EXPLORATION AND VISUALIZATION OF POVERTY-ENVIRONMENT RELATIONSHIP USING EXPLORATORY SPATIAL DATA ANALYSIS TECHNIQUES

Ferdinand J. Paraguas[1], Anton Abdulbasah Kamil[2], Kam Suan Pheng[3], Madan Mohan Dey[4], & Manik Lal Bose[5]

[1 2]School of Mathematical Sciences, Universiti Sains Malaysia, 11800, USM Pulau Pinang, Malaysia
[1 3 4 5] WorldFish Center, Jalan Batu Maung, Batu Maung 11960, Penang Malaysia
[1]f paraguas@cgiar org, [2]anton@cs usm my, [3]s kam@cgira org, [4]m dey@cgiar org, [5]m bose@cgiar org

Abstract

This paper illustrates an alternative visualization and identification methodology, namely exploratory spatial data analysis (ESDA) that allows detecting both spatial autocorrelation in the form of spatial clusters and spatial heterogeneity in the form of differentiated cluster patterns across space Specifically, the paper analyzes the spatial dimension of poverty and its linkages with environmental conditions using data from Bangladesh with the aid of ESDA
Results show significant spatial clustering of high and of low poverty rates as well as spatial heterogeneity of poverty-environment relationships – a clear indication of spatial regionalization of poverty-environment relationships The results also indicates that there is no straightforward association between poverty and any single environmental factors considered, which calls for a multivariate spatial econometric approach to simultaneously analyzes the effects of several environmental variables taking into account the effects of demographic variables as well as spatial effects In terms of policy implications, the existence of spatial regimes of poverty-environment relationships suggests a more targeted anti-poverty intervention

Keywords Spatial analysis, Exploratory spatial data analysis, spatial dependence/autocorrelation, spatial heterogeneity, spatial econometrics, Moran's *I*

1 Introduction

The identification, measurement and analysis of the inherent stochastic nature of spatial patterns (clustering) and relationships of geographical variables (e g , poverty incidence) have been a focal issue in spatial data analysis and spatial econometric modeling This paper underlines the relevance of an alternative visualization and exploration methodology, that is the exploratory spatial data analysis (ESDA), that allows detecting both spatial autocorrelation[a] in the form of spatial clusters, and spatial heterogeneity[b] in the form of differentiated cluster patterns across space Specifically, the paper illustrates the application of ESDA in the analysis of the spatial dimensions of poverty and its linkages with environmental conditions using Bangladesh data The reason for this approach stems from the drawbacks of the conventional methods such as visual inspection that "human perception is not sufficiently rigorous to assess the significant clusters and indeed tends to be biased toward finding patterns, even in spatially random data" [1] The paper also contributes to the limited number of empirical studies of poverty-environment linkages[c] It provides a holistic methodology of poverty-environment interaction analysis in a spatial context and contributes to the growing number of spatial analysis applications

The remainder of the paper is arranged as follows Section 2 provides the methodology used, the level of analysis and the sources of data Section 3 presents the results and section 4 draws important conclusions

2. Methodology

2.1. Level of Analysis, Sources of Data and Variable Used

The level of analysis is *Upazilla/Thana* level, which is a sub district administrative unit of Bangladesh The data and maps that were used are taken from the Bangladesh Country Almanac (BCA)[d] In particular, the poverty measure that is used is the headcount index (HCI), which is the proportion of households with cost of basic needs below the poverty line The set of environmental factors considered are land types (in relation to water type and inundation depth), drought severity and soil types All analyses are conducted using the GeoDa[T] software [4] ArcGIS is also used to enhance mapping of the results for GeoDa[T]

---

[a] Spatial autocorrelation is the tendency of the same variables measured in locations in close proximity to be related

[b] Spatial heterogeneity or spatial instability of geographic processes, meaning that global parameters do not well reflects processes occurring at a particular location

[c] The relationship of poverty and environment, which is regarded as the "poverty-environment nexus" is considerably debated over the last decade [see e g , [2]) The literature is abound with broad generalizations or assumptions (e g , poverty and the environment are inextricably linked in a "downward spiral") However, studies directly verifying the relationship are extremely scarce Part of this scarcity is due to the lack of established methodology

