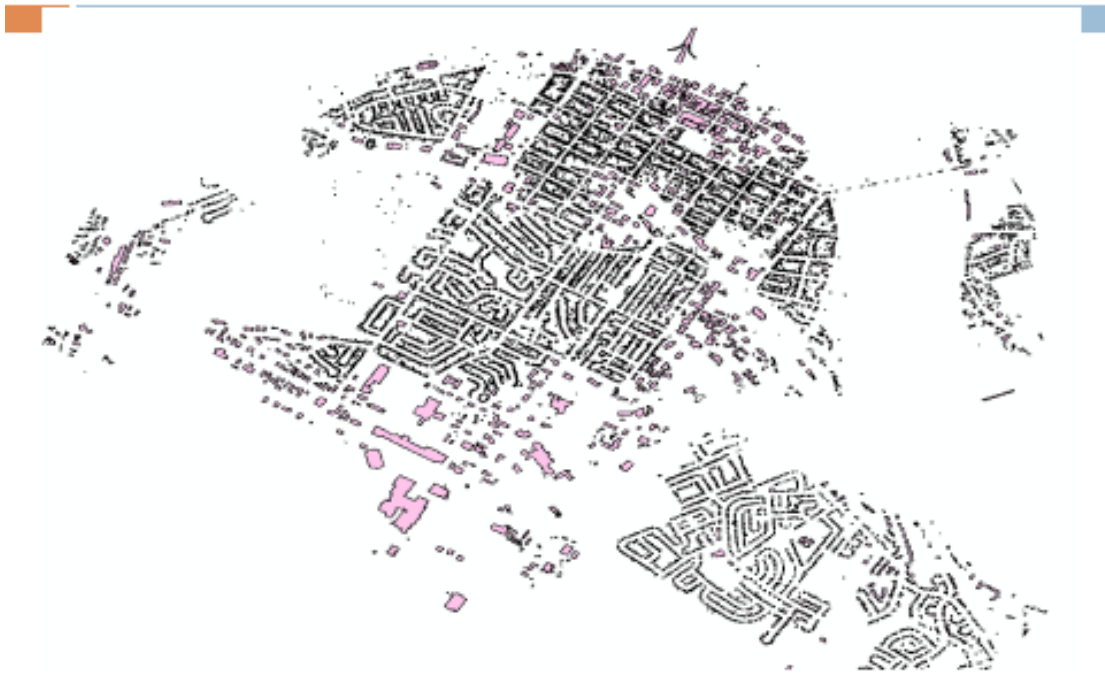


Figure 20: Extraction of buildings using LiDAR



Figure 21: Extraction of other land covers (roads and vegetation) using optical images

