

# Relationship of Population and Land Cover Pattern Based on Remote Sensing Data Using Echelon Analysis

Koji Kurihara<sup>1</sup>, Wayne L. Myers<sup>2</sup>, and Ganapati P. Patil<sup>3</sup>

<sup>1</sup>Faculty of Environmental Science and Technology

Okayama University

2-1-1 Tsushima-naka

Okayama, 700-8530, Japan

<sup>2</sup>Environmental Resources Research Institute

The Pennsylvania State University

University Park, PA 16802, USA

<sup>3</sup> Center for Statistical Ecology and Environmental Statistics

Department of Statistics

The Pennsylvania State University

University Park, PA 16802, USA

## Abstract

With continuing proliferation of human influences on landscapes, there is mounting incentive to undertake quantification of relationships between spatial patterns of human populations and vegetation. In considering such quantification, it is apparent that investigations must be conducted at different scales and in a comparative manner across regions. At the broader scales it becomes necessary to utilize remote sensing of vegetation for comparative studies against map referenced census data. This paper explores such an approach for the urbanized area in the Tokyo vicinity. Vegetation is represented by the normalized difference vegetation index (NDVI) as determined from data acquired by the *thematic mapper* (TM) sensor of the Landsat satellite. Sparseness of vegetation is analyzed in relation to density of human residence, first by regression analysis involving stratified distance zones and then by the recent *echelon* approach for characterization of surfaces. Echelons reveal structural organization of surfaces in an objective and explicit manner. The virtual

---

Prepared with partial support from the National Science Foundation Cooperative Agreement Number DEB-9524722 and the United States Environmental Protection Agency Cooperative Agreement Number CR-825506. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agencies and no official endorsement should be inferred.

surface determined by census data collected on a grid is shown to have structural correspondence with the surface representing vegetation greenness as reflected in magnitude of NDVI values computed from red and infrared bands of image data.

**Keywords:** Vegetation patterns; Human ecological influence; Remote sensing; NDVI; Echelon analysis; Spatial comparison.

## 1 Introduction

Cost effective ecological synthesis and environmental analysis issues and approaches are now in great demand (Patil, 1998). Maps of distribution patterns for plants and animals are important for land use planning (Spellerberg 1991, Miller 1994). It is useful for land use planning to evaluate the land cover patterns and monitor major changes based on remote sensing information and GIS (geographic information systems). There are several investigations of the relationship of animal richness and land cover patterns by remote sensing data (Cardillo 1999, Johnson et al. 1999). Remote sensing data are also effective for evaluating vegetative greenness (green tracts) of land cover by use of red and infrared bands. In general, there is a negative correlation between greenness and density of humans. It is also useful to investigate the structure of the urban areas based on the relationship of green tracts and human population density. Visual examination of perspective graphics is useful for investigating such mapped data. However, visual perceptions are very difficult to describe objectively. The echelon approach (Myers, Patil, and Joly, 1997; Myers, Patil, and Taillie, 1999) is designed to objectively present the areas of relative high and low values of response variables of mapped data.

The purpose of this paper is to investigate spatial structure for population and green tracts data in Tokyo metropolitan area based on the echelon approach. Population and green tracts data based on Landsat Thematic Mapper (LTM) are described in section 2. Resampling and interpolation for geometric correction were performed on LTM data for positional registration to population data. In section 3, preliminary analysis for population and LTM data are given for three sectors in the Tokyo metropolitan area. Multiple linear regression equations of population on some explanatory variables by sector and overall area are considered in section 4. Additionally, the stepwise linear regression method is applied for explanatory variables and their interactions based on AIC criterion. In section 5, echelon tables, maps and tree dendrograms of the population and green tract data are presented and the structures of these dendrograms are compared. Moreover, the echelon map for the difference of cumulative percentage between the population and green tract data is calculated to compare the structure of the differences for these data in Tokyo metropolitan area. Summary and discussion are presented in section 6.

## 2 LTM and population data of Tokyo metropolitan area

### (1) LTM data

A landscape map of Tokyo metropolitan area was constructed by use of 30m × 30m pixels (120m × 120m for band 6) Landsat Thematic Mapper (LTM) data from NASA's Home page. LTM data has seven bands with band 1 (0.45-0.52 μm, blue) wavelengths, band 2 (0.52-0.60 μm, green) wavelengths, band 3 (0.63-0.69 μm, red) wavelengths, band 4 (0.76-0.90 μm, infrared (IR)) wavelengths and bands 5,6,7 (1.55-1.75, 10.40-12.50, 2.08-2.35 μm, IRs) wavelengths, respectively. The center position of this LCM subscene is 35 ° 40' north latitude and 139 ° 40' east longitude, with 512 rows and 512 columns of pixels covering the western part of Tokyo City area on February 11 in 1989.

