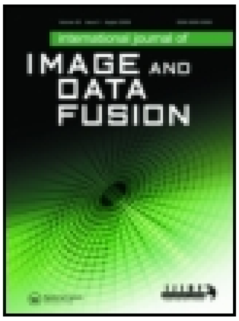

International Journal of Image and Data Fusion

ISSN: 1947-9832 (Print) 1947-9824 (Online) Journal homepage: http://www.tandfonline.com/loi/tidf20

Transformations of urban studies and planning in the big/open data era: a review

Ying Long & Lun Liu

INTERNATIONAL JOURNAL OF IMAGE AND DATA FUSION, 2016
http://dx.doi.org/10.1080/19479832.2016.1215355

REVIEW ARTICLE

Transformations of urban studies and planning in the big/open data era: a review

Ying Long[a] and Lun Liu[b]

[a]School of Architecture, Tsinghua University, Beijing, China; [b]Department of Land Economy, University of Cambridge, Cambridge, UK

ABSTRACT
As the emergence of big/open urban data in recent years, there have been lots of transformations going on in urban study and planning. In this article, we aim to provide a review of the major trends in this round of transformation. The review is started with a brief introduction to the new data environment, which has been made possible by the availability of big data and open data in recent years, as well as a review on the research progress in China, a developing country recently embracing the technology and data boom. It is followed by an analysis on the four major trends in quantitative urban study, supported by typical research cases. The four trends are (1) transformation in spatial scale from high resolution but small coverage or wide coverage but low resolution to wide coverage with high resolution, (2) transformation in temporal scale from static cross-sectional to dynamic consistent, (3) transformation in granularity from land-oriented to human-oriented and (4) transformation in methodology from conventional research group to crowdsourcing. The paper also points out that quantitative urban research is faced with problems like data bias, lack of long-term analysis, lack of linkage to planning practice etc.

ARTICLE HISTORY
Received 17 March 2016
Accepted 8 July 2016

KEYWORDS
Big data; open data; urban planning; mega-model; crowdsourcing

1. Background

As the fast development and popularity of Information and Communication Technology (ICT) and Internet of Things, there has been a boom of urban data in recent years. Such data flow with large volume, high velocity and big variety is becoming a hotspot of urban studies and planning research. Moreover, local governments, private enterprises and voluntary programmes are also expanding the data sources for urban researchers. The multisource new data environment differentiate itself from conventional survey data with its high resolution, wide coverage and timeliness, which opens up important development opportunities for urban studies, planning practice and commercial consultancy in China.

As the place of the most rapid urbanisation in the world, Chinese cities have accommodated more than 700 million people and are attracting another 30 million annually. Consequently, China's urbanisation has attracted wide attention from the academia, the government, the industry and the general public. In order to achieve a

CONTACT Lun Liu ll454@cam.ac.uk Department of Land Economy, University of Cambridge, Cambridge CB2 1 RF, UK

solid and unbiased understanding of China's urbanisation, an accurate and comprehensive measurement is needed in the current situation. Meanwhile, urban big/open data have become a hotspot of Chinese urban research, which is thought to correspond well to the human-oriented 'New Type Urbanization' in China. In the past, urban data were mainly produced and maintained by government departments, and were usually non-accessible to the public as well as the academia, which presented huge obstacles for quantitative urban research. Now things have changed a lot due to both the government's awareness of the importance of open data, and the emergence of new data sources related the development of ICT. Big/open urban data have been identified as an important complement to conventional survey data and data collected by various administrative departments in understanding urban form and functions (Jia and Jiang 2010, Goetz and Zipf 2012, Li *et al*. 2014; Crooks *et al*. 2015). Furthermore, the data infrastructure in developing countries is quite insufficient comparing with developed countries, in which situation big/open urban data are even more important as a viable and cost-efficient option for collecting urban data.

Before further discussion, we would like to provide a definition of big/open urban data to be used this report. The types of data that could be categorised under this term should meet two criteria. First, they would characterise certain aspects of urban form and functions (Crooks *et al*. 2015). Second, they are openly accessible to the public. In terms of data sources, there are generally three overlapping though different sources of such data. The first are official data portals, enabled by the recent open government initiatives that grant public access to previously non-accessible data sources. The second are big data initiatives, generating data from mobile phone activities, vehicle trajectories, public transit smart card data (SCD), business catalogues, as well as other smart city programmes (Batty 2012). Such data enable researchers to capture urban dynamics at very fine spatiotemporal scales and therefore gauge urban dynamics at finer spatial and temporal scales (Kitchin 2014, Yue *et al*. 2014). The third source is Volunteer Geographic Information (VGI) and crowdsourcing (Crooks *et al*. 2015), which allows the general public to contribute to the urban data pool in a 'bottom-up' approach. Examples of such data type include collaborative VGI mapping platforms (e.g. OpenStreetMap) and geo-tagged social media applications (e.g. Foursquare, Twitter and Flickr) (Liu *et al*. 2014).