[d] The BCA is a compilation of digital data sets, both spatial and non-spatial attributes in a CD-ROM with Mud Springs Awhere-ACT Spatial Information System tools developed using MapObjects programming It also contains the data and the report of the poverty mapping exercise conducted by the International Rice Research Institute (IRRI) and partner Institute [3]

2.2. Global and Local Measures of Spatial Association

ESDA is a set of techniques aimed at describing spatial patterns distributions in terms of spatial association patterns such as global spatial autocorrelation, local spatial autocorrelation and spatial heterogeneity One of its roles is to shed light on future possibilities in the modeling and theorizing Hence, ESDA work precedes a good spatial econometric modeling

2 2 a Univariate analysis

Among the different measures of spatial association (e g , Moran's *I*, Geary's *C* and Getis *G*), the Moran's *I* is the most popular and is used in this paper It is calculated using the following equation [5]

$$I = (n/s_o)\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right) \Big/ \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right) \quad , \qquad (1)$$

where n is the number of *Upazillas* (locations), x is a (n x 1) vector of observations x_i (e g , HCI), w_{ij} are elements of a spatial weight matrix W with order (n x n) which defines the neighborhood structure of the locations, i, j, s_0 is a normalizing factor equal to the sum of the elements of the weights matrix, i e , $s_0 = \sum_i \sum_j w_{ij}$

It is common practice to interpret Moran's *I* as a correlation coefficient, although its value is strictly speaking not restricted to the [-1,+1] interval High positive values signal the occurrence of similar attribute values over space (either high or low values), and hence spatial clustering Negative values indicate the joint occurrence of high and low attribute values in nearby locations In the absence of spatial dependence, the value of Moran's *I* approaches to (zero), an evidence of a random allocation of attribute values over space

For a row standardized W (such that the elements of a row sum up to unity[e]), the computed Moran's *I* can be interpreted as a coefficient in a regression of spatial lag of x (Wx, i e , the average poverty incidence of neighboring *Upazillas*, $\sum_{i=1}^{k} w_{ij} x_i$, against the original values of x) More formally, $Wx = a + Ix + e$, where x and Wx are as defined earlier, the parameter a is a regression intercept and I is the slope in the equation and is equivalent to Moran's *I*

Global measures of spatial autocorrelation provide a tool for testing for spatial patterning over a whole study area Corresponding local indicators of spatial autocorrelation (LISA),

---

[e] Row standardization normalizes the outside influence upon each area and facilitates easy interpretation [6]

e g , Local Moran s I_i, Local Geary C_i and Local Getis G_i) can be used to identify hot spots to indicate pockets of spatial non stationarity, and to suggest spatial regimes or outliers

The local Moran statistic for each observation i is defined as [7]

$$I_i = Z_i \sum_{j \neq i}^{n} w_{ij} Z_j \tag{2}$$

where the observations Z_i and Z_j are in standardized form (with mean of zero and variance of one) The spatial weight w_{ij} are in row-standardized form So, I_i is a product of Z_i and the average of the observations in the surrounding locations

2 2 b Bivariate analysis

The analyses extend to bivariate approach by relating poverty (HCI) of a focal location (*Upazilla*) with different environmental variables (x) of neighboring locations (Wx) against the HCI of the focal *Upazilla* These analyses do not only allow us to determine spatial relationships but also spill over effects of environmental conditions on HCI

Bivariate Moran s I of poverty and environmental variables are computed to discern global poverty-environment relationship Bivariate LISAs are also computed and mapped to visualize different spatial-environment relationships, an indication of spatial heterogeneity of relationships

2 2 c Moran Scatter Plot

A third tool in ESDA used in this paper is the univariate and bivariate Moran scatter plot A univariate Moran scatter plot for example, plots W_HCI $\left(\sum_{i=1}^{k} w_{ij} HCPI_i\right)$ against HCI, aim at visualizing four types of local spatial association between an *Upazilla* and its neighbors, each of them being localized in a quadrant of the scatter plot (Fig 1) quadrant High-High (Upper right) refers to *Upazillas* with high poverty in the neighborhood of high poverty (i e, cluster of poverty), quadrant Low-High (upper left) refers to *Upazillas* with low poverty rate surrounded by *Upazillas* with high poverty rates (i e , hot spots), etc Quadrants High-High and Low-Low indicate positive spatial autocorrelation indicating spatial clustering of *Upazillas* with similar (High or Low) poverty rates Conversely, quadrants Low-High and High-Low indicate negative spatial autocorrelation indicating spatial clustering of *Upazillas* with dissimilar poverty rates

