Quantitative Representation of Distribution and Mixture of Urban Land Use Through Spatial Autocorrelation and Information Entropy

Takashi Kyakuno
Senior Researcher, Museum of Nature and Human Activities, Japan

Abstract
The distribution and mixture of urban land use in Japanese cities was quantitatively represented by using two indexes of texture analyses, spatial autocorrelation and information entropy. The spatial autocorrelation was calculated as Moran's I, and the entropy acquired by considering not only the composition of the areas of each land use class, but also the relation of adjacent cells. Land use data in urban areas except green areas were obtained from a digital land use map, and green areas by processing remote sensing data using an Aster GDS. These data were compiled by GIS, and the two indexes of each land use were calculated in each class. It was concluded that both indexes indicate different states of land use, and their relation was different in each land use class. In some it was sufficient to consider only the entropy, but in other cases, it was not sufficient. Using these indexes together gave a more accurate portrait than one alone for grasping the distribution and mixture of urban land use.

Keywords: distribution and mixture of urban land use; spatial autocorrelation; information entropy; GIS; remote sensing

1. Introduction
Improving diversity of urban land use has become one of the most significant issues in Japan. There is a trend for relatively monotonous land use areas such as seaside industrial areas or newly developed residential areas to be given diversity by intentionally allocating various facilities or land use to such areas from the point of view of regional diversity. In many cases, the diversity of regions is discussed according to the area ratio of each land use class. It is true that the ratio is one of the most important and essential factors. However, it cannot solely express the "diversity" of land use in some areas, because it merely indicates the ratio of the considered land use by focusing the means or sums, and never represents the mixture of the distribution. It is obviously different from a case in which commercial areas are clustered into one mass, in that the areas are dispersed in a checkerboard pattern, even though the ratios of both areas are the same. Large masses may be effective for creating large open spaces or networked green areas, because big clusters of monotonous land use enable the generation of open spaces or green areas. On the other hand, dispersed land use will encourage diversity of landscape and economic circulation among various sectors, the significance of which, from the environmental standpoint, has recently been stated.

Thus, it is thought that mixing several land use classes and proposing indexes that are able to express the distribution and mixture of land use are meaningful not only for grasping regional characteristics, but also for making plans that take into consideration a diversity of land uses.

Some studies in the fields of urban planning or spatial analysis have proposed many kinds of methods for evaluating the diversity of land use, and have obtained meaningful results. One of the most popular indexes in Japan is entropy, which is based on Shannon's information entropy. Koide (1977), Tamagawa (1982) and Moon et al. (1991) implemented studies regarding the mixing of land use in large areas by employing the index \(^{1-3} \). There are also studies in which remote sensing data were analyzed by use of the index, and the environmental changes were inspected \(^{4-7} \).

Moran's I, which is an index of spatial autocorrelation, has been increasingly employed in analyses of remote sensing data or GIS data. The advantages of the index are as follows. First, it is a robust index, and is not too sensitive for slight differences of distribution. Second, the value obtained from the index can be easily interpreted and understood at a glance in analogy with the normal correlation coefficient. Third, it is able to handle both discrete data and continuous data. The precise characteristics of the
index are discussed in the next section.

Some studies have evaluated urban spaces using these indexes, and have obtained many results. However, few studies analyze the land use pattern in large areas on a detailed scale and detached land use classes, and have used both GIS data and remote sensing data to evaluate urban land use. Furthermore, the characteristic of these values in different regions, such as business districts, newly developed residential areas and seaside residential areas has hardly been discussed. The objective of this study is thus to grasp the characteristics of these indexes when adapting them to Japanese cities and to clarify the advantages of combining them with precise land use data. In addition, the author's objective was to clarify the relation of these indexes to each land use class.