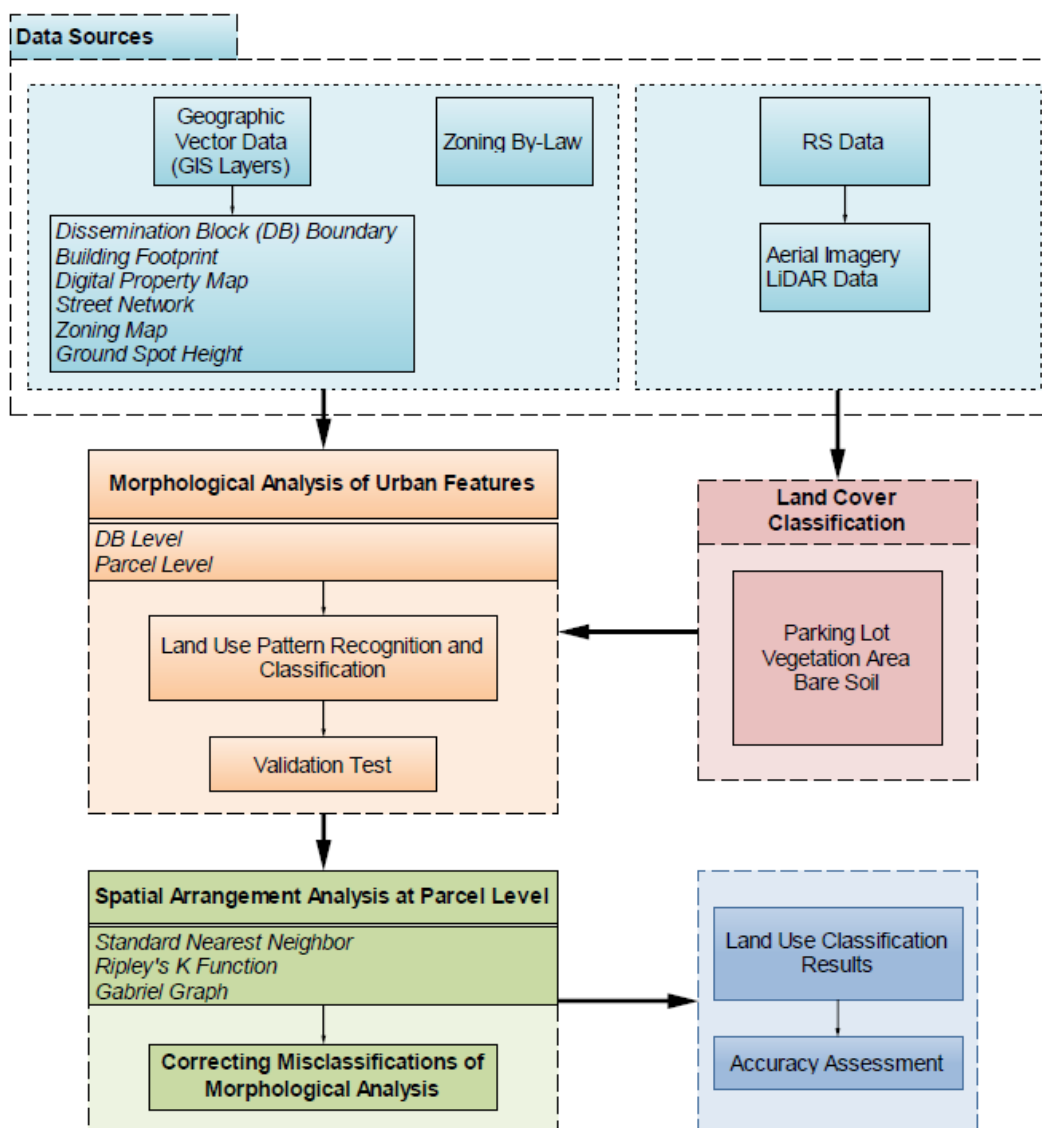


Figure 22: Flow chart of a land use extraction expert system

As to the floorspace rent and price information by the different floorspace types that is critical to the land-use calibration, we found there was no readily attainable such information. For the city of Wuhan, we had to resort to the annual reports of real estate analysis companies to collect information. The fact is that dwelling house price is the most available from the websites of real-estate agencies, next is the price/rent of office buildings, followed by the price of commercial space. There is less price/rent information on other types of floorspace, such as industrial/ governmental floorspace and warehouse. Great efforts have to be made to search and screen such information.

3.9 Lack of population/employment/land-use data at disaggregate level

The problem of lacking population/employment/land-use data at disaggregate zonal level can be mitigated by referring to census and economic survey, if they are available. However, the spatial separation between operation (blue-collar) and management (white-collar) can only be determined after the survey

practice is changed (no longer log spatially separated factories and their headquarter as one business unit, but independent ones) or dedicated data analysis is carried out. On the other hand, population, employment and space data at disaggregate levels (e.g., parcels or street blocks) can be synthesized and optimized against the targets at aggregate spatial level where such data are available (Abraham, Stefan, & Hunt, 2012), as shown in Fig.23. Possible data sources include population/employment data from census/yearbooks, mobile phone location data, and floorspace data extracted from remotely sensed imagery. They can be cross-checked for their validity. However, mixed floorspace use, which is a common practice in China, still poses a big barrier. A study about the mixed floorspace use of the city is therefore warranted and should be carried out.

