Title
Heterogeneity within Indian cities: Methods for empirical analysis

Author
Balakrishnan, Krishnachandran

Publication Date
2016

Peer reviewed|Thesis/dissertation

Heterogeneity within Indian cities: Methods for empirical analysis

by

Krishnachandran Balakrishnan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Landscape Architecture and Environmental Planning

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Louise Mozingo, Chair
Professor John Radke
Professor Isha Ray

Spring 2016

Heterogeneity within Indian cities:
Methods for empirical analysis

by

Krishnachandran Balakrishnan

Abstract

Heterogeneity within Indian cities:
Methods for empirical analysis

by

Krishnachandran Balakrishnan

Doctor of Philosophy in Landscape Architecture and Environmental Planning

University of California, Berkeley

Professor Louise Mozingo, Chair

This dissertation proposes new methods for analyzing the heterogeneity within Indian cities. The research focuses on developing methods which rely mainly on secondary datasets and demonstrates the application of the methods using data for Bangalore city. The dissertation is structured as a set of three papers, each of which focuses on a specific dimension of heterogeneity within Indian cities. The first paper uses data from the Census of India to generate sub-city typologies based on socio-economic attributes, housing quality, and access to water and sanitation infrastructure. The second paper proposes a predictive framework for high-resolution population density estimation in Indian cities. The method uses data on land-use, land-cover, street network, building height and asset ownership to predict population at a resolution of 30m. This paper also demonstrates the application of a new method for generating building height estimates at a city-wide scale using satellite stereo imagery. The third paper in this dissertation focuses on understanding heterogeneity in the volume of domestic piped water availability across parts of Bangalore city. It combines the high-resolution population mapping method of the second paper with data from the local water utility, to analyze domestic piped water availability. Using normative demand scenarios, it also estimates the deficit in domestic piped water availability and the extent of direct or indirect dependence on groundwater in a spatially disaggregated manner. The dissertation concludes with a discussion on the patterns of spatial inequality within Bangalore, as revealed by the three papers.

To my Amma and Achchan

Contents

Chapter III Densities of Bangalore

Chapter IV Domestic Piped Water Deficit in Bangalore

Chapter V Conclusion

Acknowledgements

I entered the PhD program with the intention of exploring the environmental history of urban water infrastructure in India. It appears I have ended up quite some distance from this initial intent. But all through the intervening academic wanderings I benefited greatly from the mentorship and guidance of my adviser Louise Mozingo. Her support, encouragement and incisive criticism have been invaluable to my research. It has been a privilege to work with and learn from her.

John Radke is largely responsible for making me believe that I can produce a dissertation which focuses so much on geospatial analysis. I would like to thank him for his enthusiasm, constant encouragement, thanksgiving dinners and the wide ranging discussions for which he always made time. Through her critique and guidance, Isha Ray has helped keep the dissertation grounded in the social and political realities of India. Thank you also for organizing the Water Group—it is such a great forum for discussions on water related research. Matt Kondolf always made time for detailed discussions and feedback. Despite his hectic schedule, he gave extensive comments and edits on practically every piece of writing I sent his way. His support has been invaluable to the successful completion of this dissertation.

Several other faculty members at Berkeley have encouraged and supported me over the years. In particular I would like to thank Walter Hood, Joe McBride, Peter Bosselmann, Chip Sullivan, David Meyer and Judith Stilgenbauer. Rob Thayer's enthusiasm and interest in my work has been a big source of encouragement for me. It was a privilege to meet and interact with Tom Rosin and Gail Wread during my time in Berkeley. Tom's long term work on Rajasthan helped me understand more about the social dimensions of traditional conjunctive water management practices in India.

I would like to thank my fellow Ph.D. students Allison Lassiter, Amir Gohar, Amna Alruheil, Kanokwalee Suteethorn (Mam), Tessa Beach, Wilasinee Suksawang, Kristen Podolak, Sara Carr, Raymond Wong, David de la Peña, Pedro Pinto, Zan Rubin, Manish Shirgaokar and Sharada Prasad for making the entire program so much more fun and enjoyable. Thank you in particular to Amna, Mam, Wilasinee and Amir for the surprise celebrations when I passed my qualifying exams! I am grateful to Zan for the illuminating conversations on all sorts of topics from geology and water modeling to music and dark matter. His comments and critique at various stages helped me discard a lot of tempting but distracting ideas and bring focus and clarity to this dissertation. Thank you also for helping me file this dissertation remotely.

Over the years, Tony Tieu and Sue Retta have helped me deal with the paperwork and procedures related to various administrative matters and fellowships. Jürgen Steyer helped solve all sorts of computing issues calmly and cheerfully and Mark O'Connor at the Geospatial Innovation Facility helped with customized software installations.

My initial research ideas originated from travels I undertook in India using the Geraldine Knight Scott travel fellowship. Funding for my doctoral research was provided by departmental fellowships, an International Dissertation Fieldwork Grant from the Institute of International Studies at Berkeley, a Junior Research Fellowship from the American Institute of Indian Studies and the Doctoral Completion Fellowship.

This research would not have been possible without the cooperation of officials in the Urban Development Department (Government of Karnataka), Bangalore Development Authority, Bangalore Water Supply and Sewerage Board, Karnataka State Pollution Control Board, Karnataka Census Directorate and the Office of the Registrar General of India. I would also like to thank officials in the Department of Mines and Geology, Central Groundwater Board, Geological Survey of India and Minor Irrigation Department for their cooperation.

I am grateful to Vishwanath S., Sunderrajan Krishnan, Dr. Sharad Lele, Mr.R.H. Sawkar, Dr. G.V. Hegde, Mr. K.C. Subhash Chandra, Ms. M.V. Shashirekha and Anjali Mohan for comments and suggestions at various stages in my research. Deepak Malghan generously made time for discussions on population modeling and water availability analysis. I am indebted to Carlo de Franchis for adding support for Cartosat-1 imagery in the Satellite Stereo Pipeline software and answering all my queries related to stereo image processing. Anand Sriram helped with building height estimation and with the numerous data collection trips within and outside Bangalore.

I would like to thank Aromar Revi for his unstinting support and encouragement and for enabling me to work with the Indian Institute for Human Settlements (IIHS). Somnath Sen, Thippeswamy Sir, Mohan Rao and Rajiv Raman helped me understand various aspects of water management in Bangalore. Mohan generously shared data on the boundaries of informal settlements in Bangalore. Manish Gautam put in a lot of work to help test some of the initial research ideas. I am also thankful for the wonderful hospitality provided by Pancham and Gita during my transition from Berkeley to Bangalore.

The process of dissertation research in Bangalore would have been far less enjoyable if not for the company of Siddharth Joshi, Raabiya Jayaram, Tahir Jayaram Joshi, Kapil Kaul, Neha and Rahul Sami, Kavita Wankhade, Anushree Deb, Harshvardhan Singh, Gautam Bhan, Vikrant, Jyothi Koduganti, Garima Jain, Arindam Jana, Teja Malladi, Mohan Raju, Amogh Arakali, Geetika Anand, Kumar Ray and Swastik Harish. I am grateful to Neha for her comments and suggestions, but even more so for the amazing food which she always

manages to conjure up without missing a beat in the conversations. Rahul's help was critical in implementing the stereo image processing algorithm and in refining some of the analysis methods.

The work done by Tarun Jayaram (TJ), Asim Waqif and Swastik Harish on traditional water management systems in India spurred my initial interest in the environmental history of water infrastructure in Indian cities. Thank you TJ for all the discussions and for the music. Jitender Yadav (Jitu), Sayantan Maitra (Boka), Shaily Gupta, Amlan Goswami, Aditya Narula (Adi), Amritha Ballal (Ballu), Chris Kurien and Chinglemba Chingtham (Ching) have all encouraged me in my work and hosted me during visits to New Delhi. I wish Jitu was around to see me finish this dissertation.

Thanks are also due to Stephane Paumier, Anupam Bansal, Rajesh Dongre, Ganesh Mohan, Anand Falgunan, Animesh Nayak, Amit Sina, Jassu Singh and Ching for helping me maintain one foot in the world of architecture and landscape design practice during the years of doctoral work. I owe a big debt to Rajakrishnan Rajkumar for ensuring that my Ph.D. application reached Berkeley in time—and for all the discussions, critiques, python code and arguments since then.

None of this would have been possible without the affection and encouragement of my parents. Every time I crossed an academic milestone, I suspect they were even more excited than me. To them I dedicate this dissertation. My father would have been delighted to see me finish a Ph.D., but that was not to be. My brother Rajakrishnan has been a tremendous source of support all through. Along with Swapna *chechchi* and Anindita he made sure I had a complete home away from home in California.

Most of all, I thank Shriya for her love, patience and understanding and the infinite music loops and the endless and enriching discussions, all of which I hope will keep going on and on.

I. Introduction

This dissertation proposes methods for understanding various dimensions of heterogeneity within Indian cities. It uses Bangalore as a case study to demonstrate the application of new methods to characterize heterogeneity across socio-economic parameters, population density, and domestic piped water availability. The research relies mainly on secondary datasets such as data from the Census of India, remote sensed data and other public datasets which are available from various civic agencies and planning authorities in Bangalore. The dissertation develops a set of generalizable methods that can potentially be applied to a large number of Indian cities for which these datasets are currently available or may become available in the near future.

The following sections set the context for this research by first giving a very brief overview of the India's urbanization trajectory and the growth of urban research in India. They briefly describe the various secondary data sources employed in empirical research on urban India, before discussing the datasets suitable for studies at the intra-urban scale. The last section outlines the structure of the three papers that constitute this dissertation and describes some of its key contributions.

1. Urbanization in India

According to most accounts, India is on the verge of a significant urban demographic transition. The total urban population of India which stands at 377 million as of 2011 is expected to more than double by 2050. While the percentage of urban population in India was 31.2% in 2011, in terms of absolute numbers, the urban population of India is larger than the total population of the United States in 2010. By 2050, the percentage of urban population in India is projected to cross 50%, with seven cities expected to have a population greater than 10 million (Census of India, 2011a; United Nations, 2014; United States Census Bureau, 2010).

The scale of this transition can perhaps be better understood comparing it to past trends in urbanization in India. In 1951, when the first census of independent India was completed, the total population was 361 million of which the total urban population was 62.4 million or 17.3%. Therefore, in the 60 year period from 1951 to 2011, India added 314.6 million people to its urban population (Census of India, 2011a). Current projections indicate that in the roughly 40 year period from 2011 to 2050, India will add an additional 404 million people to its urban population (United Nations, 2014).

2. Urban research and data sources in India

According to Ramachandran (1989), research on Indian cities from the perspective of social sciences was initially triggered by the work of Patrick Geddes in 1915 at the University of Bombay. Although there was some amount of research on Indian cities in the 1920s and 30s, organized urban research as we understand it today started mostly in the 1950s (Ramachandran, 1989; Mathur, 1993).

Since the 1950s, researchers from social science disciplines ranging from geography to sociology to economics have studied urbanization in India at a variety of scales. Academic sub-fields like urban geography also matured in the country by the 1980s and 1990s (Ramachandran, 1989; Mathur, 1993; Thakur & Parai, 1993;).[1] In comparison, research on cities from an environmental science perspective is relatively recent, not just in India, but globally (Shulenberger et.al., 2008; Pickett et.al., 2011).

Numerous authors from both environmental science and social science disciplines have conducted empirical research on Indian cities using primary data. But since the emphasis of this dissertation is on research using secondary data, this discussion focuses on some of the major sources of secondary data on Indian cities.

The Census of India has been the single most important source of secondary data for research on Indian cities from a social science perspective (Ramachandran, 1989). Besides this, urban researchers have also used the Economic Census and other sample surveys like the National Sample Survey and National Family Health Survey (see Mitra, 1992; Kundu & Sarangi, 2007; McKenzie & Ray, 2009).

From an environmental science perspective, remote sensed datasets have been a major secondary data source for research related to Indian cities. They have been used extensively in land-cover change and urban growth analysis at regional and city scales (see Sudhira et.al., 2004; Ramachandra et.al., 2012). Studies related to topography, drainage and surface water bodies have also applied methods which rely on remote sensed data (see Ramachandra & Kumar, 2009). Other studies related to urban India have utilized datasets compiled by various state and national level bodies like the Geological Survey of India, central and state groundwater boards and pollution control boards (see Narain, 2012).

[1] Mathur (1993) reviews urban research in India from 1960 to 1990. Thakur and Parai (1993) reviews research trends in urban geography in India, with a special emphasis on the period from 1980 to early 1990s. See Ramachandran (1989) for an overview of general trends in urban research from 1920s to 1980s.

3. Secondary data sources for research at the intra-urban scale

Although several types of secondary data exist for urban research in India, not all of them are suitable for research at the intra-urban scale. For example, a substantial number of studies which rely on census data have focused on national and regional trends in urbanization or on analysis of socio-economic aspects of specific cities or groups of cities. In comparison to this, fewer studies have focused on research at the intra-urban scale. As Ramachandran (1989) points out, this is probably because the census data at the lowest intra-urban spatial units has not always been easily accessible. Even if it is accessible, all variables are often not available for all cities. Besides, obtaining accurate boundaries of the intra-urban spatial units used has been a challenge and remains so (Hyderabad Urban Lab, 2014).

Moreover, until August 2014, the Census of India did not release Houselisting and Housing data at spatial units below the scale of the city. The Houselisting and Housing component of the census focuses primarily on housing characteristics, household ownership of assets, and access to water and sanitation infrastructure. This effectively meant that, although technically the data existed, till 2014 it was nearly impossible for researchers to systematically analyze intra-urban variation across these parameters for Indian cities.

Other datasets like the National Sample Survey and National Family Health Survey use coarse samples which, though useful for understanding regional and national patterns, cannot be used for understanding heterogeneity within Indian cities (National Sample Survey Organisation, 2001; International Institute for Population Sciences & Macro International, 2007). Similarly, data collected by various state and central environmental agencies like the pollution control boards and groundwater boards are often not at a resolution sufficient to enable intra-urban studies (see Central Groundwater Board, 2011).

On the other hand, remote sensed data from Indian and international satellites are currently available at resolutions which permit regional, city-scale and even neighborhood scale analysis. Although much of the urban research using remote sensed data has been largely conducted from an environmental science perspective, there have been several recent efforts to integrate these methods into research related to socio-economic aspects of Indian cities (see Baud et.al. 2010; Denis & Marius-Gnanou, 2011).

Researchers from both environmental science and social science backgrounds have extensively incorporated historical maps of Indian cities as another secondary data source. While geographers, urban planners and urban designers have used historic maps to analyze urban growth patterns and urban morphology (see Kosambi & Brush, 1988), environmental scientists, landscape architects and environmental planners have used it to evaluate changes in urban vegetation and urban hydrology (Nagendra et.al., 2011; Mathur & da Cunha, 2006; Mathur & da Cunha, 2009).

Apart from these sources, secondary data at the intra-urban scale has also been available from urban utilities and planning authorities. Parry (2012) uses data from urban utilities and local civic agencies to analyze disparity in access to educational and health care facilities and other amenities like parks, fire stations and solid-waste collection points in Srinagar. Several other studies have used similar sources of secondary data to analyze intra-urban variation in access to infrastructure and public utilities (see Nagne et.al., 2013; Mali et.al., 2013), concentration of industries and employment (Kalra, 2007) and incidence of vector borne diseases (Kumar et.al., 2014).

4. Outline of papers and key contributions to literature

This dissertation is structured as a set of three papers, each of which proposes a different method for examining aspects of urban heterogeneity in Indian cities. It uses Bangalore as a case study to demonstrate the application of the proposed methods.

Paper1: Sub-cities of Bangalore

The first paper focuses on heterogeneity in socio-economic status, housing conditions and access to infrastructure across different parts of Bangalore. The research draws on theories of similarity and typological analysis to identify areas within Bangalore that are similar across a range of different attributes. The research employs cluster analysis methods to delineate sub-city typologies using ward level data from Population Enumeration and Houselisting and Housing tables from the Census of India.[2] The Population Enumeration tables give information on population by sex, literacy levels, employment and social status while Houselisting and Housing tables have information mainly about asset ownership, housing conditions and access to water and sanitation infrastructure (Census of India, 2011b). [3]

In terms of data and methods, this paper advances the literature on urban heterogeneity in India. As discussed earlier, the Houselisting and Housing data for cities was never officially released at the ward level before August 2014. Hence this dataset has so far not been used to understand heterogeneity within Indian cities in a systematic manner. In addition, this is the first time that cluster analysis based methods have been used for typology delineation in the context of Indian cities.

[2] A ward is the administrative unit below the scale of a municipal corporation in Indian cities.

[3] Caste status is the main indicator of social status available from the Census. Population Enumeration data gives information on the percentage of population in each ward which is considered to come under the Scheduled Caste or Scheduled Tribes category. 'Scheduled Caste' refers to socially disadvantaged communities and 'Scheduled Tribes' refers to indigenous communities in India.

Paper 2: Densities of Bangalore

The second paper examines the concept of "density" and proposes a framework for predicting population density at 30m resolution. The proposed method utilizes data on land-use, land-cover, street network, building heights and asset ownership to predict density. Using data for the central part of Bangalore city, the research demonstrates the existence of a clear empirical relationship between the population within a 30m by 30m raster cell and its street density, building height and asset ownership values.

This paper contributes to the literature on the spatial structure of cities and high-resolution population density mapping. Although a few authors have proposed methods for high-resolution population prediction (Li & Weng, 2005; Silvan-Cardenas et.al., 2010; Bast et.al., 2015), this study is unique in the range of datasets it utilizes. In addition, this is the first time that population mapping of Indian cities has been attempted at this resolution.

The paper also demonstrates the application of a new method for generating building height estimates from satellite stereo imagery using open source software (de Franchis, 2014). Judging from literature in this field, this is the first time that satellite stereo imagery has been used for building height extraction at a city-wide scale in India. The proposed method is unique since it can be implemented without the need for surveys using Differential Global Positioning Systems (DGPS), or proprietary software, both of which are expensive and have limited the use of stereo imagery for building height extraction.

Moreover, the research demonstrates that relatively inexpensive stereo imagery from the Indian stereo satellite Cartosat-1, can be used to estimate building height with an error of approximately one floor. Since this level of accuracy should be acceptable for most city-scale applications, and since Cartosat-1 imagery is available for practically all Indian cities, the proposed method for building height extraction can be applied in the context of almost any Indian city.

Paper 3: Domestic piped water deficit in Bangalore

The third paper proposes a method for analyzing heterogeneity in domestic piped water availability across parts of Bangalore city. Building on the framework outlined in the previous paper, it first develops a population redistribution method using land-use, land-cover, street network and building heights.

This population distribution map is then used in conjunction with the water supply network map and water use data obtained from the local water utility to analyze domestic piped water availability in a spatially disaggregated manner. Using a normative demand scenario, the paper also estimates the extent of direct or indirect dependence on groundwater for domestic use within the study area in a spatially disaggregated way.

The available published literature shows that this is perhaps only the second study where inequality in domestic piped water availability in Indian cities has been assessed in a spatially disaggregated manner using secondary datasets. The only previous analysis of this type (Narain, 2012), is much coarser in spatial resolution and uses a method which is not fully explained.

Between them, the three papers in this dissertation propose methods by which existing secondary datasets can be used in new ways to analyze the heterogeneity within Indian cities. As discussed above, the dissertation contributes to the literature on urban characterization, spatial structure and density of cities, and urban water supply in developing countries.

References

Bast, H., Storandt, S., & Weidner, S. (2015). Fine-grained population estimation. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems,* Bellevue. doi:http://doi.acm.org/10.1145/2820783.2820828

Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R., & Karuppannan, S. (2010). Understanding heterogeneity in metropolitan india: The added value of remote sensing data for analyzing sub-standard residential areas. *International Journal of Applied Earth Observation and Geoinformation, 12*(5), 359-374. doi:http://dx.doi.org/10.1016/j.jag.2010.04.008

Census of India. (2011). *Instruction manuals on population enumeration and houselisting & housing*. New Delhi: Ministry of Home Affairs.

Census of India. (2011a). *Primary census abstract data tables*. New Delhi: Ministry of Home Affaris, Govt of India.

Central Groundwater Board. (2011). *Groundwater scenario in major cities of india.* (). Faridabad: Central Groundwater Board.

de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J., & Facciolo, G. (2014). An automatic and modular stereo pipeline for pushbroom images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2*(3), 49.

Denis, E., & Marius-Gnanou, K. (2011). Toward a better appraisal of urbanization in india. A fresh look at the landscape of morphological agglomerates. *Cybergeo: European Journal of Geography,*

Hyderabad Urban Lab. (2014). Will the real hyderabad please stand up? Retrieved from http://hydlab.in/blog/portfolio/data/will-the-real-hyderabad-please-stand-up/

International Institute for Population Sciences and Macro International. (2007). *National family health survey (NFHS-3), 2005–06: India: Volume II*. Mumbai: International Institute for Population Studies.

Kalra, R. (2007). *High Technology and Intra-Urban Transformations: A Case Study of Bengaluru, India,*

Kosambi, M., & Brush, J. E. (1988). Three colonial port cities in india. *Geographical Review,* , 32-47.

Kumar, D. S., Andimuthu, R., Rajan, R., & Venkatesan, M. S. (2014). Spatial trend, environmental and socioeconomic factors associated with malaria prevalence in chennai. *Malar J, 13*, 14.

Kundu, A., & Sarangi, N. (2007). Migration, employment status and poverty: An analysis across urban centres. *Economic and Political Weekly,* , 299-306.

Li, G., & Weng, Q. (2005). Using landsat ETM imagery to measure population density in indianapolis, indiana, USA. *Photogrammetric Engineering & Remote Sensing, 71*(8), 947-958.

Mali, S., Bhailume, S., & Das, S. (2013). Geoinformatics application for urban utilities information system: A case study of pune city, maharashtra, india. *International Journal of Computer Applications, 65*(22)

Mathur, A., & Da Cunha, D. (2006). *Deccan traverses: The making of bangalore's terrain* Rupa New Delhi.

Mathur, A., & da Cunha, D. (2009). *SOAK: Mumbai in an estuary* Rupa & Company.

Mathur, O. P. (1993). Responding to india's urban challenge a research agenda for the 1990s.

McKenzie, D., & Ray, I. (2009). Urban water supply in india: Status, reform options and possible lessons. *Water Policy, 11*(4), 442-460. doi:10.2166/wp.2009.056

Mitra, A. (1992). Urban poverty: A rural spill-over? *Indian Economic Review, 27*, 403-419.

Nagendra, H., Sivaram, S., & Subramanya, S. (2011). Lakes of mahadevpura constituency, bengaluru: Current status, changes in distribution, and recommendations for restoration. *Bengaluru: Report, Ashoka Trust for Research in Ecology and Environment,*

Nagne, A. D., Vibhute, A. D., Gawali, B. W., & Mehrotra, S. C. (2013). Spatial analysis of transportation network for town planning of aurangabad city by using geographic information system. *International Journal of Scientific & Engineering Research, 4*(7), 2588-2594.

Narain, S. (2012a). *Excreta matters.* (State of India's Environment: A citizens report No. 7 (Vol.1)). New Delhi: Center for Science and Environment.

National Sample Survey Organisation. (2001). *Concepts and definitions used in NSS*. New Delhi: Ministry of Statistics and Programme Implementation, Government of India.

Parry, J. A. (2012). *Spatial analysis of urban amenities in srinagar* (Master of Philosophy). Available from Department of Geography and Regional Development, University of Kashmir.

Pickett, S. T., Cadenasso, M. L., Grove, J. M., Boone, C. G., Groffman, P. M., Irwin, E., . . . Nilon, C. H. (2011). Urban ecological systems: Scientific foundations and a decade of progress. *Journal of Environmental Management, 92*(3), 331-362.

Prakasa Rao, V., & Tewari, V. (1979). Structure of an indian metropolis: A study of bangalore.

Ramachandra, T., Aithal, B. H., & Sanna, D. D. (2012). Insights to urban dynamics through landscape spatial pattern analysis. *International Journal of Applied Earth Observation and Geoinformation, 18*, 329-343.

Ramachandra, T., & Kumar, U. (2008). Wetlands of greater bangalore, india: Automatic delineation through pattern classifiers. *Electronic Green Journal, 1*(26)

Ramachandran, R. (1992). Urbanization and urban systems in india. *OUP Catalogue,*

Shulenberger, E., Endlicher, W., Alberti, M., Bradley, G., Ryan, C., ZumBrunnen, C., . . . Marzluff, J. (2008). *Urban ecology: An international perspective on the interaction between humans and nature* Springer Science & Business Media.

Silvan-Cardenas, J. L., Wang, L., Rogerson, P., Wu, C., Feng, T., & Kamphaus, B. D. (2010). Assessing fine-spatial-resolution remote sensing for small-area population estimation. *International Journal of Remote Sensing, 31*(21), 5605-5634.

Sudhira, H., Ramachandra, T., & Jagadish, K. (2004). Urban sprawl: Metrics, dynamics and modelling using GIS. *International Journal of Applied Earth Observation and Geoinformation, 5*(1), 29-39.

Thakur, B., & Parai, A. (1993). A review of recent urban geographic studies in india. *Geojournal, 29*(2), 187-196.

United Nations. (2014). *World urbanization prospects 2014: Highlights* United Nations Publications.

United States Census Bureau. (2010). *Quick facts* U.S. Department of Commerce.

II. Sub-cities of Bangalore[1]

Abstract

Urban heterogeneity could be characterized using sub-city typologies. This paper uses ward level Population Enumeration data and Houselisting and Housing data from the 2011 Census of India to construct sub-city typologies for Bangalore. Such an approach was difficult to apply in the context of Indian cities till recently, since the previous rounds of the Census did not release Houselisting and Housing data at the ward level. Nine variables from the Census were selected to represent three broad classes of attributes for each ward: housing conditions, availability of amenities and socio-economic status. Hierarchical and non-hierarchical cluster analysis methods were then used to delineate empirical typologies which permit a categorical rather than ordinal classification of the wards in Bangalore. The paper concludes with a discussion of the utility and limitations of such an approach in understanding Indian cities.

1. Introduction

1.1 Urban heterogeneity, inequality and inequity

Cities are inherently heterogeneous entities. They exhibit variation across socio-economic, political, environmental and infrastructural dimensions. "Heterogeneity" and "variation" are relatively neutral terms that describe unevenness in the distribution of an attribute in comparison to the terms "inequality" and "inequity" which have varying degrees of normative implications.

"Inequality" can technically refer to any variation in an observed attribute. On the other hand, "inequity" refers directly to injustice or unfairness arising out of variation in an attribute (CSDH, 2008; Starfield, 2011). But concern about a particular kind of inequality often implies that it could be indicative of, or lead to inequity, even when this may not be explicitly stated.

The study of urban heterogeneity assumes significance in this context since several aspects of urban heterogeneity tend to be strongly associated with urban inequality and inequity, especially in the case of developing countries (Werna, 1995; Stephens, 1996; Haddad & Nedovic-Budic, 2006; Kilroy, 2007).

[1] A version of this paper was previously published as *Balakrishnan, K. and Anand, S. (2015), The Sub-cities of Bengaluru: Understanding urban heterogeneity through empirical typologies. Economic and Political Weekly. Vol. 50, No.22.*

1.2 Studies of urban heterogeneity

Globally researchers have studied urban heterogeneity from a wide range of perspectives. There is a large body of literature that deals with intra-urban variation in public health (Bradley et.al., 1992; Songsore & McGranahan, 1993; Stephens et.al., 1997) and in urban physical/spatial aspects like land-cover, land-use and overall urban and environmental structure (Cadenasso et.al., 2007; Hoffman et.al., 2008; Baud et.al., 2010; Kostof, 1991; Forman, 2008). [2] But the studies that are most relevant to this paper deal with intra-urban variation in socio-economic aspects, access to physical and social infrastructure and housing conditions (Barnes, 2004; Jensen & Leven, 1997; Haddad & Nedovic-Budic, 2006; Martinez, 2009; Murphy, 1993; Werna, 1995; Garza, 1996; Tang & Batey, 1996; Portnov, 2002).

In India, there have been several studies since the 1950s which have focused largely on heterogeneity along social dimensions within cities. Many researchers have used population enumeration data from the census and/or other city specific surveys to generate detailed understanding of residential segregation within Indian cities (Gist, 1957; Bose, 1965; Mehta, 1968; Mehta 1969; Joy, 1975; Prakasa Rao & Tiwari, 1979; Mahadevia, 1991; Vithayathil & Singh, 2012).

But since the Census of India did not systematically release Houselisting and Housing data at the ward level till recently, studies of intra-urban variation which integrate social, economic and infrastructural aspects have been limited in India. Baud et.al. (2008 & 2009) and Sheolikar et.al. (2014) are some of the studies which have used ward level Houselisting and Housing data from the Census for studies of intra-urban variation. The former papers use it to construct multiple deprivation indices to study intra-urban variation in Delhi, Mumbai and Chennai while the latter uses it to analyze household fuel use in Bhopal to calculate ward level variation in CO2 emissions.

Most other studies which have focused on intra-urban variation of socio-economic and infrastructural aspects in Indian cities have relied on primary surveys of various kinds. Ramani et.al. (2005) examines variation in health and healthcare facilities across three wards of Ahmedabad by surveying households and healthcare facilities. Paul (2012a & 2012b) studies intra-ward disparities in access to education, healthcare and other urban amenities in Barasat and Burdwan cities using a sample survey of wards. Parry (2012) examines intra-urban disparity in access to urban amenities in Srinagar using population data from the Census of India and amenities data collected from various city agencies.

[2] See Bradley et.al. (1992) for a review of studies related to intra-urban differentials in mortality and morbidity in developing countries.

Bangalore is one of the Indian cities which has received considerable attention in terms of studies of inter-ward disparity.[3] Gist (1957) describes residential segregation within Bangalore along caste and ethnic dimensions. Prakasa Rao & Tiwari (1979) use data from a survey of Bangalore city to explore heterogeneity within the city across multiple socio-economic aspects, housing conditions and access to infrastructure. The Bangalore Master Plan-2015 (Bangalore Development Authority, 2007) identifies shadow areas—areas which are deficient in terms of access to health and education infrastructure. Organizations like Janaagraha (2013) and the Center for Sustainable Development (2012) have conducted sample surveys across all 198 wards of the city to evaluate access to urban infrastructure and overall environmental quality.

In August 2014, the Census of India released ward level Houselisting and Housing data for Indian cities for the first time. This should enable a much wider range of explorations related to intra-urban variation in India than was previously possible. This paper contributes to efforts in this direction by using ward level Census data to understand urban heterogeneity in Bangalore through sub-city typologies.

1.3 Understanding urban heterogeneity through sub-city typologies

Typology is generally understood as the study of how things can be divided into various "types" (Merriam-Webster's online dictionary, n.d.), but at a more fundamental level it can be understood as a process of categorization which enables an individual to perceive order in complex phenomena. Winch (1947) describes typologies as being created by the process of noting homogenous attributes in heterogeneous phenomena.[4] This process of identifying and grouping elements based on similarity of attributes is central to theories of perception and learning (Tversky & Gati, 1978).

The heterogeneity within a city could therefore be characterized by the creation of sub-city typologies, where each typology consists of parts of the city which are more similar to each other across multiple dimensions, than to parts which may belong to other typologies. Generating these empirical sub-city typologies using Census data can help us understand Indian cities in new ways.

Constructing sub-city typologies also provides a method of categorical classification for understanding cities, which is different from the ordinal classification one often sees in studies which use indices of ward quality or multiple deprivation. This is especially significant when dealing with the Census Houselisting and Housing data where variables like 'size of household' cannot be readily used in an ordinal classification system except

[3] In this paper I use 'Bangalore' to refer to the area within the Bruhat Bengaluru Mahanagara Palike (BBMP) boundary.

[4] For a discussion on the differences between typology and classification and empirical and heuristic types see Winch (1947).

after cross-tabulation across other relevant variables (for instance variables representative of economic status). Unfortunately, the ward level Houselisting and Housing data and Population Enumeration data released by the Census of India does not permit such cross-tabulation.

1.4 Models and analysis methods for identifying typologies

Similarity studies predominantly use geometric models (Shepard, 1962) or contrast models based on feature matching (Wittgenstein, 1953; Rosch & Mervis, 1975; Tversky, 1977). Geometric models conceptualize the objects of study as points in a coordinate space of as many dimensions as observed attributes. The similarity between objects is then inversely related to the distance between the points in coordinate space. On the other hand, contrast models conceptualize objects as collections of features and similarity is measured by analyzing their common and distinctive features. Contrast models are more appropriate for objects where qualitative aspects dominate (Tversky, 1977). This paper uses a geometric model for delineating sub-city typologies since Census data is predominantly quantitative in nature.

When using geometric models, cluster analysis or a combination of Principal Component Analysis (PCA) and cluster analysis are the main methods currently used for identifying empirical typologies.[5] In the literature, these methods have been applied in a variety of contexts and scales.

At the national scale, Bruce & Witt (1977) use cluster analysis for delineating city typologies in the United States. Orfield (2002—as cited in Hanlon 2009) and Mikelbank (2010) use cluster analysis to generate typologies of American suburban places while Hanlon (2009) uses a combination of PCA and cluster analysis for a similar study. Among studies from outside the U.S., Rovan & Sambt (2003) use cluster analysis methods to group Slovenian municipalities into four clusters based on socio-economic variables while Soares et.al. (2003) uses factor analysis and cluster analysis to illustrate socio-economic characterization and grouping of the municipalities of Portugal.

At the city scale, Portnov (2002) uses cluster analysis to determine patterns of intra-urban inequalities at the neighborhood level in Be'er Sheva, Israel. Barnes (2004) and Morenoff &

[5] Some of the early attempts at deriving urban typologies used factor analysis. Price (1942) and Hadden and Borgatta (1965—as cited in Bruce & Witt (1977)), use factor analysis to identify typologies of cities in the US by interpreting each factor as a typology. The factor loadings were used to calculate scores for each city within each factor and high scoring cities within a particular factor were interpreted to be prototypical of that typology. A somewhat similar approach has been used by Berry & Rees (1969) and Dutt et.al. (1981) to analyze heterogeneity within the city of Calcutta. It is debatable whether the results of factor analysis can be interpreted to be 'typologies' since although the factors provide dimensions of comparison for the cities, it is not possible to derive relatively homogenous groups of cities based on similarity *across* the identified dimensions. See Ramachandran (1989) for a critique of factor analysis based study of urban areas.

Tienda (1997) apply a similar approach to study urban poverty and typologies of neighborhood change in Chicago. Owens (2012) extends the latter study to all metropolitan areas in the U.S. using PCA and cluster analysis while Wyly & DeFilippis (2010) uses a similar method to identify neighborhood types using data on demographic and housing conditions for all census tracts of New York City.

2. Data and Methods

This paper uses the ward level Population Enumeration and Houselisting and Housing Data from the 2011 Census of India to generate sub-city typologies for Bangalore. [6] As described above, in the literature, PCA is often used initially to derive a small number of components on which cluster analysis is then carried out. Due to the difficulty involved in meaningfully interpreting the components generated by a PCA, this paper implements cluster analysis directly on the Census variables of interest. The full list of variables available from the 2011 Census, at ward level for Bangalore, is given in Figure1.

2.1 Variable selection, computation of indices and standardization

As a first step, all variables available from the Houselisting and Housing tables and Population Enumeration tables were mapped onto the ward boundaries of Bangalore city to explore their patterns of spatial distribution. In addition, the mean, median, standard deviation and coefficient of variation for each variable was also computed, to understand trends in their statistical distribution.

Although there are no generally accepted rules governing the relationship between sample size and number of variables used for cluster analysis, as a rule of thumb, when using *n* variables in cluster analysis, the sample size should be about 2^n (Mooi and Sarstedt, 2011). Since there are 198 wards in Bangalore, the number of variables should be seven or eight as per this thumb rule. Two rounds of analyses are presented in this paper—the first uses nine variables and the second seven variables.

The variables were selected to represent three broad classes of attributes for each ward: housing conditions, access to water and sanitation infrastructure and socio-economic status, while also capturing the variation inherent in the dataset. The full list of variables used and the attribute classes they represent is given in Figure 1. Correlation matrix for the full set of nine selected variables is given in Table 1.

[6] The Population Enumeration data is provided at the level of 'individuals', while the Houselisting and Housing data uses 'households' as the unit. For definition of 'household' see Census of India (2011). The difference in the units used across the two datasets is not significant for the analysis presented in this paper since it is conducted at the ward level using normalized ward level attributes which could be in any units. For example, a variable like 'percentage vegetation cover' at the ward level or 'ward area in sq.km.' can also be part of such an analysis.