2. Methodology
2.1 Data processing

In this study, Detached Digital Information of the Kinki area of 1996 issued by the Geographical Survey Institute of Japan was employed to grasp land use and its distribution, and six classes, Low-story Houses, High or Middle-story Houses, Industrial Facilities, Commercial or Business Facilities, Roads and Others were extracted from the digital map, where Low-story Houses means houses of one to three stories, and High or Middle-story Houses means houses higher than four stories. The digital map has a 10m resolution of land use in a large area with precise classes.

With regards to green areas, remote sensing data acquired by Aster GDS (The ID of the scene is 0309190151410310021005.) was also used to determine where green areas were located because green areas in the digital map sometimes include bare areas like baseball fields or open spaces covered by artificial pavement. The green areas were extracted by the following procedures. First, the green areas were separated from the other areas by employing the maximum likelihood method by using data of one visible band and three near-infrared bands of Aster data, in which mountainous areas, forested areas and urban areas were also included as supervisors. Second, cloud areas were removed from areas which were classified as non-green areas because they have high reflectance in all wavelength bands of the remote sensing data. If the pixels of these cloud areas corresponded to green areas on the digital map, then the pixels were redefined as green areas. Since this remote sensing image is one of a relatively clear day, the cloudy areas which correspond to green areas on the digital map were considered to be mountainous areas, because in Japan mountainous areas tend to be covered by clouds even during clear weather.

The remote sensing data was projected onto a Japanese standard coordinate system and the pixel was resized into a 17.75m grid. Corresponding to the size, the resolution of the digital map was also reformatted into this size. The domain for the analysis was set as shown in Fig.1., which has 3120 pixels by 1680 pixels.

2.2 The analysis flow

All the pixels were covered by tiles, which were sized 16 pixels by 16 pixels, and were set as basic units for the analyses. The whole domain was therefore covered by 105 by 195 tiles. If a tile had pixels which corresponded to unclassified land use, like the outside areas of the original digital maps, they were removed from the analysis samples. As a result, 15023 tiles from among 20475 tiles remained for analysis.

The ratio of area, Moran's I and the entropy of each land use class on each tile were calculated by programs which were written for calculating Moran's I and the entropy, and by the use of GIS. The value of each grid was set as 1 if the pixel corresponded to the focused on land use class, and as 0 if not. This means that the value of a pixel altered as the land use class changed. Calculations were implemented by changing the focused on land use in each tile.

2.3 Definition of areas for comparative studies

These results were separately summarized in four different types of area: a newly developed residential area (usually called "New Town" in Japanese), a central business area, a typical urban fringe area and a seaside complex area, respectively. The Senri New Town area was chosen as a newly developed residential area, and the central part of Osaka, which is one of the largest cities in Japan, was set as a central business area. As an urban fringe area, Hanshin-kan, which is between two large city centers, Osaka and Kobe, was chosen. The area has been undergoing urban sprawl for about the past fifty years. Formerly, it was principally composed of agricultural fields. The area is considered to be a typical Japanese suburban residential area, composed mainly of relatively new houses, commercial facilities and industrial facilities, but still containing the spatial frames or old houses from former agricultural villages. Seaside areas vary in form and land use composition,
but the most typical case is one dominated solely by new industrial facilities. Considering the typical case, however, may only lead to obvious results, so another case in which there are remnants of former fishing villages or agricultural villages is considered for the analyses. A rough map of the areas, indicating their locations, is shown in Fig.1.

2.4 The spatial autocorrelation and information entropy

As has been discussed, Moran's I is an index which relates to the size and number of clusters, and can be interpreted in the same way as the normal correlation coefficient, where 1 means high correlation and 0 means no relation. In the Moran's I index, there are various ways to consider the neighborhood relation. In this study, only cells adjacent to any targeted cell were considered, and cells which were placed in diagonal places on the map were not considered. The equation is as follows.

\[ I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij}} \]  

(1)

Where \( i, j \) are the running numbers of each cell (1<=i, j<=256), \( x_i \) is the value of the cell \( i \) (it is 1 if the cell corresponds to a focusing class, and 0 if not), \( x \) is the mean of \( x_i \) in the calculation domain, \( n \) is the total number of cells in a tile (\( n=256 \) in this study), \( w_{ij} \) is the weighing variable (in this study, it is set as 1 if cell \( i \) and cell \( j \) are adjacent to each other in either the vertical or horizontal direction, and 0 if not).