However, despite these positive changes, there are also several challenges of big/open data in China. First, though there have been efforts in opening up government data, data transparency still remains unsatisfying. Furthermore, data availability is also a problem for developing countries due to the lack of data collection infrastructure and tradition; therefore they are oftentimes less well charted comparing with European and North American cities (Lang 2011, Graham *et al*. 2014). Third, data from various scattered sources are usually not readily available in convenient, transferable and adjustable formats as we prefer. Data from different sources are usually produced and maintained in different formats (Reichman *et al*. 2011), and may overlap with each other, which is to some extent a waste of resources. Furthermore, the unprecedented speed of China's urbanisation posits a unique challenge to the speed of data update. For example, between the interval of two Chinese nationwide census in 2000 and 2010, about 200 million people had become new urban residents in China (Liu *et al*. 2014).

In response to the challenges mentioned earlier, we set up the Beijing City Lab (BCL; http://www.beijingcitylab.org), an online research network to produce and store data about Chinese cities. BCL is a virtual research community, dedicated to studying, but not limited to, China's capital, Beijing. The Lab focuses on employing interdisciplinary methods to quantify urban dynamics, generating new insights for urban planning and governance, and ultimately producing the science of cities required for sustainable urban development. The lab's current mix of planners, architects, geographers, economists and policy analysts lends unique research strength. Through the endeavour of the core research team led by Dr. Ying Long, BCL has been developing fast and steadily and drawing vast attention from the urban planning community both in China and overseas. It has now become one of the major gateway for foreign colleagues to learn about the latest progress in urban research in China.

BCL lays much emphasis on urban modelling and quantitative urban research in multi-scales, for instance the mega-model paradigm proposed by Dr. Ying Long and collaborators. Moreover, the research conducted by BCL centres on the living quality of human settlement in the country's 'New Type Urbanization', which aims to provide comprehensive measurement and monitor of China's urban development. Our research is expected to support related policy decision-making. Furthermore, BCL also works on spreading messages on China's quantitative urban research in the international research community. Besides publishing Chinese scholars' works and data on international journals and platforms, BCL also brings in the words of foreign scholars. For instance, an interview to Prof Michael Batty, the director of Centre of Advanced Spatial Analysis, was conducted in the name of BCL in 2013 (Liu *et al*. 2014) on the past and prospect of urban modelling, as well as another interview to the late Sir Peter Hall on China's New Type Urbanization. Prof Michael Batty has mentioned BCL several times on his blog and twitter, saying that 'BCL is one of the symbols of China's up rise'.

After several years of research boom, we think that it is now a time to take a review of the bewildering developments in this field. We aim to identify the underlying trends of paradigm shifts embedded in these research, as well as the major challenges that the field needs to focus on in the near future. Such a review will provide the research community both in China and abroad to achieve a more comprehensive and systematic understanding of this research boom in China. In this paper, we approach this topic from the perspective of four major transformations of urban studies in China under the above-mentioned technological and institutional background, which are transformation in spatial scale, in temporal scale, in granularity, and in methodology. The transformations are further illustrated below.

2. Major transformations of urban studies in China

2.1. *Transformation in spatial scale – the mega-model*

Existing Chinese urban and regional research can be generally categorised by the scale of study. The first type is in-depth research on a single city, for instance, the study on the poverty issue of Guangzhou City (Yuan *et al*. 2008) or the study of the distribution of public facilities in Beijing. The second type is analysis on the regional scale, which covers several provinces or the entire country and uses province or county as the basic unit of analysis. An example of this type of research is national macroeconomic study. Most existing research is not able to achieve both wide coverage and high spatial resolution. In other words, the wide coverage of the study area is usually achieved by sacrificing details, while in-depth studies usually cover a relatively much smaller area.

However, the emergence of the new data environment makes it possible to achieve both high spatial resolution and wide coverage, zooming in from the district/county scale in traditional data to the town/street scale. For example, open data from social websites and commercial websites are usually based on individual users, vehicles, or businesses, which provides strong support for detailed analysis. When combined with traditional data, the research potential of new open data could be further extended to reveal new phenomena that were covered in the low data resolution in the past. Urban research with high-resolution high-coverage data has been going on in the West in recent year. Ratti *et al*. (2010) re-delineated the boundaries of UK cities and regions using 12 billion cell phone call records; Becker *et al*. (2011) built an agent-based model for Leeds using geo-tagged twitter data; Rozenfeld *et al*. (2011) analysed the scale of population aggregation with high-resolution (200 m × 200 m) survey data in the United Kingdom and the United States; Sagl *et al*. (2012) investigated the human behaviour in several European cities using GPS and Flickr data.

Driven by these relevant problems, we have proposed the development of methodology capable of maintaining both a fine resolution and capable of conducting research at the national or regional scale. The mega-model is an effective tool for quantitative research, driven by the availability of big/open data and implemented through straightforward modelling approach (Long *et al*. 2014b). The resulting capability presents a new paradigm of urban and regional study (Wegener 2004, Hunt *et al*. 2005, He *et al*. 2012). The subjects of mega-modelling are usually urban systems containing several cities, but by using this modelling method, we can examine both the development of individual cities as well as the networking among cities, to achieve comprehensiveness. Mega-model is expected to bring new perspectives and interesting findings to urban studies through its integration of top-down and bottom-up approaches. Through such methodology, the physical research 'domain' of conventional urban research will be blurred and scholars will be able to conduct in-depth research on remote places. Furthermore, the imbalance between research in major cities and cities of lower tiers will also be lessened, thus more research attention will be paid to smaller and less prominent cities. Such research manner is helpful in exploring the universal rules and identifying interesting particularities in China's urbanisation by keeping the consistency in research method among cities.