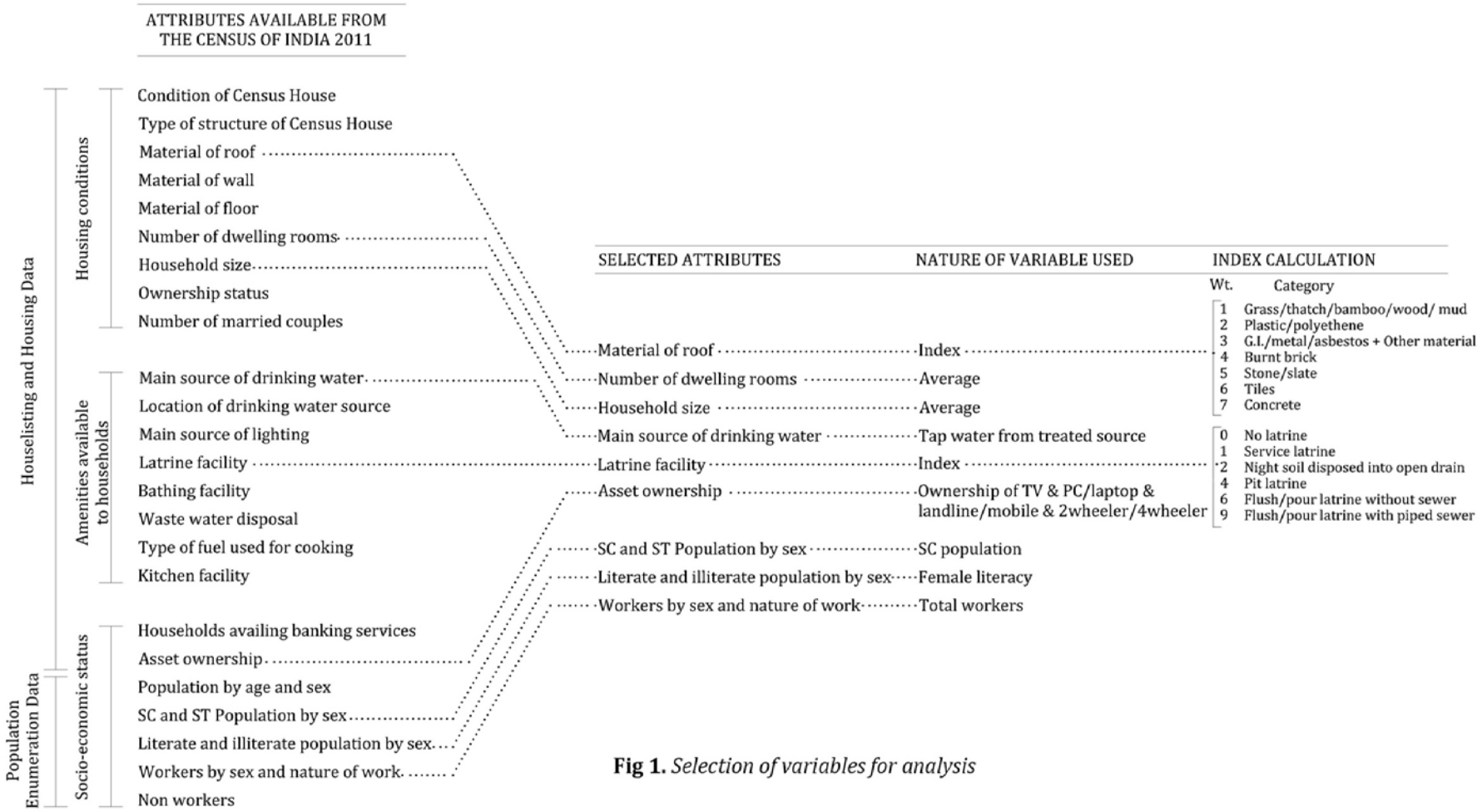


Fig 1. *Selection of variables for analysis*

All the variables used in the analysis were standardized using a z-standardization in STATA since they vary in scale and units. Out of the nine variables, two—'Material of Roof' and 'Latrine Facility'—are indices computed from the percentage of households which come under each category of possible responses to the respective questions. 'Latrine Facility' was converted into an index while 'Main Source of Drinking Water' was not, since in the former case households fall into a wide range of categories with each category having a significant number of households, while in the latter case most households fall into the categories of 'Tap water from treated source' or 'Tap water from untreated source' and hence does not need an index. [7]

Besides, the response category of 'Flush/Pour latrine with piped sewer' is highly correlated with the 'Tap water from treated source' category since both are referring to infrastructure provision by the city. By computing an index for 'Latrine Facility' the analysis captures additional variation in the data represented by households which come under other response categories like 'Septic tank', 'Service latrine', etc. Details of the response categories and the weights used are given in Figure 1.[8]

Cluster analysis was done on two sets of variables using STATA. The first round of analysis used all nine variables mentioned above while the second round excluded 'Tap water from treated source' and 'Latrine facility'. This is because, in the case of Bangalore, it is well known that there is a sharp distinction between the older parts of the city and the newer peripheries with respect to level of access to piped water supply and sewerage systems. [9] Initial exploratory analysis of the spatial distribution of these variables using the Census dataset confirms this. A second round of cluster analysis without these two variables enables us to explore other lesser known patterns of urban heterogeneity which may be obscured by them.

---

[7] In the case of 'Main Source of Drinking Water', the category of 'Borewell' also has a significant number of households. As per the Census Houselisting Manual (Census of India, 2011) definition these are households which use borewell water directly rather than through taps. But given the large number of households in this category I suspect that households which pump borewell water to tanks and then supply it through taps without treatment may also have been counted within this instead of being included in the category of 'Tap water from untreated source'. Therefore, in this paper all households which do not receive tap water from a treated source are lumped together. Since this leaves us with only two categories of responses for 'Main Source of Drinking Water', an index is not necessary.

[8] For both these variables, the computed indices displayed similar correlations with respect to individual components even when weights were scaled differently. Hence the indices computed are stable to choice of weights.

[9] Prior to 2007, the urban local body of Bangalore was the Bangalore Mahanagara Palike (BMP). The BMP boundary was expanded by government notification on 16 January, 2007 to include seven City Municipal Corporations, one Town Municipal Corporation and 110 villages which were outside the BMP boundary. This expanded entity was reconstituted as the Bruhat Bangalore Mahanagara Palike (Government of Karnataka, 2007). Most of these newly added areas are yet to be connected to piped water supply and sewerage systems. In this paper I use the term 'core' to refer to the central areas of BBMP which were within the erstwhile BMP boundary prior to 2007, and 'periphery' refers to the outer areas which were added in 2007.

2.2 Cluster analysis methods[10]

Cluster analysis methods can be grouped into hierarchical and non-hierarchical methods. In terms of implementation, one of the key differences between them is that in hierarchical methods the number of clusters is not specified at the beginning while in non-hierarchical methods the number of clusters has to be pre-specified. In this paper I first use hierarchical clustering to explore the dataset and identify the number of distinct clusters which may be present. I then use this solution to specify the number of clusters in the non-hierarchical clustering procedure. Brief descriptions of hierarchical and non-hierarchical clustering approaches and details of the specific methods used are given below.

2.2.1 Hierarchical clustering methods

Hierarchical clustering—as the name suggests—creates a hierarchy of nested clusters ranging from one cluster with all *n* objects under analysis to *n* clusters with one object each, along with measurements of dissimilarity between clusters. The creation of nested clusters could be through a process of fusion (agglomerative algorithms) which goes from *n* clusters to 1 cluster or through a process of division (divisive algorithms) which goes from 1 cluster to *n* clusters. The output can be visualized as a tree structure or dendrogram as shown in Figures 3 and 4 (Gordon, 1987; Duda et.al., 2000; Everitt et.al., 2011). This paper uses Wards linkage algorithm for agglomerative hierarchical clustering. Ward's linkage algorithm fuses clusters while minimizing the increase in within-cluster error sum of squares (Ward, 1963). [11]

Since the number of clusters is not pre-specified, the optimum clustering solution can be identified by visual inspection of the dendrogram along with statistical stopping rules. These stopping rules are statistical tests which check for the presence of underlying clusters within the data and the number of such clusters which can be identified. This paper uses the Calinski-Harabasz (Calinski and Harabasz, 1974) rule and the Duda-Hart rule (Duda et.al., 2000) as statistical methods of identifying the optimum clustering solution. [12] In the former case the cluster solution with the highest value for the Calinski-Harabasz pseudo-F statistic is selected, while in the latter procedure the cluster solution which has a combination of higher Duda-Hart index value and lower pseudo-T-squared value is selected.

---

[10] Everitt et.al. (2011) provides a comprehensive overview of cluster analysis methods. This section draws on the description given there.

[11] Refer Everitt et.al. (2011, p. 79) for a detailed description of various hierarchical cluster analysis methods and their relative advantages and disadvantages. I examined the performance of several commonly used hierarchical clustering algorithms before selecting Wards linkage.

[12] For a review of methods for determining number of clusters when using hierarchical cluster analysis see Milligan and Cooper (1985).

2.2.2 Non-hierarchical clustering methods

Non-hierarchical clustering methods partition *n* objects into a specified number of groups while optimizing a numerical criterion. In general all algorithms used for non-hierarchical clustering start with some initial partition which divides *n* objects into a specified number of groups, after which an object is moved from its original group to a new one if it helps in optimizing the clustering criterion. This process of moving objects from one group to another is repeated till no single move helps in further optimizing the clustering criterion. Non-hierarchical clustering methods are sensitive to the starting partition and hence may not give stable results.

This paper uses the *k-medians* algorithm for non-hierarchical clustering. In the *k-medians* algorithm, objects are moved to the group whose group median it is closest to, after which the group medians are recalculated. This is similar to the more commonly used *k-means* algorithm where instead of group medians, group means are used. Due to the use of medians instead of means, the *k-medians* algorithm is relatively more robust to outliers (Everitt et.al., 2011). [13]

3. Results

In this section two sets of results will be described since cluster analysis was conducted initially on the complete set of nine variables and then on a reduced set of seven variables which excluded 'Tap water from treated source' and 'Latrine facility'. While interpreting these results it is important to keep in mind that there are multiple decisions involved in arriving at a clustering solution, including choice of variables, clustering method/s, algorithms and stopping rules or starting partitions. There is no perfectly objective way of taking these decisions, as a result of which the cluster solutions should not be interpreted to be representative of some objective reality. Rather, the results should be evaluated on the basis of whether it provides an internally coherent and useful way of characterizing the city such that it provokes further questions and research.

Hierarchical cluster analysis using Ward's linkage (Ward, 1963) on nine and seven variables yielded the dendrograms shown in Figures 2 and 3. The dendrograms suggest that a four cluster solution would be optimal in both cases. This was confirmed using Calinski-Harabasz and Duda-Hart rules. The four cluster solutions have high values for the pseudo-F statistic as per the Calinski-Harabasz stopping rule (Table 2). They also have the

[13] All starting partition methods available in STATA with the k-medians algorithm were tested. Out of these, the 'p-random' method gave the most stable results. In the p-random method the dataset is partitioned into k groups (where k is the specified number of clusters) and the group medians are used as the starting points for the k-median algorithm.

best combination of high Duda-Hart Index value and low pseudo-T-squared value as per the Duda-Hart rule. [14]

After this, a four cluster solution was generated using k-medians non-hierarchical cluster analysis. This provided the results given in Table 3 and 4. Figures 9 and 10 show how they map onto the wards of Bangalore.[15] Since the results generated from the hierarchical and non-hierarchical cluster analyses were found to be similar, the following sections report only the results obtained by the final non-hierarchical cluster analyses on nine and seven variables.

3.1 Interpretation of clusters: analysis using nine variables

Cluster 1 is characterized by high socio-economic status, high levels of access to water and sanitation infrastructure and very good housing conditions. In particular it has higher means than all other clusters for 'No. of rooms', 'Asset ownership', 'Roof quality' and 'Female literacy'. It also has the lowest mean for 'Household size'. Let us call this sub-city typology the 'High-socio-economic Town' or 'High-SE Town'.

Cluster 2 has medium socio-economic status, high levels of access to water and sanitation infrastructure and medium housing conditions. It has the lowest mean for 'SC population' and highest mean for 'Latrine facility' – both by relatively thin margins. It also has the second highest mean in 'Access to treated tap water' by a very small margin. This typology could be called 'Average Town' since most of its cluster means are relatively close to the variable means for Bangalore.

Cluster 3 is characterized by low socio-economic status, very low levels of access to water and sanitation and poor housing conditions. It gets the lowest mean for 'Access to treated tap water', 'Latrine facility'and 'Roof quality'. At the same time, it has the highest mean for 'Total workers' and shares the lowest mean for 'Household size' with the High-socio-economic Town. Let us call this typology the 'Low-infra Worker Town'. [16] It is mostly the peripheral wards of Bangalore which come under this typology.

Cluster 4 has very low socio-economic status, has reasonably high levels of access to water and sanitation infrastructure, and very poor housing conditions. This cluster has the highest mean for 'SC population', 'Household size' and 'Access to treated tap water', along

[14] Although the 13 to 15 cluster solutions also have high Duda-Hart Index value and low pseudo-T-squared value, their utility is limited since solutions with more than seven to eight clusters are difficult to characterize and comprehend.

[15] Bangalore has 198 wards today, but as per the ward map available from the BBMP, two of these wards consist of non-contiguous areas. Vasanthpura (Ward No.197) consists of two proximate but distinct polygons to the south-western periphery while Hoodi (Ward No. 198) consists of four separate polygons spread across the north-eastern periphery.

[16] For the sake of brevity, I use the term 'infra' rather than 'water and sanitation infrastructure' to refer to the 'Access to treated tap water' and 'Latrine facility' variables.

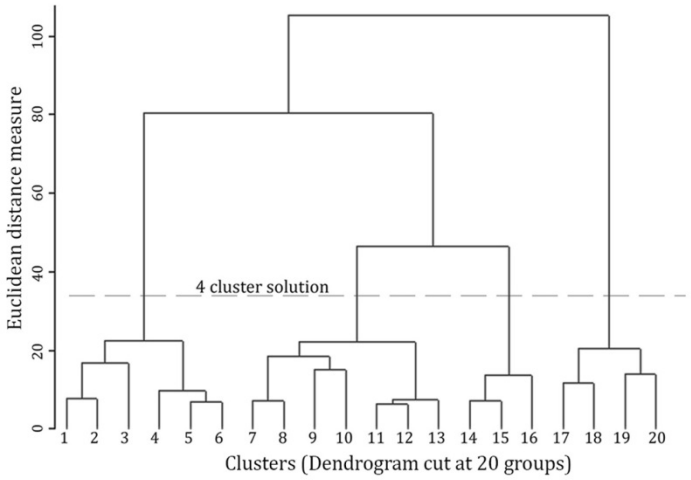


Fig 2. *Ward's linkage dendrogram for nine variables*

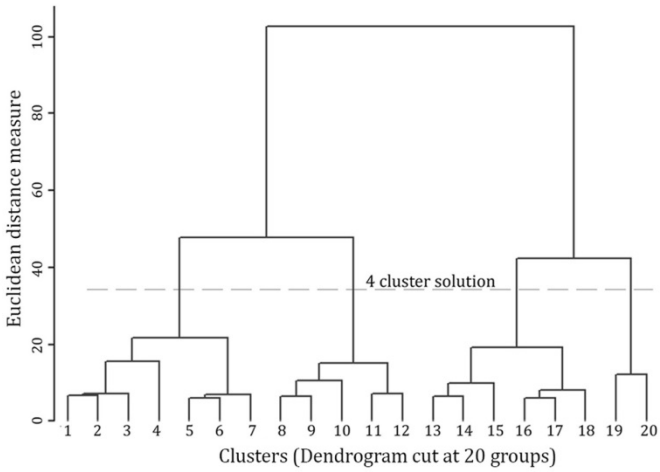


Fig 3. *Ward's linkage dendrogram for seven variables*

	Roof quality	No. of rooms	Household size	Latrine facility	Treated tap water	Asset ownership	SC population	Female literacy	Total workers
Roof quality	1.000								
No. of rooms	0.612	1.000							
Household size	-0.250	-0.230	1.000						
Latrine facility	0.553	0.225	0.073	1.000					
Treated tap water	0.437	0.152	0.331	0.761	1.000				
Asset ownership	0.672	0.819	-0.412	0.316	0.229	1.000			
SC population	-0.421	-0.263	0.381	-0.277	-0.017	-0.253	1.000		
Female literacy	0.825	0.684	-0.361	0.443	0.344	0.756	-0.332	1.000	
Total workers	-0.157	-0.176	-0.678	-0.173	-0.360	-0.027	-0.084	-0.101	1.000

Table 1. *Correlation matrix for selected variables*

	NINE VARIABLES			SEVEN VARIABLES		
	Calinski-Harabasz Rule	Duda-Hart Rule		Calinski-Harabasz Rule	Duda-Hart Rule	
No. of clusters	pseudo-F	Duda-Hart Index	pseudo T-squared	pseudo-F	Duda-Hart Index	pseudo T-squared
2	60.34	0.74	48.20	86.69	0.72	51.12
3	66.76	0.70	37.17	68.04	0.61	40.81
4	66.30	0.75	16.75	75.01	0.76	23.49
5	57.98	0.79	18.29	66.00	0.69	20.70
6	53.17	0.74	20.15	62.07	0.63	28.67
7	49.52	0.74	16.99	58.39	0.73	19.15
8	46.93	0.67	14.52	55.68	0.57	12.08
9	45.67	0.54	25.21	55.56	0.70	14.43
10	44.18	0.61	21.92	52.36	0.68	12.15
11	42.20	0.62	9.77	50.64	0.68	8.53
12	43.08	0.64	11.98	48.97	0.61	10.35
13	42.55	0.67	8.73	46.93	0.83	8.87
14	41.77	0.72	7.15	44.88	0.80	6.59
15	40.59	0.73	6.68	44.00	0.73	9.74

Table 2. *Values for Calinski-Harabasz and Duda-Hart rules*

	Cluster1	Cluster2	Cluster3	Cluster4	Bangalore
Roof quality	634.87	624.18	567.93	570.71	600.91
No. of rooms	2.52	2.07	1.84	1.71	2.03
Household size	3.93	4.12	3.94	4.54	4.12
Treated tap water	82.18	88.75	43.60	89.48	76.34
Latrine facility	852.65	876.53	668.17	826.24	808.98
Asset ownership	46.08	29.11	17.83	13.81	26.59
SC population	9.97	8.13	11.92	18.71	11.75
Female literacy	81.48	78.95	73.28	71.82	76.54
Total workers	43.53	42.00	46.14	41.38	43.22
Wards in cluster	42	64	49	43	198
Total Population	1,611,315	2,476,308	2,682,474	1,673,578	8,443,675
Typology Name	High-SE Town	Average Town	Low-infra Worker Town	Low-SE Town	

Table 3. *Nine Variables: Cluster means, city means, number of wards per cluster, population and sub-city typologies*

	Cluster1	Cluster2	Cluster3	Cluster4	Bangalore
Roof quality	638.38	615.69	565.25	562.97	600.91
No. of rooms	2.44	2.06	1.70	1.76	2.03
Household size	3.94	4.11	3.89	4.56	4.12
Asset ownership	44.41	27.81	14.17	13.94	26.59
SC population	8.01	10.27	10.63	19.48	11.75
Female literacy	81.93	77.79	72.38	71.63	76.54
Total workers	43.09	42.24	48.88	40.17	43.22
Wards in cluster	51	68	36	43	198
Total Population	1,927,643	2,842,025	1,915,266	1,758,741	8,443,675
Typology Name	High-SE Town	Average Town	Low-SE Worker Town	Low-SE Town	

Table 4. *Seven Variables: Cluster means, city means, number of wards per cluster, population and sub-city typologies*

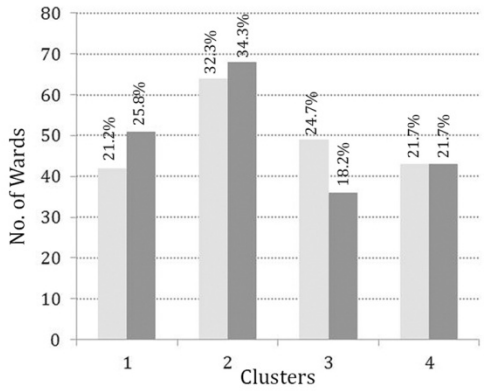


Fig 4. *Change in distribution of wards across clusters*

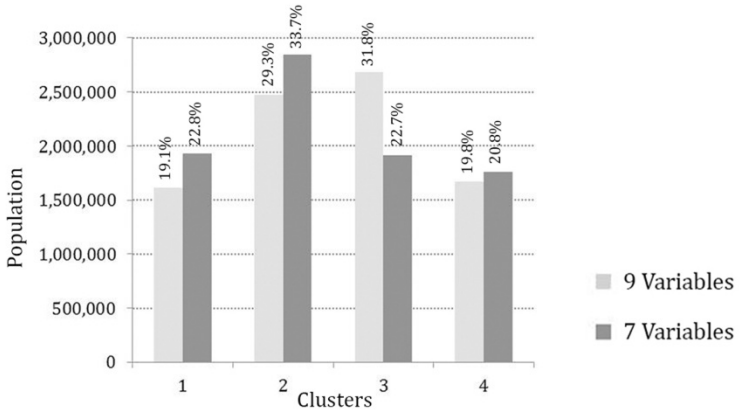


Fig 5. *Change in distribution of population across clusters*

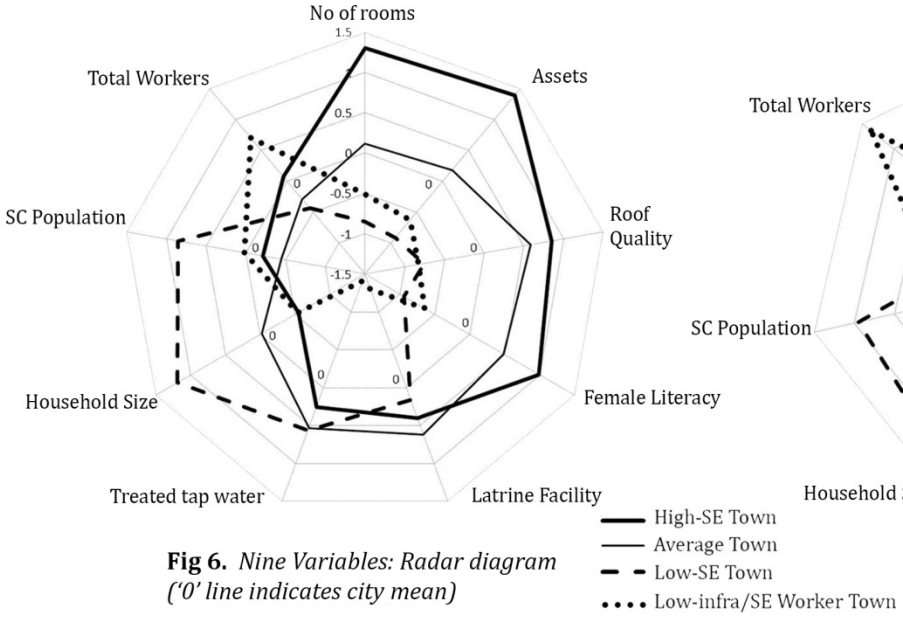


Fig 6. *Nine Variables: Radar diagram ('0' line indicates city mean)*

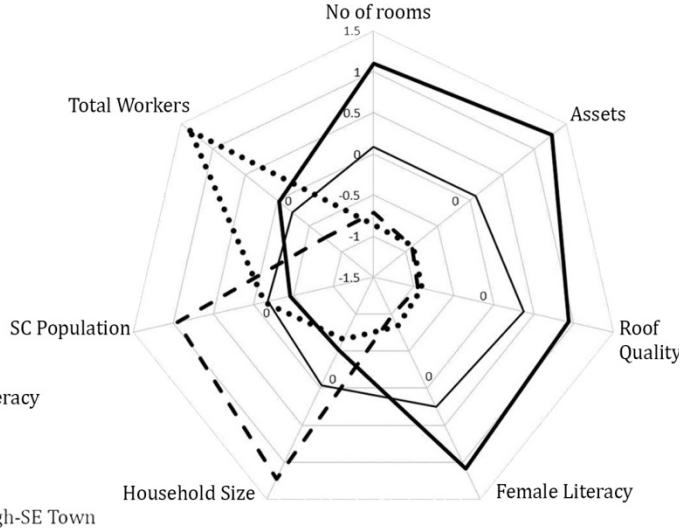


Fig 7. *Seven Variables: Radar diagram ('0' line indicates city mean)*

with the lowest means for 'Total workers', 'No. of rooms', 'Asset ownership' and 'Female literacy'. Its cluster mean for 'Roof quality' is also very close to the lowest. Let us call this the 'Low-SE Town'. Wards which come within this typology are all smaller wards from the urban core.

In summary, cluster analysis on nine variables provides us four sub-city typologies for Bangalore. These could be called the High-SE Town, Average Town, Low-infra Worker Town and the Low-SE Town. The High-SE Town and the Low-SE Town each have roughly 20% of the population of Bangalore, and the Average Town and Low-infra Worker Town each have about 30% of total population. The details of population distribution across these sub-city typologies are in Figure 5.

3.2 Interpretation of clusters: analysis using seven variables

Cluster 1 is once again characterized by high socio-economic status and high quality of housing. Like in the previous analysis using nine variables, it has the highest means for 'No. of rooms', 'Asset ownership', 'Roof quality' and 'Female literacy'. In this round it also has the lowest cluster mean for 'SC population'. This typology can retain its name of 'High-SE Town'.

Cluster 2 is also similar to what was obtained in the previous round of analysis. This time it does not have the highest or lowest mean for any variable and the means for most variables are closer to the overall means for Bangalore city, in comparison to the last round (refer Table 4 for cluster means and city means and Figure 7 for radar diagram). Once again we can retain the name given in the previous round and call this typology the 'Average Town'.

Cluster 3 again has low socio-economic status along with comparatively very poor housing conditions. But in this round of analysis it gets the highest cluster mean for 'Total workers' by a very big margin, along with the lowest cluster mean for 'No. of rooms' and 'Household size'. It comes third in all other variables except for 'SC Population' where it comes second. This time we can call this typology 'Low-SE Worker Town' since unlike in the last round, the defining feature of low access to water and sanitation infrastructure is not applicable anymore.

Cluster 4 is also similar to what we have seen in the previous round and has very low mean values on socio-economic indicators along with very poor housing. It has the highest cluster mean for 'SC population' and 'Household size' and the lowest in all other variables except 'No. of rooms' where it is third by a very small margin. Once again we have to call this typology 'Low-SE Town'.

In summary, the cluster analysis with seven variables gives us four sub-city typologies, out of which first, second and fourth typologies retain the same names as last time, while the

third typology can be called Low-SE Worker Town. This time the High-SE Town and Low-SE Worker Town each have about 23% of the total population in them while Average Town has about 33% and Low-SE Town has roughly 21%. Figure 5 gives details of the population distribution across these sub-city typologies.

4. Discussion

Cluster analysis methods help us understand Bangalore as comprising of four sub-city typologies. Across the two rounds of analysis we can observe some differences in the number of wards and total population which come under each typology. Figure 4 shows that the Low-SE Town has the same number of wards in each round of analysis. The total population within this typology is also relatively unchanged, with only a 1% difference between the two rounds. The sub-city typology maps (Fig. 9 and 10) show that in the first round the Low-SE Town comprised only of wards from the core area of Bangalore, while in the second round of analysis with seven variables, it loses some of the wards from the core areas but adds some peripheral wards, especially towards the north-east and south-west.

Figures 4 and 5 also show that the High-SE Town and Average Town both see a slight increase in the number of wards and population from the first to the second round. In the case of High-SE Town there is an increase of 4.6% and 3.7% for wards and population respectively. In comparison, the Average Town sees a change of only 2.0% in terms of number of wards, while the population increases by 4.4%. As the sub-city typology maps in Figures 9 and 10 show, in the second round of analysis, the wards which the High-SE Town gains are mostly the smaller ones to the west and south west of the High-SE Town from the first round analysis, while it loses a few wards to the east. In comparison, most of the wards the Average Town gains are the larger wards towards the north, north-west and western peripheries of the city.

The biggest shift in terms of population distribution across the two rounds of analyses occurs in the case of the Low-infra Worker Town and Low-SE Worker Town. As shown by Figure 5, the former had almost 32% of the total population (in the nine variable analysis) while the latter has only about 23% (in the seven variable analysis). This can be attributed to the presence of the two variables related to water and sanitation infrastructure in the first round of analysis which were causing the peripheral wards to cluster together since they all have very low means for these two variables—thereby generating the Low-infra Worker Town typology.

In the second round of analysis, many of these peripheral wards fall into either the Average Town or Low-SE Town typologies, leaving only two patches in the periphery where most of the wards fall within the new typology of Low-SE Worker Town. As seen in Figure 10, the first patch is around Peenya Industrial Area in the north-west part of Bangalore while the

Fig 8. *Bangalore ward map with reference place names*

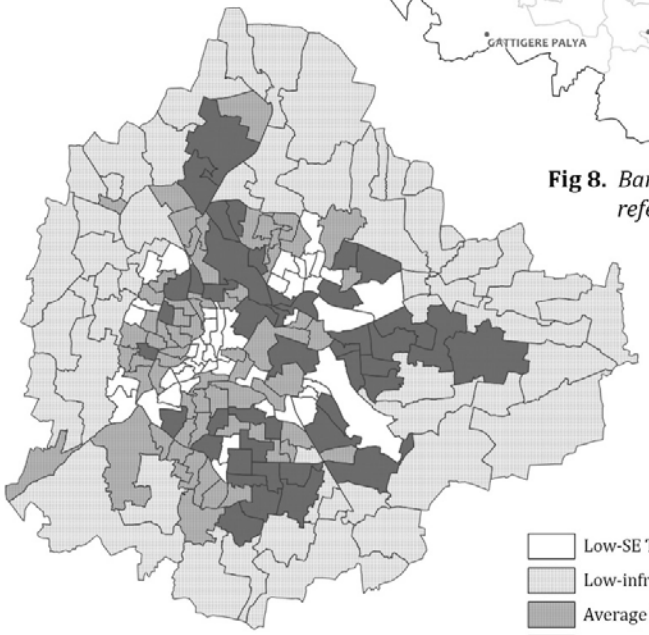


Fig 9. *Sub-city typologies map: Nine variables*

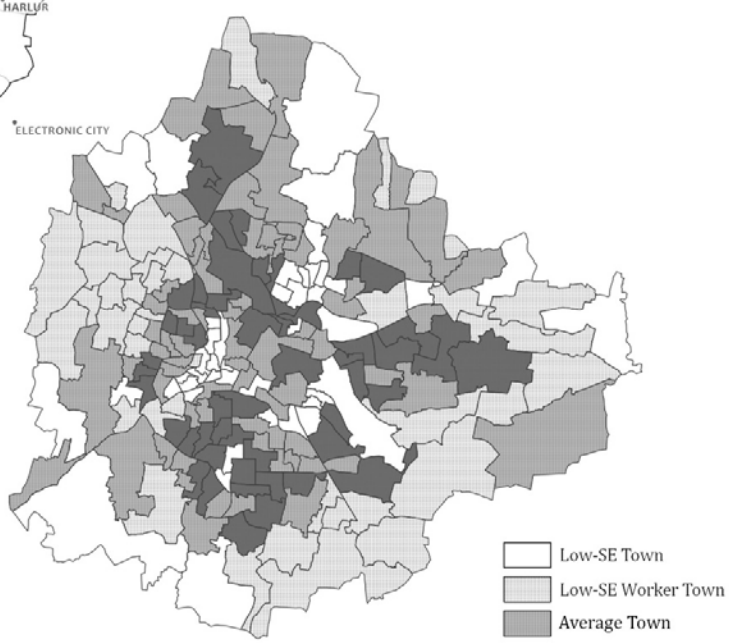


Fig 10. *Sub-city typologies map: Seven variables*

second extends from Hulimavu in the south to Whitefield in the east, past the roads leading to Electronic City. [17]

The two rounds of analyses show distinctly that there is a large section of the labor force in the periphery of the city which on average consists of small households that have very low socio-economic status, live in very poor housing conditions and have very low access to water and sanitation infrastructure. The analysis also shows that within the core of the city there are at least two very distinct groups of wards—one to the south-west and another to the north-east of the center of the city—which on average consist of large households which have very low socio-economic status and have very poor quality housing. The radar diagrams developed from the second round of analysis (Fig.7) gives striking evidence of the intra-urban heterogeneity, scale of inequality and therefore the potential inequities which may exist within Bangalore city.

While interpreting these cluster results it is important to avoid what is often termed the 'ecological fallacy' (Thorndike, 1939; Robinson, 1950; Selvin, 1958) at two levels—at the cluster level and the ward level.[18] At the cluster level, no specific ward within a sub-city typology may have the set of mean values which are used to characterize that typology. The use of cluster means to delineate the typologies runs the risk of missing extreme values across all attribute categories.

At the ward level, although a ward may fall into a particular typology, it does not mean that all neighborhoods or communities within that ward share the average characteristics which help us delineate that typology. This is especially true in the case of the peripheral wards of Bangalore which are considerably larger than the wards of the core area and have widely varying patterns of landcover and settlement structure. In the peripheral areas there are often dense high rise apartment and office clusters surrounded by open tracts of land or settlements which retain characteristics of villages. Since census data is not available below the ward level, large within-ward variations may be getting averaged out in such cases.

Due to similar reasons, I have refrained from using population density as an additional variable. As the next chapter shows, a ward level calculation of population/sq.km which does not take into account land-use, land-cover and building heights can be misleading. This can be particularly problematic when the ward sizes differ greatly like in the case of Bangalore. As discussed above, the large peripheral wards of Bangalore have scattered

[17] Electronic City is an industrial township located to the south east of Bangalore. It has a high concentration of information technology related industries.
[18] According to Susser (1973 - as cited in Subramanian et.al., 2009) it may be more appropriate to term this 'aggregative fallacy'.

pockets of dense developments and an average density calculated across the entire ward may not be representative of the density experienced by the people living in these wards.

5. Conclusion

The release of the ward level Houselisting and Housing data from the 2011 Census of India is a very welcome and long overdue step as far as Census data dissemination is concerned. This dataset in combination with the ward level Population Enumeration data opens up the potential for a wide range of new research on Indian cities at the intra-urban scale. Through the application of hierarchical and non-hierarchical cluster analysis, this paper demonstrates a method for deriving empirical sub-city typologies for Indian cities using this ward level census data.

Although there are several aspects to keep in mind while interpreting the cluster analysis results, the sub-city typologies described in this paper provide a means of characterizing Bangalore city and understanding intra-urban variation in Bangalore across multiple dimensions at the ward level. Such an understanding could be useful in city-scale vulnerability studies, or in structuring more in-depth studies—for example in stratified sampling surveys. It could also be useful in modeling various urban economic phenomena like rents and house prices.

If similar or better resolution data is available for successive or preceding census years for Indian cities, one could also examine the temporal stability, or trajectories of change of the sub-city typologies identified in this paper. Using the available dataset from the 2011 Census, this study can also be extended to other cities to examine if the sub-city typologies identified in this paper remain relevant across cities within the same size class and across size classes.

Acknowledgements

I would like to thank Arindam Jana for alerting me to the existence of BBMP wards which consist of non-contiguous areas. Shriya Anand helped refine the analysis method, implemented the analysis in STATA and gave detailed feedback on drafts of this paper. All errors in the manuscript remain my responsibility.

References

Bangalore Development Authority. (2007). *Bangalore master plan - 2015*. Bangalore: Government of Karnataka.

Barnes, S. (2004). The clustering of poor urban, chicago residents: An empirical analysis of social differentiation. *Journal of African American Studies, 7*(4), 39-58. doi:10.1007/s12111-004-1018-9

Baud, I. S. A., Pfeffer, K., Sridharan, N., & Nainan, N. (2009). Matching deprivation mapping to urban governance in three indian mega-cities. *Habitat International, 33*(4), 365-377. doi:http://dx.doi.org/10.1016/j.habitatint.2008.10.024

Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R., & Karuppannan, S. (2010). Understanding heterogeneity in metropolitan india: The added value of remote sensing data for analyzing sub-standard residential areas. *International Journal of Applied Earth Observation and Geoinformation, 12*(5), 359-374. doi:http://dx.doi.org/10.1016/j.jag.2010.04.008

Baud, I., Sridharan, N., & Pfeffer, K. (2008). Mapping urban poverty for local governance in an indian mega-city: The case of delhi. *Urban Studies, 45*(7), 1385-1412. doi:10.1177/0042098008090679

Berry, B. J., & Rees, P. H. (1969). The factorial ecology of calcutta. *American Journal of Sociology,* , 445-491.

Bose, N. K. (1965). Calcutta: A premature metropolis. *Scientific American, 213*, 90-102.

Bradley, D. J., Stephens, C., Harpham, T., & Cairncross, S. (1992). *A review of environmental health impacts in developing country cities*. Washington, D.C.: World Bank.

Bruce, G. D., & Witt, R. E. (1971). Developing empirically derived city typologies: An application of cluster analysis. *The Sociological Quarterly, 12*(2), 238-246.

