No.662

Evaluation of Some Orderings of Two-Dimensional Space in Terms of the Correlation Coefficients of Order-Differences and Distances

Yoshiaki OHSAWA
Takeshi KOSHIZUKA

Evaluation of Some Orderings of Two-dimensional Space in Terms of the Correlation Coefficients of Order-Differences and Distances

March, 1996

Yoshiaki OHSAWA
Takeshi KOSHIZUKA

Institute of Socio-Economic Planning, University of Tsukuba
Tsukuba 305, JAPAN
phone: +81-298-53-5224, fax: +81-298-55-3849
e-mail: osawa@shako.sk.tsukuba.ac.jp.

Abstract

We are concerned here with the one-dimensional orderings which assign the consecutive integers to the cells of a rectangular region. The aim of this study is to evaluate eight one-dimensional orderings according to the magnitude of their correlation coefficients of the order-differences between two points over the rectangular region and the Euclidean distances between them. The sensitivity of the shape of the study area and the sensitivity of the level of aggregation to the correlation coefficients are then examined.

Key words

correlation coefficient; ordering of two-dimensional space; spatial database; data aggregation

1. Introduction

An *one-dimensional ordering* of a region, which is divided into k spatial units, is defined as any one-to-one assignment of consecutive integer 1 through k to each spatial unit. Obviously, no one-dimensional ordering can preserve all spatial properties among members of the set of entities in the original two-dimensional space. In other words, the geometrical relationships on the original two-dimensional space cannot be perfectly duplicated, regardless of one-dimensional orderings. The first scholar to give much attention to one-dimensional orderings of two-dimensional space was Mark(1990). Mark(1990) evaluated some orderings by focussing attention on the order-differences between spatial units and their spatial neighbours. He showed that the standard row order minimizes the mean absolute difference, the root-mean-squared difference and the mean maximum absolute difference for both 8-by-8 and 256-by-256 matrices.

Some efforts have been made on the one-dimensional ordering of two-dimensional space. What seems to be lacking, however, is to clarify the relationship between the order-differences and the distances. Here is one example. Japan is divided into 47 prefectures, as shown in Figure 1. Japan comprises 660 cities, and each city is assigned to a prefecture. In general, cities are ordered firstly according to the order of its prefecture established by the Ministry of Home Affairs in Japan, as presented in Figure 1, and secondly in order of Japanese syllabary. From the geographical point of view, the orderings of the prefectures have to preserve the spatial property such that order-differences between two cities increase linearly with Euclidean distances between them. Accordingly, in order to evaluate this ordering of prefectures from this point of view, we calculate two types of correlation coefficients. The first is the correlation coefficient of the order-differences between two cities in Japan and the Euclidean distances between them. In this case, we take ${}_{660}C_2 = 217,470$ pairs of measurements. The second is the correlation coefficient of the order-differences between cities in Japan and Tokyo, and the Euclidean distances between them. In this case, 660 pairs of measurements are taken. We obtained the positions of these cities from the Ministry of Construction, Japan (1992) to calculate the Euclidean distances. Obviously, the greater correlation coefficient implies that the order is good at preserving the spatial property mentioned above. The first and second types of correlation coefficients to be computed using the positioning data are 0.826 and 0.953, respectively, showing moderately strong positive linear relationships. Hence, it may be concluded that the order of prefectures established by the Ministry of Home Affairs in Japan can preserve practically the spatial property. This example provides the starting-point of this study. The specific questions pursued are "can other orderings of prefectures produce greater correlation coefficient", "to what extent does the shape of the study area affect the correlation coefficients", and "to what extent does the levels of the aggregation have an influence upon the correlation coefficients". Similar examples will spring to mind: the assignments of telephone area codes, administrative unit codes, and postal codes. In general, the ordering of such codes are hierarchically nested. Another example occurs in the ordering of maps in an atlas. This is because maps in an atlas have to be paginated while preserving the geometrical relationships between the maps.

The aim of this study is to evaluate some one-dimensional orderings according to the magnitude of their correlation coefficients of the order-differences between two points over the rectangular region and the Euclidean distances between them. In particular, the analytical expressions of two types of the correlation coefficients which we have mentioned in the previous example are presented. The sensitivity of the shape of the study area and the sensitivity of the level of aggregation to these correlation coefficients are then examined.

This paper is organized as follows: Section 2 describes some orderings of a two-dimensional space and their properties. Section 3 gives the analytical expressions of the two types of the correlation coefficients. In Section 4, the correlation coefficients for small-sized rectangular regions are calculated and then interpreted according to the magnitude of these values. Finally, Section 5 contains our conclusions.

2. Orderings

Assumptions

This study considers only a rectangular region with sides equal m and n, denoted as Ω, for integers $m > 0$ and $n > 0$. The points to be ordered are continuously and uniformly distributed over Ω. This rectangular region is divided into mn unit squares using a rectangular grid, as shown in Figure 2. From now on, the word *cell* instead of unit square is used to avoid misunderstanding. The orderings of this study are two-phase. The orderings consist simply of the first phase in which mn cells are ordered: the second phase in which all points within each cell are ordered at ramdom, that is to say, the ordering of points within a cell is independent of their location. Recall our earlier example in which cities within a prefecture in Japan are arranged in order of Japanese syllabary. This agrees with the orderings of this second phase.