Here, we present two applications of the mega-model methodology. The first is the MVP-CA (Mega-Vector-Parcels Cellular Automata) Model which has produced detailed estimates of urban growth of all Chinese cities at the parcel scale (Figure 1). The full model simulates the growth of all 654 cities in China in the next 5 years under various developmental scenarios (Long *et al*. 2014a).

The second application of MVP-CA generated land-use maps for 297 cities using street maps and Point of Interest (POI) data (Figure 2). This project was undertaken because to the unavailability of open land use data in China and aims to provide free and open land-use data for researchers. The layouts, land-use functions, urban boundaries, densities, and degrees of land-use mix are all identified as a result of the modelling (Long and Liu 2013, Long and Shen 2015, Liu and Long 2016).

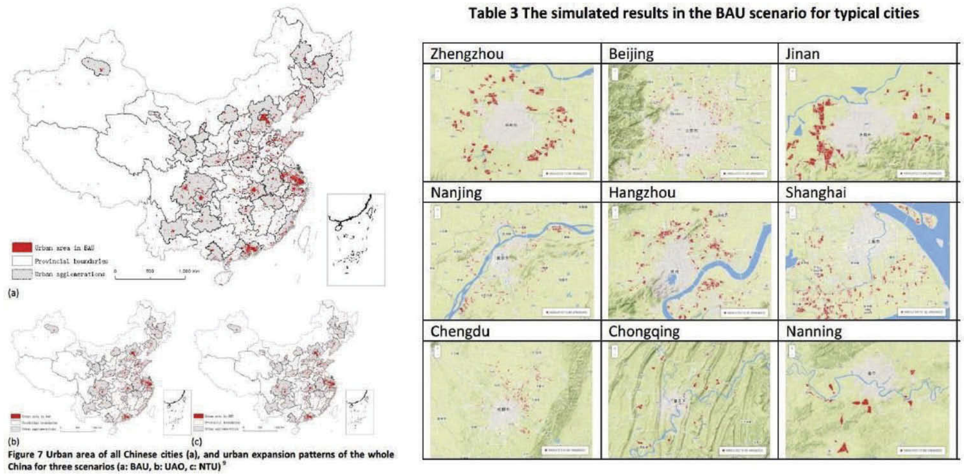

Figure 1. Urban growth simulation of all Chinese cities under various scenarios.

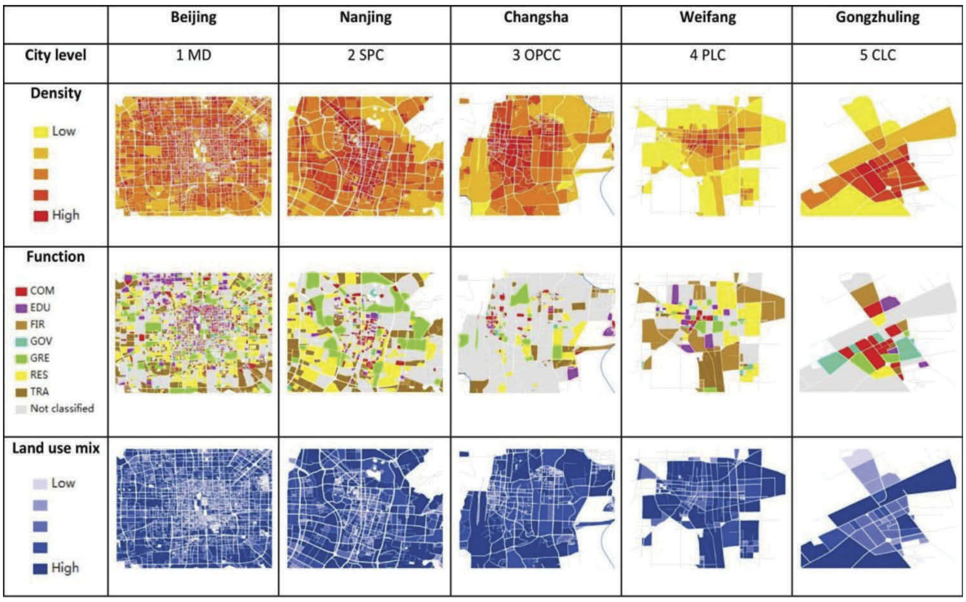

Figure 2. Derived land-use map of 297 cities using MVP-CA.

In our methodology, OSM (Open Street Map) data are used to identify and delineate parcel geometries, while POIs are gathered to infer land-use intensity, function, and mixing at the parcel-level. To be more specific, five steps are involved. First, parcel boundaries are delineated with OSM. Second, land-use density is calculated as the ratio between the counts of POIs in/close to a parcel to the parcel area and then standardised to range from 0 to 1. Third, urban parcels are identified from all generated parcels with a vector-based constrained cellular automata (CA) model. Fourth, urban function for individual parcels is identified by examining dominant POI types within the parcels, which refers to the POI type accounting for more than 50% of all POIs within the parcel.

The last step, the results are validated against both conventional manually collected parcel data and Ordnance Survey data.