Cadenasso, M. L., Pickett, S. T. A., & Schwarz, K. (2007). Spatial heterogeneity in urban ecosystems: Reconceptualizing land cover and a framework for classification. *Frontiers in Ecology and the Environment, 5*(2), 80-88.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1-27. doi:10.1080/03610927408827101

Census of India. (1991). *List of urban agglomerations with their component towns and outgrowths: Table for india.* New Delhi: Census of India.

Census of India. (2011a). *Primary census abstract – urban agglomeration table for karnataka.* New Delhi: Census of India.

Census of India. (2011b). *Percentage of households to total households by assets and amenities.* New Delhi: Census of India.

Census of India. (2011). *Instruction manual for houselisting and housing census*. New Delhi: Census of India.

Centre for Sustainable Development. (2012). *Environment report card of bangalore 2012*. Bangalore: Centre for Sustainable Development.

CSDH. (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health.* (No. Final Report of the Commission on Social Determinants of Health). Geneva: World Health Organization.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification* (Second ed.) Wiley-Blackwell.

Dutt, A. K., D'Sa, G., & Monroe, C. B. (1989). Factorial ecology of calcutta (1981) revisited. *Geojournal, 18*(2), 151-162.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (Fifth ed.). West Sussex, UK: John Wiley and Sons.

Forman, R. T. T. (2008). *Urban regions*. Cambridge, UK: Cambridge University Press.

Garza, G. (1996). Social and economic imbalances in the metropolitan area of monterrey. *Environment and Urbanization, 8*(2), 31-42. doi:10.1177/095624789600800208

Gist, N. P. (1957). The ecology of bangalore, india: An east-west comparison. *Social Forces,* , 356-365.

Gordon, A. D. (1987). A review of hierarchical classification. *Journal of the Royal Statistical Society.Series A (General), 150*(2), 119-137.

Government of Karnataka. (2007). *Notification of areas under bruhat bangalore mahanagara palike* Government of Karnataka.

Haddad, M., & Nedović-Budić, Z. (2006). Using spatial statistics to analyze Intra-urban inequalities and public intervention in são paulo, brazil. *Journal of Human Development, 7*(1), 85-109. doi:10.1080/14649880500502102

Hadden, J. K., & Borgatta, E. F. (1965). *American cities: Their social characteristics*. Chicago: Rand McNally & Co.

Hanlon, B. (2009). A typology of inner-ring suburbs: Class, race, and ethnicity in U.S. suburbia. *City & Community, 8*(3), 221-246. doi:10.1111/j.1540-6040.2009.01287.x

Hoffman, P., Strobl, J., Blaschke, T., & Kux, H. (2008). Detecting informal settlements from quickbird data in rio de janeiro using an object based approach. *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications* (pp. 531-553). Berlin: Springer. doi:10.1007/978-3-540-77058-9_29

Janaagraha. (2013). *Ward quality score databook*. Bangalore: Janaagraha Centre for Citizenship and Democracy.

Jensen, M. J., & Leven, C. L. (1997). Quality of life in central cities and suburbs. *The Annals of Regional Science, 31*(4), 431-449. doi:10.1007/s001680050057

Joy, A. (1975). *The Spatial Organization of a South Indian City: Coimbatore,*

Kilroy, A. (2009). *Intra-urban spatial inequality : Cities as "urban regions"*. Washington D.C.: World Bank.

Kostof, S. (1991). *The city shaped: Urban patterns and meanings through history* Bullfinch Press.

Mahadevia, D. (1995). *Emerging Process of Residential Segregation in Metropolitan Cities: Case Study of Mumbai and Chennai,*

Martínez, J. (2009). The use of GIS and indicators to monitor intra-urban inequalities. A case study in rosario, argentina. *Habitat International, 33*(4), 387-396. doi:http://dx.doi.org/10.1016/j.habitatint.2008.12.003

Mehta, S. K. (1968). Patterns of residence in poona (india) by income, education, and occupation (1937-65). *American Journal of Sociology,* , 496-508.

Mehta, S. K. (1969). Patterns of residence in poona, india, by caste and religion: 1822–1965. *Demography, 6*(4), 473-491.

Mikelbank, B. A. (2004). A typology of U.S. suburban places. *Housing Policy Debate, 15*(4), 935-964. doi:10.1080/10511482.2004.9521527

Milligan, G., & Cooper, M. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*(2), 159-179. doi:10.1007/BF02294245

Mooi, E. A., & Sarstedt, M. (2011). *A concise guide to market research. the process, data and methods using*
IBM SPSS statistics,. Berlin: Springer.

Morenoff, J. D., & Tienda, M. (1997). Underclass neighborhoods in temporal and ecological perspective. *Annals of the American Academy of Political and Social Science, 551*(, Globalization and the Changing U. S. City), 59-72.

Murphy, S. G. (1993). An imbalance of power: The geography of electricity provision in urban areas of south africa. *Geojournal, 30*(1), 53-62. doi:10.1007/BF00807827

Orfield, M. (2002). *American metropolitics: The new suburban reality*. Washington, D.C.: Brookings Institution Press.

Owens, A. (2012). Neighborhoods on the rise: A typology of neighborhoods experiencing socioeconomic ascent. *City & Community, 11*(4), 345-369. doi:10.1111/j.1540-6040.2012.01412.x

Parry, J. A. (2012). *Spatial analysis of urban amenities in srinagar* (Master of Philosophy). Available from Department of Geography and Regional Development, University of Kashmir.

Paul, S. (2012). Analysis of micro level disparities in urban facility-utility services: A study on barasat city, west bengal, india. *Journal of Urban and Regional Analysis, 4*(2), 173-188.

Paul, S. (2012). Factors and dimensions of inter-ward disparities in urban facility - utility services in burdwan city, india. *Archives of Applied Science Research - Scholars Research Library, 4*(3), 1376-1388.

Portnov, B. A. (2002). Intra-urban inequalities and planning strategies: A case study of be'er sheva, israel. *International Planning Studies, 7*(2), 137-156. doi:10.1080/13563470220132227

Prakasa Rao, V., & Tewari, V. (1979). Structure of an indian metropolis: A study of bangalore.

Price, D. O. (1942). Factor analysis in the study of metropolitan centers. *Social Forces, 20*(4), 449-455.

Ramachandran, R. (1992). Urbanization and urban systems in india. *OUP Catalogue,*

Ramani, K. V., Mehandiratta, S., Patel, A., Joshi, D., & Patel, N. (2005). *Urban health status in ahmedabad city: GIS based study of baherampura, kubernagar, and vasna wards*. Ahmedabad: IIM Ahmedabad.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*(3), 351-357.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573-605. doi:http://dx.doi.org/10.1016/0010-0285(75)90024-9

Rovan, J., & Sambt, J. (2003). Socio-economic differences among slovenian municipalities: A cluster analysis approach. *Metodoloki Zvezki, 19*, 265-278.

Selvin, H. C. (1958). Durkheim's suicide and problems of empirical research. *American Journal of Sociology, 63*(6, Emile Durkheim-Georg Simmel, 1858-1958), 607-619.

Sheolikar, P., Yawale, S., Puntambekar, K., & Kapshe, M. (2014). Spatial analysis of low carbon development 2050: A case of residential sector, bhopal, india. *IOP Conference Series: Earth and Environmental Science, 18*(1), 012153. doi:10.1088/1755-1315/18/1/012153

Shepard, R. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika, 27*(2), 125-140. doi:10.1007/BF02289630

Soares, J. O., Marquês, M. M. L., & Monteiro, C. M. F. (2003). A multivariate methodology to uncover regional disparities: A contribution to improve european union and governmental decisions. *European Journal of Operational Research, 145*(1), 121-135. doi:http://dx.doi.org/10.1016/S0377-2217(02)00146-7

Songsore, J., & McGranahan, G. (1993). Environment, wealth and health: Towards an analysis of intra-urban differentials within the greater accra metropolitan area, ghana. *Environment and Urbanization, 5*(2), 10-34. doi:10.1177/095624789300500203

Starfield, B. (2011). The hidden inequity in healthcare. *International Journal for Equity in Health, 10*(15) doi:10.1186/1475-9276-10-15

Stephens, C. (1996). Healthy cities or unhealthy islands? the health and social implications of urban inequality. *Environment and Urbanization, 8*(2), 9-30. doi:10.1177/095624789600800211

Stephens, C., Akerman, M., Avle, S., Maia, P. B., Campanario, P., Doe, B., & Tetteh, D. (1997). Urban equity and urban health: Using existing data to understand inequalities in health and environment in accra, ghana and são paulo, brazil. *Environment and Urbanization, 9*(1), 181-202. doi:10.1177/095624789700900115

Subramanian, S. V., Jones, K., Kaddour, A., & Krieger, N. (2009). Revisiting robinson: The perils of individualistic and ecologic fallacy. *International Journal of Epidemiology, 38*(2), 342-360. doi:10.1093/ije/dyn359

Susser, M. (1973). *Causal thinking in the health sciences*. London: Oxford University Press.

Tang, Z., & Batey, P. W. J. (1996). Intra-urban spatial analysis of housing-related urban policies: The case of liverpool, 1981-91. *Urban Studies, 33*(6), 911-936. doi:10.1080/00420989650011663

Thorndike, E. L. (1939). On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. *The American Journal of Psychology, 52*(1), 122-124.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352. doi:10.1037/0033-295X.84.4.327

Tversky, A., & Gati, I. (1978). Studies of similarity. In Eleanor Rosch, & Barbara Lloyd (Eds.), *Cognition and categorization* (pp. 1-1978) Lawrence Elbaum Associates.

Vithayathil, T., & Singh, G. (2012). Spaces of discrimination. *Economic & Political Weekly, 47*(37), 60-66.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236-244. doi:10.1080/01621459.1963.10500845

Werna, E. (1995). Urban management and intra-urban differentials in sao paulo. *Habitat International, 19*(1), 123-138. doi:http://dx.doi.org/10.1016/0197-3975(94)00060-F

Winch, R. F. (1947). Heuristic and empirical typologies: A job for factor analysis. *American Sociological Review, 12*(1), 68-75.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe Trans.). Oxford: Basil Blackwell.

Wyly, E., & DeFilippis, J. (2010). Mapping public housing: The case of new york city. *City & Community, 9*(1), 61-86. doi:10.1111/j.1540-6040.2009.01306.x

III. Densities of Bangalore

Abstract

This paper proposes a new method for high-resolution population density prediction in Indian cities. Using data for Bangalore city, the paper demonstrates that residential population density can be predicted at the scale of 30m by 30m raster cells, as a function of cell level values of street density and building height and ward level values of asset ownership. Building height data was generated from Cartosat-1 stereo imagery using an open source satellite stereo image processing software. The building height extraction was successfully carried out without the need for Differential GPS surveys and results indicate a root mean square error of approximately 3.1m. Using this building height data in conjunction with the other datasets, the paper demonstrates that a 30m resolution surface of predicted population density can be generated such that when summed to the ward level, the mean absolute percentage error between predicted population and known census population at the ward level is 10.7%. A fine-grained understanding of population densities in Indian cities, as enabled by the proposed method, can be beneficial to research, policy and practice related to urban planning.

1. Introduction

Density measures have a central role in research, policy and practice related to urban planning. In the context of cities, density is often understood as a ratio measure with the component of interest in the numerator and a unit of land area in the denominator. Researchers have examined how urban density relates to transit systems and urban resource use (see Newman & Kenworthy, 1989), urban economics and agglomeration effects (see Glaeser & Gottleib, 2009), experiential and socio-cultural aspects of cities (see Jacobs, 1961) and also to broader notions of urban sustainability (see Jenks et.al., 2003). [1]

On the policy and practice front, planning norms around the world attempt to use various forms of density regulations to control or guide urban growth (Churchman, 1999). One such tool which finds widespread use in Indian cities is Floor Area Ratio (FAR) which

[1] See Boyko & Cooper (2011) and Churchman (1999) for a detailed discussion on definitions of density and a review of research on the relationship between density and various aspects of urban life.

stipulates the ratio between total built up space and land area of a site.[2] The objective is to regulate the overall intensity of land-use in a neighborhood by limiting the total built-up area which is permitted on any site (Alexander et.al., 1988).

FAR based density regulation in Indian cities has been a cause for much debate amongst urban planning researchers and practitioners. Bertaud & Brueckner (2005) argue that FAR based regulation of job and population densities causes spatial expansion of cities which in turn leads to higher commuting and housing costs. Extending this study, Brueckner & Sridhar (2012) contend that more compact cities with closer to "international" FAR norms are beneficial from a consumer welfare perspective.

But Patel (2013) proposes that such recommendations do not take into account several factors which distinguish Indian cities from cities in developed countries. Using the example of Mumbai, Patel (2013) argues that in Indian cities more people occupy the same amount of built-up space or street space and hence increasing FAR cannot improve quality of life since it will lead to an aggravation of every form of crowding. [3] According to Shirgaokar (2013), the analysis presented by Patel (2013) could benefit from finer scale data which could help compute various metrics like office space or housing units per square kilometer.

This paper takes a step in this direction by proposing a new framework for high-resolution population density prediction. A fine-grain understanding of where the residential built-up space is and at what density people occupy it can potentially enable the computation of a whole range of spatially disaggregated metrics related to residential built-up space and per-capita access to various urban amenities and services. It can also be useful for studies related to spatial distribution of demand for various resources within a city, assessment of living conditions within and across cities and for studies related to disaster risk reduction.

The rest of this paper is structured as follows. In the next two sub-sections I provide a brief overview of various types of density, before focusing on definitions and issues related to 'measured density'. After this I discuss some of the predominant methods used for high-resolution population interpolation, redistribution and prediction. In Section 2, I describe the various datasets used in the proposed population prediction framework and the steps involved in preparing the datasets. Section 3 provides an overview of the prediction framework while Section 4 describes the results obtained for Bangalore city. I conclude with a discussion on the advantages and shortcomings of the proposed method.

[2] FAR is often referred to as Floor Space Index (FSI) also in Indian cities. If a plot has an FAR or FSI of 1, this means the maximum permissible built-up area on that plot is equal to the total area of the plot.
[3] The concept of crowding is discussed in greater detail in Section 1.2

1.1 Types of density

In spite of the centrality of the concept of density in research, policy and practice related to cities, there is much debate about its definitions and how it can be used to understand and compare across cities and neighborhoods. First, there is the question of the parameter or component whose density we are interested in. For example, it could be people, built up space or dwelling units. Second, one has to clarify what type of density we are interested in. It could be measured density, which can be understood to refer to the objective physical components of the built environment (including people) which can be measured (Alexander et.al., 1988) or perceived density, which refers to the density perceived by people (Rapoport, 1975).

Rapoport (1975) persuades us to go beyond simple ratio formulations of density and place perceived density at the core of discussions about urban density. Alexander et.al. (1988) provides a framework for this by conceptualizing perceived density as being influenced by measured density, qualitative physical features of the built environment, individual cognitive factors and other social and cultural factors.

In the context of Indian cities, measured density itself is not very well understood since the data from the census is not made available below the ward level—a ward being the administrative subdivision below that of the city corporation. My attempt in this paper is to provide a predictive framework which will enable a high-resolution understanding of measured population density in Indian cities.

1.2 Measured density: definitions and issues

Units of measured density

Measured density is usually expressed as a ratio, where the denominator is some unit of area. The numerator can be any attribute of interest like population, built-up space, households, dwelling units etc. The unit of area in the denominator encapsulates two aspects of measurement—one is the unit of measurement (sq.m., sq.km. etc.) and the other is the type of land-uses considered for measurement (Alexander et.al., 1988; Churchman, 1999; Boyko & Cooper, 2011). Both of these aspects need to be defined for the measured density to make sense. For example, population density can be measured at the city level and expressed as people/sq.km. But the sq.km. unit in the denominator could refer to sq.km. of total land area, sq.km. of land with residential land-use including minor streets, or sq.km. of land with residential land-use excluding streets and pavements. [4]

[4] Alexander et.al. (1988) defines net residential area, gross residential area, neighborhood area and city area, depending on the types of land-uses which are considered while defining the area unit in the denominator.

Types of measured density

While the denominator in the above example refers to land area outside the dwelling unit, this need not always be the case. For instance, when the numerator is population or persons, the denominator could also be area available within dwelling units. For this reason, Alexander et.al. (1988) distinguishes between two types of measured density—molecular and molar. Molecular measured density refers to people per area of dwelling space while molar measured density refers to population, built-up space, dwelling units etc. divided by the external land area of interest. [5] These are also referred to as internal and external density in the literature (Yeung, 1977).

Measured density and Modifiable Areal Unit Problem

The term Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984) refers to the following problem—the same spatial data when aggregated to different areal (spatial) units leads to very different understandings of reality. [6] For example when population data—which can be thought of as point data—is aggregated to census tracts or wards to measure population density, depending on how the tract or ward boundary is drawn, the same location may appear to have very different average population densities.

There is a version of the MAUP which comes into play when measuring population density at the city scale also. Depending on how the boundary of a city is defined, the measured population density can be dramatically different. For example a city may have several outer wards which are sparsely populated, which if included in the calculation, will bring down the average city level population density dramatically. To somewhat compensate for this, Barnes (2001) proposes the concept of weighted population density, where each areal unit with known population is weighted by the fraction of the total city population which resides in it. Barnes (2001) claims that for US cities this weighted population density is closer to what the residents of the city on average perceive.

1.3 High-resolution population interpolation, redistribution and prediction

One approach which can somewhat circumvent the MAUP involves representing population as a continuous surface which is not dependent on arbitrary areal units (Mennis, 2003). The methods which attempt to provide such a finer grain understanding of urban population distribution can be grouped into three—point interpolation based methods, population redistribution methods and methods for population prediction based on ancillary data.

---

[5] The concept of crowding is sometimes used to refer to what can be better described as molecular or internal measured density (Rapoport, 1975; Patel, 2013). But as Stokols (1972) and Alexander et.al. (1988) explain, crowding is an experiential state based on the relationship between perceived density and desirable norms.

[6] Areal unit refers to spatial unit of analysis. 'Modifiable areal unit' refers to the potential for re-aggregating or distributing spatial data from a source (original) unit of aggregation to a target (new) unit of aggregation or distribution.

Point interpolation methods

In this paper I use the term 'point interpolation methods' to refer to the application of point interpolation techniques in generating a continuous population distribution surface based on source data which is aggregated to arbitrary areal units. For example, in the case of urban population data which is available at the level of wards, the population within each ward could be assigned to its centroid. Point interpolation techniques can then be applied based on the values at the centroids to generate a continuous surface of population distribution. As pointed out by Lam (1983) such an approach has several drawbacks. Most importantly, the total population value within the boundary of the source areal unit, as depicted by the continuous surface, will differ from the original known population of the source areal unit. In the literature, interpolation methods which have this problem are referred to as 'non volume preserving' (Lam, 1983).

Population redistribution methods

In this paper I use the term 'population redistribution methods' to refer to what is otherwise variously called 'volume preserving' (Lam, 1983) or pycnophylactic (mass preserving) (Tobler, 1979; Yoo et.al., 2008) interpolation methods in the literature. Both 'volume preserving' and 'pycnophylactic' refer to areal interpolation methods where the total value of the variable within the source areal unit boundary does not change after interpolation.[7] For the purpose of this paper, I find the term 'population redistribution' to be a simpler and more straightforward way of referring to such interpolation methods since they essentially redistribute the population of a source areal unit within its own boundaries.

Population redistribution methods can be grouped into two, based on whether it uses only statistical areal interpolation techniques or it uses ancillary information also. The smooth pycnophylactic interpolation technique described by Tobler (1979) is one well-known method which can be applied without any ancillary information. Areal weighting methods which overlay target and source areal unit (or source zones) and apportion population to target areal units (or target zones) using a weighting scheme based on area of the source unit/s contained within it, also come under this category (Lam, 1983). While the areal weighting method assumes homogeneity in distribution of the variable of interest (in this case population) within the source zone, smooth pycnophylactic interpolation assumes heterogeneity in distribution of the variable within the source zone (Hawley, 2005).

Population redistribution methods which use ancillary data to create a population density surface is referred to as dasymetric mapping (Semenov-Tian-Shansky, 1928; Wright, 1936; Mennis, 2009; Petrov, 2012). The term dasymetric was invented by Benjamin Semenov-Tian-Shansky and translates approximately to 'measuring density' in Greek (Petrov, 2012).

[7] Areal interpolation refers to the transfer of information from one set of source areal units to a set of target areal units (Fisher and Langford, 1996).

Dasymetric mapping as proposed by Semenov-Tian-Shansky retains the 'volume preserving' or pycnophylactic property since the population within a known spatial unit is only redistributed to locations of concentration as inferred from the ancillary data (Petrov, 2012; Mennis, 2009). For example the known population within a ward can be redistributed to only the built-up areas within the ward, or the residential built-up areas within the ward. The concept of dasymetric mapping has also been applied to generate global scale population density maps such as Landscan (Dobson, 2000).

Numerous authors have conducted dasymetric mapping studies using ancillary data like land-cover(Langford & Unwin, 1994; Holt et.al., 2004), land-use (Maantay et.al., 2007) LIDAR based building height (Xie, 2006); and street network (Xie, 1995; Reibel and Bufalino, 2005; Long & Shen, 2014).

Population prediction based on ancillary data

In the literature, methods for high-resolution population density prediction that use ancillary data along with census population figures are also often referred to as dasymetric mapping (Mennis, 2009; Hawley, 2005). But in this paper I distinguish between dasymetric methods of population redistribution (volume preserving) and predictive frameworks in which the volume-preserving property is absent (Mennis, 2009).

Compared to the large number of studies which use dasymetric mapping methods, there has been limited work on high-resolution population density prediction. Li & Weng (2005) propose a population prediction framework using spectral values of pixels in multi-spectral remote sensed data while Silvan-Cardenas et.al. (2010) use LIDAR based building height data to predict population. Bast et.al. (2015) demonstrate a population prediction approach which uses various crowd sourced datasets from the Open Street Map project.

2. Data

This paper applies a new predictive framework for generating a 30m resolution population density surface for Bangalore city. The proposed method uses data on land-cover, land-use, building height, street network and asset ownership information to predict population. The 30m resolution is primarily dictated by the lowest resolution dataset—the Landsat data used for land-cover mapping (U. S. Geological Survey, 2015). Besides, the Shuttle Radar Topography Mission (SRTM) data (Farr et.al., 2007)—which, as described in Section 2.4, was one of the datasets used to generate building height maps—is also of 30m resolution.

The municipal corporation of Bangalore was called Bangalore Mahanagara Palike (BMP) till 2007. The BMP boundary was expanded by government notification on 16 January, 2007 to include seven City Municipal Corporations, one Town Municipal Corporation and 110 villages which were outside the BMP boundary. This expanded entity was reconstituted as

the Bruhat Bangalore Mahanagara Palike (Government of Karnataka, 2007).[8] Today it contains 198 wards and is called the Bruhat Bengaluru Mahanagara Palike or BBMP, which roughly translates to Greater Bangalore Municipal Corporation.[9]

While the initial intent was to apply the predictive framework to the region within the current administrative boundary of Bangalore (BBMP boundary), available land-use data was accurate only for the region within the 2007 administrative boundary of Bangalore (BMP boundary). As a result, the current research focuses on the areas which are within the BMP boundary. Figure1 shows all the 198 wards of BBMP (Bruhat Bengaluru Mahanagara Palike, 2016) overlaid with a population density choropleth (population in a ward/ ward area) for BBMP wards which fall within the boundary of the erstwhile BMP.

2.1 Preparation of land-cover data

Land-cover map for the BBMP area was prepared using 30m resolution Landsat 5 (U. S. Geological Survey (USGS), 2015) data from 18 January 2011, downloaded from the USGS EarthExplorer website. All seven bands of the dataset were used to conduct an unsupervised classification into 50 classes using ArcGIS 10.2. This was then reclassified into Water, Vegetation, Built-up and Vacant land-cover classes (Fig. 2) using Google Earth imagery from 11 March 2011 as a ground truth reference class (Google Earth, 2011). [10, 11]

Accuracy assessment of land-cover classification

To assess the accuracy of the land-cover classification, 400 raster cells were randomly selected (384 cells required for 95% confidence level and 5% error) converted to polygons and transferred into Google Earth. The predicted classification as obtained using ArcGIS was compared with actual land-cover classes observed in Google Earth for these 400 cells. Based on this, an error matrix was computed (Table 1) and overall accuracy was estimated to be 85% (Foody, 2002).

General problems with accuracy assessment of land-cover classification

Based on these measures, although the classification is accurate enough for the purpose of his paper, it is important to mention some of the issues faced while conducting this accuracy assessment. As pointed out by Foody (2002), there could be problems of

[8] Bangalore is the capital of the state of Karnataka

[9] In 2014, the name of the city was changed from Bangalore to Bengaluru (Times News Network, 2014). Since the analysis presented is based on Census data from 2011, I will refer to the city as Bangalore.

[10] All types of buildings (residential, industrial, commercial etc.) and streets were classified as built-up. As described in Section 2.2, the land-use map was then used to identify residential built-up cells from within this built-up class.

[11] Streets are classified as built-up. But major streets will be removed at a later stage of analysis when I intersect the built-up map with the residential land-use map, to extract residential built-up cells. This is because the land-use map represents streets as the gaps between land parcels and so during the intersection step, the cells which fall in these gaps between land-use parcels will be ignored.

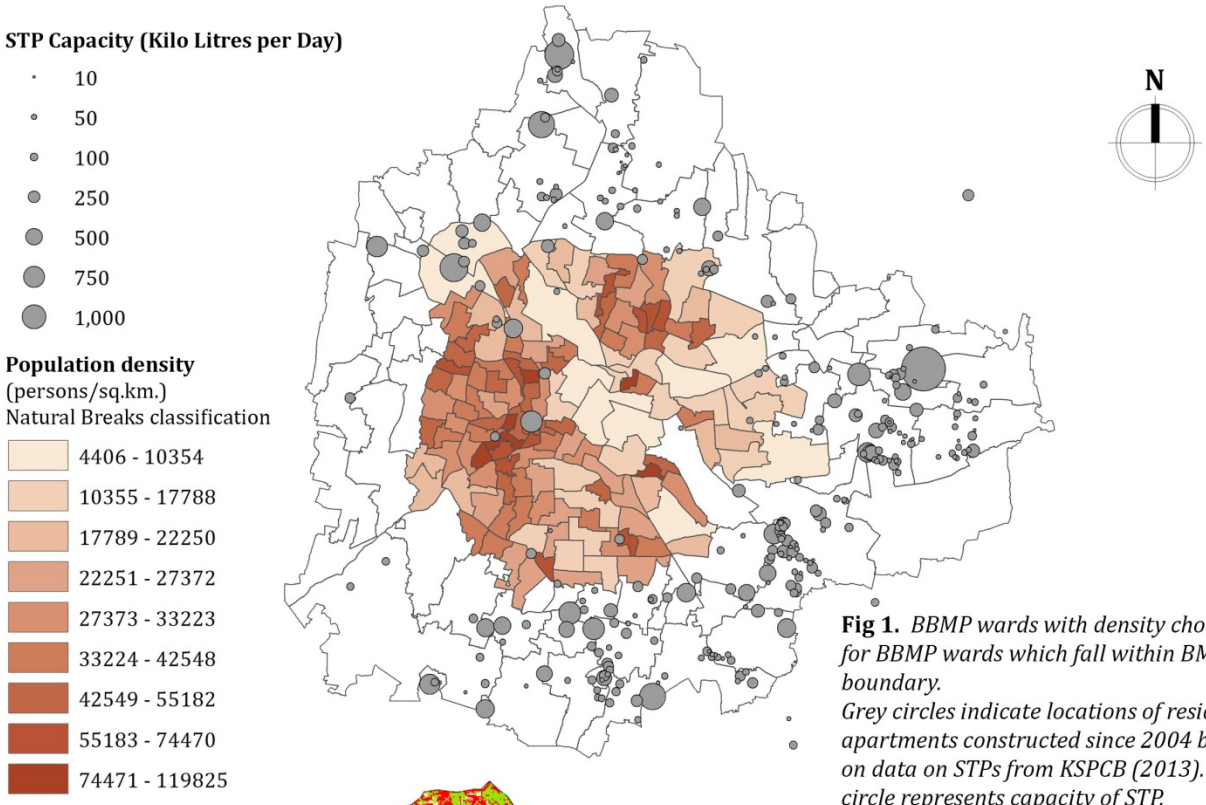


Fig 1. *BBMP wards with density choropleth for BBMP wards which fall within BMP boundary.*
Grey circles indicate locations of residential apartments constructed since 2004 based on data on STPs from KSPCB (2013). Size of circle represents capacity of STP.

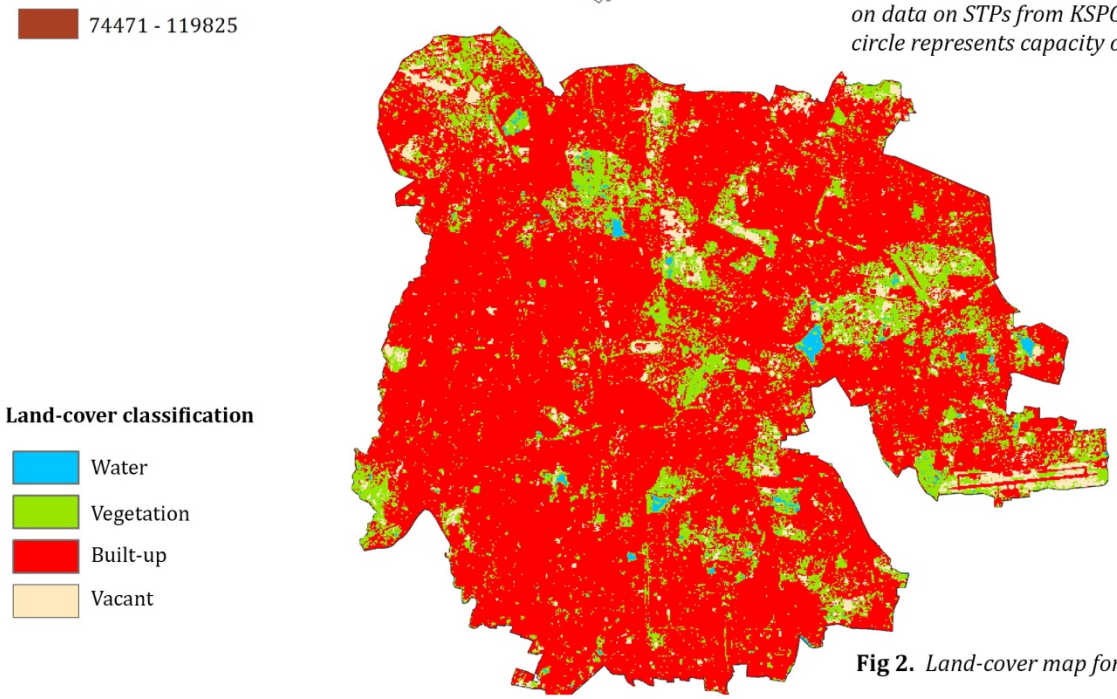


Fig 2. *Land-cover map for BMP area*

		ACTUAL				
		Water	Vegetation	Built-up	Vacant	Predicted total for each class
PREDICTED	Water	6	3	1	0	10
	Vegetation	0	122	11	5	138
	Built-up	1	11	169	6	187
	Vacant	0	12	10	43	65
	Actual total for each class	7	148	191	54	400

	Water	Vegetation	Built-up	Vacant
Agreement / Accuracy	86%	82%	88%	80%
Omission error	14%	18%	12%	26%
Commission error	40%	12%	963%	34%
Overall accuracy	85%			

Table 1. *Error matrix and accuracy assessment of land-cover classification*

mis-registration between reference imagery and the Landsat data, especially since Google Earth imagery exhibits slight lateral shifts between imagery acquired on different dates.

Besides this, since each cell in the Landsat data is 30mX30m, when conducting the accuracy assessment I was checking the land-cover visible in Google Earth imagery within each randomly selected 30mX30m cell. In many cases one such cell contained two or more land-cover classes in the Google Earth imagery. For example a cell may contain water, vegetation and vacant land. In such cases whichever land-cover class appeared to cover maximum extent of the cell was assigned as the 'true' land-cover status of that cell.

If one were to conduct an accuracy assessment by inspecting the land-cover classes within each cell by visiting the specific cell locations on ground, the problem seems even more difficult to surmount, since there will be errors related to determining the boundary of each cell on ground before even getting to issues arising from the presence of multiple types of land-cover within the same cell. Moreover, it is difficult to gain access to inspect areas within a randomly selected set of cells. As a result only cells which fall within publicly accessible areas may get surveyed leading to a biased sample.

In general, my experience with attempting to conduct accuracy assessment makes me skeptical of land-cover classification accuracy assessment results which are often reported in research papers—including mine.

2.2 Preparation of land-use data

The data on existing land-use available with the local planning authority (the Bangalore Development Authority or BDA), was from 2004 (Fig. 3). The subsequent round of existing land-use survey is currently in progress as of January 2016 and hence was unavailable for the analysis presented in this paper. While the land-use data is from 2004, the population data available from the Census of India (Census of India, 2011a), is based on population enumeration conducted in 2011 (Census of India, 2011b). This effectively means that the year of population enumeration by the Census falls in the middle of the period between two rounds of land-use survey.

One option to deal with the difference in years of Census population enumeration and land-use survey was to interpolate the population numbers for 2004 based on data from the 2001 and 2011 Census data. But this was impractical since the ward boundaries for the Census population numbers from 2001 were not available. As a result, the only feasible way of using land-use in this research was to update the 2004 land-use data to reflect the changes up till 2011. This also meant that it was more appropriate to focus the study on the BMP boundary since this area has seen relatively less land-use change compared to the peripheral parts of the city which were added to the BMP area in 2007 to create the BBMP jurisdiction.

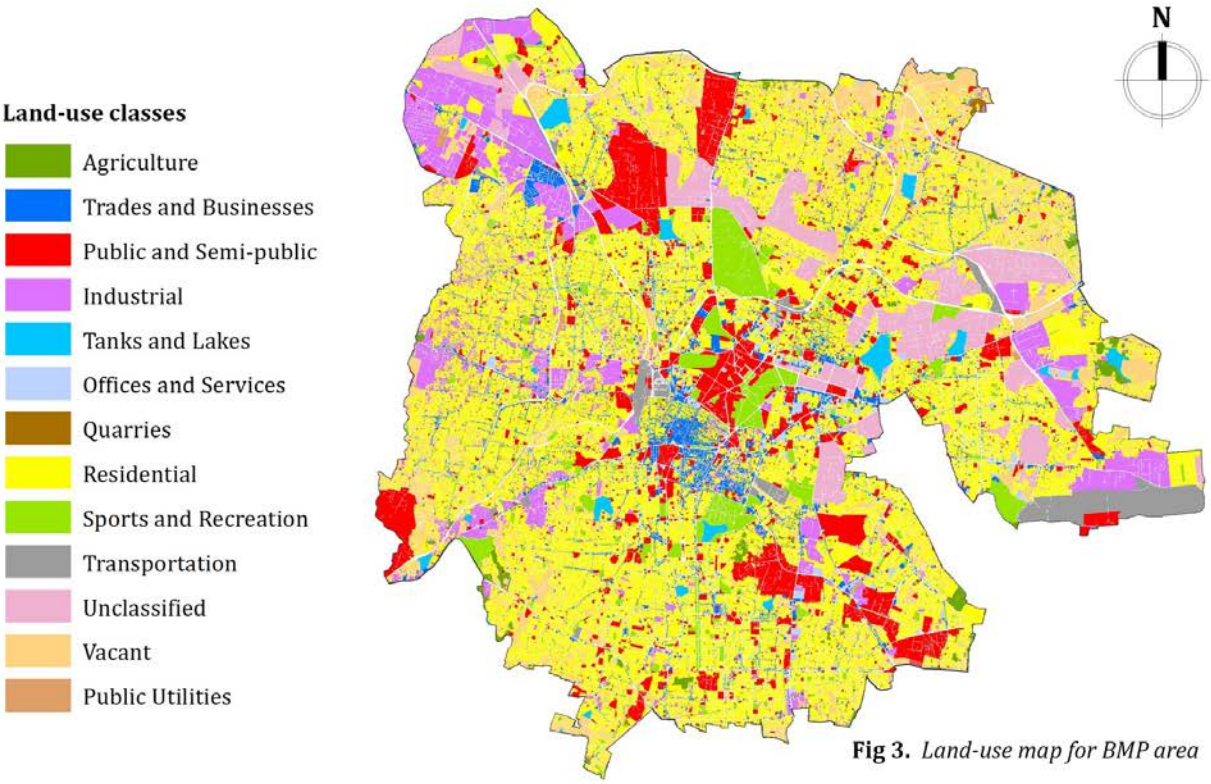


Fig 3. *Land-use map for BMP area*

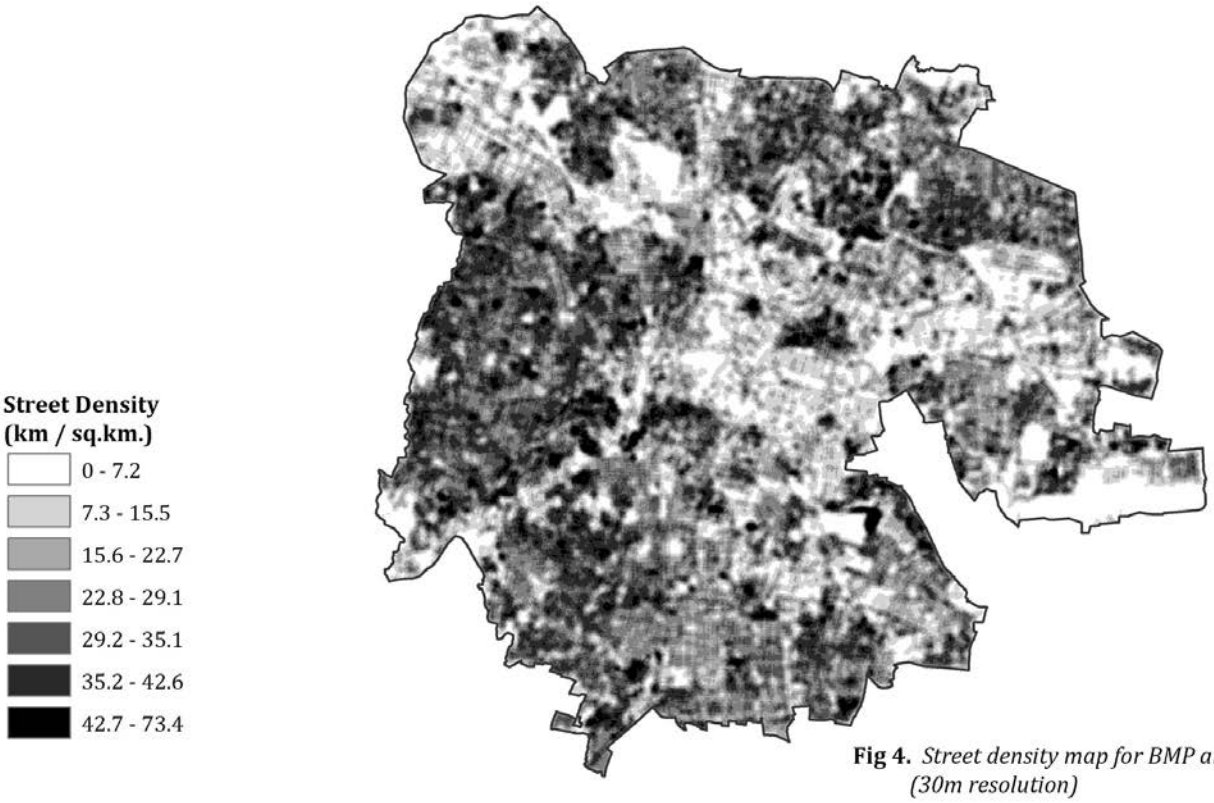


Fig 4. *Street density map for BMP area (30m resolution)*

2.2.1 Steps involved in updating land-use map

As a first step towards updating land-use data from 2004, the study assumed that all parcels which were classified as having residential land-use in 2004 remained residential in 2011 also. The next step was to systematically evaluate all changes from non-residential uses to residential over the period from 2004 to 2011. Out of these, two types of non-residential to residential land-use change are especially significant to this study—those involving the creation of apartment buildings or dense informal settlements. These are significant to high-resolution density studies since they both represent high population density raster cells, which if missed can lead to large errors.