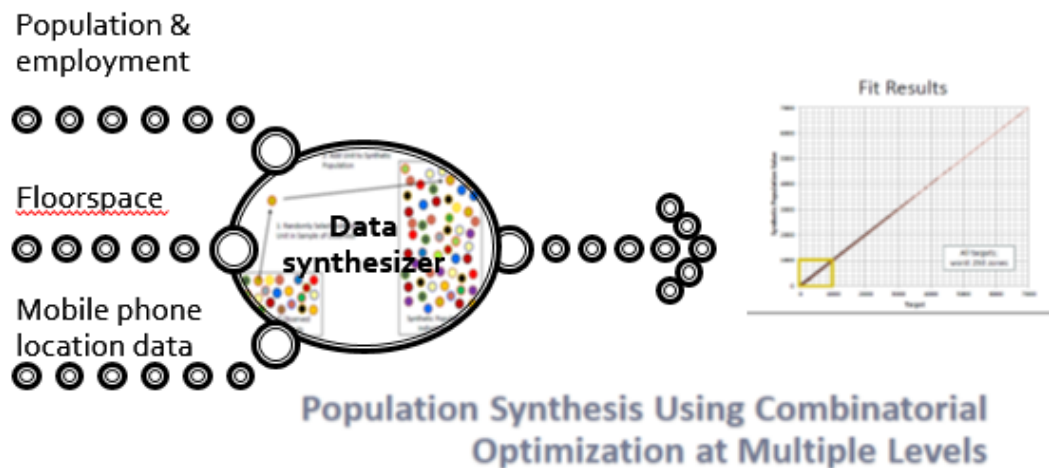


Figure 23: Schematic data synthesizer diagram

3.10 Lack of freight transport data

Lack of freight transport data, logistics flow, and trip lengths for commodities transported by different modes/vehicles of different sizes is another challenge. Flows of exchanges determined in the PECAS model from production to exchange zones and from exchanges zone to consumption zones are allocated according to exchange prices and transport (dis)utilities. These flows need to be converted to transport demands in terms of loaded trucks (and empty ones after unloaded) which will be loaded to transport networks in order to determine congested travel dis(utilities). Actually, abundant trajectory data for trucks exist, but there are not much data about what commodity a vehicle actually carries. It appears that such data are either business secret of sellers or not-a-must information to be collected by logistics operators, as the latter care most about delivering their serviced loads on time, rather than inventorying all commodities transported (they mostly categorize the goods transported into “light” and “heavy”). One possible solution is borrowing initial parameters from logistics literature, and the other is carrying out small scales of freight transport survey, the latter of which may entail a certain amount of cost and work.

3.11 Data inconsistency

Data inconsistency need to be dealt with carefully. For as complex a system as an integrated model is, large amount of data of different formats related to land use, economic activities and transportation are involved. For example, a large number of data items have been assembled for the development of Wu-

han PECAS model, including census population/employment (2010), household travel survey (2008), Yearbook (2008 onwards), input-output tables (2007 onwards), land use (current and zoned), and space quantities at TAZ level (2012), space rent/ sale price, household financial survey results, and summary statistics from Wuhan Travel Model (including time and distance skims, logsum of mode choices at TAZ level and trip length distribution by mode and trip purposes). These different types of data tend to disagree with each other. We have to use various methods to judge and cross check the validity of data from different sources. For example, when we estimated the labor consumption coefficients for different industries, we found there was inconsistency in the classification of labors in yearbooks for different years. For the years from 1994 to 1998, the labor consumptions for mining, manufacturing, electricity, gas and water production and supply, and construction were provided as labor types (technical staff, administrative staff, service staff and other). However, from 1998 on, the labor consumption by different industries are offered according to the education level of labors.