NDVI (Normalized Difference Vegetation Index) is formulated for analyzing green tracts (forest, herbaceous and so on) of land cover as follows:

$$\text{NDVI} = (\text{TM4} - \text{TM3}) / (\text{TM4} + \text{TM3}) \quad (1)$$

where TM3 is the value recorded in band3 and TM4 is the value recorded in band 4. A landscape map of NDVI in Tokyo metropolitan area is shown in Figure 1.

### (2) Population data

Population mapping of Tokyo metropolitan area is based on a national census for population of Japan taken on October 1, 1989. These data represent number of persons in rectangular cells with latitude length 30 " (about 962m in Tokyo area) and longitude length 45 " (about 1100m in Tokyo area). In general, there is a negative trend between green tracts and population density. Thus, population data (*pop*) are transformed to (30000 - *pop*) in order to have a positive correlation to green tract data based on NDVI. The areas of all cells in the population data are equal, so transformed data are considered as sparseness of population in the respective cells.

### (3) Geometric correction for LTM data

Resampling and interpolation for geometric correction of LTM data were conducted to fit the positions of population rectangular cells. Each rectangular cell for geometrically corrected LTM data has (36 pixels in a column) × (32 pixels in a row) corresponding to the rectangular cell of population data. Then there are 14 × 14 = 196 cells in the overall rectangular area, but the numbers of LTM pixels for the last three columns in the first row and the last five rows in the 14th column are zero or scant. Therefore, those cells were eliminated in data analysis, and the effective number of cells is 196 - 8 = 188. The arithmetic mean of NDVI was calculated as a representative value for each rectangular cell.

### 3 Preliminary analysis for NDVI and population data

#### (1) NDVI data

The 188 effective NDVI values are displayed on a grid in Figure 2. These values are calculated as the average of  $36 \times 32$  pixels in each rectangular cell with geometric correction for Tokyo metropolitan area. The data are rounded to two decimal places and the unit is ( $\times 10^{-2}$ ). A stem-and-leaf plot of NDVI data in Figure 3 shows that data are approximately normal with 0.0902 as an outlier.

Tokyo area is divided into two regions with the Yamanote region representing the center of Tokyo City, and the remainder as suburbs. To check for local variation in terms of distances from Yamanote region, box-whisker plots of NDVI data according to distances are shown in Figures 4, with the unit being km. There are many variations within the Yamanote region where distances are 0. On the other hand, there is a trend across the distances by which the value of NDVI increases with distance from Yamanote region.

#### (2) Population data

The 188 values of population data corresponding to the same positions of NDVI data are displayed on a grid in Figure 5. The data are rounded to one decimal place and unit is ( $\times 10^3$ ). The stem-and-leaf plot of population data in Figure 6 shows that data are not normal, but have mixed distribution. Box-whisker plots of population data according to distances are shown in Figure 7. There is evident variability, particularly with regard to the high value of population sparseness in the Yamanote region whose distance is shown as 0. There is positive correlation between distance from the Yamanote region and sparseness of population.

#### (3) The relationship of greenness (NDVI) and density of population

The scatter plot for the whole data set is shown at the top left side of Figure 8. The simple regression line of population on NDVI is

$$pop = 19080.6 + 91686.7ndvi, \quad (2)$$

where *pop* is (30000– the number of persons) and *ndvi* is the value of normalized difference vegetation index in the cell. The coefficient of determination is low ( $R^2 = 0.205$ ). The stratified scatter plots of NDVI and population data based on the distances from the Yamanote region are shown in the other parts of Figure 8. Each numerical label denoted by 0 to 10 shows the distance to the Yamanote region from the cell. It seems that there is positive correlation for the area in which distance from the Yamanote region is from 1 to 6 km.

The Tokyo area is divided into three regions which are shown in Figure 9. These are: the Yamanote region in the center of Tokyo City (east region in this map denoted by 0 with 60 cells); the region within about 6.6 km from the Yamanote region which is a high density residential area (middle region denoted by 1 with

91 cells); the region beyond 6.6 km from the Yamanote region which has moderate residential density in the Tokyo area (west region denoted by 2 with 37 cells).

There are different characteristic properties among these three regions as mentioned, so the data for these regions are analyzed separately. Identification numbers for city cells in the Tokyo area are shown in Figure 10.

#### (4) Yamanote region

The scatter plot of NDVI and population data with ID numbers for the Yamanote region is shown at the top left side of Figure 11. The simple regression line of population on NDVI is  $pop = 20800.3 + 55648.6ndvi$  with a low coefficient of determination ( $R^2 = 0.120$ ).

The Yamanote region is the center of Tokyo City in which many government and corporation offices and shopping centers are located so that few people live in this region. Moreover, there are some large parks in this region including Japanese imperial house. There are three sectors in the Yamanote region.