Adding new apartment buildings to the land-use map

A dataset related to sewage treatment plants in large housing developments, prepared by the Karnataka State Pollution Control Board (KSPCB) was used to understand where the new high-rise apartments have been built. In 2004, a new rule stipulated that all new apartment buildings of more than 20,000 sq.m. total built-up area in parts of the city with sewerage network, or apartment buildings of more than 5000 sq.m. total built-up area in parts of the city without sewerage network need to install small scale sewage treatment plants (STPs) (KSPCB, 2013). The housing developers were required to take permission from KSPCB before installing STPs. In 2013, KSPCB prepared a report which provided information about all STPs for which permission had been granted since 2004 (KSPCB, 2013).

Using this dataset, a total of 285 residential apartment buildings were geocoded. Out of these only 30 were within the BMP area (Fig. 1), which supports the earlier assertion that much of the land-use change since 2004 has happened outside the BMP area. The remaining 30 apartment buildings were visually inspected using Google Earth imagery from 23 January, 2010 (Google Earth, 2010) to check if their construction was complete by this date.[12] The completion dates of these projects were checked using data from real estate websites also. Based on this, only 8 apartment buildings were found to be complete within the BMP area by early 2010. Therefore these parcels were added to the land-use map as residential parcels.

Adding informal settlements to land-use data

According to the Karnataka Slum Development Board (KSDB), there are 542 slums or informal settlements within the Bangalore Urban District (KSDB, 2014) —the urban district being a significantly larger area than the BBMP boundary. Out of this list, 105

[12] Although the census population enumeration was conducted in February-March 2011 (Census of India, 2011b) the date of 23rd January 2010 was used as a cut-off date to determine the changes to be made to the land-use map since this was the closest date for which Google Earth imagery was available for the entire study area. Besides large apartment buildings are known to take time be fully occupied.

settlements within BMP area were identified and their boundaries demarcated and ground-truthed by a planning consultancy firm in Bangalore named INDE. [13]

All 105 settlements for which boundaries were available were georeferenced in ArcGIS and then exported to Google Earth to check for errors. Visual inspection of settlement grain in Google Earth imagery (Google Earth, 2010), was used to confirm that settlement boundaries were accurate. These polygons were then added as residential land-use parcels to the land-use map from 2004.

Checking for conversion from non-residential to residential land-use

The list of land-use classes as per the land-use map of 2004 can be seen in Figure 3. The land-cover map was used to do a preliminary assessment of land-use change in the Agriculture and Vacant classes since if these classes have not witnessed any land-use conversion, they should not have any built-up cells within them. The comparison between land-use and land-cover maps showed that there was significant land-use conversion from 2004 to 2011 within these land-use classes. As described below, several steps were then carried out to identify the residential built-up cells within these two land-use classes.

For the Agriculture category, all the parcels from the land-use map, which were greater than one cell in area (30mX30m = 900 sq.m.) were exported to Google Earth (142 polygons). Each polygon was checked for land-use conversion in Google Earth using imagery from 23 January 2010. If there was conversion to built-up, then the overall size and texture of the built-up area was used to estimate whether it was residential land-use. Based on this a new layer of polygons was generated which consisted of all the areas which had changed from agricultural to residential land-use.

Since the Vacant category consisted of 23481 parcels it was not feasible to inspect each parcel in Google Earth as described above. Instead, polygons were created from built-up cells of the land-cover map which fell within Vacant land-use class. Out of the resulting 7395 polygons, all polygons greater than 8100sq.m. in area (nine 30X30 cells), were selected and exported to Google Earth (304 polygons). Like in the case of Agriculture category, these 304 polygons were also inspected in Google Earth using imagery from 23 January 2010 and corrected based on whether they had changed to residential land-use.

Since it was infeasible to examine each of the remaining 7091 polygons (of area less than 8100 sq.m. each) which contained areas where Vacant parcels had been built upon, these were assumed to have residential built-up. To understand the extent of errors this assumption may cause, a random set of 411 cells (384 cells required for 95% confidence level with 5% error) were selected and exported to Google Earth for visual inspection.

[13] The principal of INDE, Mr. Mohan Rao, generously shared the data on these settlements for use in this research.

Based on this it was estimated that 75.8% of the cells had converted to residential built-up. This effectively meant that the remaining 24.2% cells were assigned residential land-use erroneously. Since most of the Vacant land-use parcels are to the periphery of the BMP area, this may cause peripheral wards to show significant errors in density estimation.

In the Trades & Businesses category (which consists of commercial land-use parcels), all parcels which were less than one cell in area (900 sq.m.) were considered to be residential. This is because a lot of the smaller commercial establishments have residential uses on upper levels and the land-use map does not clearly categorize such mixed-use parcels. While the assumption that all parcels less than one cell in area in this land-use class is residential is a significant approximation, it helps capture some of the residential use in mixed land-use parcels which would otherwise be completely missed. Since smaller commercial establishment were found to be somewhat uniformly distributed across all wards, the errors should also be distributed uniformly.

The categories of Sports and Recreation, Transport, Public-utilities (like water and electricity), Tanks & Lakes, Quarries, Industrial and Offices & Services categories were assumed to not have any residential built-up. While it is known that Industrial land-use does get converted into residential land-use, such land-use conversions occur mostly in the case of large apartment buildings which were already captured using the dataset on STPs described earlier.

The Public/Semi-public and Unclassified categories presented a special kind of problem. The former includes all government institutions from the legislative assembly to educational and research institutions, while the latter consists mainly of land which is controlled by the defence forces and related research institutions. Both these categories have residential areas within them, but it is not possible to identify where these residential areas are within these land-use parcels. Therefore, all the built-up cells within these two categories were assumed to have some level of residential population. Only the most obvious exceptions like the state legislative assembly campus were excluded from this. [14] Since the land-use classes of Public/Semi-public and Unclassified essentially contain government Institutions, for ease of communication, in the rest of this paper I will use the term 'Govt. land-use' to collectively refer to both these categories.

2.3 Preparation of street network data

Street network data was obtained from the local water utility as pdf files and converted to a polyline feature class in ArcGIS format. This was then georeferenced using the land-use map layer from the planning authority. The line density function in ArcGIS with search radius of 90m was used to generate a 30m resolution street density map (Fig. 4). The

[14] Even in this case there may be few residences for the legislative assembly staff – but compared to the total built up it was assumed to be very insignificant and hence neglected.

search radius of 90m was used so that the street density of any given cell is an average of the street network density in its immediate vicinity. If the search radius is made very small, then in neighborhoods with a sparse street network, a cell which has a street passing through it will get a high value, while an adjacent cell through which the street does not pass may get a considerably lower value—in other words, the decay in the influence of a street would be very rapid with increasing distance from it. Thus a larger search radius enables the influence of a street to decay less abruptly—which I propose is closer to what happens in reality.

2.4 Extraction of building height data from satellite stereo imagery

The term 'satellite stereo imagery' refers to image pairs of the earth surface generated by satellites which are equipped to capture multiple images of the same earth surface region from different points in space (Poli and Caravaggi, 2012). Given information about the camera sensor, its orientation and satellite location, it is possible to calculate the relationship between points on the earth surface and pixels on the image. Once this is done, the shift or disparity that any point located on the earth surface (or on objects on the earth surface) exhibits when viewed from two different locations in space can be computed using the stereo image pair. These disparity maps can be processed to calculate the elevation of each point, thereby yielding a model of the earth surface which includes the terrain and all other objects like buildings and vegetation which may exist on it (d'Angelo et.al., 2010). In contrast to a Digital Elevation Model (DEM) which captures only the terrain, a model of the earth surface which includes the terrain and all objects which exist on the terrain is referred to as a Digital Surface Model (DSM) (U. S. Geological Survey, 2016).

Building height extraction from satellite stereo imagery involves the following key steps:
1. Generation of the DSM,
2. Subtraction of the terrain (Digital Elevation Model or DEM) to get the normalized DSM or nDSM—which consists of all objects that extend above the terrain.
3. Removal of vegetation from the nDSM to obtain only building heights

Description of stereo imagery

The Indian Space Research Organization (ISRO) has a high-resolution satellite with stereo imaging capabilities called Cartosat-1. The satellite has two 2.5m resolution panchromatic cameras—one facing forward (Fore) by 26^{o} from the vertical and another facing backwards (Aft) by 5^{o} from vertical as shown in Figure 7 (National Remote Sensing Centre, 2015). As the satellite travels in its orbit at a height of approximately 618 km, first the Fore camera starts capturing a scene on the ground and after about 50 seconds of orbital travel, the Aft camera starts capturing the same scene from the new position of the satellite (National Remote Sensing Centre, 2015). Together the Fore and Aft images comprise the stereo pair for a single ground scene.

The National Remote Sensing Centre (NRSC) provides Cartosat-1 imagery of 2.5m resolution for a fraction of the cost of other high-resolution stereo imaging satellites. Two image pairs from January 2012 were purchased from NRSC to cover the BMP area. Each image comes with relevant metadata and a Rational Polynomial Coefficient (RPC) file. The RPC file provides information on the sensor type, sensor orientation and satellite location at the time of image capture (National Remote Sensing Centre, 2015).

Method for DSM and nDSM generation

In the literature, building height extraction using stereo imagery is usually accomplished using proprietary photogrammetry software (see Shaker et.al., 2011). These need manual intervention at various stages of processing and use highly accurate Ground Control Points (GCPs) measured using a Differential GPS (DGPS) to correct for any errors which may exist in the RPC files. In 2014, a new open source software called S2P (Satellite Stereo Pipeline) was developed which enables fully automated processing of satellite stereo imagery (de Franchis, et.al., 2014) without the need for GCPs obtained through DGPS surveys. But the lack of accurate GCPs may lead to errors in both georeferencing and height estimates in the output DSM and nDSM. Subsequent sections describe the nature of errors encountered and the calibration and validation steps which were taken to correct them.

5m resolution DSMs were generated for each Cartosat stereo image pair using S2P. Both the DSMs did not align accurately with known georeferenced layers of data like the street network. This is because RPC files are not accurate enough to enable precise geolocation of processed DSMs (C. de Franchis, personal communication, 13 January 2016). Both the DSMs were realigned using clearly visible building edges and street intersections from the street network layer. After this, a 5m DEM was generated by interpolating the 30m SRTM data using bilinear interpolation in ArcGIS 10.2.[15] This DEM was subtracted from the DSM layer to generate the nDSM. As shown in Figure 5, both the nDSMs exhibited a clear spatial bias in terms of height values. According to de Franchis (personal communication, 13 January 2016) and Titarov (2008) this is also due to RPC errors.

It was not possible to remove trees from the nDSM at this resolution due to the lack of 5m resolution land-cover or vegetation map. As described in subsequent sections, vegetation removal was achieved using the 30m land-cover map after resampling the nDSM also to 30m resolution. This is a potential source of errors in the overall population density estimation method.

[15] Bilinear interpolation uses known values from the four nearest input cells to interpolate the value at a given point location or output cell. The output value is determined as a weighted distance average of these known values (Environmental Systems Research Institute, 2011).

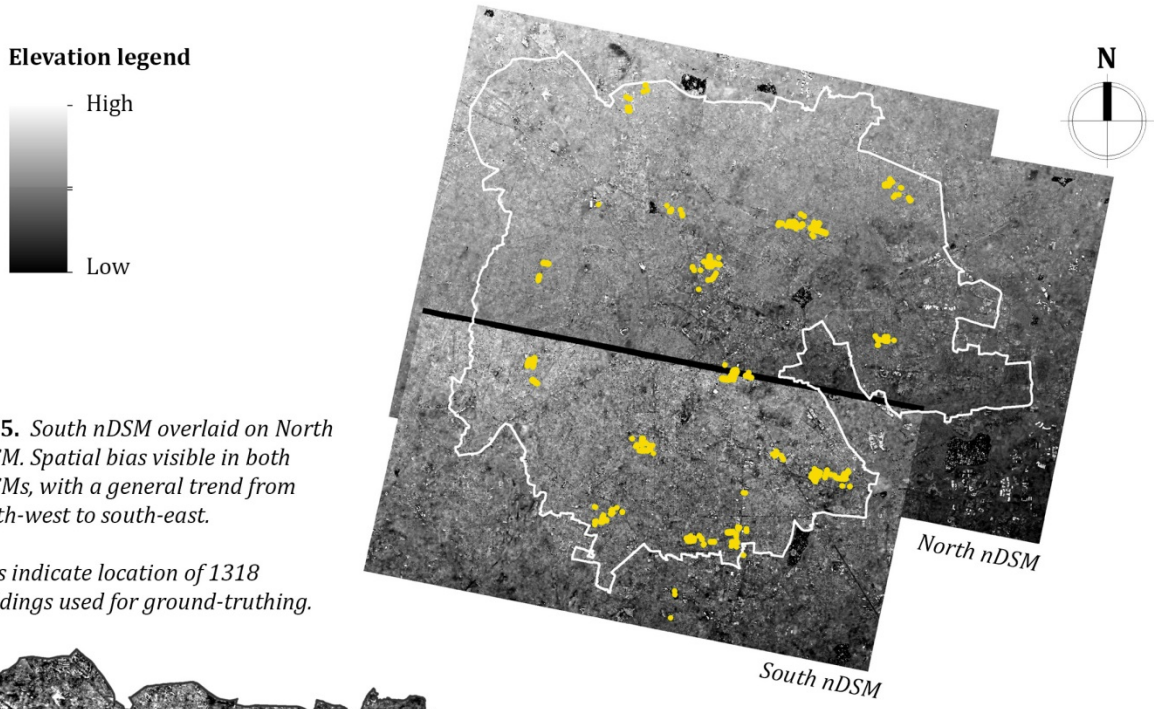
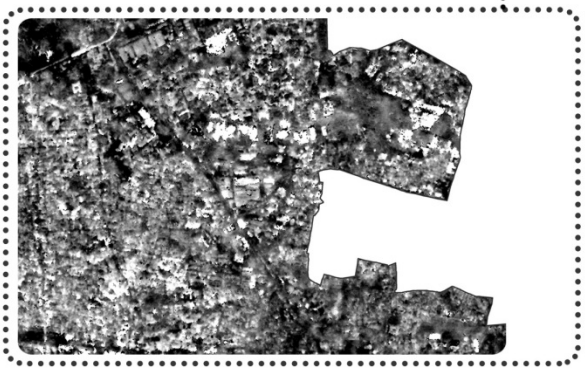


Fig 5. *South nDSM overlaid on North nDSM. Spatial bias visible in both nDSMs, with a general trend from north-west to south-east.*

Dots indicate location of 1318 buildings used for ground-truthing.

Fig 6. *Corrected nDSM for BMP area (5m resolution)*



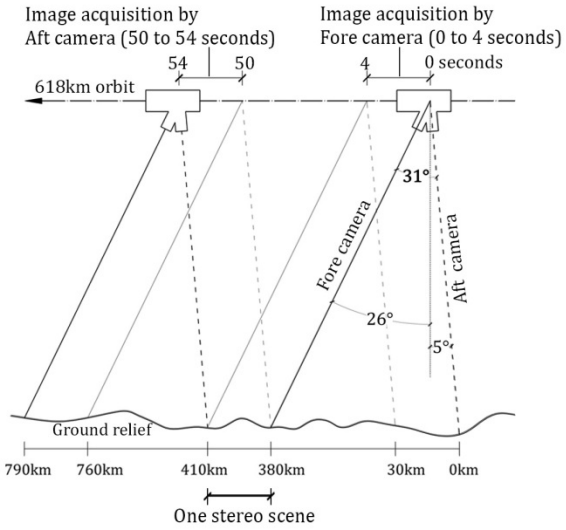


Fig 7. *Schematic diagram showing stereo image acquisition by Cartosat-1 Fore and Aft cameras. (Based on Evans et. al., 2008)*

Fig 8. *Closer view of region within dotted line in Fig.6, showing individual buildings at a resolution of 5m.*

Ground truthing

To make up for the lack of GCPs surveyed using DGPS, this paper uses a set of reference building heights to calibrate the DSM by removing the spatial bias. A different set of reference building heights were then used for accuracy assessment of the building height prediction. The reference building heights were collected by visiting 14 different neighborhoods across the city and counting the number of occupiable floors in a sample of buildings in each neighborhood. Reference building heights were collected for a total of 1318 buildings. As shown in Figure 5, the neighborhoods were chosen such that they were well distributed across the entire study area and also covered the various settlement types which are seen across the BMP area.

The height estimation was done by counting the number of occupiable floors since physical measurement of building height was not feasible for a large enough sample. All minor constructions on building roofs like water tanks, temporary roofs etc were ignored since the objective here is to estimate the volume of space available for people to live in. The number of occupiable floors was then written down for each building on a hard copy of a digitized building footprint map of the neighborhood.

A new point feature class was created by placing a point at the approximate centroid of each building footprint polygon for which the floors had been counted. Each point was then assigned the respective floor numbers counted during the building height survey. The floor count value was then multiplied by an average floor height of 3.25m to get the height estimate for each building.[16] This was deemed to be the actual height of each building.

Calibration

To remove the spatial bias and calibrate the northern nDSM, a set of 123 random points which fell within the boundary of this nDSM was selected from the point feature class mentioned above. The building height predicted by the nDSM at each of these points was extracted with bilinear interpolation using the Extract Values to Points tool in ArcGIS.

The difference between the actual building heights from the ground truth survey and the predicted building heights from the nDSM was used to calculate a 2-degree polynomial trend surface of best fit. This trend surface was then subtracted from the northern nDSM to generate a calibrated nDSM without spatial bias in elevation values. A similar procedure using a different set of 123 random points was then carried out for the southern nDSM.

Accuracy assessment

Accuracy assessment was carried out for both the nDSMs using the remaining points from the ground truthing survey. 509 points were used for the northern nDSM and 635 points

[16] Based on my experience as a practicing architect, height of a floor in most buildings in Indian cities are in the range of 3 to 3.5m. Hence 3.25 was used as average height.

were used for the southern one for this purpose. In the case of the northern nDSM, the predicted and actual building heights had a correlation coefficient (R) of 0.87 and a Root Mean Square Error (RMSE) of 3.12m. The southern nDSM had an R value of 0.89 and RMSE of 3.06m

An RMSE value of 3.06 – 3.12m means that overall the model is able to predict the number of occupiable floors with an error of about one floor. This indicates that the calibrated nDSM is performing sufficiently well for the purposes of this research. Given that this level of accuracy has been possible without the use of expensive DGPS survey, S2P output calibrated with ground truthed building heights appears to be a very cost effective means of generating nDSMs of urban areas using Cartosat-1 imagery.

3. Methods

The previous four sections (2.1 to 2.4) described how the land-cover, residential land-use, street density and building height maps for the BMP area were prepared. This section describes the method followed in developing the population prediction model.

3.1 Identify residential built-up cells within BMP area

Primary Residential Built-up Cells

Since the land-cover dataset shows us what is built-up and the land-use dataset shows us what is residential, by intersecting the two it is possible to extract the built-up cells which are residential in land-use (Fig. 9 & 10). These cells can be called primary residential built-up cells (Primary RBCs) and the population of these cells can be estimated as described in Sections 3.3 to 3.5.

Government Residential Built-up Cells

As described earlier, since it is not possible to identify which parts of the Govt. land-use parcels may have residential population, all of it is assumed to have some level of residential use. Therefore the Govt. land-use category is intersected with the built-up land-cover to extract the Govt. built-up cells which can be called Govt. RBCs (Fig. 9 & 10). Since there is no information about the number of people living in these cells, as described in Section 4, a constant population per cell is assigned to these cells.

Informal Residential Built-up Cells

Since the population in a cell is to be predicted based on the street network density and building height of a cell, it is necessary to separate out a third kind of residential built-up cells. These are cells within the known informal settlement boundaries which are built-up but happen to have very low street network density. This is necessary because, very dense informal settlements often do not have any streets which find representation in the street network dataset. As discussed in Section 3.3, lower street network density may indicate

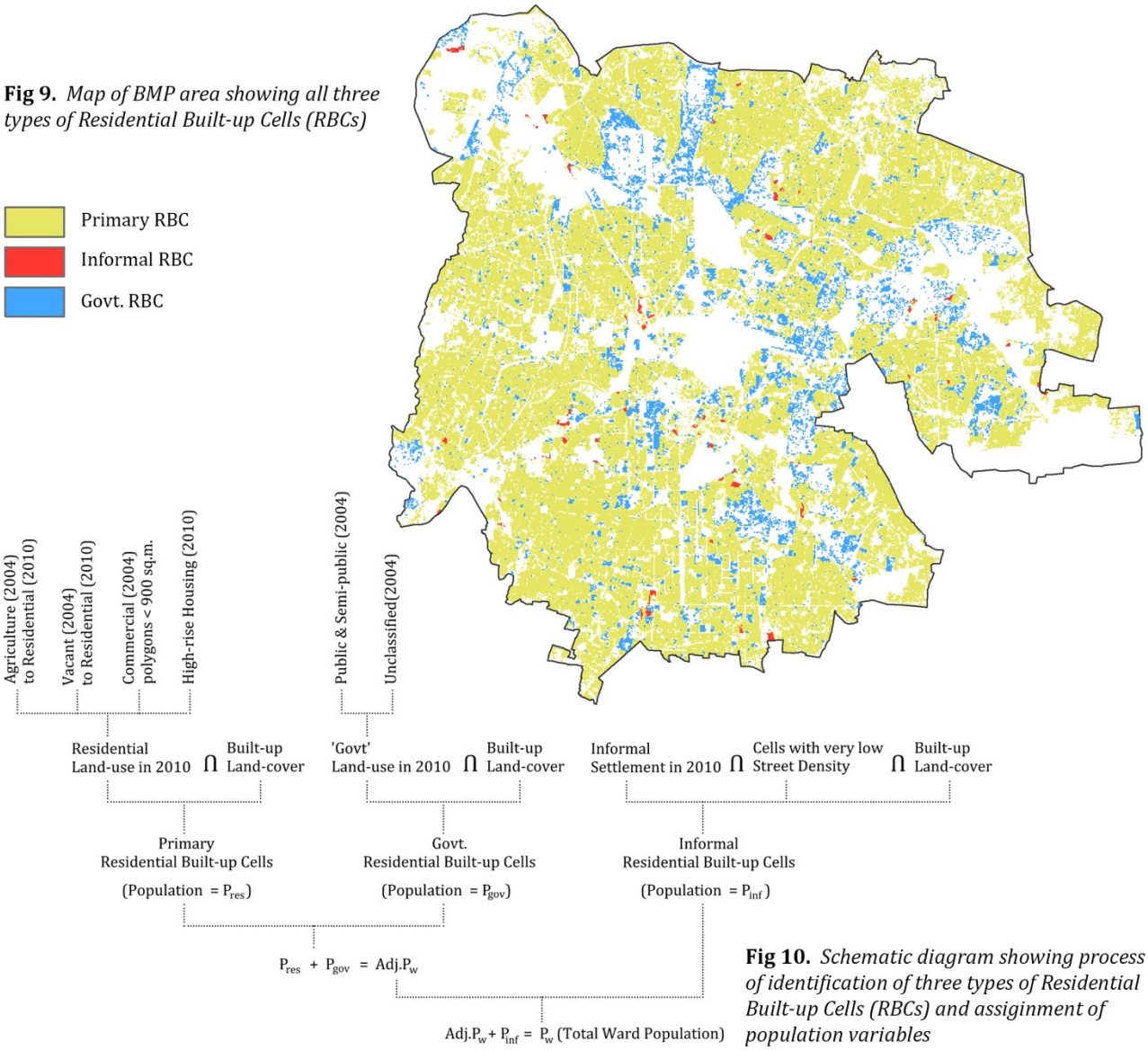


Fig 9. *Map of BMP area showing all three types of Residential Built-up Cells (RBCs)*

Fig 10. *Schematic diagram showing process of identification of three types of Residential Built-up Cells (RBCs) and assiginment of population variables*

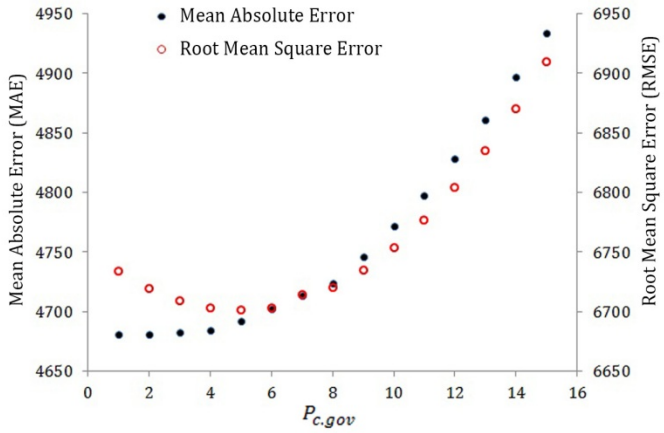


Fig 11. *Graph showing MAE and RMSE for values for $P_{c.gov}$ from 1 to 15.*

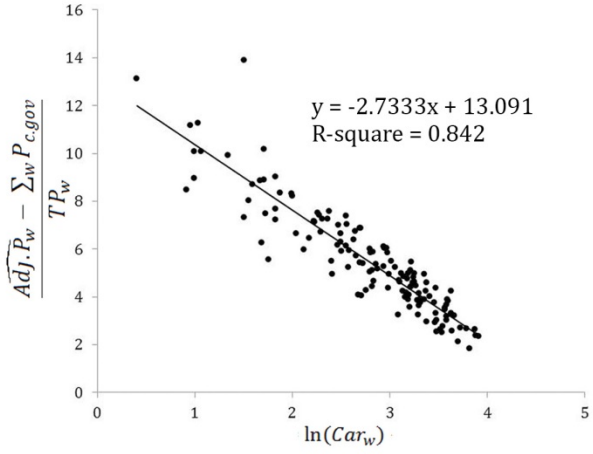


Fig 12. *Linear regression as per equation 19 with $P_{c.gov} = 4$*

lower population in a cell, as a result of which the cells in a very dense informal settlement with very low street network density may end up getting a very low predicted population.

To address this problem, first of all a definition of very low street density was required. The distribution of street density values was approximately Gaussian, and hence values more than one standard deviation below the mean (< 25.1 km/km^2) were defined as very low street density. Then all built-up cells with very low street network density within informal settlements were identified by intersecting the land-cover dataset, the informal settlement polygons and a subset of the street density dataset which contained only cells with values less than 25.1 km/km^2. We can call the cells in this layer Informal RBCs (Fig. 9 & 10).[17]

Estimating population per cell for Informal RBCs

Average population density per unit height for Informal RBCs was estimated using the population numbers from the KSDB database, the number of Informal RBCs within each settlement and the average height of such cells within each settlement. This average population density per unit height was then multiplied by the mean height of Informal RBCs within each informal settlement to arrive at an estimate of population density per Informal RBC for each informal settlement. The total population represented by these Informal RBCs within each ward was then subtracted from the Census population for each ward, such that a new adjusted Census population number was derived for every ward within the BMP area (Eq. 1, Eq. 2 & Eq. 3).

$$P_w = P_{res} + P_{gov} + P_{inf} \quad (1)$$

$$P_w = \sum_w P_{c.res} + \sum_w P_{c.gov} + \sum_w P_{c.inf} \quad (2)$$

$$Adj.P_w = P_w - P_{inf} \quad (3)$$

Where

P_w *is ward level population as per census*
$P_{res}, P_{gov}, P_{inf}$ *are population in Primary, Govt and Informal RBCs respectively in a ward*
$P_{c.res}, P_{c.gov}, P_{c.inf}$ *are cell level population in each Primary, Govt and Informal RBC*
$\sum_w$ *is summation at ward level*
$Adj.P_w$ *is adjusted population per ward*

3.2 Assign street network density and building height values

To extract street network density and building heights from the appropriate maps, the primary residential built-up cells layer was converted into a point feature class. The values

[17] It is important to bear in mind that the Informal RBC layer does not contain all RBCs within known informal settlements. Instead, it contains only those RBCs which lie within known informal settlement boundaries and have very low street density. The remaining RBCs which fall within known informal settlement boundaries but do not have very low street density are included within the Primary RBC layer.

from the street density and building height rasters were then extracted to this point feature class using the Extract Values to Points tool in ArcGIS.

3.3 Relationship between street network density, building height and population

Let us start by imagining a 30mX30m primary residential RBC which is one floor high. In this case, higher street density (as defined by total street length divided by total area of cell) can be conceptualized as increasing the number of subdivisions within that built-up cell. Increased number of subdivisions potentially translates into more number of residential units, although each would be of smaller size compared to another built-up cell which has fewer streets and hence fewer subdivisions. Therefore higher street density is potentially indicative of higher population density in primary RBCs.

In the literature, several authors have demonstrated the relationship between street density and population density. Peponis et. al. (2007) demonstrate that density of streets increases in proportion to the density of properties. They also propose that since streets within a neighborhood exist to service properties, this could be illustrative of a fundamental relationship between street morphology and parcel density. Peponis et.al. (2007) further show that there is a strong correlation between population density (per km^2) and street length per km^2.

Using a sample of 100 one km X one km street network patches from the densest areas of Greece (ranging from 1200 people/km^2 to 27000 people/km^2) Maniadakis & Varoutas (2012) show that the topological and geometric properties of street networks do have a relationship with population density. Although the relationship they find is not particularly strong, they show that as population density increases, street networks tend to have higher total length, higher number of nodes and higher number of edges.

If we assume that there exists a relationship between street density and population density within a 30mX30m cell, this relationship cannot be linear, since beyond a point the number of people occupying one cell cannot steadily increase ad infinitum. Instead, the relationship is probably better characterized by a log function or a square root function, such that, as street density increases beyond a limit, the rate of increase in population density decreases. Therefore, to start with, the population density within a primary RBC can be stated to be proportional to some function *f* of street density (Eq. 4).

$$P_{c.res} \propto f(St.den_{c.res}) \quad (4)$$

Where

$P_{c.res}$ *is the population per* $30mX30m$ *primary RBC*
$St.den_{c.res}$ *is the street density value within a* $30mX30m$ *primary RBC*

Now we will consider the case where the residential built-up cell has multiple floors. An increase in the number of floors that each cell has can be conceptualized as increasing the intensity of use of the residential built-up cell by making multiple levels available for use. In this case it is possible that the number of people who can occupy a given residential built-up cell also increases linearly as the number of floors increases. But as the number of dwelling units increases, the probability of dwelling units remaining vacant may also increase. Therefore, the relationship between population and building height in a primary RBC could be characterized by a linear, log or a square root function. So we could state that population in a primary RBC is proportional to some function g of building height.

$$P_{c.res} \propto g(Bld.ht_{c.res}) \quad (5)$$

Where

$$Bld.ht_{c.res} \text{ is building height value within a } 30mX30m \text{ primary RBC}$$

3.4 Asset ownership as a variable in understanding population density

Street density and average building height within a 30mX30m residential built-up cell is technically only giving us an understanding about the characteristics of residential space within that built-up cell. But population within such a cell would also depend on how much space each individual tends to occupy. Since residential built-up space is an expensive commodity, we can hypothesize that the amount of space that an individual or a household can afford to occupy is related to the level of affluence of the individual or household. Therefore areas with the same level of residential built-up density (as measured by the ratio of built-up space to land) can have very different population densities, if the molecular densities (Alexander et.al., 1988) of the dwelling units in these areas vary widely.

The Census of India measures several asset ownership variables, like ownership of cell phones, motorbikes, cars etc. which can indicate the level of affluence or wealth of urban households. But as of 2016, these variables are published only at the ward level. Therefore, while street network density, building height and affluence of households could potentially help us predict population at the level of 30mX30m residential built-up cell, the first two variables are available at the level of the 30mX30m cell, while the last variable is only available at the ward level. The following section describes how this problem was addressed.

3.5 Setting up the population density model

Assuming that the relationship between population and street network density and building height at the cell level are best described by Equations 4 and 5, we can combine them to arrive at Equation 6 below.

$$P_{c.res} \propto f(St.den_{c.res}) * g(Bld.ht_{c.res}) \quad (6)$$

At ward level this can be rewritten as Equation 7 since $\sum_w P_{c.res}$ is the same as P_{res} or the total population within all primary RBCs of a ward

$$\sum_w P_{c.res} \propto \sum_w [f(St.den_{c.res}) * g(Bld.ht_{c.res})] \quad (7)$$

If we were to attempt to predict population within primary RBCs at a ward level using only those variables which are available at the cell level, then this would be our final equation. So for sake of brevity, we can call $\sum_w [f(St.den_{c.res}) * g(Bld.ht_{c.res})]$ our temporary prediction at the ward level or TP_w .

This simplifies Equation 7 and gives us Equation 8.

$$\sum_w P_{c.res} \propto TP_w \quad (8)$$

But based on Section 3.4, we can assume that the population within a primary RBC has a relationship to the level of affluence within that cell. While asset ownership variables from the census can be used as a proxy for household wealth, these variables are only available at the ward level. Therefore we can state that sum of population of all primary RBCs within a ward is related to some function h of one of the ward level household asset ownership variable As_w from the census. This gives us Equation 9.

$$\sum_w P_{c.res} \propto h(As_w) \quad (9)$$

Combining Equation 8 and Equation 9 we can state that

$$\sum_w P_{c.res} \propto h(As_w) * TP_w \quad (10)$$

Or after expanding TP_w it could be written as

$$\sum_w P_{c.res} \propto h(As_w) * \left\{\sum_w [f(St.den_{c.res}) * g(Bld.ht_{c.res})]\right\} \quad (11)$$

Hence, the sum of population of all primary RBCs within a ward is proportional to TP_w multiplied by some function h of a ward level asset ownership variable As_w .

Equation 11 consists of two cell level variables and one ward level variable. The options available for the functional forms f, g & h and the options available for the census asset ownership variable As_w are described below

$f(St.den_{c.res})$	log or square root
$g(Bld.ht_{c.res})$	linear, log or square root
$h(As_w)$	unknown functional form h. As_w can be percentage of households which own personal computer, car, two-wheeler etc.