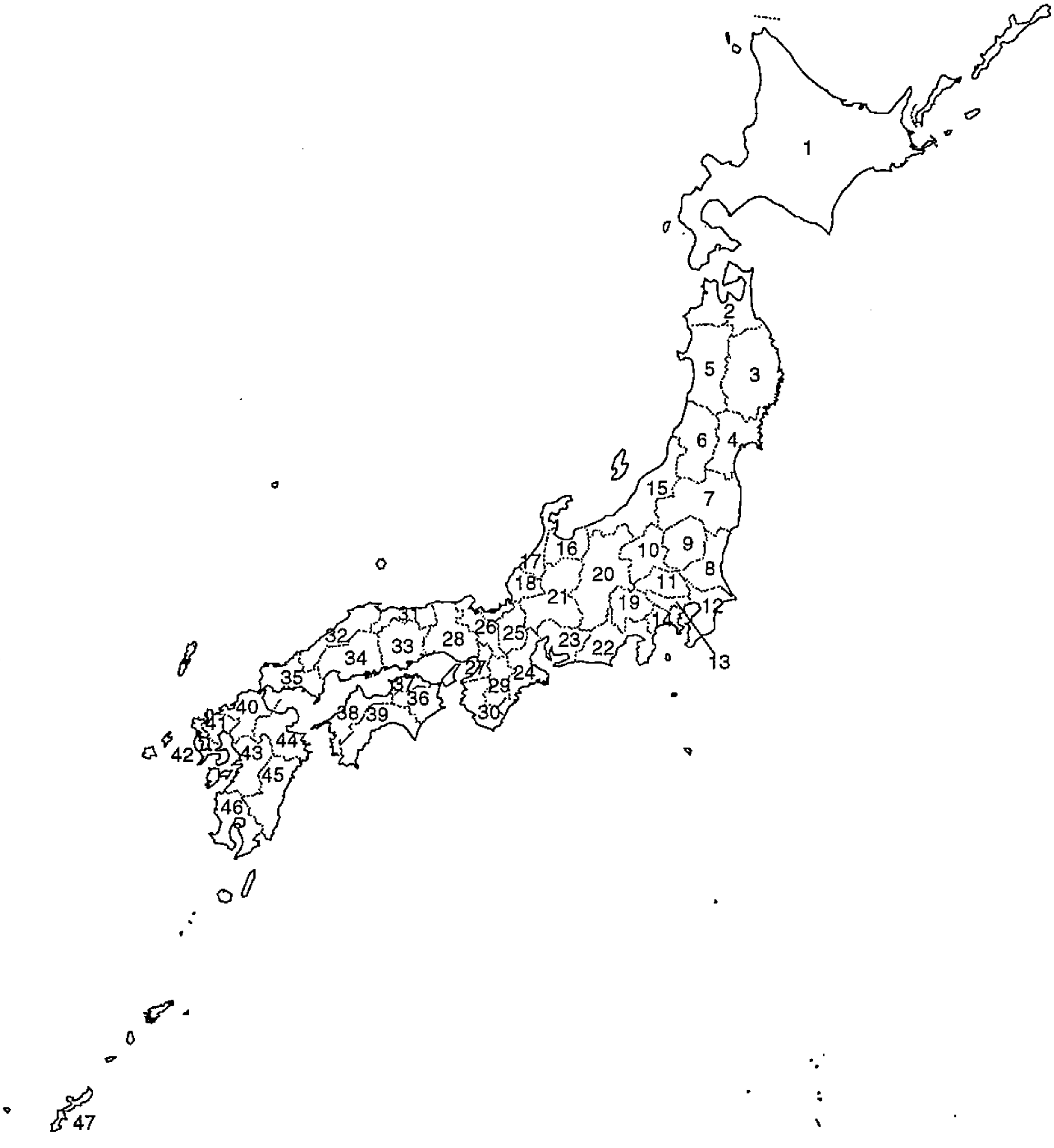

Figure 1: Ordering of Prefectures in Japan

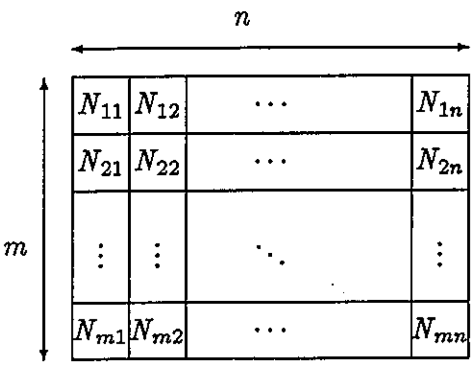

Figure 2: Rectangular Region Ω

For convenience, we may assume that the number of points within a cell is normalized into one. Consequently, we may consider that in the first phase the cells are assigned consecutive integers 1 through mn, and that in the second phase all points within the cell in the i-th row, the j-th column are numbered from $N_{ij} - 1$ to N_{ij}, where N_{ij} is the order of the (i, j)-th cell.