3.12 Complexity of developing, calibrating, validating and using ILUTMs

An ILUTM simulates the socioeconomic activities, land use and transport system of the whole city (or a region), which requires specialists of economy, transportation and modeling to work together. Even though many integrated models are available now as commercial software packages, the development, calibration and validation of the models are still daunting tasks. Therefore, as mentioned above, special training of engineers and students to understand ILUTMs should be provided, in the form of graduate courses, research programs or workshops. Furthermore, the model developer and relevant modeling team should develop a full version of calibration manual and at least automate the process, thus to make it easier for people to develop and use ILUTMs.

4 Conclusions and recommendations

In this paper, the challenges and issues in designing and developing three PECAS models in China, namely Shanghai PECAS, Wuhan PECAS and Guangzhou PECAS are presented and corresponding mitigating solutions from the development team are given. The discussions presented in this paper may be able to help guide model design, data collection and model development in developing countries. Based on our experiences with the development and implementation of the PECAS models in China, we conclude that:

1. Modeling should be promoted through all means (education, media and legislation). More investment should be put in education and training related to integrated modeling.
2. Long-term incremental and short-term agile development approach seems to be a good solution, which works with issues related to culture, institution, funding, and staffing.
3. Data issues were seen as the largest obstacle for developing such a model at first. However, the experiences to date indicate that the PECAS framework is very flexible and it can work with most spatial data collected, either aggregated or disaggregated. Of course, the aggregation effect costs the modeling accuracy, but the framework is flexible enough to allow the model development based on aggregated data, which will then be improved over time by engaged agencies and staff. Our experiences indicate that data actually is no longer an obstacle, even in the context of developing countries like China.
4. Integrated modeling requires coherent vision, plan and actions from researchers, practitioners and decision-makers. The modeling community should work together to continue improving this coherence.
5. More detailed research into the issues of scissor differences, information extraction and data synthesis are needed for improving integrated modeling.

Acknowledgements

The authors would like to thank Wuhan Transportation Planning Institute (WHTPI) for their initiatives in integrated land-use/transport modelling. This study is supported by a research fund from the WHTPI, a start-up grant from Wuhan University of Technology and a research fund from National Natural Science Foundation of China (51778510).