The optimum combination of functional forms of f, g & h, for each option of As_w was determined through trial and error.

To start with, f was set to log function and g was set to linear. Since h has an unknown functional form Equation 10 was rewritten as below

$$\frac{\sum_w P_{c.res}}{TP_w} \propto h(As_w) \tag{12}$$

But

$$\sum_w P_{c.res} = Adj.P_w - \sum_w P_{c.gov} \tag{13}$$

And

$$\sum_w P_{c.gov} = P_{c.gov} * N_{gov} \tag{14}$$

Where

$$P_{c.gov} \text{ is population in each Govt RBC in a ward}$$
$$N_{c.gov} \text{ is number of Govt RBC in a ward}$$

To initiate the estimation of function h we can start by assuming $P_{c.gov} = 0$

Therefore Equation 12 can be rewritten as

$$\frac{Adj.P_w}{TP_w} \propto h(As_w) \tag{15}$$

Equation 15 was then used to run ordinary least squares (OLS) regressions with As_w set to ward level census variables like percentage of households which own personal computer, car, two-wheeler etc. The R-square value of the regression results was used to evaluate which known functional form best approximates h in the case of each option of As_w.

This exercise was then carried out for each of the functional form options available for f & g also. The results obtained from this procedure and the final prediction framework and results is described in the next section.

4. Results

The average population per unit height for Informal RBCs was estimated to be 16.61 persons, based on the method outlined in Section 3.1. This number was multiplied by the mean height of RBCs within each informal settlement to arrive at the average population per RBC for each informal settlement. This was then multiplied by total number of Informal RBCs within each settlement to estimate the total population within each informal settlement. Total population in Informal RBCs within all settlements of a ward was then deducted from the ward population from the Census of India to arrive at a new Ward-level Adjusted Population for each ward ($Adj.P_w$) as shown in Equation 3.

Using this Ward-level Adjusted Population, the trial and error procedure revealed that almost all options for As_w had an approximately logarithmic relationship to $Adj.P_w/TP_w$, based on the initial setting of f as a log function and g as a linear function. The relationship was particularly strong when ward level 'percentage of households which own cars' variable was used for As_w.

Upon trying out all functional form options for f and g with the car ownership variable for s_w , the highest R-square value for regressions based on Equation 15 was obtained for the following combination.

$f(St.den_{c.res})$	log
$g(Bld.ht_{c.res})$	square root
$h(As_w)$	h as log and As_w as 'percentage of households which own cars'

The above combination gave an R-square value of 0.8363

Therefore Equation 15 can be rewritten as below

$$\frac{Adj.P_w}{TP_w} \propto \ln(Car_w) \tag{16}$$

Where

$$Car_w \text{ is ward level percentage of car owning households}$$

And

$$TP_w = \sum_w [\ln(St.den_{c.res}) * \sqrt{Bld.ht_{c.res}}]$$

Based on the parameters obtained from the regression equation with R-square value of 0.8363, Equation 16 can be rewritten as shown below.

$$\frac{\widehat{Adj.P_w}}{TP_w} = -2.7302 * [\ln(Car_w)] + 13.151 \quad (17)$$

Where

$$\widehat{Adj.P_w} \text{ is predicted Adjusted Population per Ward}$$

Since TP_w is known for all wards this can be rewritten as

$$\widehat{Adj.P_w} = TP_w * \{-2.7302 * [\ln(Car_w)] + 13.151\} \quad (18)$$

Equations 17 and 18 have been obtained by assuming $P_{c.gov} = 0$. If $P_{c.gov} \neq 0$ then based on Equation 16, for each value of $P_{c.gov}$ there will be a new regression equation of the form

$$\frac{Adj.P_w - \sum_w P_{c.gov}}{TP_w} \propto \ln(Car_w) \quad (19)$$

Changing $Adj.P_w$ to $\widehat{Adj.P_w}$ Equation 19 can be rewritten as Equation 20

$$\frac{\widehat{Adj.P_w} - \sum_w P_{c.gov}}{TP_w} = C_1 * [\ln(Car_w)] + C_2 \quad (20)$$

Or

$$\widehat{Adj.P_w} - \sum_w P_{c.gov} = TP_w * \{C_1 * [\ln(Car_w)] + C_2\} \quad (21)$$

Or

$$\widehat{Adj.P_w} = TP_w * \{C_1 * [\ln(Car_w)] + C_2\} + \sum_w P_{c.gov} \quad (22)$$

Or

$$\widehat{Adj.P_w} = TP_w * \{C_1 * [\ln(Car_w)] + C_2\} + P_{c.gov} * N_{gov} \quad (23)$$

For every value of $P_{c.gov}$, using Equation 19, we can obtain a new regression equation which gives parameters C_1 and C_2. These parameters can then be used to compute $\widehat{Adj.P_w}$ as shown in Equation 23. Therefore for every value of $P_{c.gov}$, Mean Absolute Error (MAE) and RMSE can also be calculated based on the difference between $Adj.P_w$ and $\widehat{Adj.P_w}$.

$P_{c.gov}$ values ranging from 1 to 15 were used to identify the optimum value of $P_{c.gov}$ which would provide the lowest values for both MAE and RMSE. As Figure 11 shows, although

RMSE was minimum (and approximately the same) for $P_{c.gov}$ = 4, 5 or 6, amongst these MAE was minimum for $P_{c.gov} = 4$. Hence this was assumed to be the optimum value of $P_{c.gov}$.

Using $P_{c.gov} = 4$ gave the regression equation shown in Figure 12 which has a higher R-square value of 0.842 in comparison to the one obtained using Equation 17. Therefore Eq. 23 can be rewritten as below, using the parameters C_1 and C_2 obtained from the regression equation shown in Figure 12 and with $P_{c.gov} = 4$.

$$\widehat{Adj.P_w} = TP_w * \{-2.7333 * [\ln(Car_w)] + 13.091\} + 4 * N_{gov} \quad (24)$$

Equation 24 gave MAE value of 4683.94 persons and RMSE value of 6702.59 while the median of absolute error was 3199.54 persons. When using absolute percentage error values for each ward, mean error was 12.44% and median error was 8.81%.

4.1 OLS regression diagnostics and Geographically Weighted Regression

However, the results of the OLS regression diagnostics indicated that the modeled relationship may be spatially non-stationary. The significant Koenker's studentized Breush-Pagan statistic (Breush & Pagan, 1979; Koenker, 1981) obtained in the OLS diagnostics indicates the existence of heteroskedasticity.[18] To check for spatial autocorrelation the standard residuals were mapped onto the wards within the study area and Morans-I (Moran, 1950) test was carried out in ArcGIS. The results show that the standard residuals are clustered in their distribution indicating a significant level of spatial autocorrelation (Fig. 13 & Fig. 14). Therefore the global solution obtained from the OLS regression is probably insufficient to model the influence of localized spatial phenomena.

Based on this, a geographically weighted version (Brunsdon et.al., 1996) of Equation19 was carried out in ArcGIS with $P_{c.gov} = 4$. Morans-I test was then conducted using the standard residuals of this Geographically Weighted Regression (GWR). The results indicate that the standard residuals from the GWR are randomly distributed (Fig. 15 & Fig.16). The GWR improved the adjusted R-squared value to 0.88 and the RMSE reduced to 5558.46 persons. The MAE was 3999.35 and median absolute error was 2785.09. Using absolute percentage error values for each ward, this works out to a mean absolute error of 10.70% and median absolute error of 7.87%.

The 'corrected Akaike's Information Criterion' (AICc) is another measure which can indicate relative goodness of fit of regression models (Akaike, 1973; Hurvich & Tsai, 1989). As the difference in AICc values between two models increases, the probability of the model with the lower value being a better approximation of the true unknown process increases. According to Burnham & Anderson (2002) if the difference between AICc values

[18] Koenkers studentized Breush-Pagan statistic is more robust to outliers in comparison to the original Breush-Pagan statistic (Koenker, 1981; Lyon & Tsai, 1996).

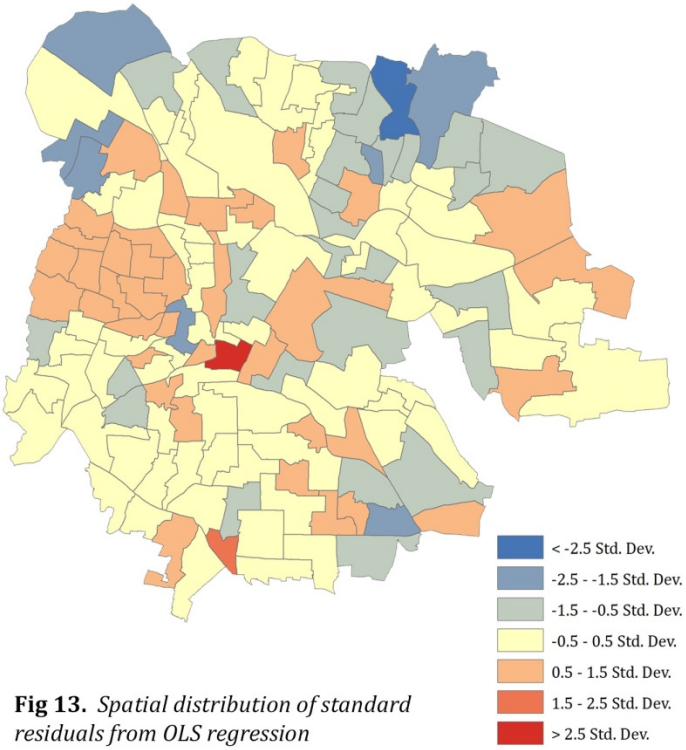


Fig 13. *Spatial distribution of standard residuals from OLS regression*

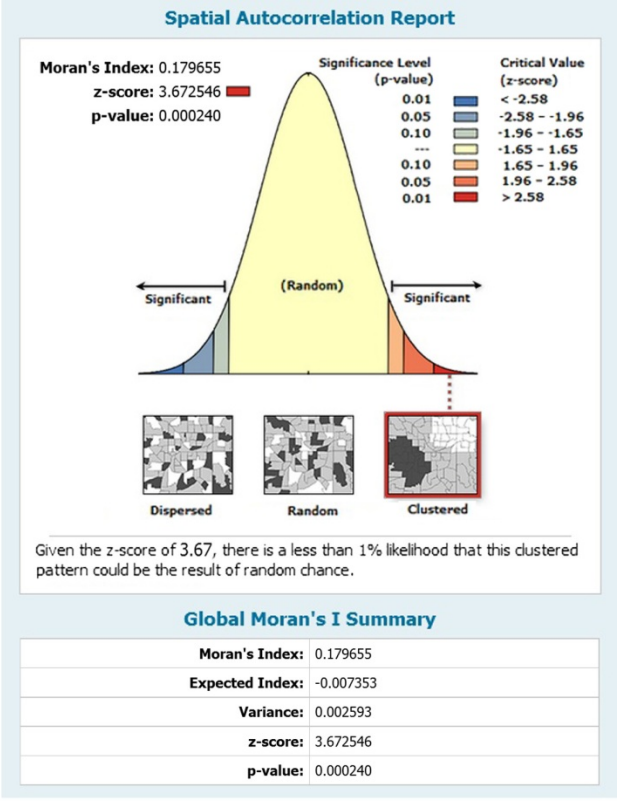


Fig 14. *Morans-I test indicating spatial clustering of errors.*

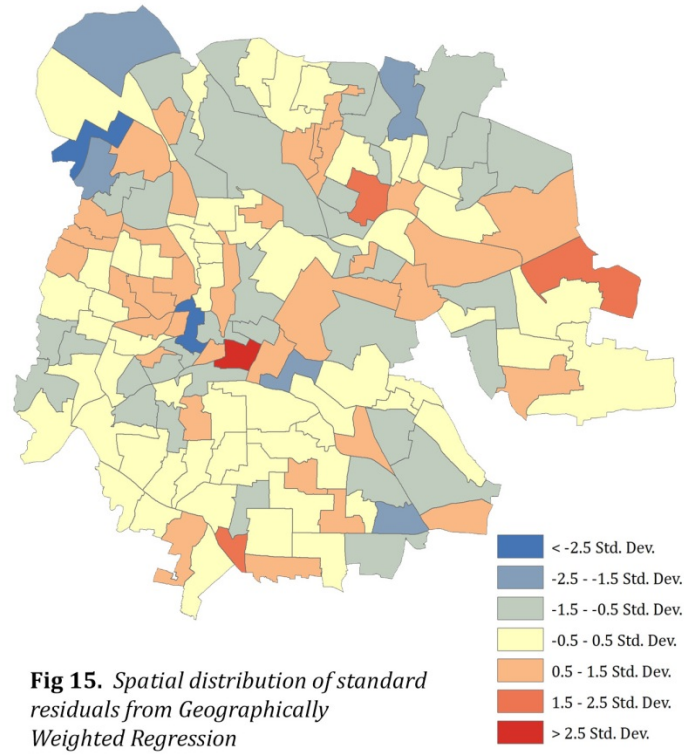


Fig 15. *Spatial distribution of standard residuals from Geographically Weighted Regression*

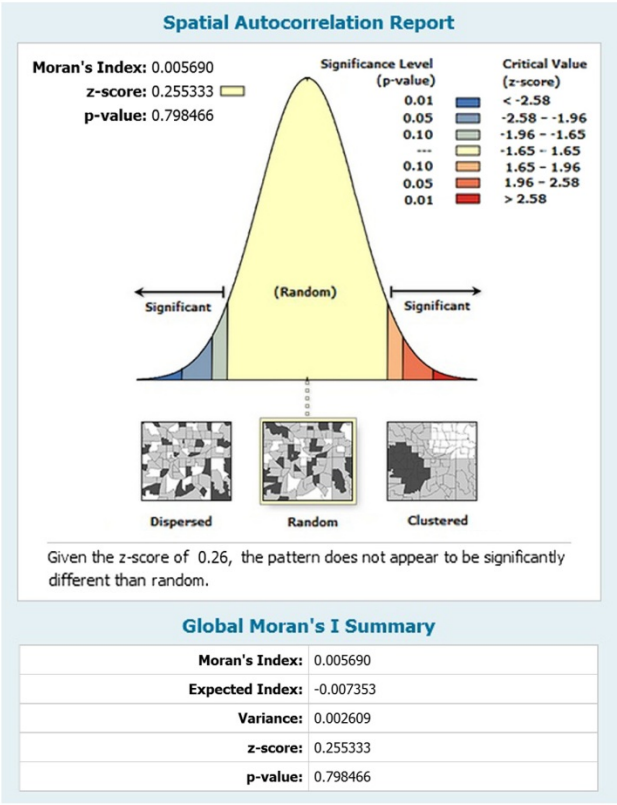


Fig 16. *Morans-I test results after Geographically Weighted Regression. Results indicate that errors are randomly distributed.*

of two models is greater than seven then we can infer that the model with the lower AICc value is a better one with a high probability. When AICc values differ by more than 10 we can infer that the probability of the model with the lower AICc value being better is very high. The AICc value obtained for the GWR was lower than that of the OLS regression model by 31. This indicates that the GWR model is a significant improvement upon the OLS regression model.

4.2 Generating the 30m resolution population density surface

As explained in Section 3.1, the population in Bangalore can be conceptualized as living within three types of RBCs—Informal RBCs which consist of built-up cells with very low street density values within known informal settlements, Primary RBCs which consists of the built-up areas whose land-use has been identified to be residential and Govt. RBCs which consist of residential areas within the Public/Semi-public and Unclassified Land-uses. Out of these, the population to be assigned to Informal RBCs was calculated as described in Section 3.1, the population to be assigned to Primary RBCs was predicted based on methods described in Sections 3.3 to 3.5 and the population to be assigned to each Govt. RBC was determined based on minimization of MAE and RMSE of $\widehat{Adj.P_w} - Adj.P_w$ as per Equation 23. Since the population to be assigned to all RBCs is now known, the results were mapped onto the RBCs in ArcGIS to generate the 30m resolution population density surface shown in Figure 18.

5. Discussion

5.1 Comparison with ward level population density map

The 30m resolution population density surface map generated by assigning the estimated population numbers to all RBCs (Fig. 18) shows much greater granularity than ward level estimates of average population density generated by dividing the total people in a ward by total area of the ward (Fig. 17). Besides identifying finer scale variation in density within a ward, it is also able to give a neighborhood level understanding of population density by identifying neighborhoods with relatively homogenous density characteristics which cut across wards which appear to have very different average population densities at the ward level.

For example, Figures 19 to 21 show the area marked as region A in Figures 17 and 18, as seen in the ward level average population density map, 30m population density surface and Google Earth. According to the ward level average population density map, wards 116 and 147 have very different average population densities. The area within the dotted line in the 30m population density surface shows us that a relatively homogenous and dense neighborhood straddles these two wards (Fig. 20). This is corroborated by the Google Earth imagery which clearly shows the neighborhood which cuts across the boundary between wards 116 and 147 (Fig. 21). Similarly, Figures 22 to 24, show the area marked as

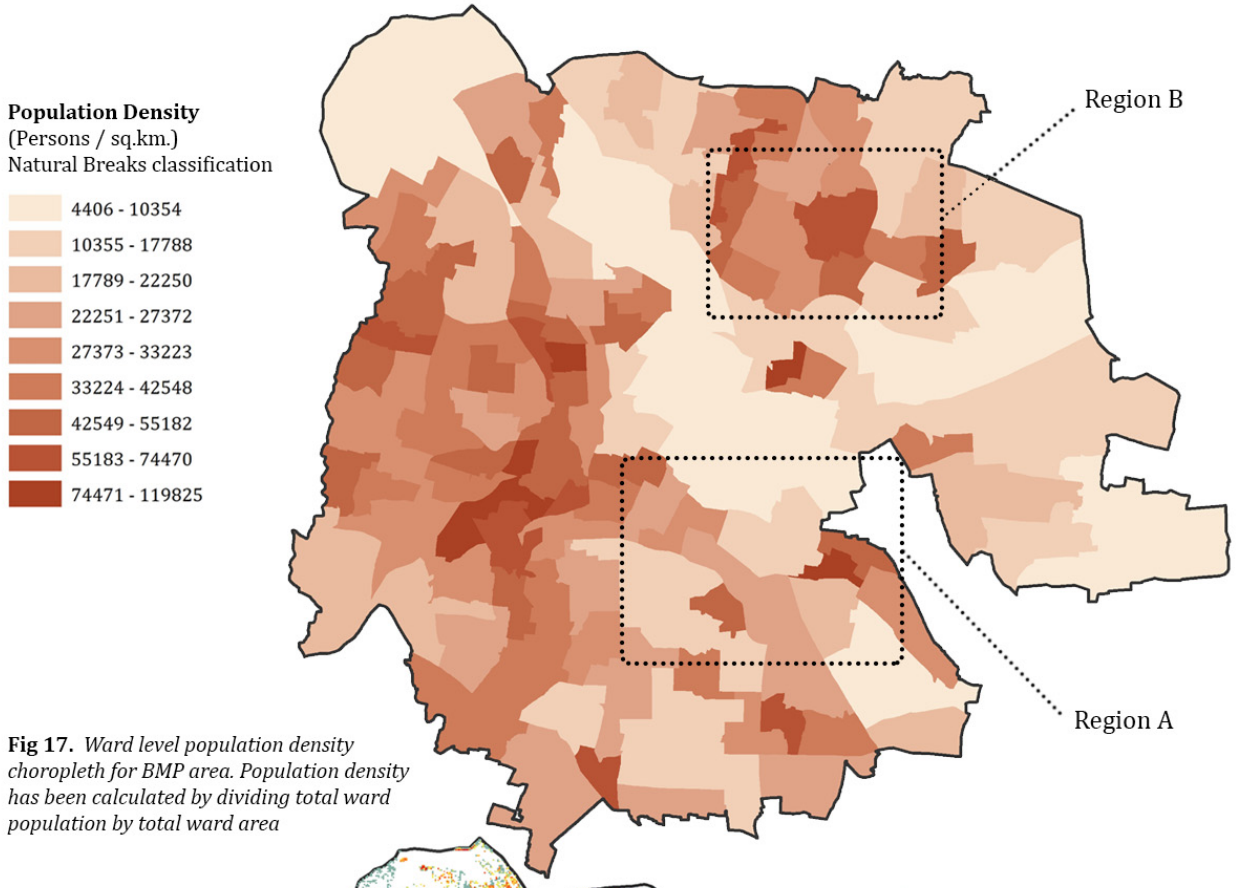


Fig 17. *Ward level population density choropleth for BMP area. Population density has been calculated by dividing total ward population by total ward area*

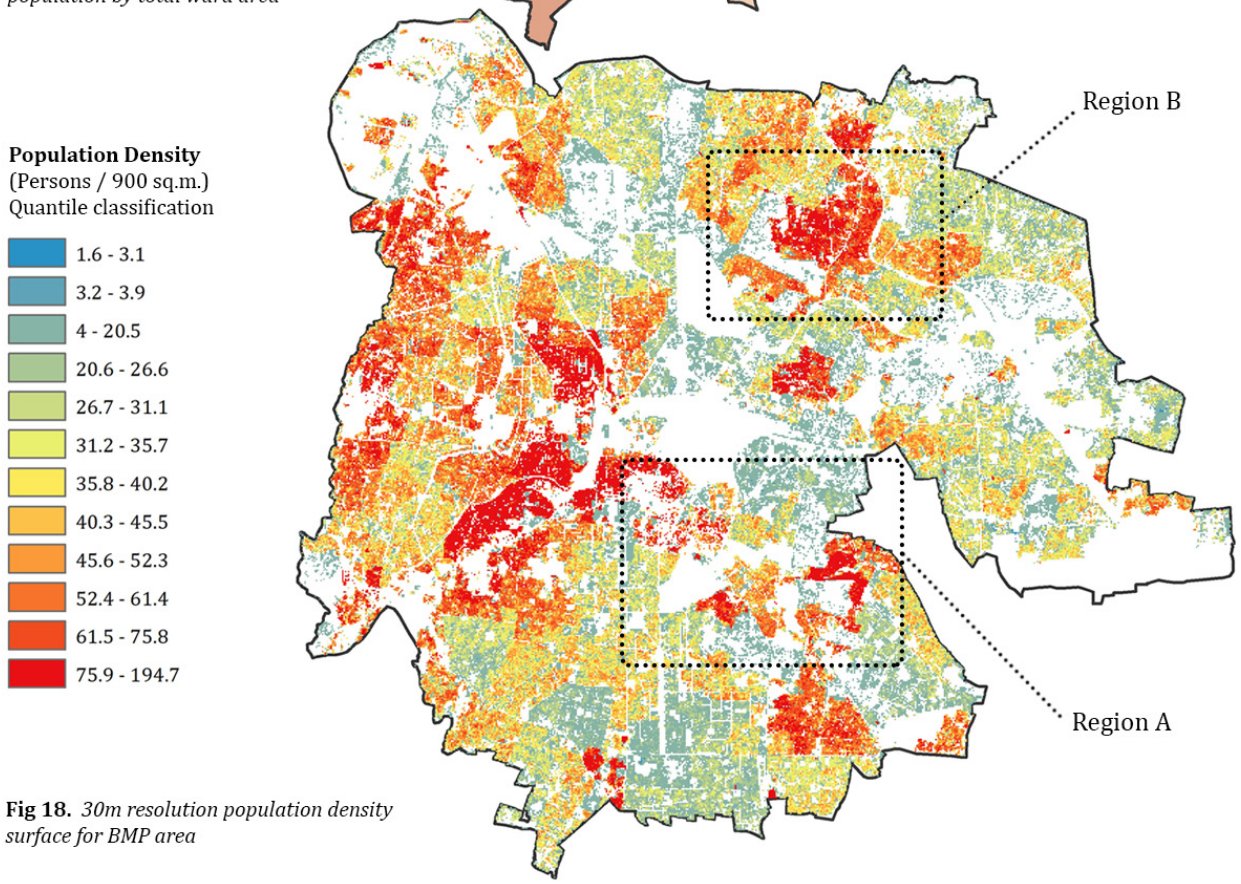


Fig 18. *30m resolution population density surface for BMP area*

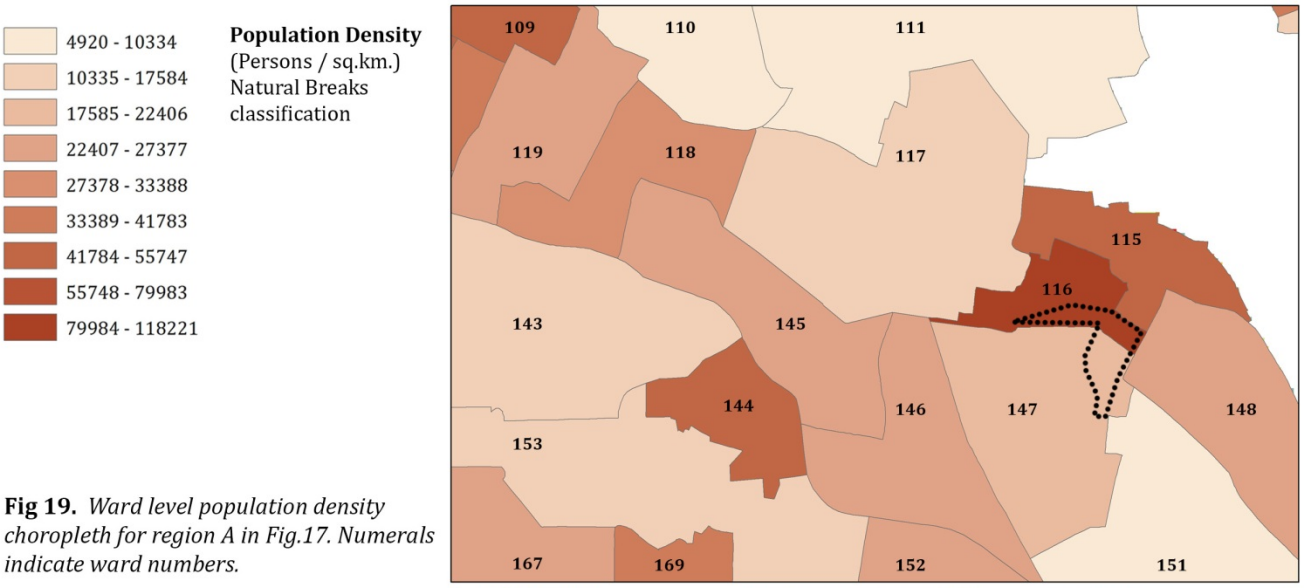


Fig 19. *Ward level population density choropleth for region A in Fig.17. Numerals indicate ward numbers.*

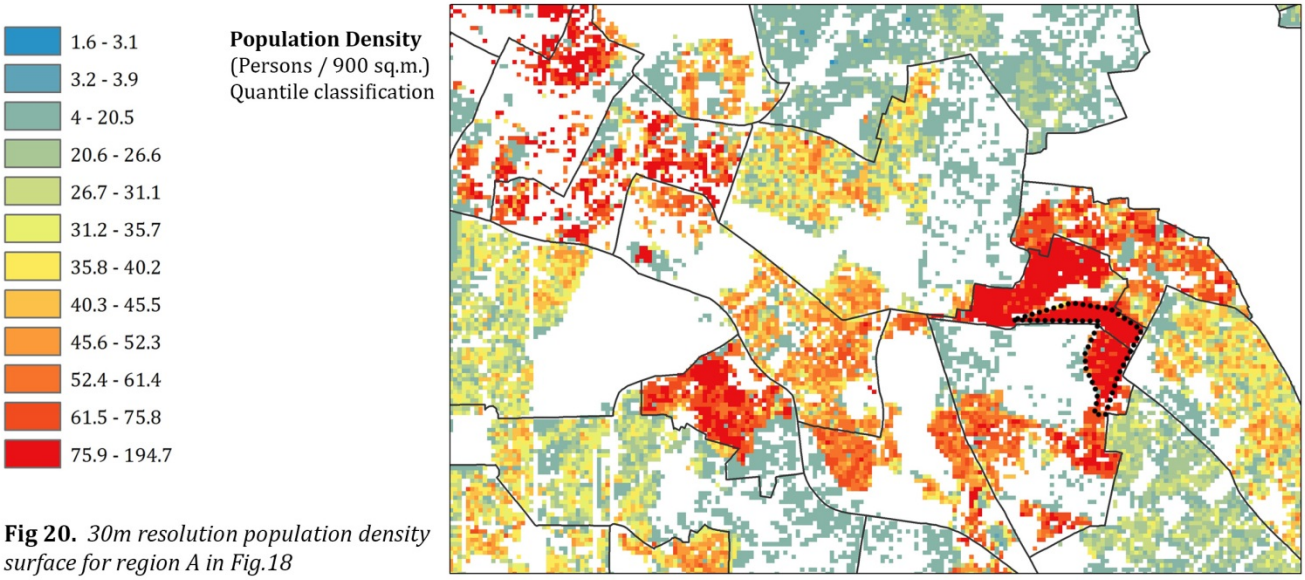


Fig 20. *30m resolution population density surface for region A in Fig.18*

Fig 21. *Google Earth imagery from 23 January 2010 for region A in Fig. 17 & 18 overlaid with ward boundaries.*

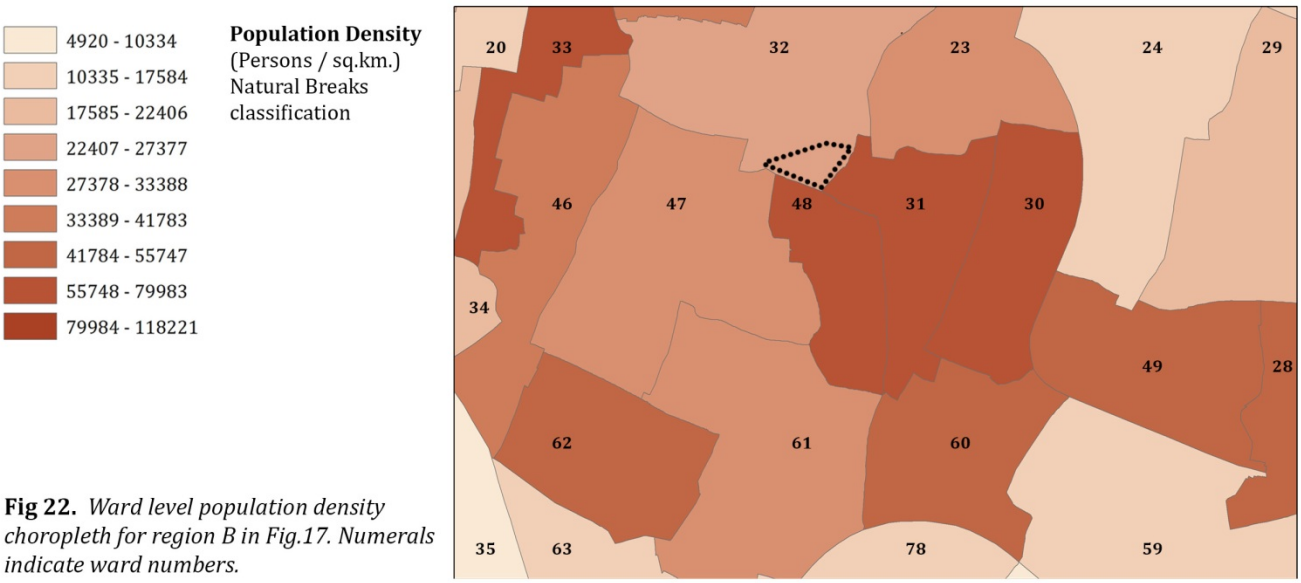


Fig 22. *Ward level population density choropleth for region B in Fig.17. Numerals indicate ward numbers.*

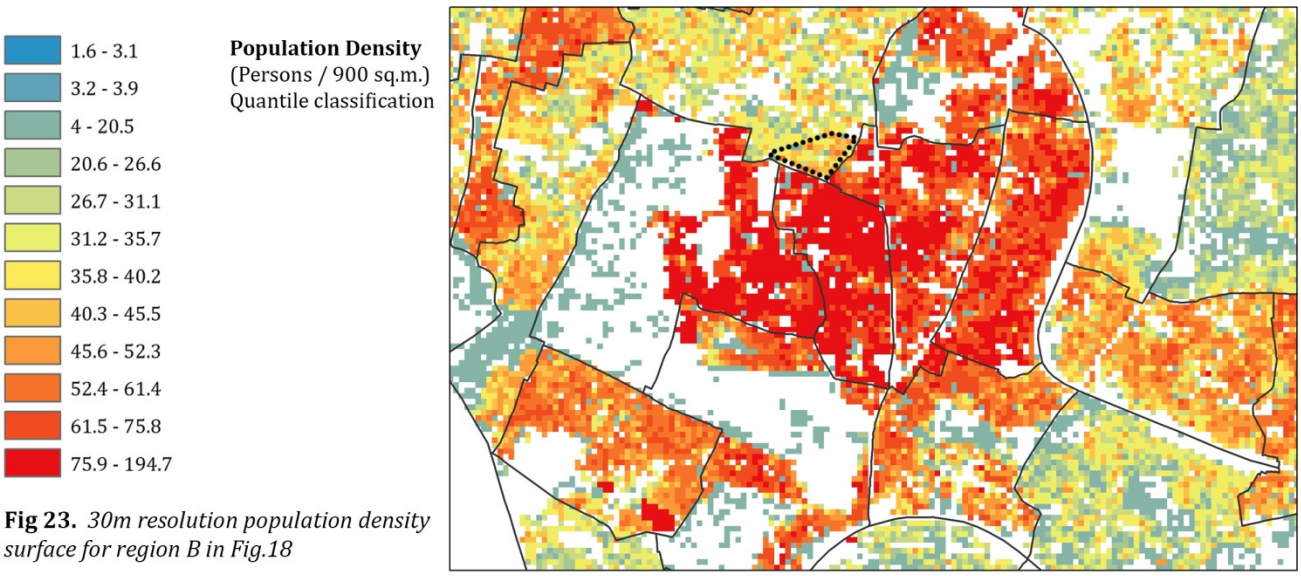


Fig 23. *30m resolution population density surface for region B in Fig.18*

Fig 24. *Google Earth imagery from 23 January 2010 for region B in Fig. 17 & 18 overlaid with ward boundaries.*

region B (Fig. 17 and Fig.18) in the northern part of BMP. Here, once again the 30m population density surface (Fig. 23) provides a much more detailed picture of relative differences in population densities within a ward. It also clearly delineates contiguous patches of similar density which cuts across wards.

However, the use of ward level car-ownership data could lead to errors since it can somewhat average out the heterogeneity in population distribution across the various neighborhoods within a ward—for example when a ward with mostly wealthy households has a small neighborhood which is relatively less wealthy. In this case since every primary RBC in the ward is being weighted on the basis of the average ward level car-ownership data, it is possible that the population in less wealthy areas within the same ward may be substantially under-predicted. This seems to be happening in the case of the area within the dotted line in ward 32 in Figure 22. Google Earth imagery (Fig. 24) shows us that this patch is may be similar in spatial characteristics to the neighborhoods in the wards which lie to its immediate south (wards 31 and 48).

On average ward 32 is considerably wealthier than wards 31 and 48 since census data tells us that 26.8% of households in ward 32 own car while only 5.6% and 2.5% of households own a car in wards 31 and 48 (Census of India, 2011c). But it is quite possible that the wealthier households are located outside the area indicated by the dotted line while, judging from the settlement characteristics which are visually apparent from Google Earth imagery (Fig. 24), the neighborhood within the dotted line may consist largely of households which are similar in wealth to Wards 31 and 48. But the use of ward level car-ownership data appears to have an averaging effect and the population density in the RBCs within the dotted line are predicted to be much lower than those of the cells within Wards 31 and 48. Similarly the 30m population density surface is probably over-predicting the population density of RBCs in the wealthier areas of ward 32 which, based on visual assessment of settlement grain, appear to be located north of the area within the dotted line.

5.2 Problems with accuracy assessment

Although the 30m population density surface provides a detailed picture of population density variation within a ward, there is no reliable method of assessing the accuracy of the predicted cell level population figures. This is because the census publishes urban population data only at the ward level. For the 2011 Census, the Census of India did publish population figures at the enumeration block level for Bangalore (Census of India, 2011b) but the enumeration block boundaries are not available making it impossible to use this data for accuracy assessment. Therefore the only practical way of assessing the accuracy of cell level population prediction is to sum them up at the ward level and compare it to known ward level census population figures as demonstrated in Section 3.5.

5.3 Comparison with sub-city typologies

The seven variable sub-city typologies presented in Balakrishnan & Anand (2015) helps understand urban heterogeneity through socio-economic typologies (Fig. 26). The mean values for various socio-economic attributes for each typology are shown in Figure 25.[19] By comparing the sub-city typologies with the 30m population density surface, it becomes apparent that most of the Low-Socio-Economic typology (Low-SE) wards, comprise largely of the densest neighborhoods within the BMP area. Figure 27 shows the 30m population density surface overlaid with the boundaries of the Low-SE wards within the BMP area.