2.2. *Transformation in temporal scale*

Another breakthrough use is the dynamic analysis of urban development. The data sources of conventional urban study and planning are mainly governmental statistic annals and self-conducted surveys, which are cross-sectional data at a single time-point. Moreover, due to the limited sampling technique, the spatial coverage of data is also limited. On the contrary, the big data, such as bus/metro card records and taxi GPS traces, are able to reflect the dynamics of the urban system in minutes and seconds, which are obviously advantageous in consistency, wide coverage and comprehensiveness. For instance, through the record of credit card transaction, the sales status in every corner of the city can be monitored and visualised in fine temporal resolution of hours, from which commercial zones can be delineated in a more reliable manner; and the further accumulation of the same data over the year will help to reveal changes in lifestyle and consumption trends, e.g. the boom of online shopping and the bust of real bookstores. Innovative works in this line include González *et al.*'s work (2008) in identifying rules in human travel behaviour with 6 months' GPS data of 100,000 people, Reades *et al.*'s work (2009) in the spatial structure of Rome with 3.5 million GPS record of 1 million cell phone users, Roth *et al.*'s work (2011) in identifying the polycentric structure of London with 11 million oyster card records, the work of SENSEable City Lab in MIT in chasing the procession of 5000 geo-tagged rubbish. By combining these big data sets, the limits of conventional urban study and research are solved (Bagchi and White 2004, Peng *et al*. 2007, Zhou *et al*. 2007, Dong *et al*. 2009, Jang 2010, Roth *et al*. 2011, Long *et al*. 2012).

After being accumulated for a certain period, the big data are also able to reflect the long-term changes and trends of urban development and life style over time. For instance, Long and Thill (2015) use 10 million passengers' bus trips from smartcards for identifying the commute pattern and the urban structure of Beijing. The research indicates that more than 95% of full-time jobs are longer than 6 h a day and 99.5% people start their daily travel from their own homes. It also demonstrated that the influential area of the Central Business District is much larger than either the Shangdi technological cluster or the 'Financial Street' (Figure 3). A further comparison of the records in 2008 and 2014 shows that the total bus trips are reducing and being replaced by the metro rail trips. Moreover, we identify lower income people from related socio-economic surveys and find that they usually spend more time on bus and transfer more; 80% of these people move their homes within 6 years and 87% change jobs, which means that they live quite an unstable life (Long and Shen 2015). Another research (Long *et al*. 2015) in underprivileged commuters finds four types of extreme transit behaviours in public transit riders who (1) travel significantly earlier than average riders (the 'early birds'), (2) ride in unusual late hours (the 'night owls'), (3) commute in excessively long distance (the 'tireless itinerants') and (4) travel over frequently in a day (the 'recurring itinerants'). It therefore reminds us to give more consideration towards this group of people in 'urban village' regeneration, public housing design and urban design.

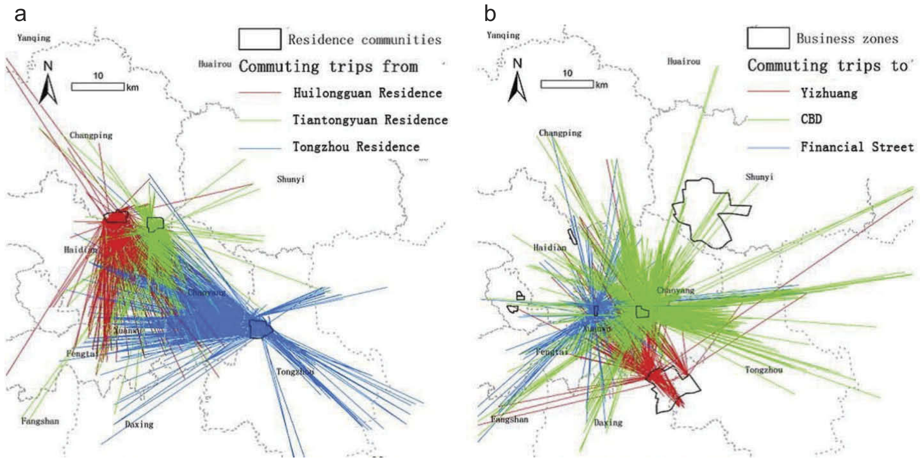

Figure 3. Commute pattern in typical areas.

2.3. *Transformation in granularity*

The new type of urbanisation is defined to be human-oriented urbanisation, which lays much emphasis on the human scale and the granularity of research. On the contrary, lots of conventional urban planning and policy making was 'building-oriented', which led to various social, economic and environmental problems caused by extensive development. In response to the situation, many Chinese cities have put forward plans of smart growth and replaced the large expansion projects with small-scale urban regeneration and redevelopment projects. In such circumstance, planning techniques that function at large spatial scale are less useful, which gives rise to the need to develop planning information at finer scales.

The new data environment again responds to such a trend by providing not only bigger data, but also a more well-rounded description of people's behaviour, feelings, emotions, experiences etc., which was very hard to obtain from conventional data. Such data could shed light upon topics on people's behaviour and activity, as well as its relation to the urban social space. Calabrese *et al*. (2010) found that individuals with stronger communication tights are more possible to show up in same physical locations, which points to a new approach for mobility prediction. Schneider *et al*. (2013) identified seventeen modes of people's daily activity with multiday cell phone data. Liu *et al*. (2014) analysed the cognitive picture of cities with Flickr photo data.