2.3. Test of Significance

A test on the null hypothesis that there is no spatial autocorrelation between observed values over the n locations can be conducted based on the standardized statistic In practice however, the test of the significance of Moran's I, and its local counterpart, LISA

(I_i) are conducted using random permutation or conditional randomization outlined in [8] and [9]

2.4. Spatial Weight Matrix (W)

The W measures the spatial linkages or proximity of the observation Formally given spatial framework of n locations, $S = \{s_i\}_{i=1}^{n}$, and a neighbor relation $N \subset SXS$, sites s_i and s_j are neighbors iff $(s_i, s_j) \in N, i \neq j$ Let $N(s_i) = \{s_j \; (s_i, s_j) \in N\}$ denotes s_i s neighborhood The elements of normalized W are $w(i,j) = 1/|N(s_i)|$ iff $(s_i, s_j) \in N$ and $w(i,j) = 0$ otherwise The form of the W can vary from a contiguity/adjacency relation to distance functions and eventually with a cut-off point (k) or k-nearest neighbor An illustration of different contiguity-base matrix, both first-order and higher-order specifications and also combinations of contiguity and distance specifications is provided in [10] The determination of the proper specification for the elements of a spatial weight matrix $\{w_{ij}\}$ is one of the difficult and controversial methodological issues in spatial analysis In this paper, we used the k-nearest neighbors weight matrix $W(k)$ of the form

$$\begin{cases} w_{ij}^*(k) = 0 & if\, i = j, \quad \forall k \\ w_{ij}^*(k) = 1 & if\, d_{ij} \leq d_i(k) \quad and \quad w_{ij}(k) = w_{ij}^*(k) \Big/ \sum_j w_{ij}^*(k) \\ w_{ij}^*(k) = 0 & if\, d_{ij} > d_i(k) \end{cases} \quad , \tag{3}$$

where $w_{ij}(k)$ is an element of the standardized weight matrix and $d_j(k)$ is a critical cut-off distance defined for each *Upazilla* i More precisely, $d_j(k)$ is the k^{th} order smallest distance between location i and all other units such that each unit i has exactly k neighbors More specifically, we set k = 6 in this paper[f]

3 Results and Discussion

The computed Morans' I for HCI is 0 55, which is statistically significant at 99% significance level This indicates that a positive spatial dependence of poverty exists in Bangladesh To delineate between spatial clustering of high and of low incidence of poverty, a Moran scatter plot on W_HCI (the spatial lag of HCI) against HCI is presented

[f] A delaunay triangulation routine is used to pick-out the three nearest *Upazillas* based on the latitude and longitude coordinates for the *Upazilla* centroid Delaunay triangulazation computes a set of triangles such that no data points are contained in any triangle's circumcircle The choice of k=6 is based on the average number of adjacent neighbors of each *Upazilla*

in Fig 1 Most of the *Upazillas* are clustered either in quadrant High-High (pockets of high poverty) or in quadrant Low-Low (pockets of low poverty)

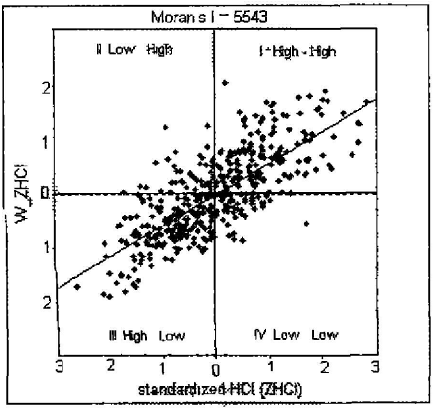

Fig. 1: *Moran Scatter Plot of HCI Based on 6 Nearest Neighbor Relation*

In order to assess the significance of the spatial clustering of *Upazillas* with similar HCI (e g , pockets of high and of low incidence of poverty), Local Moran's *I* are computed and mapped (Fig 2a) The corresponding significance map is shown in Fig 2b The colour code on the LISA map indicates the quadrant in the Moran scatter plot Among other things, the map confirmed the significance of the pockets of high incidence of poverty (pockets of High-High values of HCPI) in the northeast, which is characterized by low-lying flood-prone area Statistically significant clusters of Low-Low relationships are also found in central Bangladesh where the capital city of the Country (Dhaka City) is situated On the other hand statistically significant hot spots areas are found in coastal areas (south)