The first sector is the area of government and corporation offices and shopping centers. Thus the rate of green tracts per person in this area is lower than for other areas. Each area has some special places in which few people live. These are large shopping areas (Shinjuku 49,63, Shibuya 105,119), the government office and the Diet building (Nagata-cho 80, Kasumigaseki 95), the corporation and factory (Shiba(NEC) 123, Ohsaki(Sony) 174), the offices (Shinbashi 109), the office and the embassy (Jingu 92), the train depots (Shinagawa 162), the offices and schools (Iidabashi 53, Yotsuya 65, Ichigaya 66), and official residence, hotels and restaurants (Akasaka 94, Roppongi 108). The scatter plot for the first office and shopping area is shown at the top right side of Figure 11. The simple regression line of population on NDVI is given as  $pop = 35191.6 + 154927.5ndvi$  with coefficient of determination  $R^2 = 0.411$ . The names and ID numbers of the areas are shown in the parenthesis.

The second sector includes a large park and imperial house. Thus, the rate of green tracts per person is higher than for other areas. The area consists of imperial house (the Imperial Palace 67,81, the Shinjuku Imperial garden 64,78 the Togu Imperial garden 79,93), parks (Yoyogi Park 77,91 Natural Education Park 147,148 Toyama 36, Korakuen 39) and the graveyard (11 and Aoyama 107). The scatter plot of the second park area is shown at the bottom left side of Figure 11. The simple regression line of population on NDVI is given as  $pop = 20636.9 + 101160.7ndvi$ . This regression line is shown as dashed line with coefficient of determination  $R^2 = 0.495$ . The extent of park for the area is more than half in the cell of 91 ID number, so the value of NDVI is high. The simple regression line of population on NDVI in the second area, eliminating cell 91, is given as  $pop = 21679.2 + 147907.8ndvi$ . This regression line is shown as the solid line in the bottom left side of Figure 11. The coefficient of determination is  $R^2 = 0.614$ .

The third sector is the remaining area of the Yamanote region. The rate of green tracts per person is between the first and second areas in the Yamanote region. The scatter plot of the third area is shown in the bottom right side of Figure 11. The simple regression line of population on NDVI is given as  $pop = 19549.7 + 87787.2ndvi$ . This regression line is shown as the dashed line with coefficient of

determination  $R^2 = 0.184$ . There may be an anomaly the value of NDVI in cell of 10 ID number. The simple regression line of population on NDVI in the third area eliminating the 10th cell is given as  $pop = 22958.6 + 135587.7ndvi$ . This solid regression line is shown in the bottom right side of Figure 11 with coefficient of determination  $R^2 = 0.341$ .

### (5) The middle region

The scatter plot of NDVI and population with ID numbers in the middle region is shown in Figure 12. The simple regression line of population on NDVI in the middle region is given as  $pop=21122.7+190823.5ndvi$ . This regression line is shown as the dashed line with coefficient of determination  $R^2 = 0.603$ .

The middle region within 6.6km from the Yamanote region is mainly used for high density residence in Tokyo metropolitan area. The number of persons decreases as green tracts of town land increase. There are large parks in the areas of Yoyogi Park (90), Wadabori Park (58) and Rinshi-no-Mori (172). Thus the values of NDVI in these cells are higher than ones in other cells. The simple regression line of population on NDVI in the middle region eliminating cells 90, 58 and 172 which include the large park is given as  $pop = 22949.6 + 221390.3ndvi$ . This regression line is shown as the solid line in Figure 12 with coefficient of determination  $R^2 = 0.689$ .

### (6) The west region

The scatter plot of NDVI and population with ID numbers in the west region is shown in Figure 13. The simple regression line of population on NDVI in the west region is given as  $pop=19687.4+56932.0ndvi$ . This regression line is shown as the dashed line with coefficient of determination  $R^2 = 0.103$ .

The west region more than 6.6km from the Yamanote region is used for moderate density residence. There are large rivers in two cells (177, 178) for this region. The values of NDVI are low in such cells with water, in general. There are large parks in the areas of 97, 110, 150, 151, 152 and 163 ID numbers. Thus, the values of NDVI in these cells are higher than ones in other cells. Moreover, there may be an anomaly in the cell of 41 ID number. The simple regression line of population on NDVI in the west region eliminating the cells (41, 177, 178) for outliers and large parks (97, 110, 150, 151, 152, 163) is given as  $pop = 21217.3 + 118535.3ndvi$ . This regression line is the solid line in Figure 13. The coefficient of determination of the west region eliminating these nine cells is  $R^2 = 0.391$ .