An example of research with higher granularity is our research on the dynamics of nationwide population density at the town/sub-district scale. According to the research, one-third of the country's land is sparsely populated, due to the aggregation towards big cities and city centres (Figure 4). Besides the well-known phenomenon of decaying village (Liu *et al*. 2009), we also find a trend of decaying taking place in 180 cities of all 654 Chinese cities (Figures 5 and 6). The finding is informative for planning practice in China, which was always based on the expectation of population growth. With the recognition of the decaying trend, the goal of planning in those 180 cities should no longer be land expansion, but the enhancement of residents' living quality. Another related issue is the so-called ghost city caused by over-construction in lots of Chinese

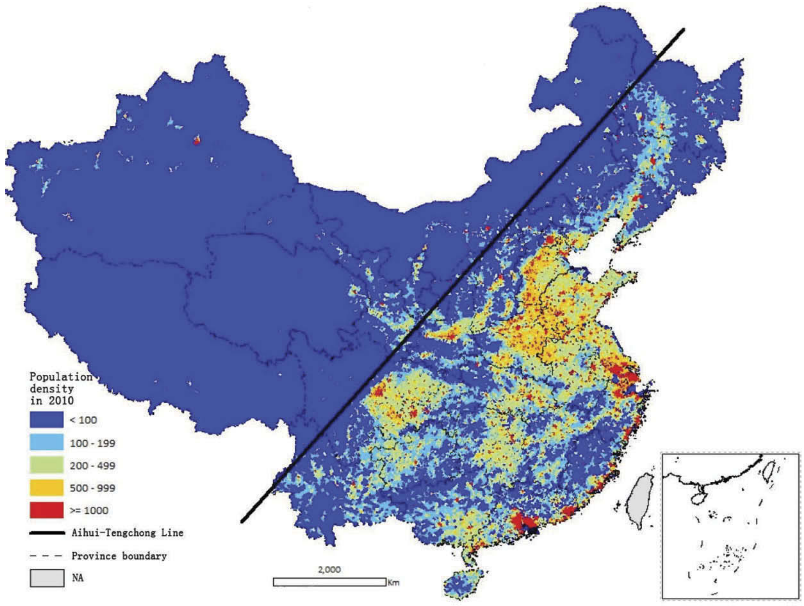

Figure 4. Distribution of population density at sub-district level in 2010.

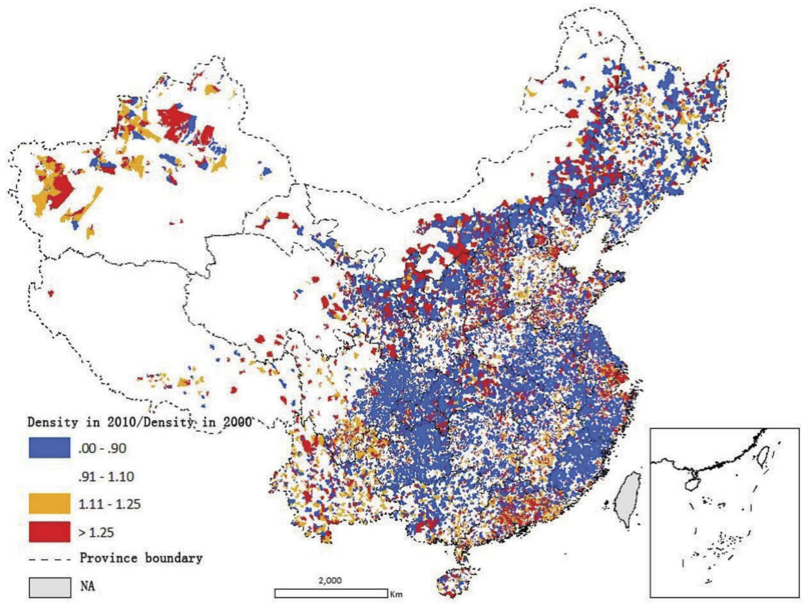

Figure 5. Population density ratio between 2000 and 2010.

cities in the past decade, which can be identified through evaluating the intensity of activity using Baidu map (Chinese version of Google map) and Weibo (Chinese twitter) data. The housing vacancy rate can be thus calculated for all cities with big data, from which the influencing factors and certain rules of urban development can be derived as reference for policy making.

Figure 6. Ratio of shrinking of Chinese cities between 2000 and 2010.

Another study on the implementation of land-use plan shows that 50% of all developments in Beijing are informal or illegal, with no planning permission; while, on the other hand, 95% of people's activity and mobility is still within the planning boundary. It indicates that the planning control is quite effective in the social sense, despite of ineffectiveness in the physical sense.

The third example is a nationwide sub-district-scale analysis on the exposure to PM2.5 pollution (Figures 7 and 8). We find that Xingtai City in Hebei Province is the most exposed city to PM2.5, and Beijing, Wuhan, and Chengdu suffer similar numbers of overexposure days.

2.4. *Transformation in methodology*

Crowdsourcing is a new form of work organisation enabled by the ICT developments, which distributes a large task to several volunteers through an online platform with higher efficiency. Similar to the trend of crowdsourcing in many other fields, there is also a trend of crowdsourcing in the field of urban research. Open research platforms such as BCL are fast emerging in recent years and demonstrating their advantages in efficiency and flexibility in terms of data collection, research collaboration and verification. For instance, it is almost impossible for a single research group to conduct detailed socio-economic and spatial transformation changes field work in all 180 shrinking Chinese cities. Therefore, BCL proposed the use of crowdsourcing as a new research paradigm through the establishment of 'China research network in shrinking cities' in 2014 (http://www.beijingcitylab.com/projects-1/15-shrinking-cities/), which is a vivid example of the transformation from conventional research group to crowdsourcing research.