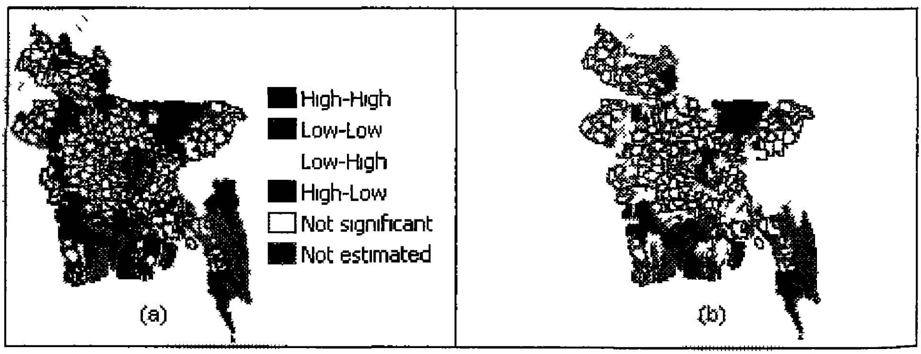

Fig. 2: *LISA Cluster Map of HCI (a) and Corresponding Significance Map (b)*

Table 1 summarizes the results of bivariate Moran scatter plot of environmental factors (x) and HCI The figures are the slopes of Moran scatter plot (i e regression of Wx against HCI) The variable "% very lowland area" exhibits the greatest magnitude (0 7787)

Upazillas in the neighborhood of high "% medium highland" area are seemed to have low incidence of poverty While the extremes (i e , highland and very lowland) are associated with high incidence of poverty Contrary to expectation, a significant negative spatial association is observed between poverty and severity of drought suggesting IRRI et al [3] argued that the relatively low concentration of poverty in the drought-prone areas particularly in the northwest region seems to have masked by the presence of infrastructure facilities such as irrigation In a review study of Heisey and Edmeades [11] in several countries, they reported a mix results on the drought stressed-poverty relationship

Table 1: *Bivariate Moran's I between Upazillas HCI and selected environmental factors of neighboring areas based on 6-nearest neighbor relation*

Biophysical characteristics	Global spatial association (Moran's *I*)	
% of High land	0 0832	***
% Medium highland 1	-0 1538	***
% Medium highland 2	-0 1872	***
% Medium lowland	0 0292	ns
% Lowland	0 1155	***
%Very lowland	0 7787	***
% area with sandy loam soil	0 1456	**
% area with clay soil	-0 1681	***
% area affected by severe drought	-0 2317	***
% area affected by moderate drought	-0 2265	***

*** significant at $\alpha = 0\ 01$ ** significant at $\alpha = 0\ 05$ ns not statistically significant

Figures 3a-c depict the significant bivariate LISA between *Upazilla's* HCI and the spatial lag of selected environmental variables Note that the results within environment group (land type, soil type and drought), supports each other, i e , the same analysis can be extracted from the result between HCI and "% lowland" and/or "% very lowland" Among others, a statistically significant concentration of poverty exists in the neighborhood of high "% of highland" in the northwest and southeast regions (a High-High relationship) and in the neighborhood of low "% of Highland" (i e , low-lying flood-prone areas) in the northeast region (a Low-High relationship) As shown in Fig 3b, the areas with low "% of highland" are also characterized by high "% of clay area" These areas are also characterized by low "% of area affected by severe drought" (Fig 3c)