## 4 Regression analysis for NDVI and population data

In the previous section, the simple regression lines of population on NDVI are given for three regions. The summaries of these simple regression lines are shown in Table 1. In this section, multiple linear regression equations of population on some explanatory variables for three divided regions are considered. The explanatory

variables are *ndvi*, *park* and *office*. The *ndvi* is the value of the normalized difference vegetation index. *Park* and *office* are the following nominal variables. The variable *office* is defined for the office area in the Yamanote region.

$$park = \begin{cases} 1 & \text{There is a large park in the cell.} \\ 0 & \text{Others.} \end{cases} \quad (3)$$

and

$$office = \begin{cases} 1 & \text{There are large office areas in the cell.} \\ 0 & \text{Others.} \end{cases} \quad (4)$$

### (1) Yamanote region

There are three explanatory variables, *ndvi*, *park* and *office*, for multiple linear regression of *pop* in the Yamanote region. Stepwise linear regression method was applied for these three explanatory variables and their interactions based on AIC criterion by eliminating the cells of outliers No. 10 and 91. The explanatory variables of *park* and interaction variables were deleted in the stepwise process. The final regression of population in the Yamanote region is given as the following formula with the coefficient of determination  $R^2=0.685$ .

$$pop = 22094.7 + 127135.8ndvi + 10909.5office \quad (5)$$

### (2) The middle region

Stepwise linear regression method was applied for explanatory variables of *ndvi*, *park* and  $ndvi \times park$  based on AIC criterion in middle region. The interaction variable  $ndvi \times park$  was deleted in the stepwise process. The final regression of population in the middle region is given as the following formula with coefficient of determination  $R^2 = 0.704$ .

$$pop = 23030.2 + 333945.5ndvi - 8628.5park \quad (6)$$

### (3) The west region

Stepwise linear regression method was applied for *ndvi*, *park* and  $ndvi \times park$  explanatory variables based on AIC criterion in west region, eliminating cells for outliers No. 41, 177 and 178. The interaction variable  $ndvi \times park$  is deleted in the stepwise process. The final regression of population in west region is given as the following formula with coefficient of determination  $R^2=0.380$ .

$$pop = 21296.9 + 121506.7ndvi - 3439.5park \quad (7)$$

#### (4) Entire area

There are three explanatory variables, *ndvi*, *park* and *office*, for multiple linear regression of *pop*. There is a different coefficient of explanatory variable *ndvi* in the middle region compared with the other two regions. Therefore, the following explanatory variable *middle* is added for predicting response variable *pop*.

$$middle = \begin{cases} 1 & \text{The cell belongs to the middle region.} \\ 0 & \text{Others.} \end{cases} \quad (8)$$

Stepwise linear regression method was applied with four explanatory variables (*ndvi*, *park*, *office*, *middle*) and their interaction variables based on AIC criterion. The cells eliminated for outliers are No. 10, 41, 91, 177 and 178. The final regression line of population for the entire area is given as the following formula with coefficient of determination  $R^2 = 0.747$

$$pop = 22298.7 + 129509.0ndvi - 2781.5park + 10892.3office - 561.1middle + 71673.6middle \times ndvi \quad (9)$$

Summaries of these multiple regression lines are shown in Table 2.

The *t* values and probabilities are also shown in this table for the coefficients of multiple regression models.

## 5 Echelon analysis and application to NDVI and population data

### (1) Echelon analysis

Echelons provide an objective approach to prospecting for areas of interest in regional monitoring of a surface variable (Myers, Patil and Joly, 1997). Echelons are derived from changes in topological connectivity with decreasing surface level. Echelons are essentially topological subsets of a surface. For the case of one horizontal dimension, Figure 14 shows a hypothetical set of hillforms and the echelon numbers. In this paper, there are two techniques based on the echelon approach. The first technique is the echelon topview map which shows the echelon cover sectors on a two dimensional base plane. The second technique is the echelon dendrogram which is a structural family tree of echelons. The echelons are connected by the relations of parent and child. The echelon table summarizes the characteristics of echelons in terms of surface structure. There is an order numbering such that the members of peaks in the map are first order echelons. Foundations for two or more first order echelons are second order echelons. Foundations for two or more second order echelons are third order echelons, and so on.

### (2) Echelon tables and maps of NDVI and population data

The information derived from the echelons of NDVI and population data are shown in Tables 3 and 4, respectively. The values (*x*) of NDVI data are transformed to

$1000x + 1000$  in these tables. Tabular information is given for each echelon number. These items are defined as follows: Order is the order of the echelon. Parent is the number of the foundation echelon. Max is the maximum surface level for the echelon. Min is the minimum surface level for the echelon. Relief is the projection above the foundation echelon in surface level units. Cells is the number of cells in the echelon. Progeny is the number of ascendants for the echelon. Family is the number of ancestral echelons in the family lineage of the echelon. Structr is structural classification for the echelon. There are five classes in this classification. Class number 1 is a tip, terminal or leaf of the tree. Class number 2 is a twig, being a second order echelon with only first order progeny. Class number 3 is a stem, being induced by only first order progeny but having order greater than two. Class number 4 is a branch being induced by progeny that have order greater than 1. Class number 5 is a crotch or fork where the order increases.