Orderings of Cells

We deal with the following eight orderings of cells:

(o1) row order

(o2) column order

(o3) maximum value distance spiral order

(o4) rectilinear distance spiral order

(o5) N-shaped order

(o6) U-shaped order

(o7) cross order

(o8) Hilbert order

Figure 3 is a series of eight diagrams illustrating examples corresponding to $m = n = 4$. The definition of these orderings except rectilinear distance spiral order, and their properties of relevance to spatial data-handling are given in Mark(1990).

Row order (column order) is the row-by-row (column-by-column) traversal of a rectangular grid. Clearly, when $m = n$, the column order also can be obtained by rotating the row order by 90°. These two orderings belong to the category of *positive monotonic order*, i.e., the orders always increase with increasing value of one coordinate when the other is held constant. The orderings in remote sensing and matrix ordering within computer programmings are typical examples of the row order.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16

7	6	5	16
8	1	4	15
9	2	3	14
10	11	12	13

14	6	5	13
7	1	4	12
8	2	3	11
15	9	10	16

1	3	9	11
2	4	10	12
5	7	13	15
6	8	14	16

1	4	13	16
2	3	14	15
5	7	9	11
6	8	10	12

1	3	9	11
4	2	12	10
13	15	5	7
16	14	8	6

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Figure 3: Eight Orderings of 4-by-4 Rectangular Region

The maximum value distance spiral order (rectilinear distance spiral order) simply begins at the central cell, moving in a counter-clockwise direction around the central point of the region, denoted as O. All cells are arranged in increasing order of the maximum value distance , i.e., Chessboard metric (rectilinear distance, i.e., Manhattan metric) from O to the center of the cells; they eventually spiral from O. These two spiral orderings are *continuous*, i.e., the cells with consecutive orders are always first-order neighbours in terms of rook's case (queen's case). The maximum value distance spiral order (rectilinear distance spiral order) will be referred to here as *M-spiral order* (*R-spiral order*). A good example of the spiral orderings are the ordering of wards in Paris, France.

The rest are *quadrant-recursive.* The original rectangle region is divided into four quadrants. Each is then further subdivided until the cell is reached: see Samlet(1984) for a tutorial survey. Because of the symmetric property, the ordering of four quadrants can be classified into three groups as shown in Figure 4. They are called N-shaped, U-shaped and cross templates. *N-shaped order*, which is a sequence now known as the Morton order, can be obtained by applying a N-shaped template recursively without rotation. In the same way, *U-shaped order* can be gained by applying a U-shaped template recursively without rotation. *Cross order* can be obtained by applying a cross template recursively without rotation. Hilbert order is based on the U-shaped order template, but the template is rotated systematically

1	3		1	4		1	4
2	4		2	3		3	2

Figure 4: N-shaped, U-shaped and Cross Templates

by a recursion procedure such that the ordering is continuous: see Goldschlager(1981). The quadrant-recursive orderings are important for data-structure in image analysis and GIS.

It should be noted that the first four orderings can be defined for any natural numbers m and n. However, the last four orderings are only valid for square regions where $m = n$ and, both m and n are any natural numbers of the powers of two.

3. Correlation Coefficients

Correlation Coefficients

Let Z be the random variable of the order-difference between any two points over the study area Ω, and let R be the random variable of the Euclidean distance between them. The first measure investigated in this study is the correlation coefficient of Z and R. This correlation coefficient is formulated as follows:

$$\rho_{Z,R} = \frac{E[ZR] - E[Z]E[R]}{\sqrt{E[Z^2] - E[Z]^2}\sqrt{E[R^2] - E[R]^2}}. \tag{1}$$

It is obvious from the definition that $E[Z]$, $E[Z^2]$, $E[R]$ and $E[R^2]$ are independent of orderings. Some simple calculations show that

$$E[Z] = \frac{1}{m^2n^2}\int_0^\infty\int_0^\infty | s - t | \, dsdt = \frac{mn}{3}, \tag{2}$$

$$E[Z^2] = \frac{1}{m^2n^2}\int_0^\infty\int_0^\infty (s-t)^2 dsdt = \frac{m^2n^2}{6}. \tag{3}$$

Moreover, we have

$$\begin{aligned} E[R] &= \frac{1}{m^2n^2}\int_0^m\int_0^n\int_0^m\int_0^n \sqrt{(u_1-v_1)^2+(u_2-v_2)^2} dv2_1 dv_1 du_2 du_1 \\ &= \frac{1}{30m^2n^2}\{2(m^5+n^5) - 2(m^4 - 3m^2n^2 + n^4)\sqrt{m^2+n^2} \\ &\quad +5mn(m^3 \log\frac{\sqrt{m^2+n^2}+n}{m} + n^3\log\frac{\sqrt{m^2+n^2}+m}{n})\}, \end{aligned} \tag{4}$$

$$\begin{aligned} E[R^2] &= \frac{1}{m^2n^2}\int_0^m\int_0^n\int_0^m\int_0^n (u_1-v_1)^2 + (u_2-v_2)^2 dv_2 dv_1 du_2 du_1 \\ &= \frac{m^2+n^2}{6}. \end{aligned} \tag{5}$$

These two integrations above were done by Ghosh(1951).