References

- Abraham, J. E., Gordon, R., & Hunt, J. D. (2005). *The Sacramento PECAS model*. Paper presented at the Transportation Research Board 84th Annual Meeting, Washington, DC.
- Abraham, J. E., Stefan, J. K., & Hunt, J. D. (2012). *Population synthesis using combinatorial optimization at multiple levels*. Paper presented at the Transportation Research Board 91st Annual Meeting, Washington, DC.
- Abraham, J. E., Hunt, J. D., & Wang, W. (2014). *Generating PECAS base year built form for Clayton County in Atlanta*. Paper presented at the 5th TRB Conference on Innovations in Travel Modeling, Baltimore, MD.
- Bacharach, M. (1970). *Biproportional matrices and input-output change*. Cambridge, UK: Cambridge University Press.
- Batty, M. (1979). Progress, success and failure in urban modeling. *Environment & Planning A*, 11(8), 863–878.
- Beykaei, S. A. (2013). *Development of a hybrid GIS/RS expert system for extracting subzonal land-use information*. PhD Dissertation, University of New Brunswick, Fredericton, NB.
- Boyce, D. E., Chon, K. S., Ferris, M. E., Lee, Y. J., Lin, K.T., & Eash, R. W. (1985). *Implementation and evaluation of combined models of urban travel and location on a sketch planning network*. Paper presented at the Transportation Research Board 64th Annual Meeting, Washington, DC.
- de la Barra, T. (2012). Mathematical description of TRANUS. Retrieved from <http://www.modelistica.com>
- Echenique, M. D., & Lindsay, W. (1969). A spatial model of urban stock and activity. *Regional Studies* 3(3), 281–312.
- Echenique, M., Flowerdew, A., Hunt, J., Mayo, T., Skidmore, I., & Simmonds, D. (1990). The MEPLAN models of Bilbao, Leeds and Dortmund. *Transport Reviews*, 10(4), 309–322.
- Fuenmayor, G. (2016). *Building a spatial economic model for Caracas using PECAS*. Ph.D. dissertation, Department of Civil Engineering, University of Calgary, Calgary, AB.
- Gao, S., Lehmer, E., Wang, Y., McCoy, M., Johnston, R. A., Abraham, J. & Hunt, J. D. (2010). *Developing California integrated land-use/transportation model*. Paper presented at the Transportation Research Board 89th Annual meeting, Washington, DC.
- Hunt, J. D., & Simmonds, D. C. (1993). Theory and application of an integrated land-use and transport modelling framework. *Environment and Planning B*, 20, 221–244.
- Hunt, J. D., Johnston, R., Abraham, J. E., Rodier, C. J., Garry, G. R., Putman, S. H., & de la Barra, T. (2001). Comparisons from Sacramento model test bed. *Transportation Research Record*, 1780, 53–63. doi: <http://dx.doi.org/10.3141/1780-07>
- Hunt, J. D., & Abraham, J. E. (2003). *Design and application of the PECAS land-use modelling system*. Retrieved from http://www.scag.ca.gov/Documents/Design_Application%20PECAS.pdf
- Hunt, J. D., Kriger, D. S., & Miller, E. J. (2005). Current operational urban land-use transport modelling frameworks: A review. *Transport Reviews*, 25(3), 329–376.
- Hunt, J. D., Abraham, J. E., De Silva, D., Zhong, M., Bridges, J., & Mysko, J. (2008). *Developing and applying a parcel-level simulation of developer actions in Baltimore*. Paper presented at the Transportation Research Board 87th Annual Meeting, Washington, DC.
- Hunt, J. D., & Abraham, J. E. (2009). PECAS theoretical formulation. Retrieved from <http://www.hbaspecto.com/pecas/downloads/>
- Hunt, J. D., Weidner, T., Knudson, R., Bettinardi, A., & Wardell, E. (2010). *Applying the SWIM2 integrated model in freight planning in Oregon*. Paper presented at the Transportation Research Board 90th Annual Meeting, Washington DC.