Figures 15 and 16 show the echelon topview maps stratified by the echelon numbers for each cell based on NDVI and population data, respectively.

In Figure 15 of NDVI data, cells having small numbers have high values of NDVI, with more green area of land cover. Conversely, cells having large numbers have low values of NDVI, with little green area of land cover. The high value area of NDVI consists of the cells that include a large park and the edge parts of the west region. The low NDVI area consists of government and corporation offices and shopping centers in the Yamanote region and the main parts of the middle region.

In Figure 16 of population data, cells having small numbers signify sparse area with low density of population. The sparse area consists of offices, factories, imperial areas in the Yamanote region, and the edge parts of west regions. Cells having large numbers signify areas with high density of population. The dense area encompasses the east parts of the middle region.

The top sides of Figure 17 show the distributions of echelon order for NDVI and population data. The bottom sides of Figure 17 show the distributions of number of cells in the echelons for NDVI and population. Figure 18 shows the distributions of numbers of ancestral echelons in the family for the top sides and the distributions of the structural classification of echelons in the bottom sides for NDVI and population data.

The distributive properties of NDVI and population data are quite similar based on the echelon order, numbers of cells in the echelon, numbers of ancestral echelon in the family, and the structural classification.

### **(3) Echelon dendrograms of NDVI and population data**

Echelon dendrograms represent the surface topology for the structure of the map data. The echelon dendrogram has peaks as terminals, foundation of peaks as distal nodes, foundations of foundations as interior nodes, and the lowest level as the root.

Figures 19 and 20 show the dendrograms for echelon based on NDVI and population data, respectively. Letters are used to represent the orders of echelons, that is A means the first order and B means the second order, and so on. The number of letters represents the distances from their foundations. The structures of Figure 19 and 20 are quite similar. The lengths of some first order (peak) echelons (denoted

A) for NDVI in Figure 19 are quite long. There is a large park and many green tracts in the cells of such prominent first order echelons. The first peak area consists of the cells whose echelon numbers are (1, 2, 4, 17, 25). The positions of these cells are shown in the top left side of Figure 21. This area belongs to the middle part of the Yamanote area. The second peak area consists of the cells for echelon numbers (3, 6, 11, 18, 20). These cells are shown in the top right side of Figure 21. This area belongs to the south west part of the map.

There are also some prominent peaks for population data in Figure 20. For the first peak area, there are the cells whose echelon number are (1, 3, 4, 6, 18, 19, 20). These cells are shown in the bottom left side of Figure 21. This corresponds to the main part of the Yamanote area. There are four first order echelons in this area, but the reason for the sparseness of population in each cell is different. These are due to a large park, large shopping area and government office. The second peak area consists of the cells for echelon numbers (2, 7, 8, 21, 23). These cells are shown in the bottom right side of Figure 21. This area belongs to the south west part which is far from the center of Tokyo City.

#### **(4) Comparison between the echelons of NDVI and population data**

In order to compare the differences of structure between NDVI and population data in each cell, the echelon numbers of NDVI and population data are translated to the values of cumulative percentage for these data, respectively. Figure 22 shows the absolute differences of cumulative percentage between the echelon numbers of NDVI and population data.

In general, there is a tendency that the sparseness of population increases as the value of NDVI increases. There are two patterns for having large values of absolute differences of cumulative percentage. One is due to sparsely populated areas which has few green tracts. The other is due to densely populated areas which have many green tracts. Such patterns include large shopping areas (Shinjuku 49(37), 63(72), Shibuya 105(59), 119(65)), the government office (Kasumigaseki 95(78)), the corporation and factory (Shiba(NEC) 123(75), Ohsaki(Sony) 174(50)), the office and the embassy (Jingu 92(47)), the train depots (Shinagawa 162(56)), the offices and schools (Iidabashi 53(61)), the official residence, hotels and restaurants (Akasaka 94(70)), and the Tama river side (Tamagawa 177(41), 178(57)).

Structures of NDVI and population data over the whole area can be compared in terms of four echelon measures. These are based on a pruning process for the tree of pattern information. The echelon tree is recursively partitioned into an inner set of limbs and outer set of boughs. The limbs and boughs are obtained by cyclic calculations as follows:

1. The first cycle
  - (a) Find the maximum path (number of nodes) from first order echelons to the root. Find the common nodes for all paths having this maximum length. These nodes become the first limb of echelon tree. The path is determined by analysis of the echelon data in Table 3.