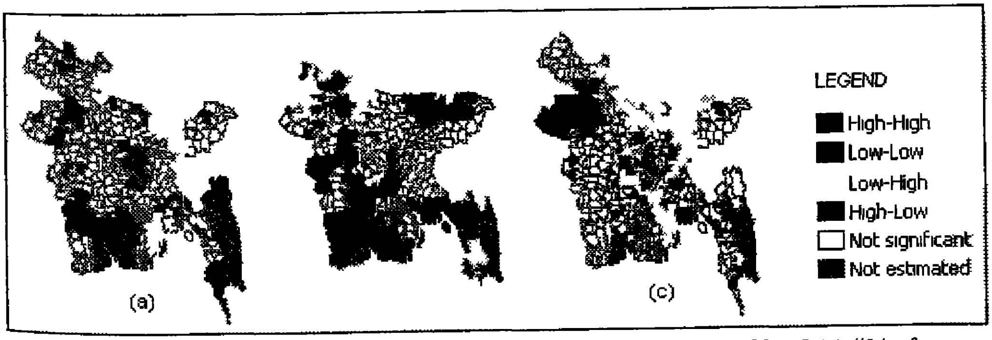

Fig. 3: *Bivariate LISA Cluster Map of Selected Environmental Variables [(a) "% of Highland', (b) "% Clay Area" and (c) "% Area Affected by Drought"] Against HCI*

4 Summary and Conclusion

In this paper, we illustrate an alternative visualization and identification methodology, namely exploratory spatial data analysis (ESDA) that allows detecting both spatial autocorrelation, in the form of spatial clusters and spatial heterogeneity in the form of differentiated cluster patterns of poverty and its linkages with environmental conditions using data from Bangladesh The use of such analysis highlights results that would escape notice in a standard visual nature of poverty maps

Results show significant spatial clustering of high and of low poverty rates and spatial heterogeneity or instability of poverty-environment relationships More meaningfully, this indicates that a spatial regionalization of poverty-environment relationships exists Poverty exists both in low-lying flood-prone areas as well as in high altitude areas However, significant clusters of low poverty are also observed in high altitude areas Secondly, certain level of drought severity and soil type also characterizes high- and low-altitude areas These results suggest that there is no straightforward and clear-cut association between poverty and any single environmental factors considered Therefore, a multivariate approach of analysis which combines all these environmental variables taking into account the effects of non-environmental factors (e g , demographic variables) is deemed necessary Moreover, the presence of spatial patterns of poverty incidence and poverty-environment relationships as evident in the analysis suggests a multivariate spatial econometric modeling to account for spatial effects due to these spatial patterns [see 6 for details]

In terms of policy implications, the spatial of spatial regime or regionalization of poverty and poverty-environment interactions suggest a more targeted anti-poverty intervention

References

1 S F , Messner, L Anselin, R D Baller, D F Hawkins, G Deane and S F Tolnay, "The Spatial Patterning of Country Homicide Rates an Application of Exploratory Spatial Data Analysis", *Journal of Quantitative Criminology* **15** (1999), 423-450
2 K S Parikh, "Poverty and Environment Turning the Poor Into Agents of Environmental Regeneration" (UNDP Working Paper Series, 1998)
3 IRRI, BARC, LGED, BBS, "Geographical Concentration of Rural Poverty in Bangladesh", [Final Report submitted to The Consortium of Spatial Information (CSI) and the Food and Agricultural Organization (FAO), 2004]
4 L Anselin, I Syabri and Y Kho, "GeoDa an Introduction to Spatial Data Analysis", (Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics University of Illinois, Urbana-Champaign, Urbana, IL, USA, 2004)
5 A D Cliff and J K Ord, *Spatial Autocorrelation* (Pion, London, 1973)
6 http //www spatial-econometrics com Accessed on August 5, 2004
7 L Anselin, "Local Indicators of Spatial Association-LISA", *Geographical Analysis* **27** (1995), 93-115
8 Bailey, T C and Gatrell, A C , *Interactive Spatial Data Analysis* (Longman, Harlow, 2004)

9 L Anselin, "The Moran Scatter plot as an ESDA Tool to Assess Local Instability in Spatial Association", in *Spatial Analytical Perspectiveson GIS in Environmental and Socio-Economic Sciences*, (eds) M Fischer, H Scholten & D Unwin, (Taylor and Francis, London, 1996) 111–125

10 H H Kelejian and D P Robinson, "Spatial Correlation a Suggested Alternative to the Autoregressive Model", in *New Directions in Spatial Econometrics*, (eds) L Anselin & R Florax, (Springer-Verlag, NY, 1995)

11 P W Heisey and G O Edmeades, "Maize Production in Drought-Stressed Environments Technical Options and Research Resource allocation", (World Maize Facts and Trends, Centro International de Mejoramiento de Maiz y Trigo (CIMMYT), 1999)