An example of crowdsourcing research is the verification of open-data-generated land-use map. To tackle the problem of highly constrained access to land-use map in China, Liu and Long (2016) developed an algorithm for generating land-use map

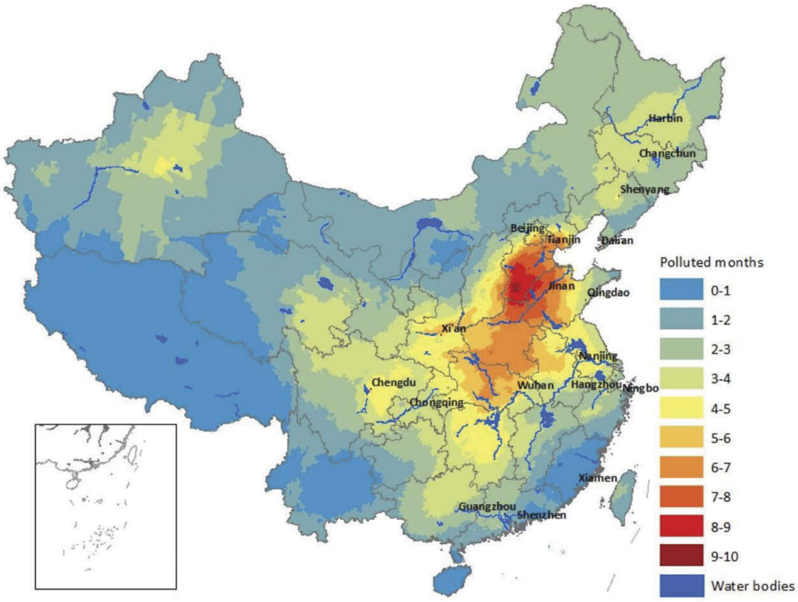

Figure 7. Urban environment at fine spatial scales: the number of polluted months in a year for each Chinese sub-district violating national PM2.5 standard.

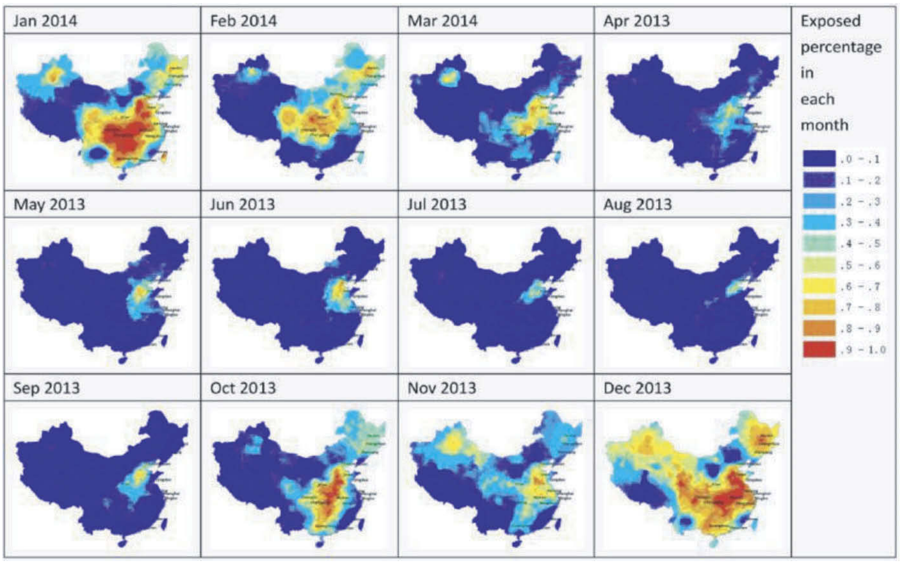

Figure 8. Exposed days in each month for each sub-district.

(including function, density, land-use mix etc.) for 297 Chinese cities using OSM and POI data. We then made the maps fully open on the web and obtain the real land-use map of 10 cities through voluntary contribution, from which the validity of our methodology is further testified. The simulation results of MVP-CA model also received lots of feedbacks from those with local knowledge through online platforms, which is of great help in improving model performance.

The attempts by Chu (2016) is also a vivid example of crowdsourcing urban study. They invited the public to put colour pins on a huge Beijing map to indicate their preferences for the walking environment with different colours. A total number of 1600 pins, representing 1600 pieces of ratings are collected. Based on the crowdsourced comments, the researchers further analysed the detailed built environment features at the pinned locations and found the influencing factors.

3. Challenges in the big/open data era

Several Chinese urban research and planning institutes have started to conduct quantitative urban research, including the works byBCL. Recognising that the new data are able to cover large geographic area in fine resolution, we proposed the mega-model, a new regional and urban research paradigm. Meanwhile, we have also identified four transformations in quantitative urban research, namely transformation in spatial scale, in temporal scale, in granularity and in methodology, which are all cantered on the improving people's quality of life. However, there are also several issues that we need to pay attention to.