- (b) All other components of the tree are pruned away to form a set of subtrees that are called boughs.
- 2. Recursive cycles
  - (a) The process of pruning is conducted separately for each of the boughs. Each bough yields a limb and a set of boughs.
- 3. The cycle is ended if there are no remaining boughs.

The four measures, Divergence (Di) , Scope (Sc), Bunching (Bu) and Stacking (St), are defined as follows.

The divergence is defined by the percent of limb nodes at the cycle. This measure provides a structural characterization of surface complexity. The scope is defined by the percent of the cells in limbs at the cycle. The value of this measure indicates whether surface complexity is concentrated in certain sectors. The bunching is defined by the number of boughs as a percent of first order echelons at the cycle. This measure is particularly sensitive to the depth of structure and its consistency among the branches. The stacking is defined by the percent of first order echelons in limbs at the cycle. This measure is indicative of whether echelons sharing a common foundation (parent) tend to have the same or different orders.

### (i) Calculations of four measures for NDVI data

The numbers of total nodes, total cells and first order echelons are calculated as the values of 31, 188 and 16 respectively based on Table 3.

1. The first cycle

The number of nodes in the maximum path is 8. This path is connected from first order echelon 3 or 6. Then, the number of limb nodes is 8. This consists of 8(3), 20(6), 21(17), 26(10), 27(10), 28(4), 29(7) and 31(72), where the leading value is the echelon number and the value in the parenthesis is the number of cells in this echelon. The number of boughs is 9. These are 3, 6, 11, (19, 7, 8), (23, 9, 13), (25, 17, 1, 4, 2), (24, 5, 14), 15 and (30, 22, 10, 12, 16). The number of first order nodes in limbs is 0. Thus the four measures are given by

$$Di = \frac{8}{31} = 0.258, Sc = \frac{129}{188} = 0.686, Bu = \frac{9}{16} = 0.563, St = \frac{0}{16} = 0.000. \quad (10)$$

2. The second cycle

The number of limb nodes is 8+11. The new nodes are 3(8), 6(4), 11(1), 19(5), 23(3), 17(2), 25(4), 24(3), 15(1), 30(2) and 22(3). The number of boughs is 12. These are 7, 8, 9, 13, 1, 4, 2, 5, 14, 10, 12 and 16. The number of first order nodes in limbs is 4. These are 3, 6, 11 and 15. Thus the four measures are given by

$$Di = \frac{19}{31} = 0.613, Sc = \frac{165}{188} = 0.878, Bu = \frac{12}{16} = 0.750, St = \frac{4}{16} = 0.250. \quad (11)$$

3. The third cycle

The number of limb nodes is  $8+11+12$ . The new nodes are  $7(3)$ ,  $8(2)$ ,  $9(2)$ ,  $13(1)$ ,  $1(4)$ ,  $4(2)$ ,  $2(2)$ ,  $5(1)$ ,  $14(1)$ ,  $10(1)$ ,  $12(1)$  and  $16(2)$ . The number of boughs is 0. The number of first order nodes in limbs is  $4+12$ . The new nodes are 7, 8, 9, 13, 1, 4, 2, 5, 14, 10, 12 and 16. Thus the four measures are given by

$$Di = \frac{31}{31} = 1.000, Sc = \frac{188}{188} = 1.000, Bu = \frac{0}{16} = 0.000, St = \frac{16}{16} = 1.000. \quad (12)$$

**(ii) Calculations of four measures for population data**

The numbers of total nodes, cells and first order echelons are 33, 188 and 17 respectively based on Table 4.

1. The first cycle

The number of nodes in the maximum path is 9. This path is from first order echelon sets (1 and 3) or (4 and 6). Then, the number of the limb nodes is 8. These are  $20(19)$ ,  $26(3)$ ,  $27(4)$ ,  $29(3)$ ,  $30(2)$ ,  $31(27)$ ,  $32(16)$  and  $33(27)$ . The number of bough is 9. They are (1, 3, 8), (19, 4, 6), (25, 5, 13), 12, 14, 15, (28, 24, 23, 21, 2, 8, 7, 22, 9, 11, 10), 16, 17). The number of first order nodes in limbs is 0. Thus the four measures are given by

$$Di = \frac{8}{33} = 0.242, Sc = \frac{101}{188} = 0.537, Bu = \frac{9}{17} = 0.529, St = \frac{0}{17} = 0.000. \quad (13)$$