Joint Probability Density Function

Knowing $E[ZR]$ enables $\rho_{Z,R}$ to be calculated. To derive the joint probability density function of Z and R, denoted as $h_{Z,R}(z,r)$, we concentrate on the spatial relationships between any two cells Let Z_{ijkl} be the random variable of the order-difference between the (i,j)-th and the (k,l)-th cells. Let R_{ijkl} be the random variable of the Euclidean distance between the (i,j)-th and the (k,l)-th cell. From the assumption mentioned earlier, as the random variables Z_{ijkl} and R_{ijkl} are stochastic independent, the joint probability density function of Z_{ijkl} and R_{ijkl}, denoted as $h_{Z_{ijkl},R_{ijkl}}(z,r)$, can be given as:

$$h_{Z_{ijkl},R_{ijkl}}(z,r) = f_{Z_{ijkl}}(z)g_{R_{ijkl}}(r).$$

where, $f_{Z_{ijkl}}(z)$ is the probability density function of Z_{ijkl}, and $g_{R_{ijkl}}(r)$ is the probability density function of R_{ijkl}. Therefore, we have

$$\begin{aligned} h_{Z,R}(z,r) &= \frac{1}{m^2n^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n} h_{Z_{ijkl},R_{ijkl}}(z,r) \\ &= \frac{1}{m^2n^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n} f_{Z_{ijkl}}(z)g_{R_{ijkl}}(r). \end{aligned}$$

Hence, we have

$$\begin{aligned} E[ZR] &= \int_0^\infty\int_0^\infty zrh_{Z,R}(z,r)drdz \\ &= \frac{1}{m^2n^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\int_0^\infty zf_{Z_{ijkl}}(z)dz\int_0^\infty rg_{R_{ijkl}}(r)dr. \end{aligned}$$

It follows from the definition of N_{ij}'s that the random variable Z_{ijkl} can be regarded as the difference of the random variable with uniform distribution over the interval $[N_{ij}-1, N_{ij}]$ and the random variable with uniform distribution over the interval $[N_{kl}-1, N_{kl}]$. Therefore, the probability density function of Z_{ijkl} can be given as follows:

1) $i = k$ and $j = l$

$$f_{Z_{ijkl}}(z) = \begin{cases} 2-2x & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

2) otherwise

$$f_{Z_{ijkl}}(z) = \begin{cases} x - \mid N_{ij}-N_{kl}\mid +1 & \mid N_{ij}-N_{kl}\mid -1 \le x \le \mid N_{ij}-N_{kl}\mid \\ \mid N_{ij}-N_{kl}\mid +1-x & \mid N_{ij}-N_{kl}\mid \le x \le \mid N_{ij}-N_{kl}\mid +1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, we have

$$\int_0^\infty z f_{Z_{ijkl}}(z)dz = \begin{cases} \frac{1}{3} & i=j \text{ and } k=l; \\ \mid N_{ij}-N_{kl} \mid & \text{otherwise.} \end{cases} \quad (6)$$

Moreover, $\int_0^\infty r g_{R_{ijkl}}(r)dr$ can be analytically expressed. As the solving of the integration is rather tedious and its analytical expression is very complicated, they are given in the Appendix. After all, $E[ZR]$ can also be described analytically. This is an important fact to stress.

Two remarks are in order on the other orderings. First, there is another type of row-by-row ordering such as the row-prime order: see Mark(1990). Row-prime order is row-by-row, but in reversing direction after each row. It is interesting to note that, because of symmetry, both row order and the row-prime order have the same values of $\rho_{Z,R}$, irrespective of m and n. Second, both inward and outward versions of the two spiral orders also have the same values of $\rho_{Z,R}$, irrespective of m and n.

Central and Peripheral Subregions

Here, we consider the case where the study area Ω is divided into two subregions. The subregion including the central point O is called *central*. The other subregion is called *peripheral*. For example, the shading in the square in Figure 5 designates the central subregion. Let $\tilde{Z}$ be the random variable of the order-difference between any point within the central subregion and any point within the peripheral subregion. Let $\tilde{R}$ be the random variable of the Euclidean distance between them. The second measure examined in this study is the correlation coefficient of $\tilde{Z}$ and $\tilde{R}$. This correlation coefficient is defined as follows:

$$\rho_{\tilde{Z},\tilde{R}} = \frac{E[\tilde{Z}\tilde{R}] - E[\tilde{Z}]E[\tilde{R}]}{\sqrt{E[\tilde{Z}^2] - E[\tilde{Z}]^2}\sqrt{E[\tilde{R}^2] - E[\tilde{R}]^2}}. \quad (7)$$