3.1. *Dealing with data bias*

This issue has been repeatedly discussed since the emergence of new data. For instance, the studies on urban residents' happiness using geotagged Weibo are suffering from data bias on several aspects, including the duplicity of Weibo senders, the limitations of natural language processing technology, the representativeness of Weibo senders, and the black box of Weibo API, all of which bring doubts to the reliability of such studies. There are a few strategies to tackle this problem. The first strategy is to make use of the data bias. For instance, recognising that low-income people are more likely to travel by bus frequently, we studied the travel behaviour and the change of residence and work locations of low-income people from 2008 to 2010 with SCD. The second strategy is to study the behaviour of special groups, such as our study on the travel behaviour of university students and four extreme socio-economic groups in Beijing. The third strategy is to combine these data with other data on studying the same issue to improve the stability of research results. In another research, we combined SCD, travel survey data, social website check-in data, and taxi GPS data in evaluating the planning implementation of Beijing, which indicates that more than 95% people conduct their daily life within the planning boundary. The last strategy is to use more than one data set to complement each other, thus depicting the whole urban system.

3.2. *Short-term data visualisation versus long-term data exploration*

Most data used in current research are collected in less than 1-week time instead of years. Moreover, some research is merely data visualisation. Comparing with the new data, conventional data, such as yearbook data, can reflect the transformation of the urban system over the years. However, the situation would change a lot with the accumulation of new data, which could lead to quite different research results. For example, the 1-day record of credit card can be applied to identify the patterns of consumption, while 1-

month record can help identify the influence of festivals and years' records can further manifest the impacts of technological progress on consumption, which will change the current situation of lack of long-term research and theoretical breakthrough. Such a research trajectory is reflected in our research with SCD from 2008 to 2014.

The limitation on the time span of data also results in the lack of inter-city connection information, which further undermines the capacity of the mega-model approach. For instance, in terms of cell phone data, usually only 1–2 months' record is accessible for researchers, which is not ideally sufficient to identify the passenger flow, and thus the socio-economic connection between cities. In the long term, we expect the mega-model to go beyond a large-scale collection of independent models, but an integrated systematic view of cross-regional and cross-temporal urban transformations.

3.3. *Current situation analysis versus future planning support*

Up to now, there is more existing research aiming at analysing the current situation of urban systems than evaluating their future development, which needs to be changed. In order to provide effective guidance to urban planning and design with the new data and research methods, my collaborators and I have proposed a new methodology named Data Augmented Design (DAD).

DAD is a planning and design method based on quantitative urban analysis, which provides whole-process tools for field survey, information processing, design, and short-term and long-term evaluation. DAD aims at enhancing the scientific base of design, as well as evoking the creativity of planners and designers. To be more specific, DAD is not betrayal from art and design, but a new design method that emphasises the inspiration power of quantitative analysis. We expect DAD to reduce the working load of designers and thus let them focus on creation instead of repetitive work, and at the same time improve the measurability of design. Moreover, DAD is simple and straightforward, which makes it convenient to be generalised but also sensitive to the specialty of each project.

4. Conclusions

In this short report, we present several major trends and challenges in assembling big/open urban data sets for Chinese cities, and showcase our attempt at applying big/open urban data to understanding China's urbanisation. We identify four major transformations of urban studies in China, which are the transformations in spatial scale, in temporal scale, in granularity, and in methodology. Despite the potentials presented by these new data, we also ask for attention to three major challenges in the related field.

The new data environment has been drawing more and more attention from both researchers and planners, since it enables detailed observation on individuals' activities in the urban space. These detailed data could be applied to provide helpful information for the decisions on heated topics such as urban regeneration, shrinking city, public participation etc., as well as provide new developing opportunities for urban study, planning and design, construction and commercial consultancy, which correspond to the human-oriented development strategy to the central government's New Type Urbanization policy.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Bagchi, M. and White, P., 2004. What role for smart-card data from bus systems? *Proceedings of the ICE – Municipal Engineer*, 157 (1), 39–46. doi:10.1680/muen.2004.157.1.39

Batty, M., 2012. Smart cities, big data. *Environment and Planning B: Planning and Design*, 39 (2), 191–193. doi:10.1068/b3902ed

Becker, R.A., *et al*., 2011. A tale of one city: using cellular network data for urban planning. *IEEE Pervasive Computing*, 10 (4), 18–26. doi:10.1109/MPRV.2011.44

Calabrese, F., *et al*., 2010. The geography of taste: analyzing cell-phone mobility and social events. *In*: *Proceedings of the 8th international conference on pervasive computing*, Helsinki. Berlin: Springer-Verlag, 22–37.

Chu, Y., 2016. Research on the walking environment of Beijing: a case study of pin map. *Beijing Planning Review*, 3, 130–140.

Crooks, A., *et al*., 2015. Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 29 (5), 720–741. doi:10.1080/13658816.2014.977905

Dong, X., *et al*., 2009. Data processing and analyzing system for bus IC card based on GIS. *Geospatial Information*, 7 (5), 124–126.

Goetz, M. and Zipf, A., 2012. Towards defining a framework for the automatic derivation of 3D CityGML models from volunteered geographic information. *International Journal of 3-D Information Modeling (IJ3DIM)*, 1 (2), 1–16. doi:10.4018/IJ3DIM

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

Graham, M., *et al*., 2014. Uneven geographies of user generated information: patterns of increasing informational poverty. *Annals of the Association of American Geographers*, 104 (4), 746–764. doi:10.1080/00045608.2014.910087

He, L., Song, Y., and Dai, S., 2012. Research on the paradigm of urban planning responding to uncertainty. *City Planning Review*, 7, 15–22.