2. The second cycle

The number of limb nodes is  $8+12$ . The new nodes are  $18(2)$ ,  $19(1)$ ,  $25(1)$ ,  $12(2)$ ,  $14(1)$ ,  $15(2)$ ,  $16(1)$ ,  $17(1)$ ,  $21(11)$ ,  $23(3)$ ,  $24(11)$  and  $28(12)$ . The number of bough is 11. These are 1, 3, 4, 6, 5, 13, 2, 8, 7, (22, 9, 11) and 10. The number of first order nodes in limbs is 5. They are 12, 14, 15, 16 and 17. Thus the four measures are given by

$$Di = \frac{20}{33} = 0.606, Sc = \frac{149}{188} = 0.793, Bu = \frac{11}{17} = 0.647, St = \frac{5}{17} = 0.294. \quad (14)$$

3. The third cycle

The number of limb node is  $8+12+11$ . The new nodes are  $1(5)$ ,  $3(1)$ ,  $4(1)$ ,  $6(1)$ ,  $5(4)$ ,  $13(2)$ ,  $2(7)$ ,  $8(1)$ ,  $7(2)$ ,  $22(3)$  and  $10(7)$ . The number of boughs is 2. These are 9 and 11. The number of first order nodes in limbs is  $5+10$ . The new nodes are 1, 3, 4, 6, 5, 13, 2, 8, 7 and 10. Thus the four measures are given by

$$Di = \frac{31}{33} = 0.939, Sc = \frac{183}{188} = 0.973, Bu = \frac{2}{17} = 0.118, St = \frac{15}{17} = 0.882. \quad (15)$$

#### 4. The fourth cycle

The number of limb nodes is  $8+12+11+2$ . The new nodes are  $9(2)$  and  $11(3)$ . The number of boughs is 0. The number of first order nodes in limbs is  $5+10+2$ . The new nodes are 9 and 11. Thus the four measures are given by

$$Di = \frac{33}{33} = 1.000, Sc = \frac{188}{188} = 1.000, Bu = \frac{0}{17} = 0.000, St = \frac{17}{17} = 1.000. \quad (16)$$

The features of these four measures for NDVI and population data are shown in Figure 23.

## 6 Results and discussion

In this paper, the relationships of population and green tracts in the Tokyo area based on remote sensing data are discussed. The green tracts of land cover are evaluated by NDVI. The population data in rectangular cells are based on a national census for population of Japan.

Firstly, the relationships between NDVI and population data of cells are analyzed for stratified areas by their positions. The values of NDVI and sparseness of population increase as the distance from the Yamanote region increases, except for the Yamanote region itself as shown in Figures 4 and 7. There is a positive correlation in the area where distance from the Yamanote region is from 1 to 6 km shown as Figure 8. Thus, the Tokyo area is divided into three regions of the Yamanote, the middle, and the west zones.

In the Yamanote region, the explanatory variables *ndvi* and *office* are important to predict the population of this region. The *office* is the index variable for presence of government and corporation offices and shopping centers. The coefficient of determination based on multiple regression of population on *ndvi* and *office* is fairly high ( $R^2 = 0.685$ ). The index variable, *park* of a large park is not an important variable in the Yamanote region.

In the middle region, the explanatory variables *ndvi* and *park* are selected to predict the population of this region. The coefficient of determination for the multiple regression of population on *ndvi* and *park* is a fairly high ( $R^2 = 0.704$ ).

In the west region, the explanatory variables *ndvi* and *park* are selected to predict the population of this region. The coefficient of determination for the multiple regression of population on *ndvi* and *park* is low ( $R^2 = 0.380$ ).

For the Tokyo area as a whole, the explanatory variable *middle* is added as an index for the middle of the region. The explanatory variables *ndvi*, *park*, *office*, *middle* and *ndvi*  $\times$  *middle* are selected to predict the population across the entire region. The coefficient of determination for the multiple regression line of population on these five explanatory variables is a high ( $R^2 = 0.747$ ). The coefficient of *ndvi* for simple regression in the middle region is different from the other two regions, so the interaction variable *ndvi*  $\times$  *middle* is included.

Relationships between NDVI and population spatial data are also analyzed for their topological connectivity by the echelon approach.

The echelon topview maps stratified by the echelon numbers are given for the population and green tract data. The information derived from the echelons is summarized in echelon tables. Echelon tree dendrograms of the population and green tract data are constructed to represent the surface topology for the structure of the map. The echelon map for the difference between the population and green tract data is formulated to compare the structure of the difference.

There are different properties in the Yamanote, the middle and the west regions. The Yamanote region is the center of Tokyo City. This region is a mixture of many types of areas. Although this region consists of sparsely populated areas for government and corporation offices and shopping centers, there are some fairly dense areas in northern parts. This corresponds to the following features in the echelon map for population data of Figures 16 and 21. There are peaks for echelon numbers 1, 3, 4, 6 and 20, along with foundations as echelon numbers 31, 32 and 33. There are also some imperial houses and gardens. In echelon map for NDVI data of Figure 15, these correspond to peaks as echelon numbers 1, 2 and 4, along with foundation echelon number 31.