To calculate $\rho_{\tilde{Z},\tilde{R}}$, we shall show the joint probability density function of $\tilde{Z}$ and $\tilde{R}$, denoted as $\tilde{h}_{\tilde{Z},\tilde{R}}(z,r)$. As is the case with $h_{z,r}(z,r)$, we have

$$\begin{aligned} \tilde{h}_{\tilde{Z},\tilde{R}}(z,r) &= \frac{1}{\mid I_C \mid\mid I_P \mid} \sum\sum_{(i,j)\in I_C} \sum\sum_{(k,l)\in I_P} h_{Z_{ijkl},R_{ijkl}}(z,r) \\ &= \frac{1}{\mid I_C \mid\mid I_P \mid} \sum\sum_{(i,j)\in I_C} \sum\sum_{(k,l)\in I_P} f_{Z_{ijkl}}(z) g_{R_{ijkl}}(r). \quad (8) \end{aligned}$$

where, I_C is the set of the indices of the cell in the central subregion, and I_P is the set of the indices of the cell in the peripheral subregion. Therefore, we can get the following expected

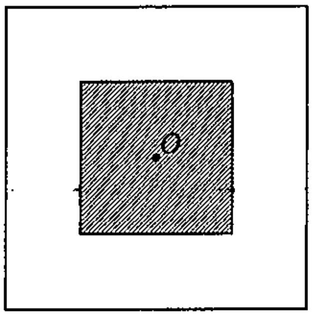

Figure 5: Central and Peripheral Subregions

values:

$$
\begin{aligned}
E[\tilde{Z}] &= \frac{1}{|I_C||I_P|}\sum_{(i,j)\in I_C}\sum_{(k,l)\in I_P}\int_0^\infty z f_{Z_{ijkl}}(z)dz;\\
E[\tilde{R}] &= \frac{1}{|I_C||I_P|}\sum_{(i,j)\in I_C}\sum_{(k,l)\in I_P}\int_0^\infty r g_{R_{ijkl}}(r)dr;\\
E[\tilde{Z}\tilde{R}] &= \frac{1}{|I_C||I_P|}\sum_{(i,j)\in I_C}\sum_{(k,l)\in I_P}\int_0^\infty z f_{Z_{ijkl}}(z)dz\int_0^\infty r g_{R_{ijkl}}(r)dr;\\
E[\tilde{Z}^2] &= \frac{1}{|I_C||I_P|}\sum_{(i,j)\in I_C}\sum_{(k,l)\in I_P}\int_0^\infty z^2 f_{Z_{ijkl}}(z)dz;\\
E[\tilde{R}^2] &= \frac{1}{|I_C||I_P|}\sum_{(i,j)\in I_C}\sum_{(k,l)\in I_P}\int_0^\infty r^2 g_{R_{ijkl}}(r)dr.
\end{aligned}
$$

The expectation $\int_0^\infty z f_{Z_{ijkl}}(z)dz$ is given in (6). Moreover,

$$
\int_0^\infty z^2 f_{Z_{ijkl}}(z)dz = \begin{cases} \frac{1}{6} & i=j \text{ and } k=l; \\ |N_{ij}-N_{kl}|^2+\frac{1}{6} & \text{otherwise.} \end{cases} \tag{9}
$$

The expectation $\int_0^\infty r g_{R_{ijkl}}(r)dr$ can be expressed analytically, as we have already pointed out. In addition, the expectation $\int_0^\infty r^2 g_{R_{ijkl}}(r)dr$ can also be analytically described: see Appendix. It follows from the expression (7) that $\rho_{\tilde{Z},\tilde{R}}$, can also be given analytically.

4. Computational Results

Correlation Coefficients $\rho_{Z,R}$

As computing $\rho_{Z,R}(\rho_{\tilde{Z},\tilde{R}})$ by the expressions mentioned previously usually requires tedious calculations, we carry out the calculation by means of a personal computer. Broadly speaking, as the running time of computing $\rho_{Z,R}(\rho_{\tilde{Z},\tilde{R}})$ is proportional to the total number of cell pairs,

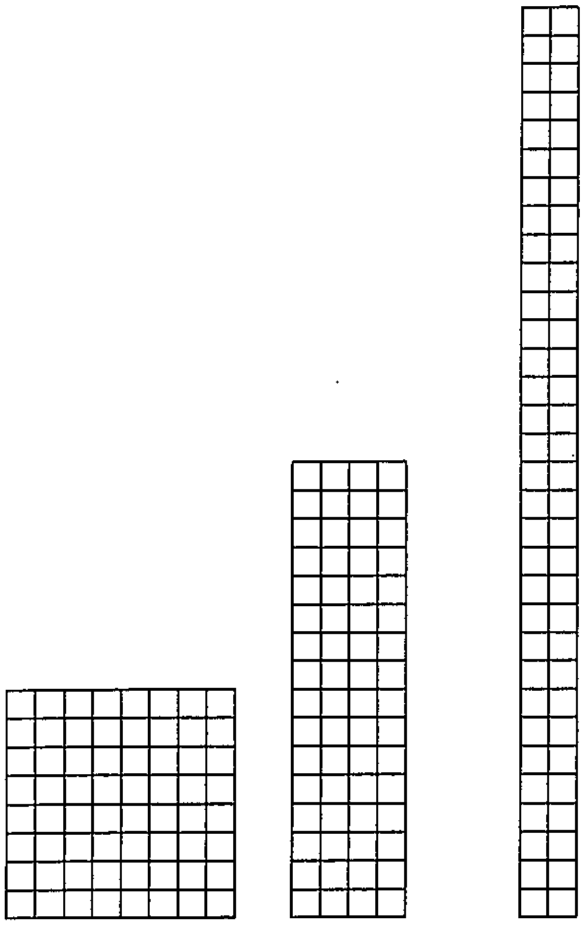

Figure 6: 8-by-8, 16-by-4 and 32-by-2 Regions

i.e., m^2n^2, it is impossible to compute $\rho_{Z,R}(\rho_{\tilde{Z},\tilde{R}})$ for large m and n. Therefore, we take up small-sized rectangular regions.

First, effects resulting from the shape of the study area Ω on $\rho_{Z,R}$ are examined. We deal with the following four rectangular regions:

(d1) 8-by-8 region

(d2) 16-by-4 region

(d3) 32-by-2 region

(d4) 64-by-1 region

The spatial configurations of these rectangular regions except the 64-by-1 region are illustrated in Figure 6. These four regions are filled with the integer 1 through 64, using each of the eight orderings mentioned previously. It should be noted that in the last three regions, the quadrant-recursive orderings cannot be defined. In addition, in the case of the 8-by-8 region, the correlation coefficients of the row order agree with that of the column order. Furthermore, in the cases of the 32-by-2 and the 64-by-1 regions, the two spiral orderings are the same. Computational results on the correlation coefficients $\rho_{Z,R}$'s are provided in Table 1.

Second, we would like to focus attention on the effects due to aggregation of cells on $\rho_{Z,R}$.

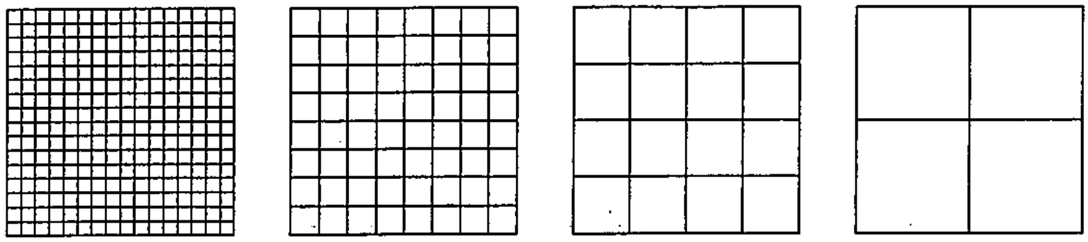

Figure 7: Successive Spatial Aggregations of Cells

A brief overview of the aggregation problems in spatial analysis can be found in Fotheringham et al.(1995). Here we deal with four zonal aggregations starting from the base level 16-by-16 cells, as shown in Figure 7. In other words, we consider the four cases where the sets consist of one cell, four cells, sixteen cells and sixty four cells are ordered instead of cells, respective. Clearly, the correlation coefficients satisfy the property of scale invariance, that is to say, they are independent of the measurement units employed. Therefore, to calculate their correlation coefficients, we may take up the following three rectangular regions, in addition to the region (d1):

(d5) 16-by-16 region

(d6) 4-by-4 region

(d7) 2-by-2 region

Therefore, the cells of the 16-by-16, the 4-by-4 and the 2-by-2 regions are assigned the integer 1 through 256, 16, 4, respective, to using each of the eight orderings. It should be noted that in the case of the 2-by-2 region, all the orderings except the cross order are identical. Moreover, in the case of square regions, the results of the column order is the same with that of the row order. The correlation coefficients to be computed are givin in Table 2.

Correlation Coefficients $\rho_{\tilde{Z},\tilde{R}}$

In order to explore the effects due to the shape of the study area on $\rho_{\tilde{Z},\tilde{R}}$, we deal with the regions (d1), (d2) and (d3). For these regions, we define the central 2-by-2 subregion as a central subregion. We are not concerned here with the 64-by-1 region where the central 2-by-2 subregion could not be defined. The results are presented in Table 3.

Next, to examine the zonal aggregation problem, we take up the region (d7). For comparative purposes, in addition to the location, the area of the central subregion have to be

Table 1: **Shape of Rectangular Regions and Correlation Coefficients $\rho_{Z,R}$'s**

ordering	8-by-8 region	16-by-4 region	32-by-2 region	64-by-1 region
row order	0.677	0.982	0.997	0.999
column order	0.677	0.154	0.256	0.999
M-spiral order	0.129	0.225	0.246	0.250
R-spiral order	0.118	0.214	0.246	0.250
N-shaped order	0.614	n.a.	n.a.	n.a.
U-shaped order	0.578	n.a.	n.a.	n.a.
cross order	0.268	n.a.	n.a.	n.a.
Hilbert order	0.572	n.a.	n.a.	n.a.