5.4 Potential sources of errors

One of the major sources of errors in population estimation could be the discrepancy in year of acquisition of the various data layers. While the population enumeration for the 2011 Census was conducted in 2010, the Landsat Data is from 2011 and the Cartosat data is from 2012 since this was the best cloud-free data available. Moreover, the land-use data was available for the year of 2004 and it had to be updated to 2011 (the year of Census enumeration) based on the procedure described in Section 2.2.2.

Although the Cartosat imagery has a resolution of 2.5m, the vegetation removal from the nDSM had to be conducted by using the land-cover map which was generated from the 30m Landsat data. This could potentially lead to incomplete removal of vegetation which may have contributed to errors in building height estimation. The lack of information about residential areas within the land-uses classes of Public/Semi-public and Unclassified, could also be contributing to errors in population estimation.

Most significantly it is important to note that although the paper describes a predictive framework, it is not possible to validate it at the ward-level based on the available data. This is because, as demonstrated earlier, spatial non-stationarity of the OLS model needs to be addressed for using Geographically Weighted Regression (GWR) and GWR works effectively only if there are several hundred features (Environmental Systems Research Institute, 2016). Since there are only 137 wards in the dataset, splitting it into a calibration and validation datasets would have rendered the GWR procedure ineffective. Hence further research with a larger dataset is required to fully validate the approach described in this paper. Further research is also required to show whether the proposed predictive framework is applicable to other Indian cities.

[19] It is important to note that the sub-city typologies in Balakrishnan & Anand (2015) were generated using all wards within the BBMP boundary. If sub-city typologies were to be generated using only the BMP wards, the results may be quite different.

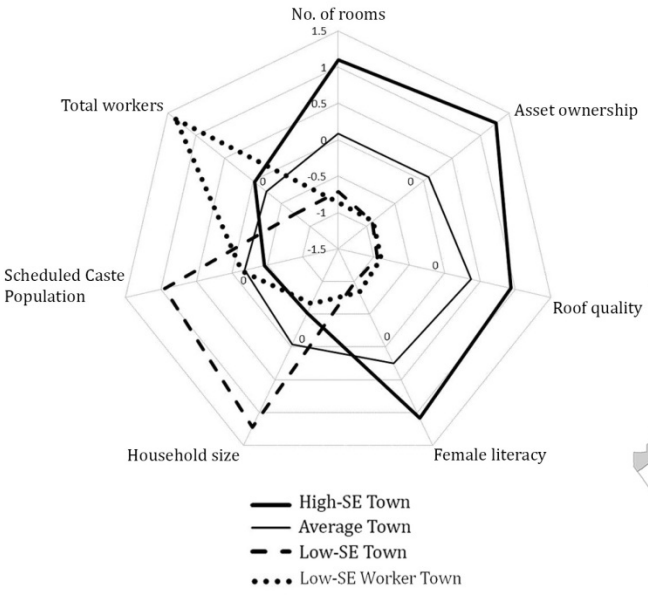


Fig 25. *Radar diagram showing cluster/typology means for each of the seven variables. Values are normalized, '0' line indicates city mean.*

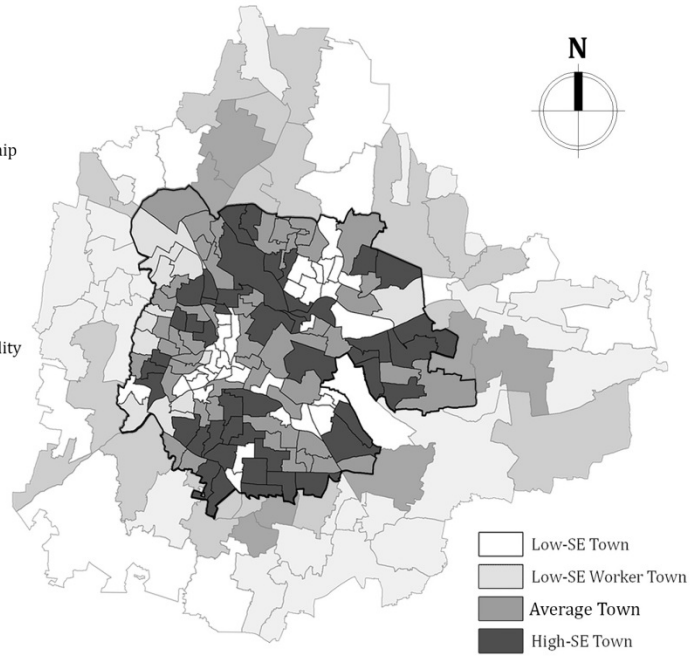


Fig 26. *Sub-city typologies map using seven variables. BMP area is highlighted.*
Note: Typologies were generated using all BBMP wards

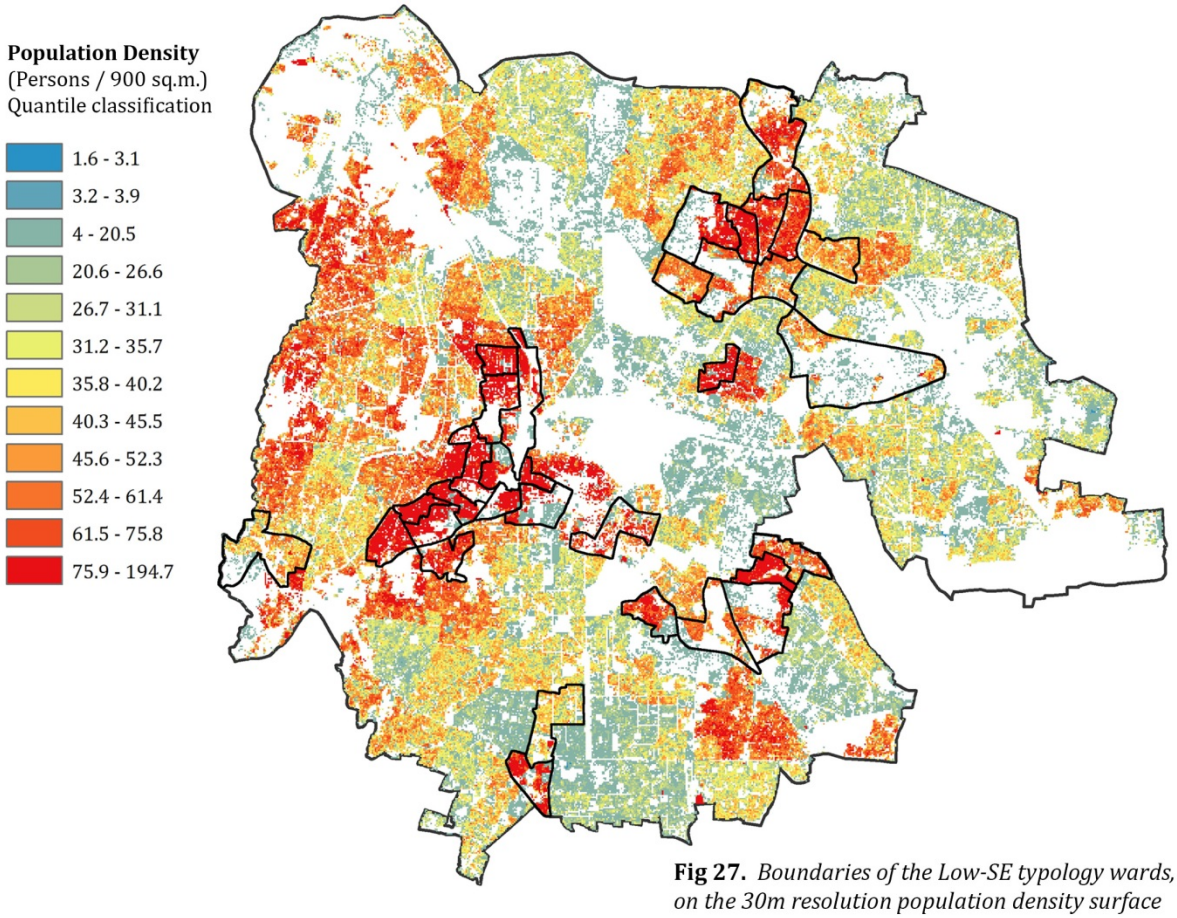


Fig 27. *Boundaries of the Low-SE typology wards, overlaid on the 30m resolution population density surface*

6. Conclusion

This paper describes a method for predicting population density at the scale of 30mX30m raster cells. Using data on land-cover, land-use, street network, building height and ward level data on car ownership, the proposed method is shown to predict cell level population density such that, when summed up at the ward level, the mean absolute percentage error between predicted ward level population and known ward level population from the census is 10.7%.

The paper demonstrates that population within a cell is proportional to log of street density value and square root of building height value within a cell. It also shows that as the average wealth of households within a ward decreases, the molecular density (internal density) within residential units increases. The research also demonstrates that the relationship between population density, street density, building height and ward level car ownership is spatially non-stationary. As pointed out by Patel (2013), the notion of internal density within residential units has significant implications for urban planning paradigms which emphasize regulation of Floor Area Ratio (FAR). While planners attempt to regulate intensity of land-use by stipulating FAR values, it is important to bear in mind that this regulates only the volume of built-up space. As the method outlined in this paper shows, the internal density of the same quantum of built-up space can vary widely across different parts of the same city.

A fine-grained understanding of population density can be a very useful tool in examining the fundamental structure of Indian cities and in exploring the heterogeneity within them. It can also contribute to improving access to urban services and amenities, resource allocation, infrastructure planning and disaster risk reduction in cities.

Acknowledgements

I would like to thank Carlo de Franchis for adding support for Cartosat imagery in S2P software and patiently answering all my questions specifically related to the details of the software and more generally about DSM generation from stereo imagery. Officials in the Urban Development Department (Govt. of Karnataka), Bangalore Development Authority and Bangalore Water Supply and Sewerage Board provided access to various datasets. Rahul Sami helped complete the stereo image processing successfully and Mohan Rao provided data related to the boundaries of informal settlements in Bangalore. Anand Sriram helped in conducting the building height survey in an efficient manner. I would also like to thank Shriya Anand for detailed discussions and feedback on drafts of the paper. All errors in the manuscript remain my responsibility.

Part of the work described in this paper was enabled by a Junior Research Fellowship awarded by the American Institute of Indian Studies.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. *Selected papers of hirotugu akaike* (pp. 199-213) Springer.

Alexander, E. R., Reed, K. D., & Murphy, P. (1988). *Density measures and their relation to urban form* Center for Architecture and Urban Planning Research, University of Wisconsin--Milwaukee.

Balakrishnan, K., & Anand, S. (2015). Sub-cities of bengaluru. *Economic & Political Weekly, 50*(22), 63.

Barnes, G. (2001). Population and employment density and travel behavior in large US cities.

Bast, H., Storandt, S., & Weidner, S. (2015). Fine-grained population estimation. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems,* Bellevue. doi:http://doi.acm.org/10.1145/2820783.2820828

Bertaud, A., & Brueckner, J. K. (2005). Analyzing building-height restrictions: Predicted impacts and welfare costs. *Regional Science and Urban Economics, 35*(2), 109-125.

Boyko, C. T., & Cooper, R. (2011). Clarifying and re-conceptualising density. *Progress in Planning, 76*(1), 1-61.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society,* , 1287-1294.

Brueckner, J. K., & Sridhar, K. S. (2012). Measuring welfare gains from relaxation of land-use restrictions: The case of india's building-height limits. *Regional Science and Urban Economics, 42*(6), 1061-1067.

Bruhat Bengaluru Mahanagara Palike. (2016). BBMP ward information. Retrieved from http://bbmp.gov.in/en/ward-info

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis, 28*(4), 281-298.

Burnham, K. P., & Anderson, D. (2003). Model selection and multi-model inference. *A Pratical Informatio-Theoric Approch.Sringer,*

Burton, E., Jenks, M., & Williams, K. (2003). *The compact city: A sustainable urban form?* Routledge.

Census of India. (2011a). *Primary census abstract data tables*. New Delhi: Ministry of Home Affaris, Govt of India.

Census of India. (2011b). Directorate of Census Operations, Karnataka.

Census of India. (2011c). *Percentage of households to total households by amenities and assets*. New Delhi: Ministry of Home Affairs, Govt of India.

Churchman, A. (1999). Disentangling the concept of density. *Journal of Planning Literature, 13*(4), 389-411.

d'Angelo, P., Uttenthaler, A., Carl, S., Barner, F., & Reinartz, P. (2010). Automatic generation of high quality DSM based on IRS-P5 cartosat-1 stereo data. *Special Publication SP-686,* , 1-5.

de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J., & Facciolo, G. (2014). An automatic and modular stereo pipeline for pushbroom images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2*(3), 49.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing, 66*(7), 849-857.

Environmental Systems Research Institute. (2011). How is bilinear interpolation calculated? Retrieved from http://support.esri.com/fr/knowledgebase/techarticles/detail/31810

Environmental Systems Research Institute. (2016). Geographically weighted regression. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/geographically-weighted-regression.htm

Evans, G. A., Ramachandran, B., Zhang, Z., Bailey, G. B., & Cheng, P. (2008). An accuracy assessment of CartoSat-1 stereo image data-derived digital elevation models: A case study of the drum mountains, utah. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37*

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., . . . Roth, L. (2007). The shuttle radar topography mission. *Reviews of Geophysics, 45*(2)

Fisher, P. F., & Langford, M. (1996). Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping∗. *The Professional Geographer, 48*(3), 299-309.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*(1), 185-201.

Glaeser, E. L., & Gottlieb, J. D. (2009). *The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States,*

Google Earth 7.1. (2010). Bangalore, 12°58'23.81"N 77°35'57.04"E.

Google Earth 7.1. (2011). *Bangalore, 12°58'23.81"N 77°35'57.04"E*

Government of Karnataka. (2007). *Notification of areas under bruhat bangalore mahanagara palike* Government of Karnataka.

Hawley, K., & Moellering, H. (2005). A comparative analysis of areal interpolation methods. *Cartography and Geographic Information Science, 32*(4), 411-423.

Holt, J. B., Lo, C., & Hodler, T. W. (2004). Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science, 31*(2), 103-121.

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297-307.

Jacobs, J. (1961). *The death and life of great american cities* Vintage.

Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics, 17*(1), 107-112.

KSDB. (2014). *District-wise / town-wise slum ownership details* Karnataka Slum Development Board.

KSPCB. (2013). In Shankar R. D. N., Yathish M. G.(Eds.), *A report of monitoring committee on performance of wastewater treatment plants in the apartments and other establishments in BBMP limits of bangalore*. Bangalore: Karnataka State Pollution Control Board.

Lam, N. S. (1983). Spatial interpolation methods: A review. *The American Cartographer, 10*(2), 129-150.

Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal, 31*(1), 21-26.

Li, G., & Weng, Q. (2005). Using landsat ETM imagery to measure population density in indianapolis, indiana, USA. *Photogrammetric Engineering & Remote Sensing, 71*(8), 947-958.

Long, Y., & Shen, Z. (2015). Population spatialization and synthesis with open data. *Geospatial analysis to support urban planning in beijing* (pp. 115-131) Springer.

Lyon, J. D., & Tsai, C. (1996). A comparison of tests for heteroscedasticity. *The Statistician,* , 337-349.

Maantay, J. A., Maroko, A. R., & Herrmann, C. (2007). Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). *Cartography and Geographic Information Science, 34*(2), 77-102.

Maniadakis, D., & Varoutas, D.Structural properties of urban street networks of varying population density. *Proceedings of the 10th European Conference on Complex Systems (ECCS'13) Barcelona,* 1-6.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping∗. *The Professional Geographer, 55*(1), 31-42.

Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass, 3*(2), 727-745.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1/2), 17-23.

National Remote Sensing Centre. (2015). *Cartosat-I: A global IRS mission for large scale mapping and terrain modelling applications.* Hyderabad: National Remote Sensing Centre.

Newman, P. W., & Kenworthy, J. R. (1989). Gasoline consumption and cities: A comparison of US cities with a global survey. *Journal of the American Planning Association, 55*(1), 24-37.

Openshaw, S., & Openshaw, S. (1984). The modifiable areal unit problem.

Patel, S. (2013). Life between buildings: The use and abuse of FSI. *Economic and Political Weekly, 48*(36), 68-74.

Peponis, J., Allen, D., French, S., Scoppa, M., & Brown, J. (2007). Street connectivity and urban density. *6th International Space Syntax Symposium Istanbul: ITU Faculty of Architecture,*

Petrov, A. (2012). One hundred years of dasymetric mapping: Back to the origin. *The Cartographic Journal, 49*(3), 256-264.

Poli, D., & Caravaggi, I. (2012). Digital surface modelling and 3D information extraction from spaceborne very high resolution stereo pairs. *JRC Scientific and Technical Reports, Ispra,*

Rapoport, A. (1975). Toward a redefinition of density. *Environment and Behavior, 7*(2), 133.

Reibel, M., & Bufalino, M. E. (2005). Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A, 37*(1), 127-139.

Semenov-Tian-Shansky, B. (1928). Russia: Territory and population: A perspective on the 1926 census. *Geographical Review,* , 616-640.

Shaker, I. F., Abd-Elrahman, A., Abdel-Gawad, A. K., & Sherief, M. A. (2011). Building extraction from high resolution space images in high density residential areas in the great cairo region. *Remote Sensing, 3*(4), 781-791.

Shirgaokar, M. (2013). Limitations of the anti-floor space index position. *Economic and Political Weekly, 48*(29), 123-125.

Silvan-Cardenas, J. L., Wang, L., Rogerson, P., Wu, C., Feng, T., & Kamphaus, B. D. (2010). Assessing fine-spatial-resolution remote sensing for small-area population estimation. *International Journal of Remote Sensing, 31*(21), 5605-5634.

Stokols, D. (1972). On the distinction between density and crowding: Some implications for future research. *Psychological Review, 79*(3), 275.

Times News Network. (2014, November 1). It is official: Bangalore becomes bengaluru. *The Times of India*

Titarov, P. S. (2008). Evaluation of cartosat-I geometric potential. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 37*, 841-846.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association, 74*(367), 519-530.

U.S. Geological Survey. (2015). *Landsat—Earth observation satellites: U.S. geological survey fact sheet 2015–3081* doi:http://dx.doi.org/10.3133/fs20153081

U.S. Geological Survey. (2016). *What is the difference between a DEM and DSM*

Wright, J. K. (1936). A method of mapping densities of population: With cape cod as an example. *Geographical Review, 26*(1), 103-110.

Xie, Y. (1995). The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems, 19*(4), 287-306.

Xie, Z. (2006). A framework for interpolating the population surface at the residential-housing-unit level. *GIScience & Remote Sensing, 43*(3), 233-251.

Yeung, Y. (1977). High-rise, high-density housing: Myths and reality. *Habitat International, 2*(5-6), 587-594.

Yoo, E., Kyriakidis, P. C., & Tobler, W. (2010). Reconstructing population density surfaces from areal data: A comparison of tobler's pycnophylactic interpolation method and Area-to-Point kriging. 面状数据的人口密度面重构： Tobler's pycnophylactic 插值法和面到点克里金插值法的对比. *Geographical Analysis, 42*(1), 78-98.

IV. Domestic Piped Water Deficit in Bangalore

Abstract

This paper proposes a method for evaluating deficit in domestic piped water availability within Indian cities in a spatially disaggregated way. The proposed method uses data on land-use, land-cover, street density and building height to first generate a 30m resolution population dasymetric map for the central part of Bangalore city. This dasymetric map is then used in conjunction with the water supply network map and water use data to generate a 30m resolution map which estimates the volume of water available per capita per day in households across the study area. Results indicate that almost half the population within the study area could be receiving less than 100 liters per capita per day of domestic piped water. The analysis also shows that 31.65% of the total domestic water use within the study area could be dependent on some form of groundwater. The paper concludes with a discussion on inequality in access to domestic piped water supply within central Bangalore, as revealed by the proposed method.

1. Introduction

Domestic piped water supply in Indian cities exhibits a high degree of heterogeneity. Access to the supply network, frequency of supply and volume of water available to individuals vary widely within and across cities (Narain, 2012). Data from the 2011 Census shows that, in cities with more than 100,000 people, 72% of urban households have access to tap water from a treated source on average (Wankhade et.al., 2014).[1] But the level of access reduces significantly for cities with smaller population and the average across all urban areas is 62%.

The level of access to the piped network gives only a partial picture about the state of service provisioning since the volume of water available to individuals and the frequency of supply is usually inadequate. Data from 1405 Urban Local Bodies (ULBs) obtained through a service level status analysis exercise undertaken in 2010-11 (Ministry of Urban Development, 2012), indicates that average domestic water availability (measured at the consumption end) is 69.2 liters per capita per day (lpcd). [2] The average duration of

[1] The Census of India does not specifically ask whether households have access to municipal piped water supply. The data about access to 'tap water from a treated source' can be considered to be a reasonable approximation to municipal supply.

[2] Urban Local Bodies refers to the municipal governance institutions in India.

continuous supply, which meets standards of minimum supply pressure, across all the ULBs was only 3.2 hours a day.

Similar exercises conducted in 2011-2012 (1202 ULBs) and 2012-2013 (1067 ULBs) indicate average domestic water availability to be 78.5 lpcd and 100 lpcd respectively, while the median values were 74 lpcd and 85 lpcd. The data from 2012 and 2013 show that average duration of continuous water supply across all ULBs was 4.1 and 5.4 hours per day respectively (CEPT University, 2014).

1.1 Standards for water supply in India

In comparison to the service levels achieved, the Ministry of Urban Development sets the benchmark for average volume of domestic water supply (measured at the consumption end) at 135 lpcd and for duration of continuous supply at 24 hours a day (Ministry of Urban Development, 2012). On the other hand, standards for volume of supply set by the Central Public Health and Environmental Engineering Organization (CPHEEO), recommend a *maximum* of 150 lpcd (CPHEEO, 1999) for metropolitan and mega cities which have sewerage systems.[3, 4] The CPHEEO standard includes water requirements for commercial, institutional and minor industries, but adds that bulk supply to such establishments should be assessed separately.

Meanwhile, the Bureau of Indian Standards recommends 150 – 200 lpcd as the *minimum* water requirement for residents of urban areas with a population greater than 100,000. Out of this 45 lpcd is suggested as the volume of water required for toilet flushing purposes. It also recommends that the minimum water requirement standard can be reduced to 135 lpcd for households which are in Lower Income Group category or Economically Weaker Section of Society category (Bureau of Indian Standards, 1993).

1.2 Variation in water demand estimates and volume of water supplied across cities

While no Indian city manages to provide 24 hours continuous domestic water supply for all consumers (Ministry of Urban Development, 2012), there is wide variation in the volume of water supplied across Indian cities and in the water demand estimates made by ULBs. Based on a survey of 71 cities Narain (2012) points out that in spite of the multitude of existing standards related to volume of water requirement in Indian cities, every ULB appears to come up with its own benchmark for per capita water requirement. By

[3] The Census of India defines 'metropolitan' cities as those with a population more than 1 million (Census of India, 2011a). While it is unclear what CPHEEO means by 'megacities', the United Nations defines megacities as those with a population more than 10 million (United Nations, 2014).

[4] The CPHEEO does not define what percentage of the households in a city needs to be connected to a sewerage network for this standard to be applicable. The CPHEEO standard excludes Unaccounted for Water (UFW) which can be an additional 15% maximum. UFW refers to water which is supplied into the network, but is lost due to leakages, illegal connections etc. (CPHEEO,1999).

comparing official urban water demand estimates to city population, Narain (2012) shows that the per capita domestic water demand assumed by ULBs ranges from 120 to 389 lpcd.

Using data from a survey of 2734 households across seven large Indian cities, Shaban & Sharma (2007) show that actual water use depends more on volume of supply (availability at consumption end), rather than consumer demand.[5] Of the seven cities in their study, only Kolkata (115.6 lpcd) had average water consumption greater than 100 lpcd. The other six cities ranged from 77.1 lpcd in Kanpur to 96.2 lpcd in the case of Hyderabad with the average consumption across all seven cities being 92 lpcd. Meanwhile, the percentage of households which considered the level of water availability to be adequate ranges from a high of 82% in Madurai to a low of 49% in Hyderabad with an average of 71% across all seven cities.

1.3 Spatial inequality in domestic piped water availability within Indian cities

As described earlier, it is possible to understand extent of domestic access to piped water supply network in an aggregated manner at the ward level using Census data (Census of India, 2011b). Variation in volume of supplied water across Indian cities and average hours of continuous supply at a city level can also be reasonably well understood based on data from the service level status analysis (see Ministry of Urban Development, 2012) and the 71 city survey described in Narain (2012) apart from other smaller surveys like Shaban & Sharma (2007). But what is less clear is the heterogeneity or spatial inequality in domestic piped water availability within Indian cities.

While there are several studies which show the extreme inequality in access to water supply faced by slums within cities (see Gronwall et.al., 2010; Narain, 2012), studies which reveal spatial heterogeneity in volume of water received by households within a city are fewer in number. Narain (2012) shows that different parts of Delhi gets widely varying volumes of water per capita, but the method of analysis used is unclear. Other surveys in Delhi (Zerah, 2000) and Chennai (Srinivasan, 2008) also give an indication of the extent of heterogeneity in volume of piped water received by households with access to the supply network in different parts of these cities.

Using Bangalore as a case study, this paper proposes a method for improving our understanding of the spatial heterogeneity in volume of piped water received by households within Indian cities. Section 2 describes the data and methods used, while Section 3 describes the results obtained. The paper concludes with a discussion on some of the benefits and shortcomings of the proposed method.

[5] The cities surveyed in Shaban & Sharma (2007) are Delhi, Mumbai, Kolkata, Hyderabad, Kanpur, Ahmedabad and Madurai.

2. Methods and Data

Conceptually the proposed method consists of the following five steps:

a) Mapping the spatial distribution of population within the city at 30m resolution
b) Mapping the population with access to the water supply network
c) Estimation of the total population with access to the supply network within each spatial zone for which water supply / consumption data is available
d) Calculation of average water availability within each service station
e) Generation of deficit scenario maps using normative lpcd demand estimates

The subsections below describe the study areas, methods and datasets used in greater detail.

2.1 Extent of study areas

The municipal corporation of Bangalore was called Bangalore Mahanagara Palike (BMP) till 2007. The BMP boundary was expanded by government notification on 16 January, 2007 to include seven City Municipal Corporations, one Town Municipal Corporation and 110 villages which were outside the BMP boundary. This expanded entity was reconstituted as the Bruhat Bangalore Mahanagara Palike (Government of Karnataka, 2007).[6] Today it contains 198 wards and is called the Bruhat Bengaluru Mahanagara Palike or BBMP, which roughly translates to Greater Bangalore Municipal Corporation.[7]

While the initial intent of this paper was to analyze the population distribution and domestic piped water availability in the entire region within the current administrative boundary of the city of Bangalore (BBMP boundary), available land-use data was accurate only for the region within the 2007 administrative boundary of Bangalore (BMP boundary). As a result, the study area for population distribution comprises of parts of Bangalore city within the erstwhile BMP boundary. As discussed in Section 3, the spatial units of the water utility based on which water supply data is available, does not coincide with the ward boundaries. Hence a subset of these spatial units of water supply which cover much of the area within the BMP boundary is used as the study area for domestic water availability analysis.

Figure1 shows all the 198 wards of BBMP (Bruhat Bengaluru Mahanagara Palike, 2016) overlaid with a population density choropleth for BBMP wards which fall within the boundary of the erstwhile BMP which forms the study area for population distribution analysis. Section 3 describes in detail the process of delineating the boundary of the study area for domestic water availability analysis.

[6] Bangalore is the capital of the state of Karnataka
[7] In 2014, the name of the city was changed from Bangalore to Bengaluru (Times News Network, 2014). Since the analysis presented is based on Census data from 2011, I will refer to the city as Bangalore.

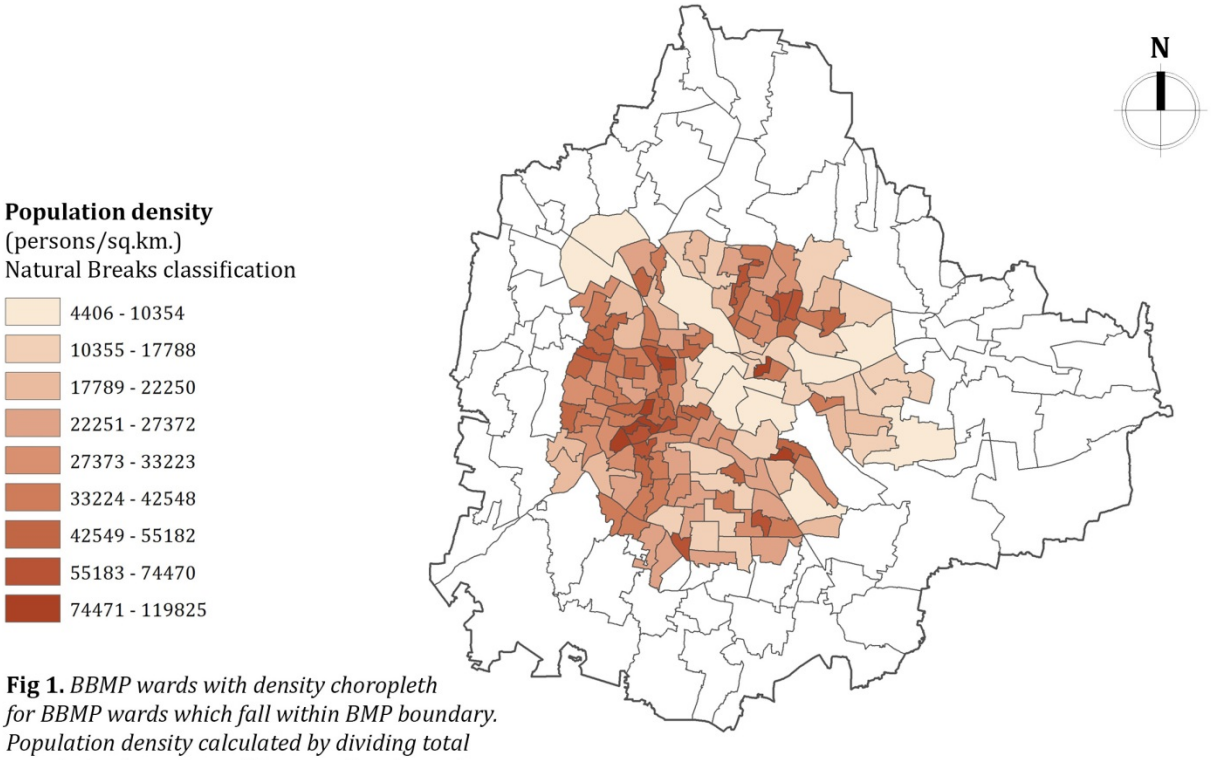


Fig 1. *BBMP wards with density choropleth for BBMP wards which fall within BMP boundary. Population density calculated by dividing total population in each ward by area of each ward. Population data is from Census of India (2011a)*

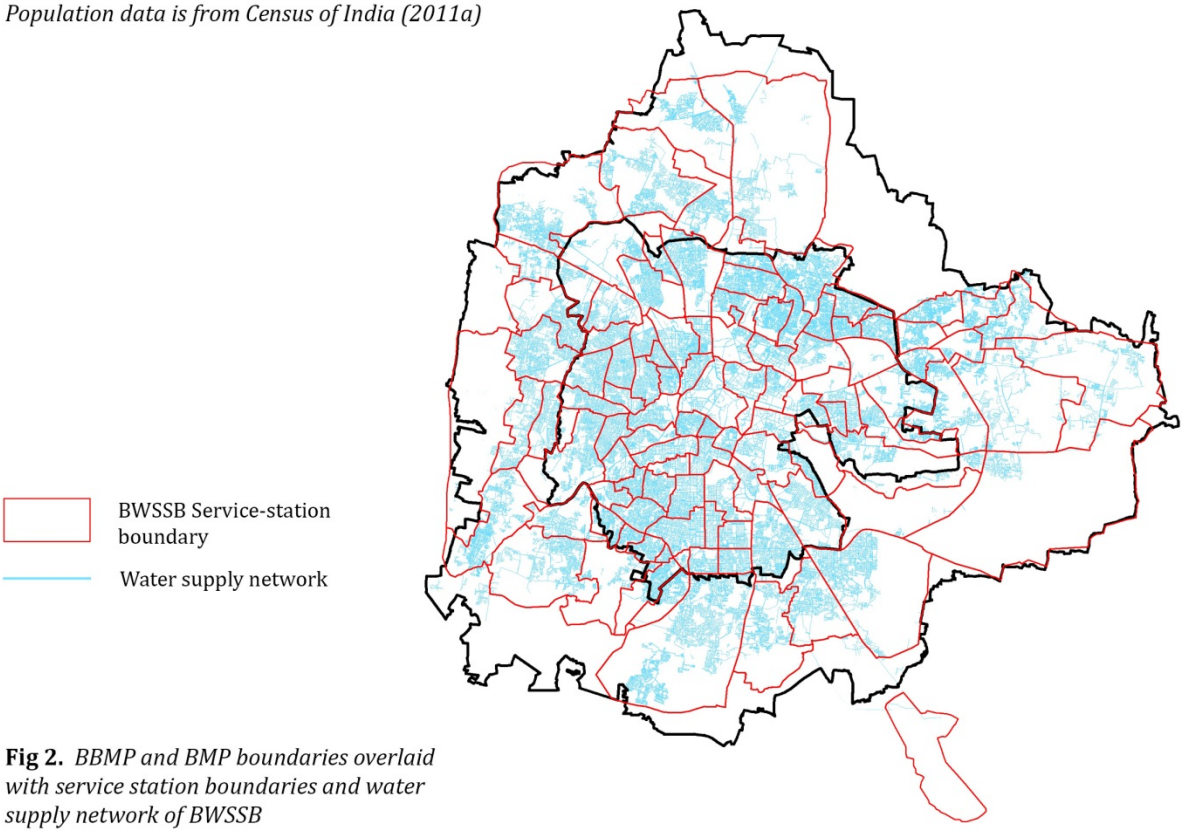


Fig 2. *BBMP and BMP boundaries overlaid with service station boundaries and water supply network of BWSSB*

2.2 Mapping population at 30m resolution

Balakrishnan (2016), demonstrates a framework for predicting population density at 30m resolution in Indian cities. The method outlined in this paper uses a similar approach to redistribute population (rather than predict population), such that the 'pycnophylactic' property is maintained (Tobler, 1979). Such population redistribution methods which use ancillary data to create a population density surface is referred to as dasymetric mapping (Semenov-Tian-Shansky, 1928; Wright, 1936; Mennis, 2009; Petrov, 2012).The proposed method uses data on land-use, land-cover, building heights and street density to redistribute ward level census population numbers to generate a 30m resolution dasymetric map such that the total population within each ward boundary remains accurate after redistribution.

Since the details regarding preparation of the various datasets are already described in Balakrishnan (2016), this section focuses on describing the population redistribution method. The description given builds on the overall predictive framework outlined in Balakrishnan (2016).

As described in Balakrishnan (2016), the residential population within a ward can be conceptualized as being distributed over three types of Residential Built-up Cells (RBCs) – Primary, Govt. and Informal.[8]

Let

$$P_w = P_{res} + P_{gov} + P_{inf} \quad (1)$$

$$P_w = \sum_w P_{c.res} + \sum_w P_{c.gov} + \sum_w P_{c.inf} \quad (2)$$

Where

P_w *is ward level population as per census*

$P_{res}, P_{gov}, P_{inf}$ *are population in Primary, Govt and Informal RBCs respectively in a ward*

$P_{c.res}, P_{c.gov}, P_{c.inf}$ *are cell level population in each Primary, Govt and Informal RBC in a ward*

$\sum_w$ *is summation at ward level*

Based on the method outlined in Balakrishnan (2016), $P_{c.gov}$ and $P_{c.inf}$ are known for all 30m cells. Therefore for each ward P_{gov} and P_{inf} are also known and hence P_{res} can be calculated as per Equation 3.

$$P_{res} = P_w - P_{gov} - P_{inf} \quad (3)$$

[8] Refer Balakrishnan (2016), for a description of Primary, Govt. and Informal RBCs.

Therefore only the known P_{res} values of a ward have to be redistributed across the primary residential built up cells (RBCs) within the ward. This can be achieved as described below.

Based on the results from Balakrishnan (2016), it is known that the population within a primary RBC is proportional to log of street density value of the cell and square root of building height value of the cell. This gives us Equation 4.

$$P_{c.res} \propto \ln(St.den_{c.res}) * \sqrt{Bld.ht_{c.res}} \tag{4}$$

Where

$$P_{c.res}\ is\ the\ population\ per\ 30mX30m\ Primary\ RBC$$

$$St.den_{c.res}\ is\ the\ street\ density\ value\ within\ a\ 30mX30m\ Primary\ RBC$$

$$Bld.ht_{c.res}\ is\ building\ height\ value\ within\ a\ 30mX30m\ Primary\ RBC$$

Since $P_{res} = \sum_w P_{c.res}$ we can derive Equation 5 from Equation 4

$$P_{res} \propto \sum_w [\ln(St.den_{c.res}) * \sqrt{Bld.ht_{c.res}}] \tag{5}$$

Therefore

$$P_{res} = k * \sum_w [\ln(St.den_{c.res}) * \sqrt{Bld.ht_{c.res}}] \tag{6}$$

Where

$$k\ is\ a\ constant\ of\ proportionality$$

Therefore

$$P_{c.res} = k * \ln(St.den_{c.res}) * \sqrt{Bld.ht_{c.res}} \tag{7}$$

For every ward, k can be calculated as shown below

$$k = \frac{P_{res}}{\sum_w [\ln(St.den_{c.res}) * \sqrt{Bld.ht_{c.res}}]} \tag{8}$$

Therefore within every ward,

$$P_{c.res} = \frac{P_{res} * \ln(St.den_{c.res}) * \sqrt{Bld.ht_{c.res}}}{\sum_w [\ln(St.den_{c.res}) * \sqrt{Bld.ht_{c.res}}]} \tag{9}$$

Using Equation 9 the population to be assigned to every 30mX30m primary RBC ($P_{c.res}$) within any ward can be calculated, such that the total population within all primary RBCs within the boundaries of a ward is equal to the known P_{res} value of that ward.Once the respective $P_{c.res}, P_{c.gov}$ & $P_{c.inf}$ values are assigned to all Primary, Govt. and Informal RBCs, the residential population dasymetric mapping can be completed.

2.3 Mapping the population with access to the water supply network

The Bangalore Water Supply and Sewerage Board (BWSSB) is the organization which is responsible for providing water supply and sewerage services within the Bangalore urban region. The area within which BWSSB is mandated to provide services is partitioned into zones called 'divisions'. Each division consists of multiple zones called 'subdivisions', each of which in turn consists of several smaller zones called 'service-stations' (Fig. 2). The water supply network data for each subdivision was obtained from the Bangalore Water Supply and Sewerage Board as a separate pdf file. The overall network map was generated by digitizing these pdf files and georeferencing and merging them in ArcGIS 10.2 (Fig. 2).

Water supply data (measured at the consumption end) and data on number of connections was collected from the BWSSB at the subdivision and service-station levels. But this data is given for four different categories of connections—Domestic, Partial Non-Domestic, Non-Domestic and Industrial. Partial Non-Domestic refers to residences which may have other small scale establishments attached where water may be used for non-domestic purposes (eg: dental clinics). As per the water accounting system followed by BWSSB, the volume of water supplied to the residential population within Govt. RBCs is subsumed within the Non-Domestic category. Therefore the water availability analysis in this paper focuses only on Primary and Informal RBCs since the total water available to these two types of RBCs can be calculated by combining data on the Domestic and Partial-Non Domestic categories.[9] The sum of volume of water supplied to Domestic and Partial-Non Domestic categories (measured at consumption end) is used for the domestic piped water availability analysis in the rest of this paper.

Conceptually, a buffer zone of stipulated width can be created around the water supply network in ArcGIS to identify the regions where RBCs potentially have access to the network. BWSSB officials stated that land parcels upto a maximum of 80'-90' from the supply network could potentially access the network (S. Narahari, personal communication, 29 October 2015). But since the residential plot sizes in some areas are considerably bigger, using a buffer distance of 90' would have missed several of the buildings which were set back from the adjoining streets in these large plots. Therefore after measuring the sizes of some of the largest residential parcels from the land-use

[9] The assumption that total domestic water availability is the sum of Domestic and Partial Non-Domestic categories, could lead to an overestimation of the water which is available for domestic purposes. But given the nature of data available, this appears to be the most reasonable way of proceeding.

dataset, 90m was used for the buffer distance (Fig. 4). This means that the extent of access to piped water network as indicated by Figure 4 may be an overestimate, since in most regular neighborhoods RBCs more than 90' from the network may not have access to piped water supply.

2.4 Estimation of population with access to supply network in each service-station

Once the buffer zone was created, it was used to clip the population dasymetric map and extract only the Primary and Informal RBCs which fell within the buffer. The Zonal Statistics as Table tool in ArcGIS 10.2 was then used to calculate the total residential population with access to supply network within each service-station by using the service-station boundaries as the 'zone' boundaries.

2.5 Calculation of water availability within each 30m cell

Monthly average water availability data (measured at the consumption end) for the calendar year 2015 was collected from BWSSB for each service-station. This was used to calculate an average daily water availability figure for each service-station for the entire year. But water supply measured at the consumption end is based on meter readings and could be an underestimate of total water available to residential users since there could be several users whose meters are not working or are using illegal connections.[10] The BWSSB estimates that across the city, an average of 11.28% of supplied water (measured at distribution end) could be 'unaccounted for' due to these reasons—of which 10% is attributed to malfunctioning meters and 1.28% to illegal connections.

But volume of supply at distribution end is known only at the subdivision level and that too in an aggregated manner for all categories of consumers. Based on the data available it is known that 92% of the total measured water consumption across the service-stations in the study area can be attributed to the Domestic and Partial Non-Domestic category users. If we assume that the 11.28% of supplied water which is 'unaccounted for' is actually consumed by all categories of users in proportion to their measured consumption, then we can infer that the proportion of supplied water consumed by users in the Domestic and Partial Non-Domestic categories—but unaccounted for—is 10.38% (92% of 11.28%). Therefore 10.38% of total water supplied (measured at distribution end) within each subdivision was calculated and this volume was allocated as share of 'unaccounted domestic water consumption' to each service-station within it. This allocation was done in proportion to the total number of Domestic and Partial Non-Domestic category connections within each service-station.[11]

[10] Some of the domestic users could be relying on public taps also. BWSSB estimates that 0.28% of total supplied water (measured at distribution end) is consumed via public taps. Since this a comparatively small figure it was ignored in the analysis.

[11] This allocation could not be done on the basis of population within each service-station since large parts of many subdivisions fall outside the study area for high-resolution population mapping.

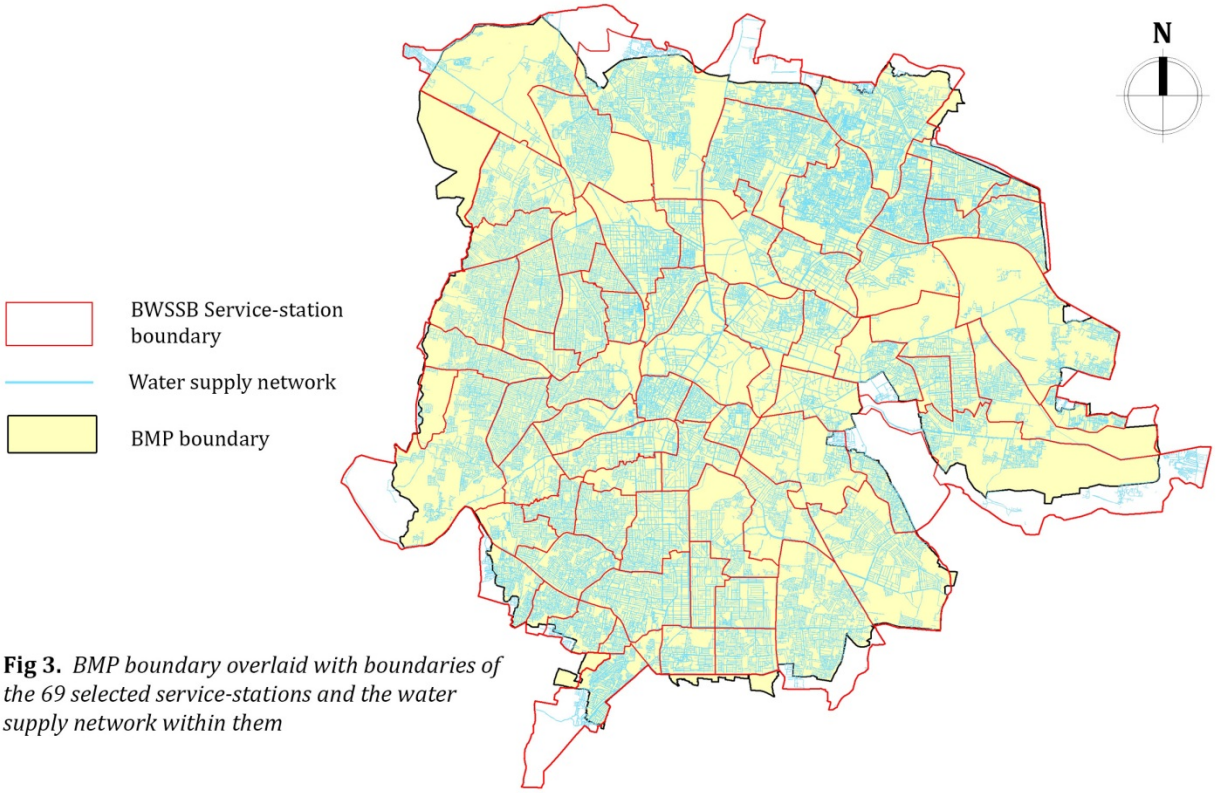


Fig 3. *BMP boundary overlaid with boundaries of the 69 selected service-stations and the water supply network within them*

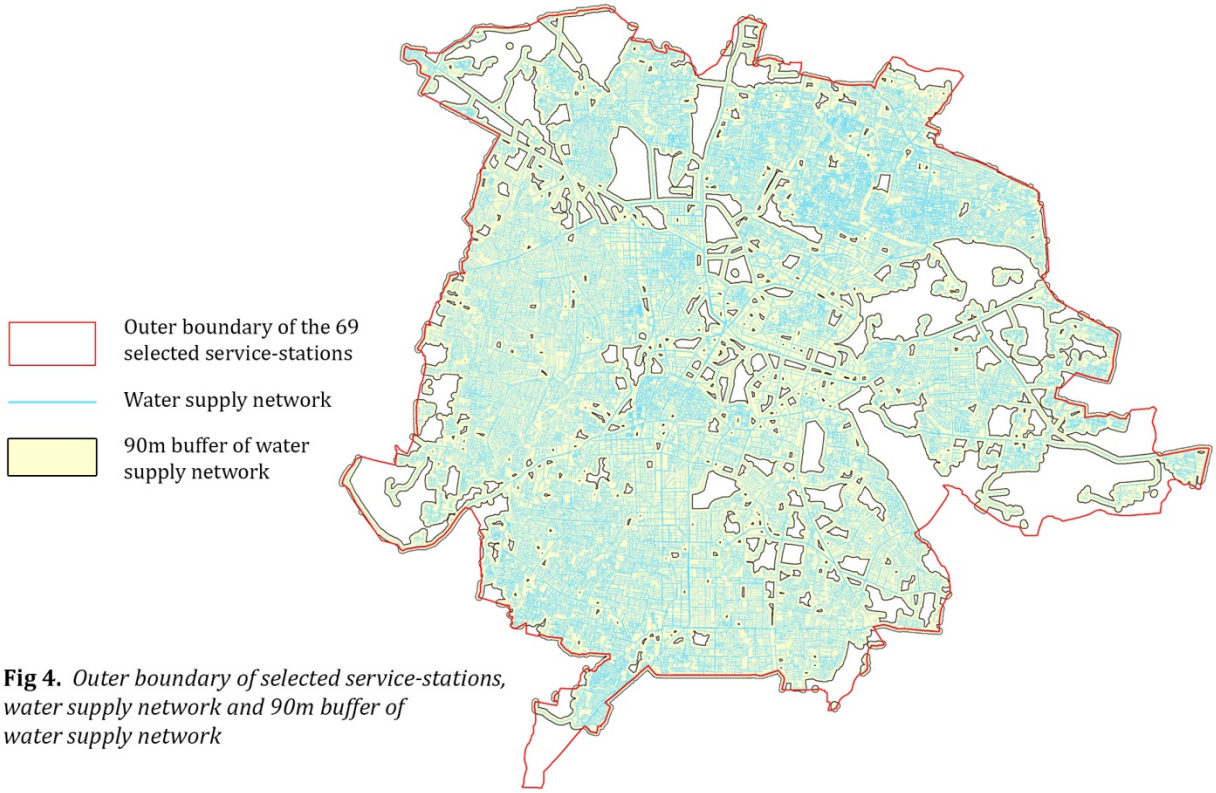


Fig 4. *Outer boundary of selected service-stations, water supply network and 90m buffer of water supply network*

Since we have no information on heterogeneity in supply within a service-station, there was no choice but to assume that the total water available for domestic use within each service-station is equally divided amongst the population which has access to the piped network. This gives us an average lpcd figure for each service-station (Fig. 7 & 8).

2.6 Generation of deficit scenario maps using normative demand estimates

Due to supply constraints, in Indian cities, the volume of water used by households is not dependent on demand, but on availability of water (Shaban & Sharma, 2007). Based on a 2005 survey of 2734 households across seven cities, Shaban & Sharma (2007), show that while there is variation in water use across socio-economic categories, the difference between the highest and lowest categories is around 20 lpcd. Their results indicate that average per capita water use ranges from 78.9 to 102 lpcd across socio-economic categories and from 77.1 to 115.6 lpcd across the seven cities.

According to the World Health Organization (WHO), average availability of 100-200 lpcd is considered the 'optimal access', while average availability of 50 lpcd is considered 'intermediate access' which can meet all basic consumption, hygiene and sanitation requirements (Howard & Bartram, 2003; World Health Organization, 2003). Gleick (1996) also suggests 50 lpcd as the minimum water required to meet basic human needs.

As discussed in Section 1.1, in India, the various norms for water supply set by government agencies are in the range of 135 to 200 range (for cities with population more than 100,000) with the Ministry of Urban Development stipulating 135 lpcd, measured at the consumption end, as the benchmark for volume of water availability. (Ministry of Urban Development, 2012).

Based on the above stated information, for this study I have chosen 100 lpcd as a conservative demand norm, which can be used to identify water deficit. While this may not be perfect, it can be considered as a demand scenario which is helping us understand a deficit scenario. Spatially disaggregated water deficit can then be estimated by first calculating demand within each 30mX30m RBC and then calculating the gap between water availability and normative water demand.

3. Results

Figure 5 shows the population dasymetric map for the BMP area prepared on the basis of the method outlined in Section 2.2, after removal of the Govt. RBCs. The Govt. RBCs have been removed since, as discussed in Section 2.3, the analysis in this paper focuses on Primary and Informal RBCs only. As Figure 2 shows, since the ward boundaries do not coincide with BWSSB service-station boundaries, it was necessary to select a subset of service-stations which provide a satisfactory coverage of the BMP region. Figure 3 shows the 69 selected service-station boundaries overlaid on the BMP boundary.

As can be seen from Figure 3, some service-stations which extend beyond the BMP boundary also had to be included in the study since these covered a significant part of the city within the BMP boundary. These service-stations were selected on the basis of the extent of their piped network which exists outside the BMP boundary. Only those service-stations where the extent of piped network outside the BMP boundary was relatively minor compared to the piped network extent inside the BMP boundary were selected.

The selected subset of 69 service-stations which define the study area was then used to clip the population dasymetric map yielding the Primary and Informal RBCs within the study area which was used for further analysis (Fig. 6). The buffer of piped water supply network shown in Figure 4 was then used to clip these RBCs, thereby providing the map of all Primary and Informal RBCs which potentially have access to piped water network. After this, based on the method described in Section 2.5, the per capita water availability within each service-station was estimated yielding the map shown in Figure 7.

This water availability estimate was then used to calculate the percentage of population which fall within various categories of water availability. Six water availability categories each of 50 lpcd bandwidth were created since 50 lpcd can be assumed to be the minimum requirement for basic human needs(Howard & Bartram, 2003; World Health Organization, 2003; Gleick, 1996) and I have chosen 100 lpcd as the normative demand for deficit scenario generation. Based on this, we can see that 46.93% of the population within Primary and Informal RBCs in the study area are water deficient (Fig. 8). Out of this, 13.5 percent of people do not get the 50 lpcd which is considered the basic minimum for human needs. Figure 8 shows a simplified map of the RBCs within the study area using these six categories of per capita water availability.

It is important to remember that the water deficit scenarios mentioned above refers only to deficit in availability of piped water supply. As studies in Indian cities have shown, households rely on a spectrum of water supply options to meet their water requirement. This can range from purchase of packaged water for drinking and potable uses to reliance on various forms of private self-supply based on wells or bore-wells, to purchasing water from private water tankers (Srinivasan, 2008; Ranganathan, 2014; Narain, 2012). Almost all of these alternatives to piped water supply depend on some form of groundwater (Narain, 2012). As a result, while estimates of availability in piped water supply as described above may not indicate absolute deficit, it may indicate some form of groundwater dependence (Narain, 2012; Foster and Mandavkar, 2008).

The method described in this paper could therefore also be used to generate overall estimates of domestic piped water supply and groundwater dependence. The total daily volume of domestic piped water supplied (measured at consumption end) to the 4.96 million people in the Primary and Informal RBCs within the study area is 442.21MLD.

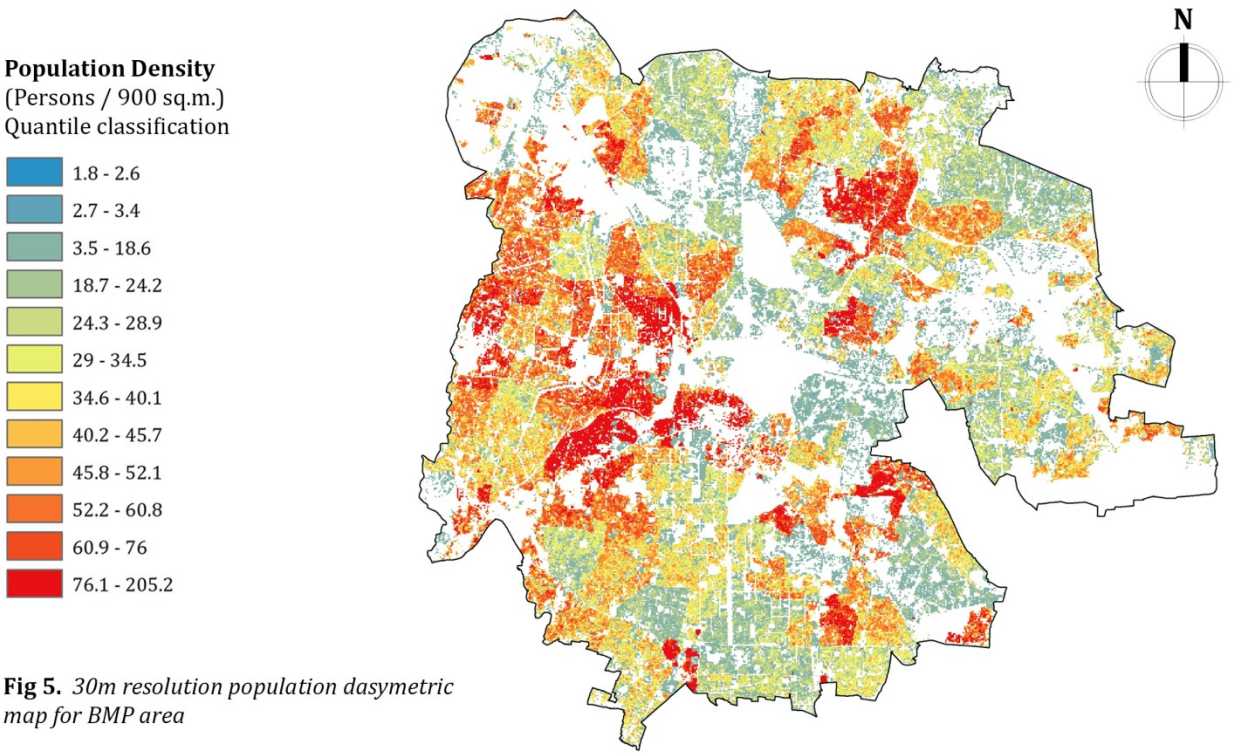


Fig 5. *30m resolution population dasymetric map for BMP area*

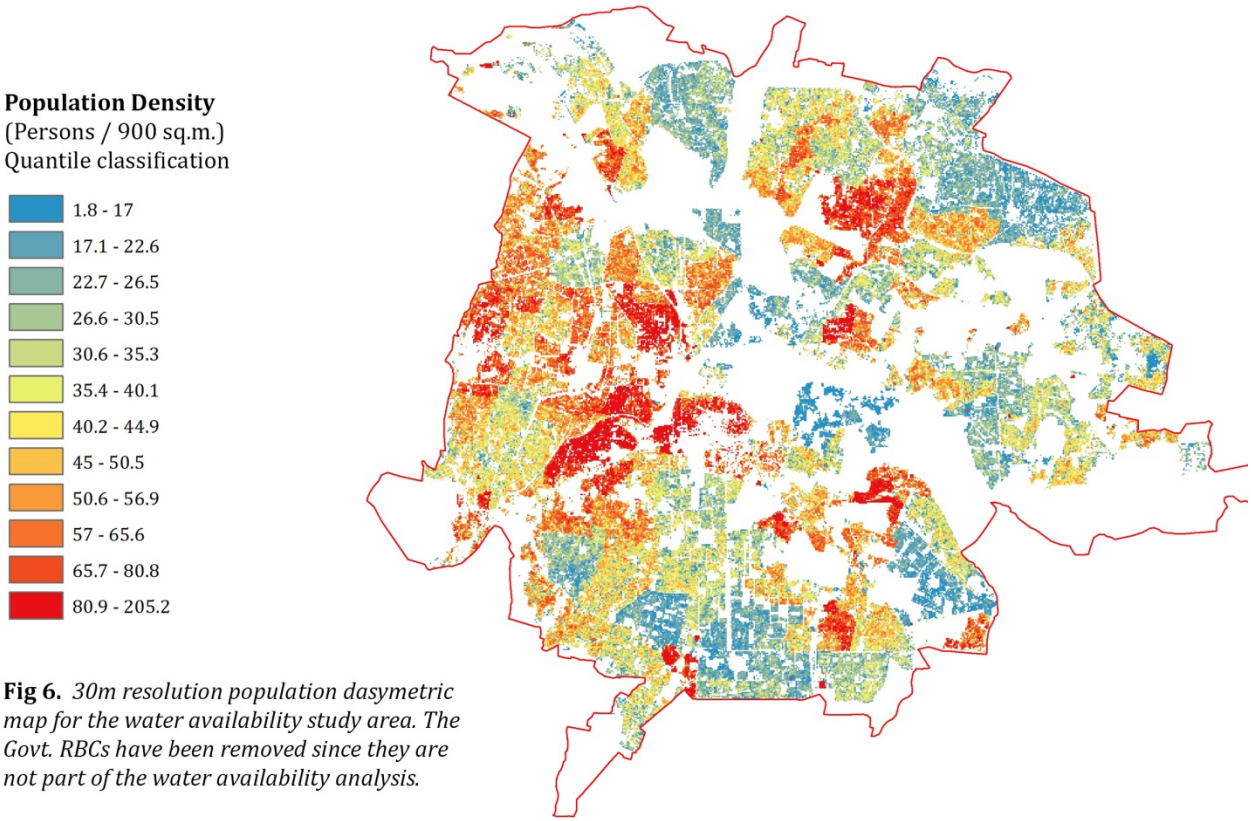


Fig 6. *30m resolution population dasymetric map for the water availability study area. The Govt. RBCs have been removed since they are not part of the water availability analysis.*

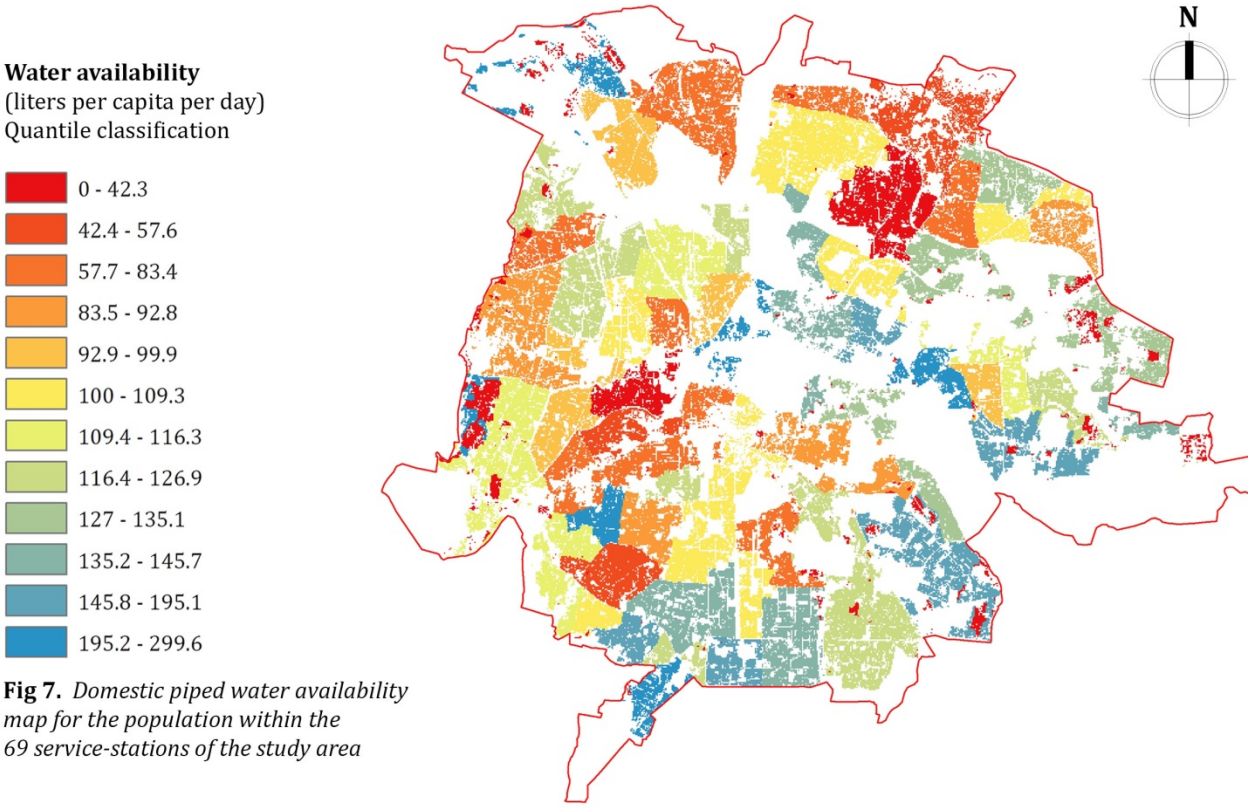


Fig 7. *Domestic piped water availability map for the population within the 69 service-stations of the study area*

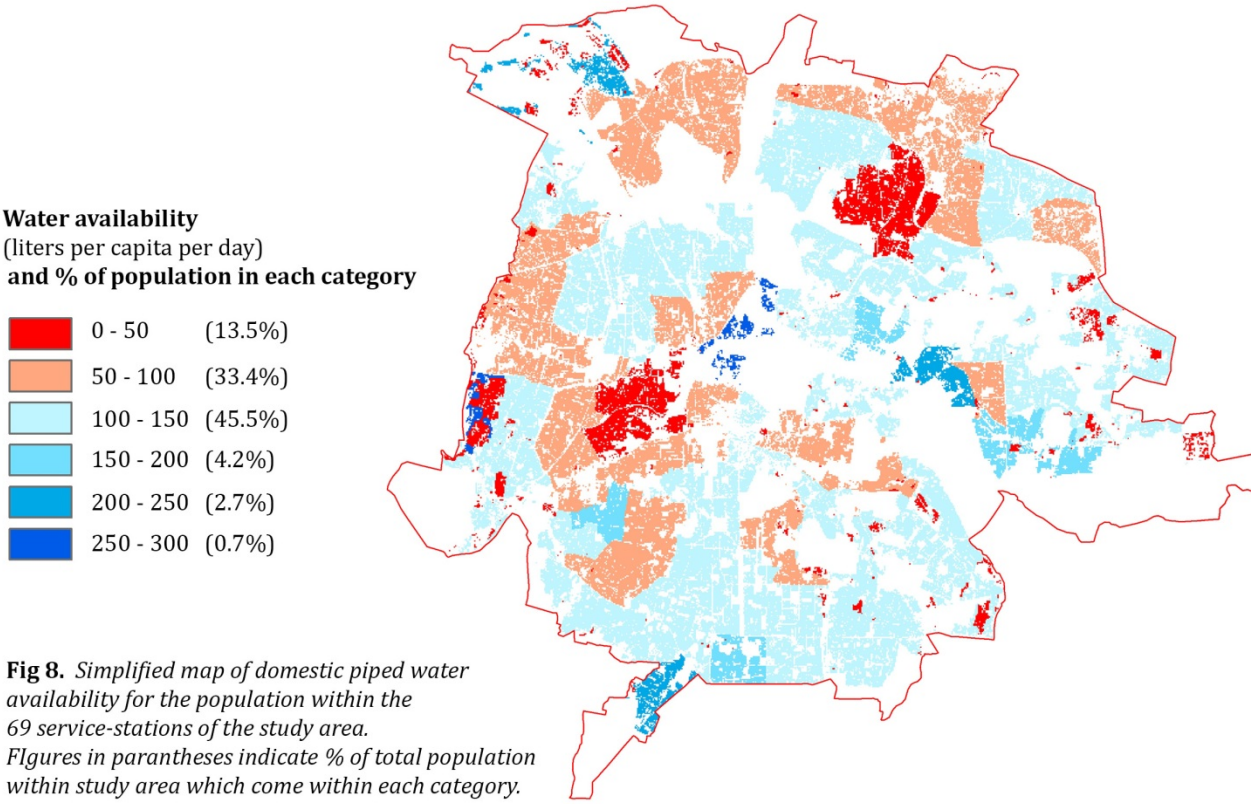


Fig 8. *Simplified map of domestic piped water availability for the population within the 69 service-stations of the study area. FIgures in parantheses indicate % of total population within study area which come within each category.*

The total domestic piped water deficit figure for the Primary and Informal RBCs within the study area can be calculated using the population dasymetric map shown in Figure 6 and the water availability map shown in Figure 7. Based on these datasets and using a normative demand estimate of 100 lpcd, the total domestic piped water deficit works out to 204.81 MLD.

If we assume that this deficit of 204.81MLD is somehow met using the various alternate water supply options discussed earlier, then the total domestic water use within Primary and Informal RBCs of the study area can be estimated to be 647.02MLD (442.21 MLD of piped water + 204.81 MLD from other sources). If, as discussed earlier, the deficit of 204.81MLD is met from alternate supply options which directly or indirectly depend on groundwater, then the percentage of total domestic groundwater use within the Primary and Informal RBCs of the study area can be estimated to be 31.65% of the total domestic water use described above.

4. Discussion

The results obtained indicate that almost half of the population in Primary and Informal RBCs within the service-stations selected for study are potentially experiencing deficit in domestic piped water availability. Figure 7 shows that significant disparity in piped water availability across different parts of the study area. While some of the residential areas in the central part of the city receive almost 300 lpcd of domestic piped water, many other parts of the city receive close to 50 lpcd or less.

4.1 Comparison with ward level asset ownership map

Figure 9 shows the domestic piped water availability map with ward boundaries overlaid on it. This enables us to compare the domestic piped water availability with average ward level wealth estimates as indicated by car ownership (Fig. 10). Comparing Figures 9 and 10, we can see that most of the wealthier wards (> 32.4% households owning cars) in the center of the study area and to the south and south-east have higher level of domestic water availability also. But some of the wards with similar levels of car ownership which are located to the south-west, west, north and east have significantly lower levels of water availability in comparison.

Although several of the wealthier wards appear to receive high levels of domestic piped water supply, based on the available data it is not possible to conclusively say whether wealthier neighborhoods in general tend to have higher levels of water availability. This is because we neither have any estimates for wealth below the scale of the ward, nor do we have any data on heterogeneity in water availability within a ward.

On the other hand, it seems reasonably clear that some of the poorest wards within the study area do have the lowest levels of per capita domestic piped water availability. As

Figure 9 shows, there are two clusters of wards—to the south-west and north-east of the center of the study area—which tend of have very low levels of water availability. The ward level car ownership map (Fig. 10) suggests that these are probably some of the poorest wards within the study area as well.

4.2 Comparison with sub-city typologies

By comparing the per capita water availability map with the sub-city typologies map (Fig. 11) for Bangalore (Balakrishnan & Anand, 2015), it is evident that most of the areas within the Low-SE typology wards also happen to have the lowest levels of per capita domestic piped water availability. In particular, the two clusters—to the south-west and north-east of the center of the study area—appear to be the worst off. These two sets of wards appear to have not only the lowest levels of water availability but also on average have very low housing quality, high percentage of Scheduled Caste population and very low scores on several other socio-economic indicators as shown in Figure 12.

4.3 Comparison with census data on access to domestic piped water supply

Figure 13 shows the percentage of households in BBMP wards who indicated that their primary source of water is 'tap water from a treated source' as per the Census of India (2011c). By comparing this to Figure 2 it is apparent that the wards within the BMP area which are well covered by the piped water supply network, have the highest percentage of households with access to 'tap water from a treated source', while the wards outside the BMP area score very low on this metric.

Upon comparing Figures 13 and 9 we can see that, in general, the wards towards the south in Figure 9 which have higher domestic piped water availability show up as wards where more than approximately 90% of the households state 'tap water from treated source' to be their primary source of domestic water supply. Similarly, the wards to the north, north-east and east in Figure 9 which have low levels of domestic piped water availability have lower values in Figure 13 also. But the somewhat surprising trend is that the wards to the west and south-west, including some of the poorest wards which as per Figure 11 come within the Low SE typology, and which as per Figure 9 have the lowest levels of domestic piped water availability, appear to have very high values in Figure 13. I do not have a satisfactory explanation for why this could be happening.

The analysis presented in this paper shows that even within the wards which as per the Census have the highest levels of access to 'tap water from treated source', and which as per the BWSSB piped water supply network data have the highest levels of network coverage, 31.65% of the total domestic water is actually dependent on some form of groundwater. If one were to conduct a similar analysis for the entire BBMP area, then the percentage of total domestic water use which is dependent on some form of groundwater is bound to be significantly higher.