The information entropy is an index which numerically indicates the variation of areas and their dispersion. It shows the maximum number when the areas of each class are equivalent and the cells are adjacent to cells of each class in the same probability. On the other hand, it shows 0 if the domain is occupied by only one class. The definition is as follows:

\[ E = -\sum_{j} p_i \log_2 p_i \]  

(2)

Where \( i, j \) is the running number of each land use class (defined as 1 in the case that the land use class is focused on, but 0 in the other land use classes), \( p_i \) is the ratio of an area of land use class \( i \) in a target tile, \( q_{ij} \) is the probability of which land use class \( j \) is adjacent when considering land use class \( i \). (\( \sum w_{ij}=1 \) each \( i \)).

The relation of these two indexes is shown in Fig.2., in which typical patterns of dispersion are shown as points in the plane consisting of these two indexes. For example, the checker board pattern of a land use indicates -1 in Moran's I and 1 in entropy. Then, the relation is plotted at the bottom of the plane.

The relation is also plotted in the upper right hand side when some clusters are formed. Random dispersion patterns are indicated as points in the right hand side, and cases in which a few of the grids of a land use are dispersed are shown as points of the left hand side. Some actual samples are illustrated.

3. Results and Discussion

3.1 Land use patterns of each class

Seven thousand one hundred and one tiles, which
contain more than 20% of the urbanized area, are chosen from all tiles. An urbanized area refers to classes other than water zones and green zones. In each tile, the spatial autocorrelation and entropy of each land use class were calculated and are shown as points in Fig.3. This illustrates a rough tendency towards dispersion of each land use class in urban areas.

Low-story Houses shows widely spread points in the plane, and is densely accumulated on the right hand side, which means that such houses are placed in various forms and show various dispersion patterns in urban areas. But, most of them tend to be dispersed randomly in urban areas, as can be seen by comparing the figure with the theoretical model shown in Fig.2. High or Middle-story Houses shows more monotonous dispersion patterns than Low-story Houses. In particular, there are few randomly dispersed patterns or stripe patterns unlike with Low-story Houses. This means that High or Middle-story Houses tend to aggregate in urban areas. Industrial Facilities show a similar tendency as High or Middle-story Houses.

Commercial or Business Facilities show as widely spread plots in the plane, as do Low-story Houses. However, there are no densely accumulated domains in the plane as in Low-story Houses, which means that Commercial or Business Facilities do not always assume a particular dispersion pattern but quite varied ones, including random, clustered and isolated patterns.

The patterns of dispersion of Roads represent the basic framework of urban areas. The graph shows that various dispersion patterns can be taken. In particular, a lot of points are concentrated on the right hand side of the vertical middle place of the plane, which means roads tend to form striped patterns in urban areas.

Comparing these graphs, it is supposed that the locations of Commercial or Business Facilities show diverse patterns and are unlikely to concentrate in particular patterns, and it is therefore generally difficult to define their typical dispersion patterns of the Commercial or Business Facilities. If business facilities and commercial facilities are, however, analyzed separately, other results will be acquired because business facilities are likely to form more clusters than commercial facilities due to the zoning system of Japanese urban planning law. It is difficult to precisely distinguish between business facilities and commercial facilities, because many facilities serve both functions.

The Japanese urban planning system does not usually exclude small commercial facilities, even in zones for residential areas, therefore such facilities can be built in various forms among various land use classes. They are sometimes isolated in residential areas or industrial areas, but are sometimes clustered together. The exception is the case of the so-called New Town area, because land use plans are strict and commercial facilities are encouraged to aggregate at town centers in advance.

### 3.2 Land use characteristics in different types of area

Figs.4.-1, 5.-1, 6.-1 and 7.-1 show land use maps of the four different types of areas and Figs.4.-2, 5.-2, 6.-2 and 7.-2 are graphs showing the composition of each land use class in the areas. The ratio of green areas to the whole area is also displayed as a number. The reason why the green areas are dealt with separately from the other land use classes is that green areas are found in all the land use classes mentioned in this study, and places which contains green areas sometimes also have houses, roads and commercial facilities. Figs.4.-3, 5.-3, 6.-3 and 7.-3 show the spatial autocorrelation and entropy of each land use class at each tile of each area. A rough tendency can be grasped through an
overview of the compositions of the areas. In the 'Newly Developed Residential Area', the area of Roads, Low-story Houses and High or Middle-story Houses are large as are Others. Others are thought to be mainly green areas. In the 'Central Business Area', Commercial or Business Facilities and Roads prevail because the land use of these areas is simplified, and several classes characterize the whole area. In the 'Urban Fringe Areas', there are more Low-story Houses and Roads than other classes since the area is typical of suburban residential areas in Japan, and mainly consist of a variety of houses. In the 'Seaside Complex Area', various land use classes exist almost equally, and there are no conspicuous characteristics.

In order to grasp the mixing of land use, it is effective to refer to the spatial autocorrelation and entropy. Figs.3.-3, 4.-3, 5.-3 and 6.-3 show a different tendency of distribution of each land use class. Land use classes sharing more than 5 percent in each area are plotted in each figure. In the 'Newly Developed Residential area', Low-story Houses tends to show a vermicular cluster pattern or to form small clusters, whereas High or Middle-story Houses are likely to form big clusters, compared with the figure in the theoretical model in Fig.2. Roads are distributed in various ways, including line-like patterns in the area. It is interesting that Green shows a similar pattern to Low-story Houses. In the 'Newly Developed Residential Area', the lot size of each low story house is large and there are generally relatively large gardens in them. Furthermore, in the area there are some large parks, including a mass of green areas in addition to many trees on the street. Green areas sometimes form big clusters, and sometimes form vermicular clusters like Low-story Houses. On the other hand, High or Middle-story Houses have an obvious tendency to form clusters. In the 'Newly Developed Residential Area', high story buildings like apartment houses were separated from other types of buildings according to precise land use plans, and therefore High or Middle-story Houses were regulated in designated zones. There are some points which are located in the left hand side of the vertical middle place in Fig.4.-3, which means that High or Middle-story Houses are isolated in some tiles but not clustered. There are no isolated Low-story Houses, which means that they are never isolated in such a planned area. Vermicular clusters of Low-story Houses are the result of the existence of narrow streets running through a Low-story Houses area. Roads show a variety of values in the figure. One characteristic is that roads make some clusters, which are indicated by points on the upper right hand side of Fig.4.-3. This means that there are very wide roads in the area. Each type of land use tends to cluster in the 'Newly Developed Residential Area', and the types of land use that are not Roads are clearly separated from each other.

In the 'Central Business Area', Commercial or Business Facilities can be regarded as forming vermicular clusters by referring to Fig.5.-3, which is also shown in the land use map. Commercial or Business Facilities are connecting until roads cut them off. Roads shows similar patterns as Commercial or Business Facilities, and can be thought to form vermicular cluster patterns as well. Different from 'Newly Developed Residential Areas', narrow networked roads which are indicated as points around x=1.8-2.0 and y=0 are never seen. This indicates that these two land use classes form many clusters in micro perspective, and these clusters are mixed in macro perspective. In the area, the Roads areas tend to be wider than most roads in usual residential areas, and the density of Roads is high. There are only two main land use classes and the land use patterns can be almost regarded as simplified in small scale but mixed in large scale.

In the 'Urban Fringe Area', Low-story Houses show a variety of patterns as can be seen from the dispersed pattern in Fig.6.-3. Some Low-story Houses form clusters vermiculated by Roads, and other Low-story Houses form line patterns. Several Low-story Houses are isolated. As discussed before, this area is a typical suburban area of Japan, in which some zones have been developed using deliberate plans, but most zones were created without any total plan so that urban sprawl occurred. The former land use and periods of development of each zone are also different, which cause inconsistent land use patterns of the area. This tendency also relates to the pattern of the roads. Generally, narrow networked roads prevail over other types. The spatial frames of former agricultural villages can still be seen in these forms of road. It is a typical phenomenon of the area that Commercial or Business Facilities are not clustered, but form line-like patterns. This is because the sizes of lots in the area are generally small and the roads are narrow. Therefore, it is prohibited by the building standard law in Japan to build large-scale buildings in these areas. It is a matter of course that Commercial or Business Facilities form lines along roads.

In the 'Seaside Complex Areas', each land use forms various patterns of distribution, the same as the 'Urban Fringe Area'. Low-story Houses are located in various patterns as well as Commercial or Business Facilities. They sometimes form lines, clusters, vermicular clusters, or are isolated. High or Middle-story Houses form either large clusters or small clusters, the same as Industrial Facilities. Their sizes may depend on the sizes of the lots in which the facilities are located. Roads also show various forms; some are narrowly networked, others are wide. The crossing points of wide roads can be regarded as a kind of "cluster" of Roads pixels, and this is indicated as points in the upper right hand side in Fig.7.-3.
Fig. 4.-1 Land Use Map of 'Newly Developed Residential Area'

Fig. 5.-1 Land Use Map of 'Central Business Area'

Fig. 4.-2 Percentages for Areas of Each Land Use Class and Ratio of Green Area to the Whole Area in 'Newly Developed Residential Area'

Fig. 5.-2 Percentages for Areas of Each Land Use Class and Ratio of Green Area to the Whole Area in 'Central Business Area'

Fig. 4.-3 Plots by Entropy and Spatial Autocorrelation of Each Land Use Class in 'Newly Developed Residential Area'

Fig. 5.-3 Plots by Entropy and Spatial Autocorrelation of Each Land Use Class in 'Central Business Area'
Fig. 6.-1 Land Use Map of 'Urban Fringe Area'

Fig. 7.-1 Land Use Map of 'Seaside Complex Area'

Fig. 6.-2 Percentages for Areas of Each Land Use Class in 'Urban Fringe Area'

Fig. 7.-2 Percentages for Areas of Each Land Use Class in 'Seaside Complex Area'

Fig. 6.-3 Plots by Entropy and Spatial Autocorrelation of Each Land Use Class in 'Urban Fringe Area'

Fig. 7.-3 Plots by Entropy and Spatial Autocorrelation of Each Land Use Class in 'Seaside Complex Area'
Using these two indexes led to interesting results. Planes spanned by these indexes express various forms of the mixture of land use. For example, big clusters and vermiculated big clusters can be distinguished from each other by referring to the spatial autocorrelation. Using the two indexes enables us to clearly represent the difference in the features of land use patterns between *Low-story Houses* and *High or Middle-story Houses* in the 'Newly Developed Residential Area' well. Furthermore, one cluster can be distinguished from several small clusters by referring to the spatial autocorrelation. On the other hand, vermicular clusters can be distinguished from small clusters or line-like patterns by referring to the entropy. Differences in the road pattern of the 'Central Business Area' and 'Urban Fringe Area' are well represented by using entropy. The land use patterns of *Low-story Houses* of different areas can also be distinguished by employing entropy. In 'Newly Developed Residential Areas', *Low-story Houses* tend to form vermicular clusters, but are sometimes isolated or form small clusters in 'Seaside Complex Areas'. Entropy plays an important role in distinguishing these differences.