Table 2: **Zonal Aggregation and Correlation Coefficients $\rho_{Z,R}$'s**

ordering	16-by-16 region	8-by-8 region	4-by-4 region	2-by-2 region
row order	0.687	0.677	0.633	0.441
column order	0.687	0.677	0.633	0.441
M-spiral order	0.094	0.129	0.251	0.441
R-spiral order	0.112	0.118	0.151	0.441
N-shaped order	0.616	0.614	0.592	0.441
U-shaped order	0.581	0.578	0.561	0.441
cross order	0.269	0.268	0.265	0.234
Hilbert order	0.572	0.572	0.561	0.441

identical. Therefore, in this case, the central 4-by-4 subregion is defined as a central subregion. We are not concerned here with the regions (d5) and (d6). This is because, in both, the central subregion which coincide with the central subregion of the 8-by-8 region, could not be defined. The correlation coefficients are displayed in Table 4.

Interpreting the Results of Correlation Coefficients

In this study, eight systematically generated ordering are compared. Based on the numerical results, the following three properties have been drawn:

(p1) The most striking feature of Tables 1 and 2 is that without exception the row order maximizes $\rho_{Z,R}$'s. Moreover, generally speaking, except for the column order, the next four best according to the magnitude of $\rho_{Z,R}$'s are quadrant-recursive orderings. The two worst are the spiral orderings. In contrast to the feature above, we see from Tables 3 and 4 that either R-spiral order or the row order maximizes $\rho_{\bar{Z},\bar{R}}$'s. In addition, the three best according to the magnitude of $\rho_{\bar{Z},\bar{R}}$ are the two spiral orderings and the row order. The worst four are the recursive orderings. Tables 1, 2 3 and 4 also reveal that

Table 3: **Shape of Rectangular Regions and Correlation Coefficients** $\rho_{\tilde{Z},\tilde{R}}$'s

ordering	8-by-8 region	16-by-4 region	32-by-2 region
row order	0.599	0.962	0.991
column order	0.599	0.073	0.166
M-spiral order	0.706	0.936	0.985
R-spiral order	0.790	0.936	0.985
N-shaped order	0.408	n.a.	n.a.
U-shaped order	0.351	n.a.	n.a.
cross order	0.118	n.a.	n.a.
Hilbert order	0.344	n.a.	n.a.

Table 4: **Zonal Aggregation and Correlation Coefficients** $\rho_{\tilde{Z},\tilde{R}}$'s

ordering	16-by-16 region	8-by-8 region
row order	0.641	0.599
column order	0.641	0.599
M-spiral order	0.770	0.706
R-spiral order	0.818	0.790
N-shaped order	0.413	0.408
U-shaped order	0.355	0.351
cross order	0.119	0.118
Hilbert order	0.347	0.344

among the quadrant-recursive orderings examined, N-shaped order, i.e., Morton order is the best. This is followed by U-shaped order and Hilbert order, but these two are numerically almost equal. Whereas cross order has the lowest correlation coefficient. Furthermore, comparing Table 1 with Table 3, and Table 2 with Table 4, we see that, except for the two spiral orderings, $\rho_{Z,R} > \rho_{\tilde{Z},\tilde{R}}$. On the other hand, in the case of the two spiral orderings, $\rho_{Z,R} < \rho_{\tilde{Z},\tilde{R}}$.

(p2) Analyzing the magnitude of correlation coefficients in Tables 1 and 3 indicates that except for the column order, as the rectangular region becomes more oblong in shape, both $\rho_{Z,R}$'s and $\rho_{\tilde{Z},\tilde{R}}$'s increase, as we would expect. In particular, in the case of the 8-by-8 rectangular region, the largest $\rho_{Z,R}$ (the largest $\rho_{\tilde{Z},\tilde{R}}$) is about 0.677 (0.790). However, in the case of the 32-by-2 rectangular region, the largest correlation coefficient is 0.997 (0.991).

(p3) An observation of Tables 2 and 4 shows that, except for the two spiral orderings, the correlation coefficients become smaller as the cells are aggregated. To put it another way, the zonal aggregation exerts a bad influence upon correlation coefficients. In contrast to this, if the spiral orderings are adopted, the correlation coefficients become larger as the cells are aggregated. That is to say, the zonal aggregations exert a favorable influence upon correlation coefficients.

5.Conclusions

To explore the one-dimensional orderings of two-dimensional space is an important subject from the perspective of spatial database and GIS. In this study, attention was focused on the correlation coefficients of the order-differences and the distances. To begin with, the analytical expressions of the two types of correlation coefficients by means of a set of mathematical expressions were demonstrated. Next, eight systematically generated orderings of small-sized regions are evaluated according to the magnitude of the correlation coefficients. Based on the numerical experiences, it was shown that without exception the standard row order is the best according to the magnitude of $\rho_{Z,R}$'s, and that either the R-spiral order or the row order maximizes $\rho_{\bar{Z},\bar{R}}$'s. Thus, it can be concluded that the standard row order performs very well. This generally conforms with Mark's(1990) conclusion. It was also shown that both the shape of the study area and the zonal aggregation have a great influence on correlation coefficients.

