Discussion Paper No. 72R

PERCEPTION OF SPATIAL DISPERSION IN THE DISTRIBUTION OF POINT OBJECTS

Yukio Sadahiro *

MARCH, 1998

*Research Center for Advanced Science and Technology,
University of Tokyo
4-6-1, Komaba, Meguro-ku, Tokyo 153, Japan

March 12, 1998

Perception of Spatial Dispersion in the Distribution of Point Objects

Yukio Sadahiro
Department of Urban Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Phone: +81-3-3812-2111 (ext. 6273)
Fax: +81-3-5800-6965
E-mail: sada@okabe.t.u-tokyo.ac.jp

Perception of Spatial Dispersion in the Distribution of Point Objects

Abstract

The present paper analyzed the perception of spatial dispersion in the distribution of point objects. Spatial dispersion is one of the major concepts communicated by dot maps. To promote efficient communication of this concept, two experiments were conducted for investigating the relationship between the perception of spatial dispersion and the following map characteristics: spatial arrangement of points, number of points, symbol size, and map scale. Regression models were then built on the basis of the experiment results to describe the relationship quantitatively. The obtained models enable map authors to predict the degree of spatial dispersion perceived by map readers. From the models and the results of a computer-assisted simulation, the following conclusions were obtained: 1) map scale greatly affects the perception of spatial dispersion, 2) spatial arrangement of points and number of points are also influential, 3) the size of point symbols do not significantly affect the perception of spatial dispersion.

Introduction

Maps showing the distribution of point objects, which are often called dot maps, are fundamental tools for visualizing the spatial pattern of points. They reveal a great amount of information about the distributions, and convey various spatial concepts to map readers. Thus researchers who study spatial point patterns, such as cartographers, geographers, and epidemiologists, use dot maps to analyze point distributions visually.

Spatial concepts, however, cannot be always successfully communicated if map authors including GIS do not understand the perception of spatial concepts by map readers. To improve the communication of spatial concepts, the perception of spatial concepts on dot maps has been studied in the literature. Spatial cluster, for instance, is one of the basic concepts communicated by dot maps, and its perception was analyzed by Sadahiro (1997). He discussed the principles causing cluster perception, and proposed a statistical method for predicting spatial clusters perceived on dot maps. The experiments were performed to test the validity of the model, and it was reported that the model successfully predicted perceptual clusters. The concepts of spatial region and spatial regionalization have been studied by Jenks (1973) and McCleary (1975). They examined spatial regions perceived on dot maps by experiments. Spatial dispersion, however, has not yet studied in its perceptual aspect though it is a very important concept in geography (King, 1969), GIS (Fotheringham and Rogerson, 1994), spatial statistics (Cressie, 1993), and other related fields (Pielou, 1977; de Lepper, et al., 1995). The pattern of point distributions is often described by the degree of their spatial dispersion: "tightly clustered," "clustered," "dispersed," and "uniform" distributions. Though Sadahiro (1997) investigated local clusters perceived on a map, the paper did not deal with the perception of global spatial dispersion. Hence, to fill the gap of the research, we study the perception of spatial dispersion in the distribution of points and build mathematical models representing the relationship between the perception and map characteristics. This enables map authors to predict the degree of perceived spatial dispersion, and to display point objects appropriately for communication of spatial dispersion.

In the following sections, we first discuss map characteristics that affect the perception of spatial dispersion. We then conduct two experiments to investigate how the characteristics affect the perceived spatial dispersion in point distributions, and build mathematical models to represent the perception on the basis of the results of the experiments. The models enable us to predict the degree of spatial dispersion perceived by map readers and provides design guidelines for map authors to communicate the concept of spatial dispersion. We then perform computer-assisted Monte-Carlo simulations to illustrate the relationship between map characteristics and the perception of spatial dispersion.

Map Characteristics Affecting Perception of Spatial Dispersion

A variety of variables affect the perception of spatial concepts on dot maps. We classify them into two types: visual variables and spatial variables. Visual variables determine the appearance of point symbols, for instance, symbol color, symbol size, and background color. Spatial variables are related to the spatial distribution of points, including the spatial arrangement of points, the number of points, map orientation and map scale. Among the above, we focus on four variables that are considered to be most important and influential in the perception of spatial dispersion, in order to avoid difficulties arising from the correlation among variables.

From visual variables, we choose the size of map symbols. Symbol size is an important variable in dot maps, which greatly affects the visibility of spatial pattern (Robinson et al., 1995) and the perception of spatial clusters (Sadahiro, 1997). