2014－C44

Some notes on the spatial representation

Economic Division in British India

Souknilanh Keola and Kenmei Tsubota

March, 2015

Institute of Developing Economies
Japan External Trade Organization

調査研究報告書 (Chosa Kenkyu Report)
［Interdisciplinary Studies Center］2014－C44
［Economic Division in British India］

Interim Report for *Economic Division in Colonial India*, IDE-JETRO, 2014

Some notes on the spatial representation*+

Souknilanh Keola and Kenmei Tsubota

Institute of Developing Economies, JETRO

March, 2015

Abstract

This paper shows two notes on the representation methods of spatial data. Firstly, with the use of remote sensing technique and data, it proposes one representation of cities in space. Population density grid data shows the distribution of human residents. Focusing on the densely populated regions, it shows the continuous shape of the cities. The shape of the cities can be compared inter temporally across space at different population size. Some examples are shown with Indian data. Secondly, for the representation of geographical units, c.f. administrative area, it is pointed out some measurement errors stemming out from the construction of data. Specifically, centrality of geographical units such as polygons of districts would have certain deviations from the true centre of the districts. With Indian districts data, taking the true centre at the location of district headquarters, we examine the distribution of the measurement errors and correlations with other variables. Results show that there are measurement errors and they correlate with size of the districts.

---

* This work is a part of the project on "*Economic Division in Colonial India*". As the title of this report shows, the project is still in progress and this paper includes very preliminary results and analysis. Its contents may be modified based on the updated available data.

+ We would like to acknowledge Arif Khan and his team for the data entry and Sai Hihi for her excellent research assistance.

1. Introduction

Remote sensing technique is a useful tool for the analysis on distributions of certain activities which not only include human but also nature itself. Changes of such distributions are particular interests of the geographers. By applying remote sensing technique, this paper offers some attempts of examinations that are relevant to economic geography, urban geography, and population geography.

First, it provides an attempt to show cities in a continuous function of population density. In urban economics, cities are assumed to be monocentric or polycentric in terms of population density and land rent. These shapes can be examined by employing the remote sensing technique and grid data. However, the representation and the comparison of the shapes have some difficulties. For example, the decline in the gradients of the population density from the center is different among the cities with respect to the population size and other characteristics and cities are not in one dimension but distribute in all the directions with height of population density meaning 3 dimension. Since the nature of the city is the concentration of people, first note examines how it differs among the size of cities. While the shape of population density itself is traditional question or is rather familiar in urban economics, the use of remote sensing data allows us to compare systematically among different cities. At the current stage of the analysis, the comparison of the shapes among different cities is not completed. So the main analysis is still left. However the procedure and the scope are discussed.

Second, it shows how representation of geographical units may contain certain spatial measurement errors. One of the simplest representations of geographical units is the centrality of the units. The centrality of the geographical units can be obtained easily and shows relative position in space. However, there should be a caveat how this centrality represent the centre of the variable of interests. If the variable is uniformly distributed over the geographical unit, the centrality is exactly the same as its centre. However, if the variable is not distributed uniformly, there should be a certain gap between the centrality of the unit and the true centre. For example, population may not be uniformly distributed in it and there should be different between the centrality of population and that of the unit. This gap can cause measurement errors when we employ the centrality of the geographical units for any quantitative analysis. With using the Indian district data, we find there is a significant correlation between the measurement error and area size of the geographical units. The size of geographical units is normally determined by political and/or administrative purposes. For example, population size affects political pressure to split the geographical units into much smaller for better access to public facilities and facilitates effective provision of public services. Thus smaller geographical regions tend to be densely populated.

2. Representation of cities by population density functions

In Urban economics, the shape of the population density function is to be monotonically decreasing. This is because of the distance from the city center where there are employments. Suppose there are homogeneous workers in terms of skill and wages. As the distance from the city center increases, the land price may decrease the associated lot size of each house increase and the population density decrease (c.f. Fujita (1989)). The shape of the population density function was an empirical question in 1960's to early 1990's. For example, McDonald (1989) reviews articles from 1950s to late 80's. The initiation of such studies is done by Clark (1951). Such studies of population density had not been the patented article only for urban economics. In the literature of remote sensing, the recognition technique and transformation of information which have multiple dimensions had been well advanced. For example, see Brivio and Zilioli (2001) and Batty and Kim (1992). In this literature, most of the studies employ negative exponential functions or inverse power functions. It is important to note that most of the analyses estimate the decreasing function from CBD to its fringe as half of the city. Having the fact that city spreads to 360 degrees, it should be emphasized that the differences among the degrees also explain the characteristics of the shape of the city and such features are needed to be included.

With using the remote sensing grid data of population density, it is a straightforward procedure to have comparative data across cities. As long as there is one centre in the city, negative exponential function would fit to the population density function. However, if there are more than one centres, the errors of the fit would increase and classification of the urban configurations may be required. The system of cities normally has regularity such as Zipf's law. Corresponding city shapes may also have certain regularities suggesting the fractal structure which was pointed out by Batty and Longley (1994).

As an application of the studies on the shape of cities, India offers vast number of cities (Figure 1). From the Census in 2011, there are three megacities which have 10 million populations; Mumbai, Bengaluru and Hyderabad. There are 46 large cities which have more than 1 million populations. The definition of cities varies among countries. Having different definition of cities by different minimum population does not affect the analysis of population density functions. It is because that as long as we can detect the centres of regions, we can draw the shape of the agglomeration of people. One of the difficulties may be the recognition of regional centres. In particular, when the boundaries of each region are connected, the choice of centres among the regions may be the source of difficulties. Simple principle may be needed for such analysis.

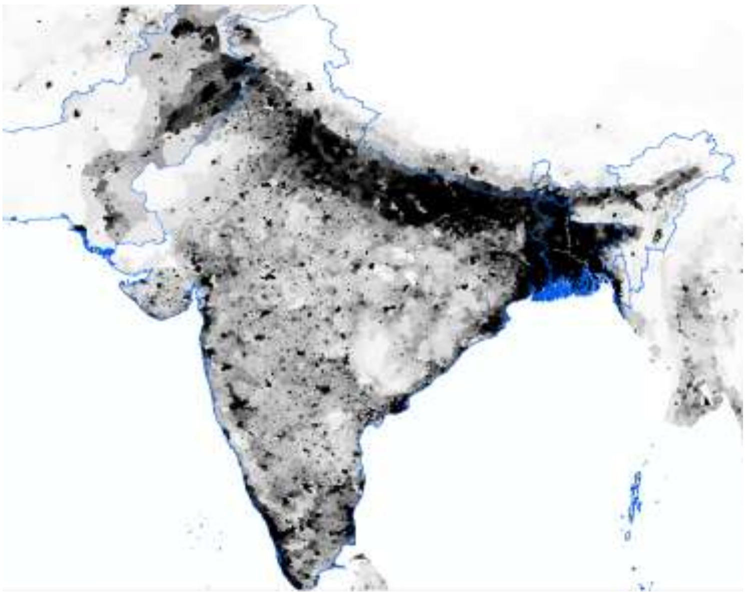

Figure 1. Population density of South Asia

By constructing the grid data around the centre of each city, it is straightforward to estimate the population density function for each. Since the satellite images from Landsat are available from early 1990s, it can be examined how population geography has evolved over time. One important point which is mostly neglected in the literature is the directional difference of the urban expansion. As an example, Figure 3 shows the shape of a city.

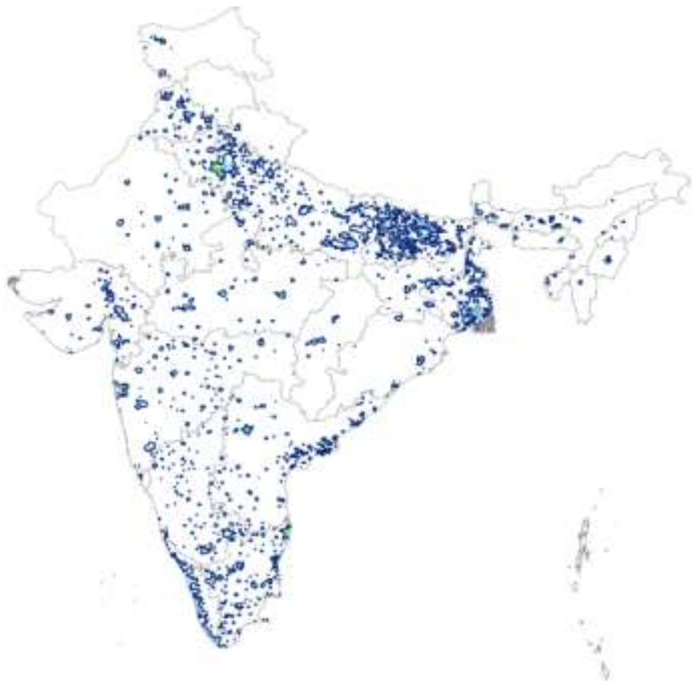

Figure 2. Highly populated areas in India

Figure 3. Contours at different thresholds (Kolkata)

There are some subcentres of the cities. The decision of the inclusion to the largest city or isolation as independent cities is still pending at the current stage of our analysis and clear decision rule is needed. It may be one idea to change the decision rules and compare the different outcomes under different decision criteria. If we separate out all of the unconnected cities, we can take it as a conservative definition of city and may lose some commuting behaviors and the effects of subcentres. On the other hand, distance based inclusion rule, any agglomerations within 30km buffer from the highest populated area, can be recognized as the definition close to metropolitan style. Appropriate distance for each city is not obvious. Such distance is needed to be justified from the data itself.

3. Representation bias of centrality

This section shows the systematic bias stemming out from the use of centrality of polygons. When the precise representative locations of administrative boundary are not available or costly, centralities of geographical units are frequently employed as alternatives. Centrality of geographical units represents the unweighted centre of it. The use of this centrality implicitly assumes that the variable of interest is uniformly distributed. Thus, if the target variable is not uniformly distributed, the measurement errors are always associated with this representation.

With using Indian districts data, this section shows how such errors are systematically distributed (c.f. such errors are relatively larger when the districts have larger size). There are 592 districts. All of the centres are obtained from India Place Finder[1], which is approximately the centre of highly populated areas. After obtaining the centrality, we calculate the Vincenty-style calculation of distance between the true centre and the centrality of districts. The summary statistics is shown in Table 1. Graphical representation of this measurement error is shown in Figure 4. Taking the Vincenty-style distance as vertical axis and the area size as horizontal axis, scattered plots of shows positive correlations of these variables.

	25%	50%	75%	mean	Min	Max
measurement error	9.686	16.603	27.11	31.298	0.848	471.888

Table 1. Measurement error as the distance between centrality and district centre

[1] http://india.csis.u-tokyo.ac.jp/

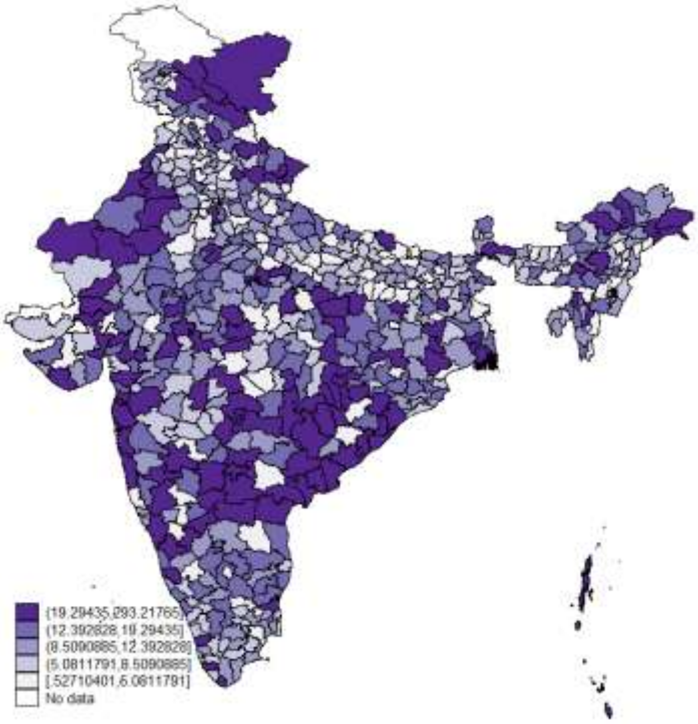

Figure 4. Distance between centrality of polygons and actual centre

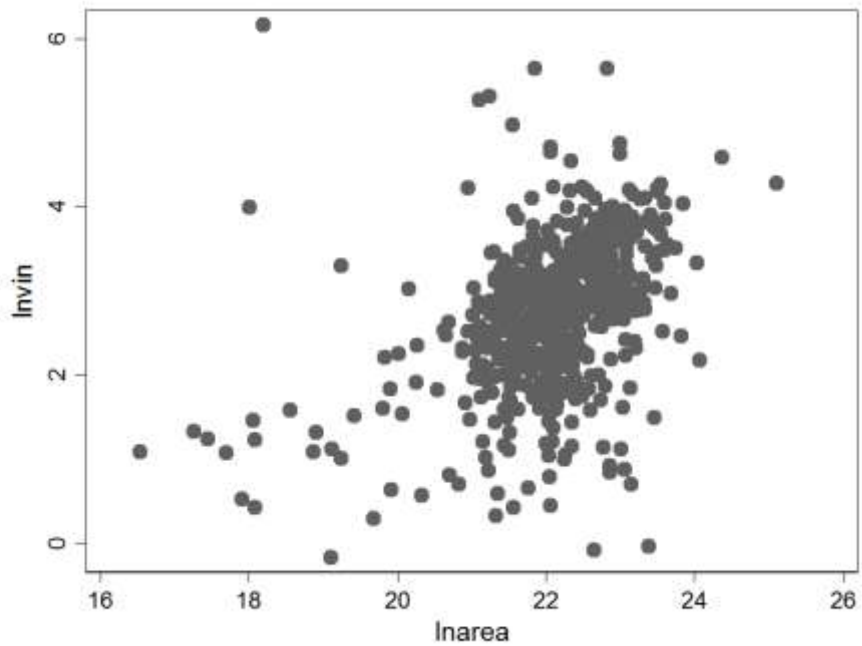

Figure 5. Distance between centrality of polygons and actual centre and its area size

Taking logarithm of area size and population size, we estimate simple OLS regression with some variables. Area size is positive significant. Square of area size is also positive significant when we include it. Inclusion of total population shows negative significant and enlarge the coefficient of area size. However, this significance of total population suggests that the area size is correlated with population size. The inclusion of population density is also in the same direction.

	(1)	(2)	(3)	(4)	(5)	(6)
ln(Area)	0.367***	-0.868	0.399***	-0.857	0.288***	-0.969
	[0.0336]	[0.637]	[0.0347]	[0.632]	[0.0410]	[0.632]
ln(Area)^2		0.0291*		0.0296**		0.0296**
		[0.0150]		[0.0149]		[0.0149]
ln(Total Population)			-0.111***	-0.112***		
			[0.0337]	[0.0336]		
Population Density					-0.111***	-0.112***
					[0.0337]	[0.0336]
Constant	-5.335***	7.708	-4.493***	8.785	-4.493***	8.785
	[0.741]	[6.759]	[0.778]	[6.710]	[0.778]	[6.710]
Observations	592	592	592	592	592	592
Adjusted R-squared	0.167	0.171	0.181	0.185	0.181	0.185

Table 2. Estimation Results: log-log

For a different specification of variables, without taking log, we attached in the Appendix. The results are consistent with the one shown above.

4. Concluding remarks

This paper put two notes on the representation of the spatial units with the use of geographical tools. In the first note, we started the discussion of the traditional literature of urban economics. The revival of the analysis of urban density functions with using satellite images would bring further understandings of the regularities within cities. Since the system of cities has the regularity, it can be mostly observed that the structure of within city and the shapes among cities may also have certain regularities. However, natural features such as rivers and mountains within city boundaries would certainly affect the expansion of cities and the distribution of population. Such natural environmental features are needed to be controlled for the estimation.

Second note states that the possibility of systematic measurement errors by the use of centrality of geographical units. The measurement error comes from the difference between the unweighted and weighted centrality of geographical units. If the centrality of population is the concerning, the gap between the true centrality and the one obtained has systematic errors. The larger units have larger errors. However, even at units with small area size, if the regions are composed from islands, it also generates huge errors.

References

Brivio , Pietro Alessandro and Eugenio Zilioli (2001) "Urban Pattern Characterization through Geographical Analasis of Satellite Images" in Jean-Paul Donnay, Mike J. Barnsley and paul A. Longley (ed.) *Remote sensing and Urban Analysis*, Taylor & Francois, London.

Batty, Michael and Kwang Sik Kim (1992) "Form Follows Function= Reformulating Urban Population Density Functions", *Urban Studies*, Vol. 29 7, pp1043-1070.

Batty, Michael and Paul Longley (1994) *Fractal Cities: A Geometry of Form and Function*, Academic Press, London.

Clark, Colin (1951) "Urban Population Densities", *Journal of the Royal Statistical Society (Series A)*, Vol. 114, pp490-496.

Fujita, Masahisa (1989) *Urban Economic Theory: Land Use and City Size*, Cambridge University Press, Cambridge.

McDonald, John F. (1989) "Econometric studies of urban population density: A survey", *Journal of Urban Economics*, Vol. 26, pp361-385

Appendix

	(1)	(2)	(3)	(4)	(5)	(6)
Area	1.06e-09***	1.43e-09***	1.10e-09***	1.57e-09***	9.99e-10***	1.31e-09***
	[2.42e-10]	[4.05e-10]	[2.46e-10]	[4.23e-10]	[2.45e-10]	[4.15e-10]
Area^2		-8.78E-21		-1.07E-20		-7.12E-21
		[7.64e-21]		[7.86e-21]		[7.73e-21]
Total Population			-7.54E-07	-0.00000107		
			[0.000000963]	[0.000000989]		
Population Density					-540.8	-489.6
					[351.7]	[356.1]
Constant	17.52***	16.01***	18.64***	17.26***	18.34***	17.04***
	[1.816]	[2.241]	[2.314]	[2.523]	[1.892]	[2.362]
Observations	592	592	592	592	592	592
Adjusted R-squared	0.03	0.03	0.029	0.031	0.032	0.032

Table A1. Estimation Results

調査研究報告書 (Chosa Kenkyu Report)
［Interdisciplinary Studies Center］2014－［C44］
［Economic Division in British India］

2015 年 3 月 31 日発行
発行所　独立行政法人日本貿易振興機構
アジア経済研究所
〒261-8545　千葉県千葉市美浜区若葉 3-2-2
電話　043-299-9500