Finally, there are some possible modifications to our analysis for future research. In this study, we limitted the discussion to the standard correlation coefficients given by (1) and (7). The strict requirement of a linear relationship may seem too demanding. It may be sufficient to know, for example, as the value of one coordinate increases so the other increases, and, as the the value of one coordinate decreases so the other decreases. Spearman's rank correlation coefficient is the familiar measures of correlation that takes account of this extra flexibility. Future research will be needed to explore the orderings in terms of the other criteria such as Spearman's rank correlation coefficients.

Literature Cited

[1] Fotheringham, A.S., P.J.Densham and A.Curtis (1995): The Zone Definition Problem in Location-Allocation Modeling, *Geographical Analysis, 27*, pp.60-77.

[2] Ghosh, B.(1951): Random Distances within a Rectangle and between two Rectangles, *Bulletin of Calcutta Mathematical Society, 43*, pp.17-24.

[3] Goldschlager, L.M.(1981): Short Algorithms for Space-Filling Curves, *Software - Practice and Experience, 11*, pp.99-100.

[4] Mark, D.M. (1990): Neighbour-based Properties of Some Orderings of Two-dimensional Space, *Geographical Analysis, 22*, pp.145-157.

[5] Ministry of Construction, Japan (1992): *Coordinates of Localities in Japan.* Nippon Chize Center. (in Japanese).

[6] Ohsawa Y., T.Koshizuka and O.Kurita (1991): Errors Caused by Rounded Data in Two Simple Facility Location Problems *Geographical Analysis, 23*, pp.56-73.

[7] Samet, H. (1984): The Quadtree and Related Hierarchical Data Structures, *ACM Computing Surveys, 16*, pp.187-260.

Appendix

We consider a rectangle with sides equal a and b. Two points are uniformly and independently distributed within this rectangle. Let R denote the random variable of the Euclidean distance between them. The probability density function of R, denoted by $\phi_R(r : a, b)$ is represented as

$$\phi_R(r : a, b) = \begin{cases} \frac{2r}{a^2b^2}\{ab\pi - 2(a+b)r + r^2\} & 0 \leq r \leq b \\ \frac{2r}{a^2b^2}\{2ab \arcsin \frac{b}{r} + 2a\sqrt{r^2 - b^2} - 2ar - b^2\} & b < r \leq a \\ \frac{2r}{a^2b^2}\{2ab(\arcsin \frac{b}{r} - \arccos \frac{a}{r}) + 2a\sqrt{r^2 - b^2} + 2b\sqrt{r^2 - a^2} - r^2 - a^2 - b^2\} & a < r \leq \sqrt{a^2 + b^2} \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

The expectations of R and R^2 are given as follows:

$$\int_0^{\sqrt{a^2+b^2}} r\phi_R(r : a, b)dr = \frac{1}{30a^2b^2}\{2a^5 + 2b^5 - 2(a^4 - 3a^2b^2 + b^4)\sqrt{a^2 + b^2} + 5a^4b \log \frac{\sqrt{a^2 + b^2} + b}{a} + 5ab^4 \log \frac{\sqrt{a^2 + b^2} + a}{b}\}, \quad (11)$$

$$\int_0^{\sqrt{a^2+b^2}} r^2\phi_R(r : a, b)dr = \frac{a^2 + b^2}{6}. \quad (12)$$

The probability density function (10) and the expectations (11) and (12), which are equivalent to (4) and (5), were derived by Ghosh(1951).

Making use of the indirect method, which Ghosh(1951) showed, the probability density function $g_{R_{ijkl}}(r)$ can be represented as follows:

1) $i = k$ and $j = l$

$$g_{R_{ijkl}}(r) = \phi_R(r : 1, 1);$$

2) $i = k$ and $j \neq l$

$$g_{R_{ijkl}}(r) = \frac{1}{2}\phi_R(r:1,|j-l|+1) + \frac{1}{2}\phi_R(r:1,|j-l|-1) - \phi_R(r:1,|j-l|);$$

3) $i \neq k$ and $j \neq l$

$$\begin{aligned} g_{R_{ijkl}}(r) &= \frac{1}{4}\phi_R(r:|i-k|+1,|j-l|+1) + \frac{1}{4}\phi_R(r:|i-k|-1,|j-l|-1) \\ &-\frac{1}{2}\phi_R(r:|i-k|+1,|j-l|) - \frac{1}{2}\phi_R(r:|i-k|,|j-l|+1) \\ &+\frac{1}{4}\phi_R(r:|i-k|+1,|j-l|-1) + \frac{1}{4}\phi_R(r:|i-k|-1,|j-l|+1) \\ &-\frac{1}{2}\phi_R(r:|i-k|,|j-l|-1) - \frac{1}{2}\phi_R(r:|i-k|-1,|j-l|) \\ &+\phi_R(r:|i-k|,|j-l|). \end{aligned}$$

Hence, we can recognize that $g_{R_{ijkl}}(r)$ can be analytically given with help of already known expressions $\phi_R(r:\cdot,\cdot)$'s. It follows from (11) and (12) that both $\int_0^{\sqrt{a^2+b^2}} r g_{R_{ijkl}}(r)dr$ and $\int_0^{\sqrt{a^2+b^2}} r^2 g_{R_{ijkl}}(r)dr$ can also be analytically presented by the expressions derived so far.