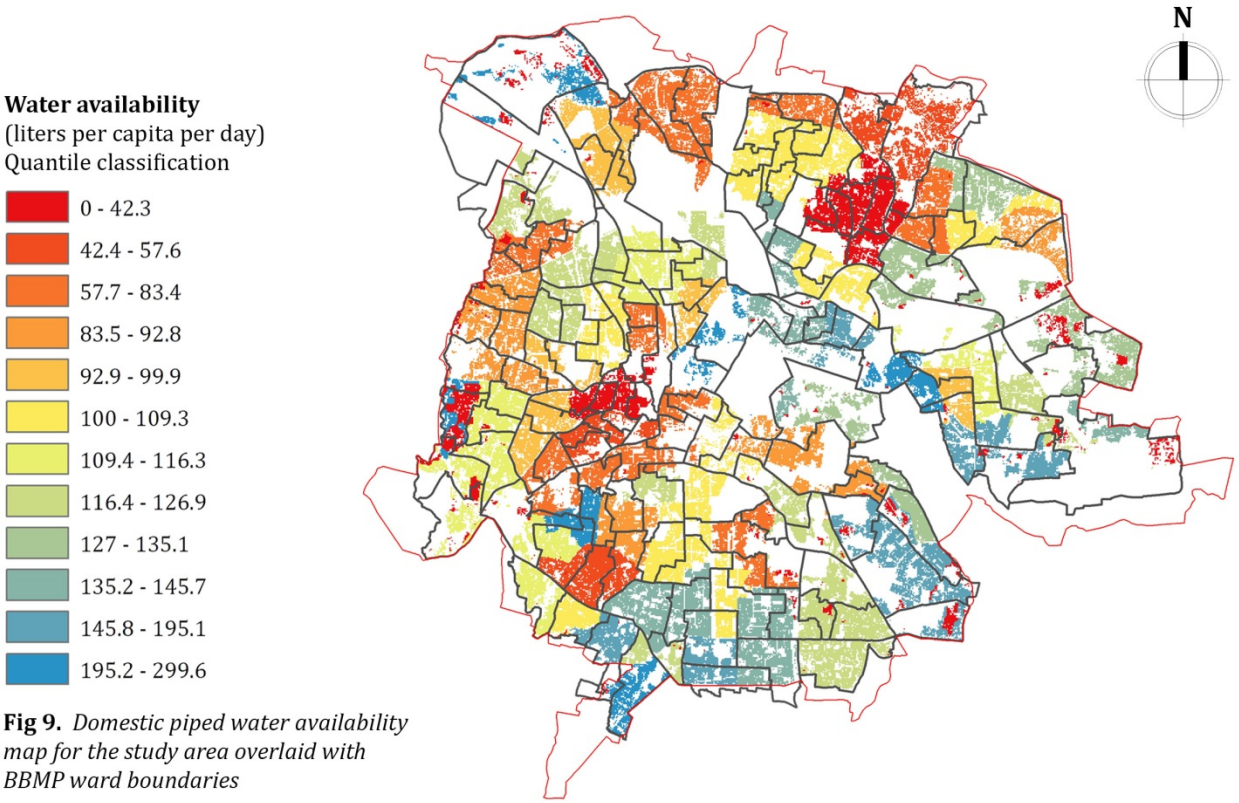


Fig 9. *Domestic piped water availability map for the study area overlaid with BBMP ward boundaries*

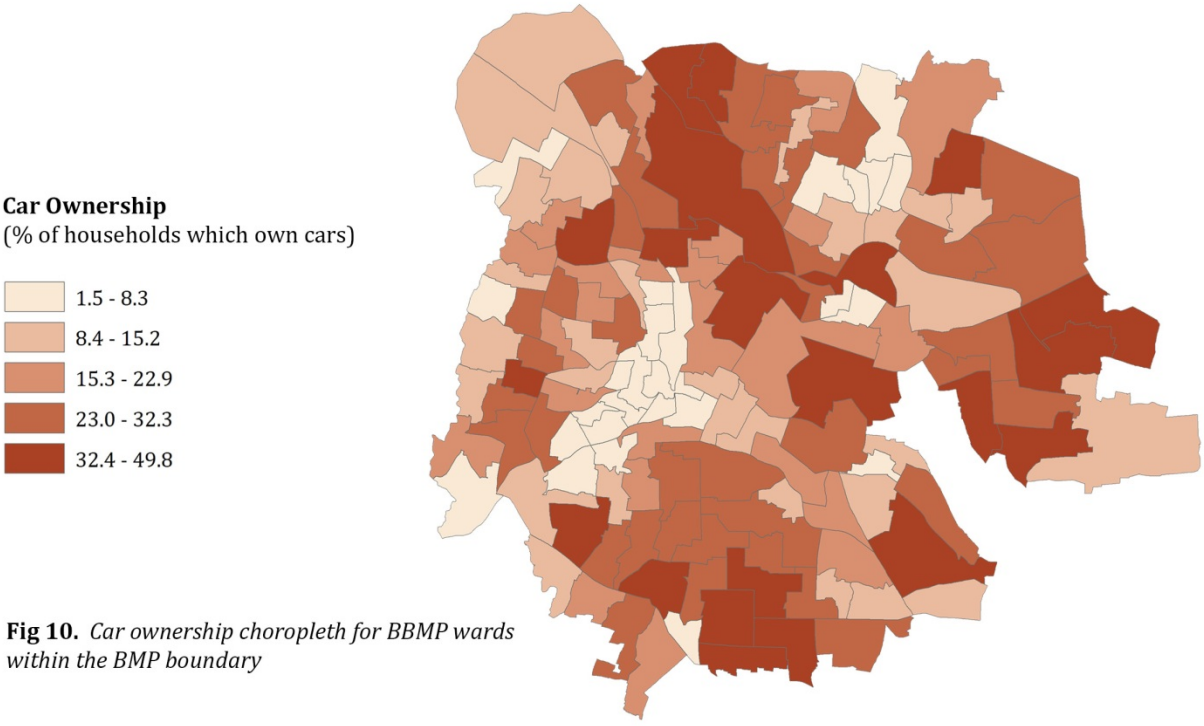


Fig 10. *Car ownership choropleth for BBMP wards within the BMP boundary*

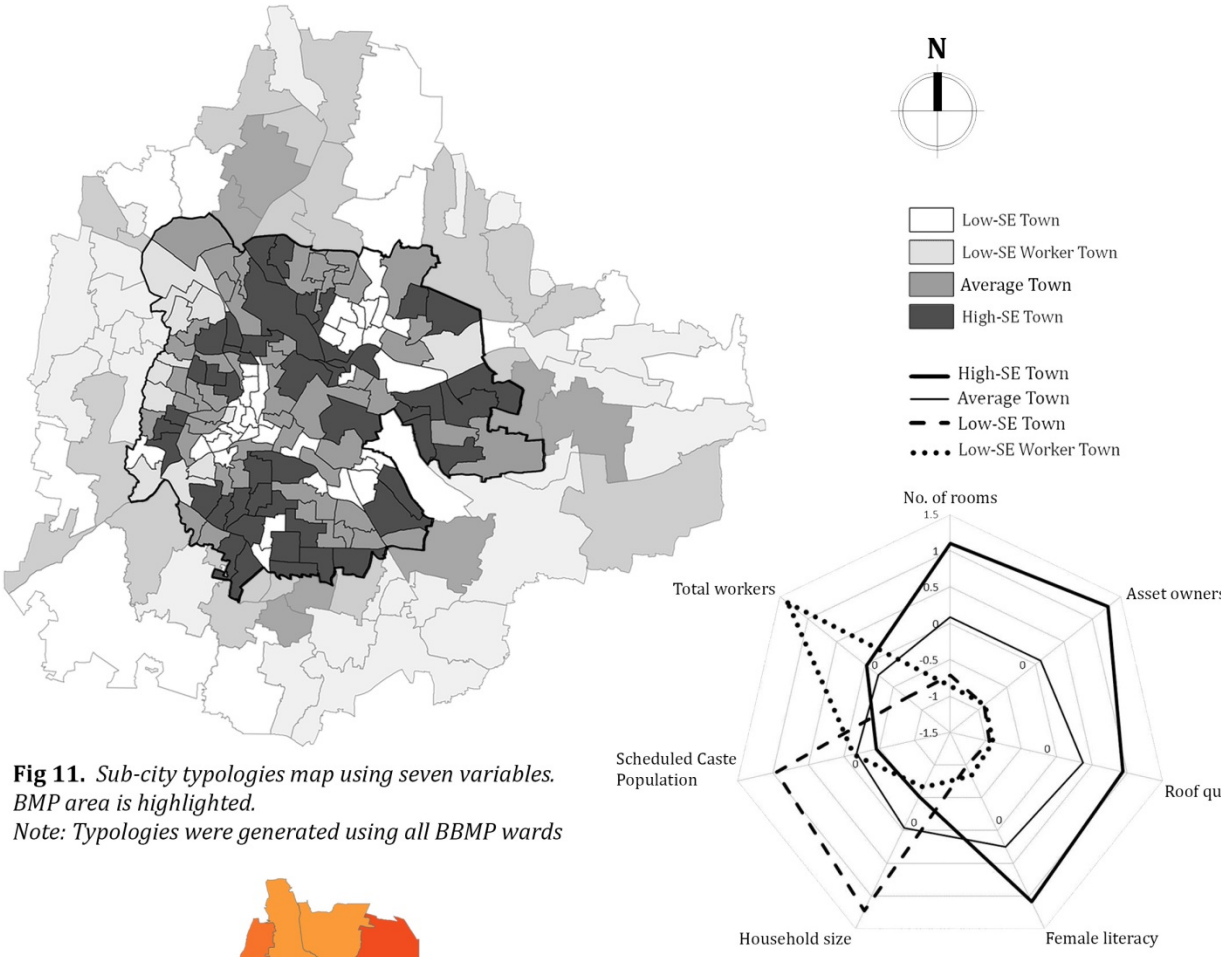


Fig 11. *Sub-city typologies map using seven variables. BMP area is highlighted.*
Note: Typologies were generated using all BBMP wards

Fig 12. *Radar diagram showing cluster/typology m for each of the seven variables. Values are normalize line indicates city mean.*

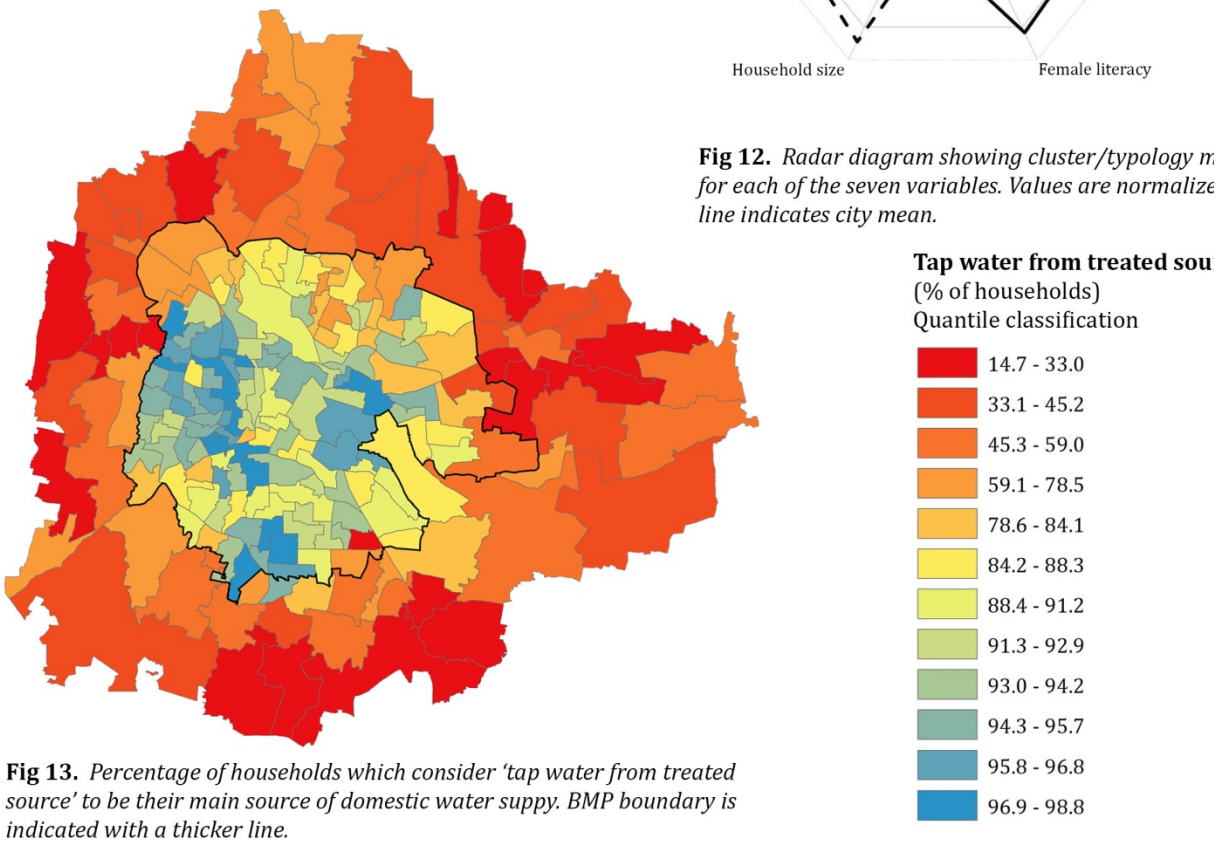


Fig 13. *Percentage of households which consider 'tap water from treated source' to be their main source of domestic water suppy. BMP boundary is indicated with a thicker line.*

Besides, this study has only analyzed domestic water use, whereas industrial and commercial uses are also very significant in terms of total volume. At the least, the wards outside the BMP area are bound to show high dependence on some form of groundwater for industrial and commercial uses also. Therefore it seems very likely that if we were to do a water availability analysis for the entire BBMP area taking into account domestic, commercial and industrial water uses, the total share of groundwater use is bound to be high—probably significantly higher than the 50% mark.

4.4 Shortcomings of the analysis

The analysis presented here should be interpreted only as a water availability scenario which is based on a normative demand scenario of 100 lpcd. Since water use in Indian cities is constrained by level of supply, it is not possible to have a clear understanding of what the actual water demand of various socio-economic groups or neighborhoods are. Therefore in some sense it could be claimed that the "actual deficit" may be lower in many areas if people have just worked their lives around the reality of low water availability and therefore make do with less than 100 lpcd.

But at the other end of the spectrum, it is also possible that there are neighborhoods which get 150 or 200 lpcd but find that insufficient because of the kind of lifestyle these neighborhoods or households aspire to. So there could be overestimation of "actual deficit" at the lower end of the water availability spectrum and underestimation of "actual deficit" at the upper end of the spectrum. But what is very clear from the analysis presented, is the spatial inequality in levels of domestic piped water availability.

The total domestic water availability has been calculated by adding up the total water consumption figures for the Domestic and Partial Non-Domestic categories for all service-stations. This could lead to an overestimation of the total water which is actually available for domestic uses, since some of the water from the Partial Non-Domestic category may be used for small establishments which are operating out of residences (eg: health clinics). Moreover, as described in Section 2.5, an adjustment factor has been applied to account for water that is used but not accounted for in the consumption figures due to faulty meters or illegal connections. This adjustment factor is based on estimates by the BWSSB for their entire service area and its applicability for the study area could be debated.

As discussed in Section 3, there are a few service-stations which have some amount of piped network extending outside the BMP boundary. For the purpose of analysis, all the water consumed within these service-stations were assigned to the RBCs within the BMP area since the land-use data required to prepare the population dasymetric map was complete only for the BMP area. Therefore in some of the outer edges of the study area there could be RBCs whose water availability has been overestimated.

There could also be errors in the dasymetric mapping at the ward level since the $P_{c.gov}$ and $P_{c.inf}$ values have been taken from Balakrishnan (2016) which uses a population prediction framework that cannot as yet be fully validated at the 30m resolution due to lack of data at the appropriate level. Besides, the Census population figures used in the dasymetric mapping are from 2011 (Census of India, 2011a), while the water availability data is from 2015. Although it is possible that much of the population growth and land-use change since 2007 has happened outside the BMP boundary (Balakrishnan, 2016), there could be errors which arise from the use of datasets from different years.

Another potential issue with the analysis is that the network maps obtained from BWSSB could be incomplete as isolated pipeline fragments were noticed in several areas, especially towards the outer edges of the BMP area. In addition, the assumption that all RBCs within 90m of any pipeline has access to the water supply network may not be very robust since according to BWSSB officials, only land parcels which are within 80' – 90' of the network can potentially access it (S. Narahari, personal communication, 29 October 2015). But since the analysis was based on built up cells and not land parcels, it was necessary to use a 90m buffer distance such that RBCs in the center of very large plots which had pipe lines adjacent to the plot boundary but not within 90' of the RBC, would not show up as RBCs without water supply network connectivity.

5. Conclusion

This paper describes a method for generating population dasymetric maps of Indian cities and using them to understand resource demand / deficit scenarios. Using data for the central part of Bangalore city, the paper shows how this method can be used to understand domestic piped water availability within Indian cities in a spatially disaggregated way.

The dasymetric mapping method demonstrates a potential application of the empirical relationship between population density of a 30mX30m Primary RBC and its street density and building height values as described in Balakrishnan (2016). The analysis using data on water availability illustrates the stark spatial inequalities in availability of domestic piped water supply, apart from potential dependence on various forms of groundwater

The water availability analysis also shows that almost half the city could be experiencing some level of deficit in domestic piped water availability, while 31.65% of the total domestic water use within the study area could be dependent on some form of groundwater. The analysis also highlights the extreme inequality faced by some of the wards within the BMP area, which not only have very low availability of piped water supply, but also have very low scores on various socio-economic and housing quality indicators.

Acknowledgements

I would like to thank Mr. Thippeswamy for patiently answering all my queries related to the BWSSB water supply system in Bangalore. Mr. Narahari, Mr. Maheshwarappa and Ms. Lakshmi Yadav provided access to various datasets and also answered my questions regarding the functioning of BWSSB. Manish Gautam assisted in implementing an initial version of the population dasymetric mapping presented in this paper. I would also like to thank Dr. Deepak Malghan for discussions on population modeling and water availability analysis. All errors in the manuscript remain my responsibility.

Part of the work described in this paper was enabled by a Junior Research Fellowship awarded by the American Institute of Indian Studies.

References

Balakrishnan, K. (2016). Densities of bangalore. *Manuscript in Preparation,*

Balakrishnan, K., & Anand, S. (2015). Sub-cities of bengaluru. *Economic & Political Weekly, 50*(22), 63.

Bruhat Bengaluru Mahanagara Palike. (2016). BBMP ward information. Retrieved from http://bbmp.gov.in/en/ward-info

Bureau of Indian Standards. (1993). *Indian standard code of basic requirements for water supply, drainage and sanitation* (Fourth Revision, Fourth Reprint ed.). New Delhi: Bureau of Indian Standards.

Census of India. (2011a). *Primary census abstract data tables*. New Delhi: Ministry of Home Affaris, Govt of India.

Census of India. (2011b). *Percentage of households to total households by assets and amenities.* New Delhi: Census of India.

Census of India. (2011c). *Percentage of households to total households by amenities and assets*. New Delhi: Ministry of Home Affairs, Govt of India.

Central Public Health and Environmental Engineering Organization. (1999). *Manual on water supply and treatment*. New Delhi: Ministry of Urban Development.

CEPT University. (2014). *Improving urban water supply and sanitation outcomes: Service level benchmarking databook*. Ahmedabad: CEPT University.

Foster, S., & Mandavkar, Y. (2008). *Groundwater use in aurangabad – A survey and analysis of social significance and policy implications for a medium-sized indian city.* (Case Profile Collection No. 21). Washington D.C.: The World Bank.

Gleick, P. H. (1996). Basic water requirements for human activities: Meeting basic needs. *Water International, 21*(2), 83-92.

Government of Karnataka. (2007). *Notification of areas under bruhat bangalore mahanagara palike* Government of Karnataka.

Gronwall, J. T., Mulenga, M., & McGranahan, G. (2010). *Groundwater, self-supply and poor urban dwellers A review with case studies of bangalore and lusaka.* (Human Settlements Working Paper Series No. Water and Sanitation - 26). London: International Institute for Environment and Development. Retrieved from http://pubs.iied.org/pdfs/10584IIED.pdf

Howard, G., & Bartram, J. (2003). *Domestic water quantity, service level, and health* World Health Organization Geneva.

Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass, 3*(2), 727-745.

Ministry of Urban Development. (2012). *Service levels in urban water and sanitation sector: Status report (2010 - 2011)* (First Edition ed.). New Delhi: Ministry of Urban Development, Government of India.

Narain, S. (2012a). *Excreta matters.* (State of India's Environment: A citizens report No. 7 (Vol.1)). New Delhi: Center for Science and Environment.

Petrov, A. (2012). One hundred years of dasymetric mapping: Back to the origin. *The Cartographic Journal, 49*(3), 256-264.

Ranganathan, M. (2014). 'Mafias' in the waterscape: Urban informality and everyday public authority in bangalore. *Water Alternatives, 7*(1)

Semenov-Tian-Shansky, B. (1928). Russia: Territory and population: A perspective on the 1926 census. *Geographical Review,* , 616-640.

Shaban, A., & Sharma, R. (2007). Water consumption patterns in domestic households in major cities. *Economic and Political Weekly,* , 2190-2197.

Srinivasan, V. (2008). *An integrated framework for analysis of water supply in a developing city: Chennai, india.* (Unpublished PhD). Stanford University, Retrieved from http://www.indiawaterportal.org/node/7846.

Times News Network. (2014, November 1). It is official: Bangalore becomes bengaluru. *The Times of India*

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association, 74*(367), 519-530.

United Nations. (2014). *World urbanization prospects 2014: Highlights* United Nations Publications.

Wankhade, K., Balakrishnan, K., & Vishnu, M. J. (2014). *Urban water supply and sanitation in india: Sustainaing policy momentum.* (). Bangalore: Indian Institute for Human Settlements.

World Health Organization. (2003). *The right to water.* (Health and Human Rights Publication Series No. 3). Geneva: World Health Organization.

Wright, J. K. (1936). A method of mapping densities of population: With cape cod as an example. *Geographical Review, 26*(1), 103-110.

Zerah, M. (2000). *Water: Unreliable supply in delhi*. New Delhi: Manohar - CSH.

V. Conclusion

The research presented in this dissertation proposes methods for understanding heterogeneity within Indian cities. The methods described in the three papers, use existing datasets from secondary sources for this purpose. The results obtained by applying these methods to Bangalore city reveal significant heterogeneity across multiple dimensions.

1. Summary of results

Figures 1 to 5 show heterogeneity in socio-economic and housing conditions, residential population density and domestic piped water availability across Bangalore. The map of sub-city typologies (Fig. 1) indicates that there are several clusters of wards within the BMP boundary which have very low values for socio-economic and housing indicators (Low SE Town). Two large clusters of wards—to the south-west and north-east of the center—are especially prominent. We can also see that there are large clusters of wards which belong to the High-SE category, to the north, south and eastern parts of the area within the BMP boundary.

The paper on density prediction demonstrates a new method for generating building heights at a city-wide scale with a root mean square error of 3.06 to 3.12m. Although other authors have used Cartosat-1 stereo imagery to estimate building heights for smaller neighborhoods in urban India (see Saha, 2014), this is the first time that building height extraction and accuracy estimation has been conducted at this scale for an Indian city using Cartosat-1 stereo imagery.

Figure 3 shows that most of the wards which come within the Low-SE Town also tend to have neighborhoods with very high levels of population density. Since the highest population density class in Figure3 has a very wide range (76.1 – 205.2 persons/900sq.m.), Figure 4 shows the variation within this density class. While many neighborhoods within the High SE Town wards have population densities in the range of 20 to 25 persons/900sq.m. or less (Fig. 3), several areas within the Low SE Town wards have population densities which are more than five times this value (Fig. 3 & 4).

Figure 5 shows the variation in access to domestic piped water availability across Bangalore. Once again we can see that there are several areas which get less than 50 liters per capita per day (lpcd) of water while many other parts of the study area receive more than five times that volume. The analysis presented in this dissertation also shows that approximately 31.65% of domestic water use, within the study area, is dependent on direct

or indirect sources of groundwater. It is important to bear in mind that the study focused only on domestic piped water supply in the central part of Bangalore which has the highest levels of piped network connectivity. Therefore, if a similar analysis were to be carried out for all types of water uses for the entire area within the current administrative boundary of Bangalore city, the total extent of groundwater dependence is bound to be much higher.

2. Heterogeneity, inequality, inequity

Figure 1 and Figure 2 together indicate the extent of residential segregation of scheduled caste population at the ward level in Bangalore. As Figure 2 shows, the Low-SE Town wards have very high proportion of Scheduled Caste population while having very low values for other housing and socio-economic indicators. Since the 1950s, several researchers have highlighted the extent of residential segregation by caste, religion, region of origin and ethnicity which is prevalent across Indian cities (Gist, 1957; Bose, 1965; Mehta, 1968; Mehta 1969; Joy, 1975; Prakasa Rao & Tiwari, 1979; Mahadevia, 1991; Vithayathil & Singh, 2012; Sidhwani, 2014). In particular, Mehta (1969) uses data from 1822 to 1965 to show how residential segregation on the basis of caste and religion has remained relatively stable over almost 150 years in the city of Pune in western India.

The paper on sub-city typologies shows that apart from ward level residential segregation on the basis of caste, the wards which have high scheduled caste population tend to also be spatially clustered and have low values for other socio-economic and housing quality indicators on average (Fig. 1 & Fig. 2). The results from population density prediction and dasymetric mapping show that these wards which have a high proportion of scheduled caste population also tend to have the highest population densities (Fig. 3). The final paper on domestic piped water availability further shows that many of these areas receive the lowest levels of piped water supply while being connected to the supply network (Fig. 5).

The term “heterogeneity”, used liberally in this dissertation, refers only to variation in an attribute—including instances where the variation is purely random. As discussed in the first paper, the word “inequality” begins to convey a certain normative implication. The results for Bangalore, obtained using the methods proposed in this dissertation, calls for a reassessment of the appropriateness of these terms. The systematic and severe spatial inequalities evident in Bangalore are nothing short of serious “inequities.”

Significantly, there could also be a substantial level of residential segregation on the basis of religion in Bangalore. But the available data does not permit us to understand this dimension of residential segregation and the potential spatial inequalities and inequities it engenders.

Fig 1. *Sub-city typologies map using seven variables. BMP area is highlighted.*
Note: Typologies were generated using all BBMP wards

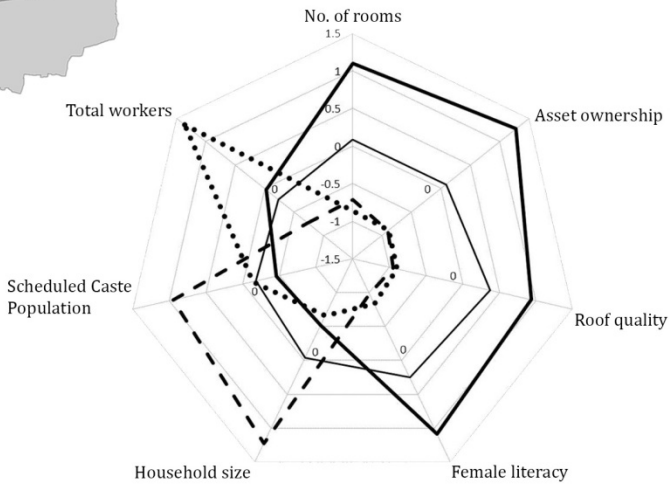
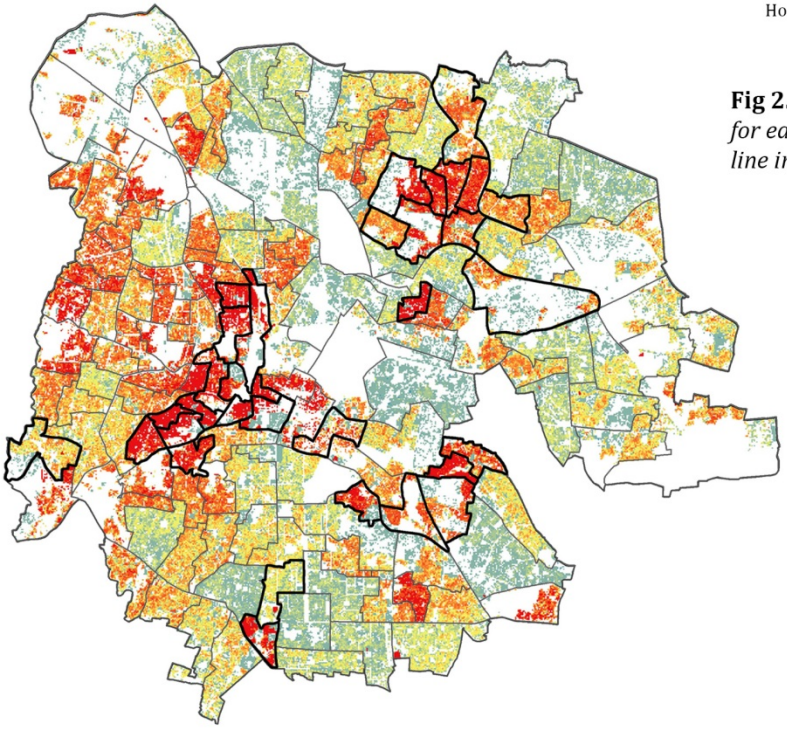


Fig 2. *Radar diagram showing cluster/typology means for each of the seven variables. Values are normalized, '0' line indicates city mean.*



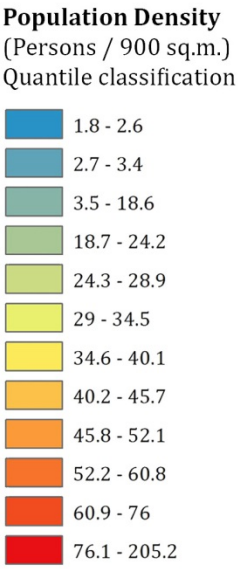


Fig 3. *30m resolution population dasymetric map for BMP area, overlaid with ward boundaries. Thicker lines indicate boundaries of Low-SE Town wards*

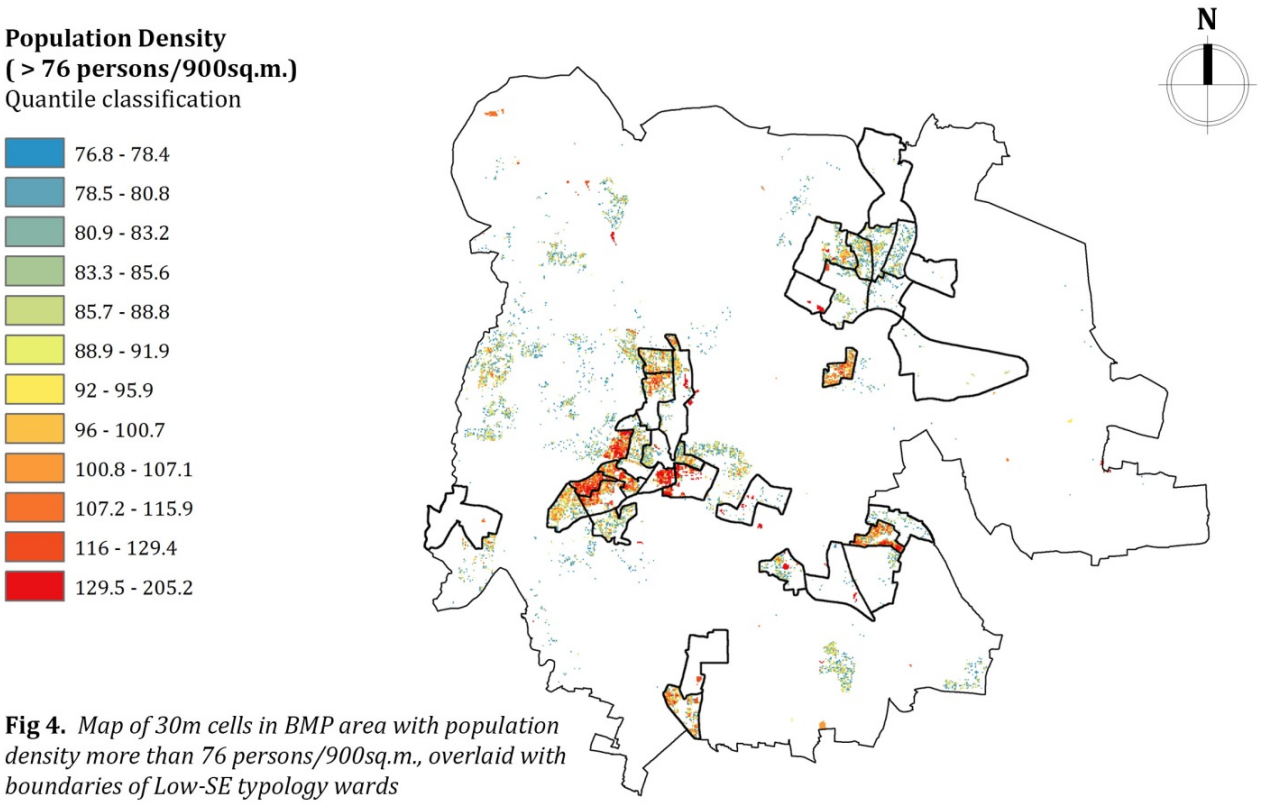


Fig 4. *Map of 30m cells in BMP area with population density more than 76 persons/900sq.m., overlaid with boundaries of Low-SE typology wards*

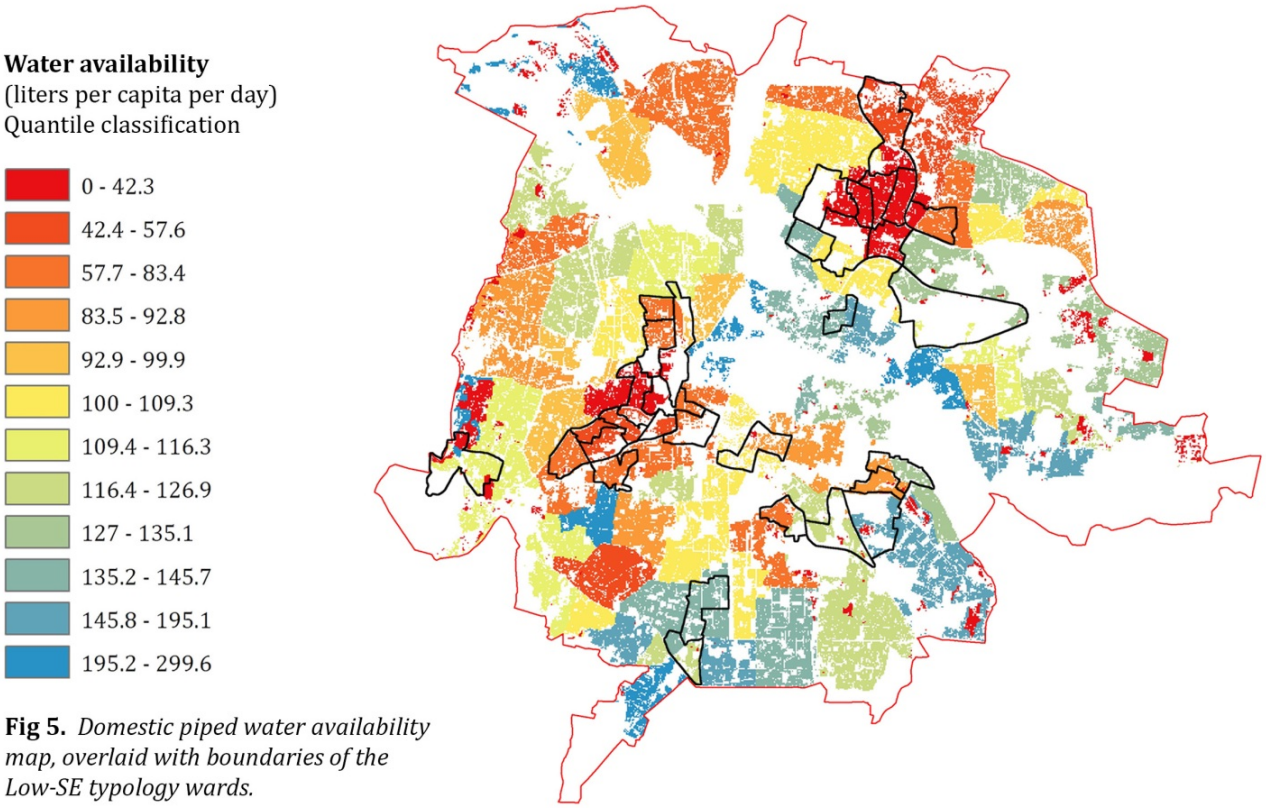


Fig 5. *Domestic piped water availability map, overlaid with boundaries of the Low-SE typology wards.*

3. Directions for further research

Each of the methods presented in this dissertation could be extended to several other cities in India for which the required data is currently available. For example, the census data used for the analysis presented in the first paper on sub-city typologies is available for thousands of cities across the country since August 2014. Census data for these cities could be used to understand whether the sub-city typologies identified are stable or variable across size classes and city types.

The method for high-resolution density prediction illustrates a spatially non-stationary relationship between population density, street density, building height and ward level asset ownership indicators. A sound theoretical foundation for this is currently lacking and could be a direction for future research. The building height estimation method developed in this paper on density prediction can be applied to almost any city in India, since it uses stereo satellite data which is readily available for practically the entire country. As noted elsewhere, the building height estimation method does not need expensive DGPS surveys or proprietary stereo image processing software.

The high-resolution population density prediction and dasymetric mapping methods can potentially be used to analyze living conditions, access to urban amenities and demand and availability of a wide range of resources across any city. In short, the proposed methods can be taken forward into a variety of intra-urban research areas where the focus is on per capita demand/availability/access to any resource or amenity. Besides, high-resolution population distribution information can be combined with various other datasets to understand extent of exposure to hazards such as urban flooding.

In comparison to the methods proposed in the first two papers, the analysis of domestic piped water availability can be extended to relatively fewer cities in India at present since information on water supply network may not be readily available for many cities.

Although the analysis presented in the first paper helps us understand the heterogeneity within Bangalore at the ward level, the population density prediction and dasymetric mapping presented here shows the potential variation which exists at even finer scales. Wards which appear to belong to one typology in the first paper can be seen to have neighborhoods of many different spatial types within them as per the 30m resolution population prediction and redistribution maps. This should make us very cautious about even ward level generalizations within Indian cities. It also shows that it is important to understand the heterogeneity within Indian cities at the scale of neighborhoods.

References

Bose, N. K. (1965). Calcutta: A premature metropolis. Scientific American, 213, 90-102.

Gist, N. P. (1957). The ecology of bangalore, india: An east-west comparison. Social Forces, , 356-365.

Joy, A. (1975). The Spatial Organization of a South Indian City: Coimbatore,

Mahadevia, D. (1995). Emerging Process of Residential Segregation in Metropolitan Cities: Case Study of Mumbai and Chennai,

Mehta, S. K. (1968). Patterns of residence in poona (india) by income, education, and occupation (1937-65). American Journal of Sociology, , 496-508.

Mehta, S. K. (1969). Patterns of residence in poona, india, by caste and religion: 1822–1965. Demography, 6(4), 473-491.

Prakasa Rao, V., & Tewari, V. (1979). Structure of an indian metropolis: A study of bangalore.

Saha, D. K. (2014). DSM extraction and evaluation from cartosat-1 stereo data for bhopal city, madhya pradesh. International Journal of Scientific and Research Publications, 4(5)

Sidhwani, P. (2015). Spatial inequalities in big indian cities. Economic & Political Weekly, 50(22), 55.

Vithayathil, T., & Singh, G. (2012). Spaces of discrimination. Economic & Political Weekly, 47(37), 60-66.