- Iacono, M., Levinson, D., & El-Geneidy, A. (2008). Models of transportation and land-use change: A guide to the territory. *Journal of Planning Literature* 22, 323–340.
- Jin, Y., Echenique, M., & Hargreaves, A. (2013). A recursive spatial equilibrium model for planning large-scale urban change. *Environment and Planning B: Planning and Design*, 40, 1027–1050.
- Jin, Y., Denman, S., Deng, D., Rong, X., Ma, M., Wan, L.,...Long, Y. (2017). Environmental impacts of transformative land use and transport developments in the Greater Beijing region: Insights from a new dynamic spatial equilibrium model. *Transportation Research, Part D, Transport and Environment*, 52, 548–561. doi: <http://dx.doi.org/10.1016/j.trd.2016.12.009>
- Long, Y. & Shen, Z. J. (2015). *Geospatial analysis to support Urban Planning in Beijing*. New York: Springer.
- Lowry, I. S. (1964). *A model of metropolis*. Santa Monica, CA: Rand Corp.
- Mao, J., & Yan, X. (2004). Corridor effects of the urban transport artery on land use — a case study of the Guangzhou Avenue. *Geography and Geo-Information Science*, 20(5), 58–61. (in Chinese)
- Mao, J., & Yan, X. (2005). A study to the impacts on urban land use of transport system in highly-densely developed city — a case study of Guangzhou. *Economic Geography*, 25(2), 185–188, 210. (in Chinese)
- Miernyk, W. H. (1965). *The elements of input-output analysis*. New York: Random House.
- Miller, E. J., & Salvini, P. A. (2001). The integrated land use, transportation, environment (ILUTE) microsimulation modelling system: Description and current status. In D. Hensher & J. King (Eds.), *Travel behavior research: The leading edge* (pp. 711–724). Amsterdam: Elsevier.
- Miller, R. E., & Blair, P. D. (2009). *Input-output analysis: Foundations and extensions* (Second Edition). Cambridge, UK: Cambridge University Press.
- Mir, M. S, Krishna Rao, K. V., & Hunt, J. D. (2010). Space development modeling of urban regions in developing countries. *Journal of Urban Planning and Development*, 136(1), 75–85.
- Pan, H. (2007). Rail transit and optimization of spatial structure in metropolitan. *Shanghai Urban Planning Review*, 06, 37–43. (in Chinese)
- Pan, H., & Zhang, M. (2008). Rail transit impacts on land use. *Transportation Research Record*, 2048, 16–25. doi: 10.3141/2048-03.
- Pan, H., & Chen, G. (2009). Rail transit and residential location choice. *Urban Planning Forum*, 5, 71–76. (in Chinese)
- Putman, S. H. (1983). *Integrated urban models*. London: Pion.
- Putman, S. H. (1991). *Integrated urban models 2*. London: Pion.
- Putman, S. H. (1994, January). *Integrated land use and transportation models: An overview of progress with DRAM and EMPAL, with suggestions for further research*. Paper presented at the Transportation Research Board 73rd Annual Meeting, Washington, DC.
- Putman, S. H. (1996). Extending DRAM model: Theory–practice nexus. *Transportation Research Record*, 1552, 112–119. doi: <http://dx.doi.org/10.3141/1552-16>
- Shi, J., Tong, X., Zhang, H., & Tao, D. (2012). Building “new towns” from industrial zones: An Urban-Sim application in Yizhuang, Beijing. *Urban Studies*, 19(02), 98–10, 24. (in Chinese)
- Shiravi, S., Zhong, M., Beykaei, S. A., Hunt, J. D., & Abraham, J. E. (2015). An assessment of the utility of LiDAR data in extracting base-year floorspace and a comparison with the census-based approach. *Environment and Planning B: Planning and Design*, 42, 708–729.
- Stone, R. (1961). *Input-output and national accounts*. Paris: Organization for Economic Cooperation and Development.
- Tang, Q. (2017). *Scissor difference of socioeconomics travel and space consumption behavior of rural and urban households and its impact on modeling accuracy and data demand of integrated models*. Master's

- thesis (in Chinese). Wuhan University of Technology, Hubei Sheng, China.
- Tobben, J. (2017). On the simultaneous estimation of physical and monetary commodity flows. *Economic Systems Research*, 29(1), 1–24. doi: 10.1080/09535314.2016.1271774
- Waddell, P. (2000). A behavioral simulation model for metropolitan policy analysis and planning: Residential location and housing market components of UrbanSim. *Environment and Planning B*, 27(2), 247–264.
- Waddell, P., Peak, C., & Caballero, P. (2004). *UrbanSim: Database development for the Puget Sound region*. Seattle: Center for Urban Simulation and Policy Analysis, University of Washington.
- Wegener, M. (2004). Overview of land use transport models. In D. A. Hensher & K. J. Button (Eds.), *Transport geography and spatial systems* (pp.127–146). Kidlington, UK: Pergamon/Elsevier Science.
- Weidner, T., Hunt, J. D., Abraham, J. E., Moeckel, R., Gregor, B., & Knudson, B. (2008). *Final calibration of the Statewide Oregon2TM integrated land use-transport model*. Paper presented at the Transportation Research Board 87th Annual Meeting, Washington, DC.
- Zheng, S., & Huo, Y. (2012). Spatial integrated modeling of Beijing's office market: Model calibration and simulations based on UrbanSim. *Urban Studies*, 19(02), 116–124. (in Chinese)
- Zhong, M., Hunt, J. D., & Abraham, J. E. (2007). Design and development of a statewide land-use transport model for Alberta. *Journal of Transportation Systems Engineering and Information Technology*, 7(1), 79–89.
- Zhong, M., Hunt, J. D., Abraham, J. E., Li, Y. Q., Li, J. Z., Yang, W., & Wang, X. Z. (2016). *A summary of design and development strategies of Wuhan PECAS model*. Paper presented at the World Conference on Transport Research (WCTR), Shanghai, China.