Hunt, J.D., Kriger, D.S., and Miller, E.J., 2005. Current operational urban land-use–transport modelling frameworks: a review. *Transport Reviews*, 25 (3), 329–376. doi:10.1080/0144164052000336470

Jang, W., 2010. Travel time and transfer analysis using transit smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, 2144, 142–149. doi:10.3141/2144-16

Jia, T. and Jiang, B., 2010. Measuring urban sprawl based on massive street nodes and the novel concept of natural cities. arXiv preprint arXiv:1010.0541.

Kitchin, R., 2014. The real-time city? Big data and smart urbanism. *GeoJournal*, 79 (1), 1–14. doi:10.1007/s10708-013-9516-8

Lang, T., 2011. Advancing global health research through digital technology and sharing data. *Science*, 331 (6018), 714–717. doi:10.1126/science.1199349

Li, D., *et al*., 2014. From digital earth to smart earth. *Chinese Science Bulletin*. doi:10.1007/s11434-013-0100-x

Liu, L., Long, Y., and Batty, M., 2014. A retrospect and prospect of urban models: reflections after interviewing Mike Batty. *City Planning Review*, 24 (4), 8–14.

Liu, X. and Long, Y., 2016. Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning & Design*, 43, 341–360. doi:10.1177/0265813515604767

Liu, Y., Liu, Y., and Zhai, R., 2009. Geographical research and optimizing practice of rural hollowing in China. *Acta Geographica Sinica*, 64 (10), 1193–1202.

Long, Y., Gu, Y., and Han, H., 2012. Spatiotemporal heterogeneity of urban planning implementation effectiveness: evidence from five urban master plans of Beijing. *Landscape and Urban Planning*, 108 (2–4), 103–111. doi:10.1016/j.landurbplan.2012.08.005

Long, Y. and Liu, X., 2013. Featured graphic: how mixed is Beijing, China? A visual exploration of mixed land use. *Environment and Planning A*, 45, 2797–2798. doi:10.1068/a130162g
Long, Y., *et al*., 2015. Early birds, night owls, and tireless/recurring itinerants: an exploratory analysis of extreme transit behaviors in Beijing, China. Available from: http://arxiv.org/abs/1502.02056 [Accessed 7 August 2016].
Long, Y. and Shen, Z., 2015. *Geospatial analysis to support urban planning in Beijing*. Berlin: Springer-Verlag.
Long, Y. and Thill, J.-C., 2015. Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing. *Computers, Environment and Urban Systems*, 53, 19–35. doi:10.1016/j.compenvurbsys.2015.02.005
Long, Y., Wu, K., and Mao, Q., 2014a. Simulating parcel-level urban expansion for all Chinese cities. arXiv preprint arXiv:1402.3718.
Long, Y., *et al*., 2014b. Big models: from Beijing to the whole China. *In*: Y. Long and Z. Shen, eds. *Geospatial analysis to support urban planning in Beijing*. Berlin: Springer-Verlag, 255–272.
Peng, H., Zhao, Q., and Zhao, S., 2007. Transfer matrix construction method based on bus IC card data processing. *Computer and Communications*, 4, 32–34.
Ratti, C., *et al*., 2010. Redrawing the map of Great Britain from a network of human interactions. *PLoS One*, 5 (12), e14248.
Reades, J., Calabrese, F., and Ratti, C., 2009. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36 (5), 824–836. doi:10.1068/b34133t
Reichman, O., Jones, M.B., and Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. *Science*, 331 (6018), 703–705. doi:10.1126/science.1197962
Roth, C., *et al*., 2011. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS One*, 6 (1), e15923.
Rozenfeld, H.D., *et al*., 2011. The area and population of cities: new insights from a different perspective on cities. *The American Economic Review*, 101 (5), 2205–2225. doi:10.1257/aer.101.5.2205
Sagl, G., *et al*., 2012. From social sensor data to collective human behaviour patterns: analysing and visualising spatio-temporal dynamics in urban environments. *In*: *Proceedings of the GI-forum 2012: geovisualization, society and learning*, Salzburg. Berlin: Wichmann Verlag, 54–63.
Schneider, C.M., *et al*., 2013. Unravelling daily human mobility motifs. *Journal of the Royal Society Interface*, 10 (84), 1–8.
Wegener, M., 2004. Overview of land-use transport models. *Handbook of Transport Geography and Spatial Systems*, 5, 127–146.
Yuan, Y., Xu, X., and Xue, D., 2008. Spatial types and differentiation mechanism of new urban poverty of Guangzhou City in transitional China. *Geographical Research*, 27 (3), 672–682.
Yue, Y., *et al*., 2014. Zooming into individuals to understand the collective: a review of trajectory-based travel behaviour studies. *Travel Behaviour and Society*, 1 (2), 69–78. doi:10.1016/j.tbs.2013.12.002
Zhou, T., Zhai, C., and Gao, Z., 2007. Approaching bus OD matrices based on data reduced from bus IC cards. *Urban Transport of China*, 5 (3), 48–52.