4. Conclusion

Applying both entropy and spatial autocorrelation in order to understand urban land use is a more effective method for comprehending land use classes than using either index alone. Applying both indexes offers a new way for describing the morphology of urban land use. Entropy is targeted at diversity and complexity, while spatial autocorrelation is targeted at representing spatial similarity on a local scale. Using both indexes enabled the author to grasp urban land use from the perspective of diversity and similarity.

There are still some problems with this scheme. First, some spatial situations which look different from each other have close values in both indexes. This tendency became conspicuous in line-like dispersion patterns. The second problem concerns scale dependency. In this study, the scale of the basic grid was set to 17.75m due to the pixel size of the remote sensing data, and the analysis tiles were set to 16 pixels by 16 pixels. The reason why this tile size is employed is that in future studies these results will be compared with results by the fractal analysis in which it is desirable that the analysis tiles have a size of 2^n pixels and the tile size is not excessively large for usual Japanese cities. The result will differ if these basic values are different. In particular, making the grid size smaller will lead to a change of the result. The third problem originates in the way in which adjacent cells were considered. In this study, for the purpose of reducing computing time, diagonal cell areas were not regarded as adjacent. When taking this into account, different results may occur. Furthermore, the Euclidian system is not the only way to address urban space. Another grid system like a pentagon system or triangle system may lead to another result. This study is a first step for quantitatively describing urban land use pattern. Based upon the fact that many patterns of dispersion were well comprehended by the scheme of this study, these problems will be discussed with various cases and various methodologies.

Acknowledgement

The author is grateful for the support by Grant-in-Aid for Scientific Research (18700239) of Japan.

References

1) Koide, O. (1977) 'Tochiriyoukonngoudo no Tekiyou Narabini Sono Kentei' (in Japanese), City planning review, 12, pp.79-84 (in Japanese).
2) Tamagawa, H. (1982) 'Tochirinoyou no Titsujoisei no Suuriteki Hyogen ni Kansuru Kenkyu' (in Japanese), City planning review, 17, pp.73-78 (in Japanese).
3) Moon, T. et al. (1999) A study on the Indicator of Mixed Degree of Land Use, City planning review, 26, pp.505-510 (in Japanese).
4) Tsunekawa, A. et al. (1991) Quantitative Analysis Methods of Mixed Land Use, Environmental Information Science, 20(2), pp.115-120 (in Japanese).
5) Kumagai, J. (2000) Development of a Spatial Analysis Method for Land Cover taking into account the Complexity of Local Distributions, Environmental Information Science, 14, pp.1-5 (in Japanese).
6) Kyakuno, T. and Sotoma, M. (2004) Study on the change of land use and the degree of mixture of it focusing on distance from coastal line in waterfront area. Journal of Architecture and Planning, 579, pp.75-80 (in Japanese).
7) Kyakuno, T. and Miyazaki, H. (2005) Texture analysis of SAR data and its relation to land coverage in urban area. Journal of Architecture and Planning, 590, pp.111-116 (in Japanese).
8) Zhang, C. (1999) Development of a Spatial Analysis Tool for Irregular Zones Using the Spatial Data Framework. Graphical Review of Japan, 72A-3, pp.166-177 (in Japanese).
9) Yoon, J. et al. (2004) A Study on the Planning Utilization Based by the Spatial Cognition of Mental Evaluation, Journal of Architecture and Planning, 585, pp.61-67 (in Japanese).
10) Myint, S. W. (2003) Fractal approaches in texture analysis and classification of remotely sensed data - comparisons with spatial autocorrelation technique and simple descriptive statistics-, Int. J. Remote Sensing, 24(9), pp.1925-1947.
11) Myint, S. W. et al. (2004) Wavelets for Urban Spatial Feature Discrimination -Comparisons with Fractal, Spatial Autocorrelation, and Spatial Co-occurrence Approaches-, Photogrammetric & Remote Sensing, 70(7), pp.803-812.
12) Cliff A.D. and Ord J.K.: Spatial processes -models & applications-Pion, 1981.