The middle region is residential area with high density in Tokyo metropolitan area. Population decreases with increasing distance from the Yamanote region. NDVI increases with increasing distance from the Yamanote region. There is a gentle slope which consists of echelon numbers 24, 28, 31 32 and 33 in echelon map for population data of Figure 16. In case of the echelon map for NDVI data of Figure 15, the slope consists of echelon numbers 21, 26, 28, 29 and 31.

The west region is a mixture of many types of areas like the Yamanote region. There are some dense areas which are near the railroad station. These areas are in the west parts of echelon numbers 28 and 31 in Figure 16. The sparse areas are in the southwest part of this region which consists of echelon number 2 in Figure 16. These areas are alongside the river. There are some gardens and green tract areas, corresponding to the peaks of echelon numbers 3, 7 and 9 in Figure 15. There are also a few green tract areas corresponding to echelon numbers 26, 27 and 30.

Based on the echelon topview maps and tree dendrograms, the structures of the population and green tract data are quite similar. The peaks of these maps are also similar. The difference between the population and green tract data is mainly caused by the buildings where few people live in the Yamanote region and along the river in the west region.

The structure of NDVI and population data in Tokyo metropolitan area is likewise quite similar based on echelon tree measures. One of the most different structural aspects between these data is the order of the root echelon. The root echelon is order three for NDVI data, whereas the root echelon order is order four for population data. However, this is not an important indicator of the difference in their structure. Because, if there were not a peak consisting of echelon numbers 9, 11 and 22 for population data in Figure 20, the root echelon order would be changed to three like NDVI data.

This remote sensing data was taken on February 11 in winter. The value of NDVI in summer is larger than in winter. For instance, the value of NDVI is between 0.45

and 0.65 in summer and between 0.2 and 0.4 in winter for evergreen trees. Thus each regression line of population on NDVI has different regression coefficients based on the time of sensing data. Although the regression coefficients are different for the seasons, the coefficient of determination for fit may be not so different.

## References

- [1] M. Cardillo, D. W. Macdonald and S. P. Rushton (1999). Predicting mammal species richness and distributions : testing the effectiveness of satellite-derived land cover data. *Landscape Ecology*, 14, 423-435.
- [2] D. Carr (1991). Looking at large data sets using binned data plots. pp. 7-39, A. Buja and P. Tukey eds. *Computing and Graphics in Statistics*, New York: Springer-Verlag.
- [3] D. Carr, R. Littlefield W. Nicholson and J. Littlefield (1987). Scatterplot matrix techniques for large N. *Journal of the American Statistical Association*, 82, 424-436.
- [4] E. Isaaks and R. M. Srivastava (1989). *An introduction to applied geostatistics*, New York: Oxford University Press.
- [5] J. R. Jensen (2000). *Remote sensing of the environment: An earth remote perspective*. Upper saddle river: Prentice Hall.
- [6] G. D. Johnson, W. M. Myers, G. P. Patil T. J. O'Connell and R. P. Brooks (1999). Predictability of bird community-based ecological integrity using landscape measurements. Technical Report 99-0601, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, The Pennsylvania State University, University Park, PA.
- [7] Kurihara, K. (1995). An approach to statistical software for exploratory data analysis. *Proceedings of International Conference on Statistical Methods and Statistical Computing for Quality and Productivity Improvement*, 435- 441.
- [8] Kurihara, K. (1997). On the Properties of a Test Process Composed of a Test for Outliers and t Test. *Proceeding of the Sixth China-Japan Symposium on Statistics*, 103-106.
- [9] R. I. Miller (1994). *Mapping the diversity of nature*. London: Chapman and Hall.
- [10] W. L. Myers, G. P. Patil and K. Joly (1997). Echelon approach to areas of concern in synoptic regional monitoring. *Environmental and Ecological Statistics*, 4, 131-152.
- [11] W. L. Myers (1999). Remote sensing and quantitative geogrids in PHASES. Technical Report ER9901, Environmental Resources Research Institute, The Pennsylvania State University, University Park, PA. pp. 1-57.
- [12] W. L. Myers, G. P. Patil, and Charles Taillie (1999). Conceptualizing pattern analysis of spectral change relative to ecosystem status. *Ecosystem Health*, 5(4), 285-293.

- [13] G. P. Patil (1998). Statistical ecology and environmental statistics for cost-effective ecological synthesis and environmental analysis. In *Modern Trends in Ecology and Environment*, R. S. Ambasht, ed. Backhuys Publ., The Netherlands. pp. 5-36.
- [14] F. F. Sabins (1996) *Remote sensing : Principles and interpretation*, Third Edition. New York: Freeman and Company.
- [15] I. F. Spellerberg (1991). *Monitoring ecological change*, Cambridge: Cambridge University Press.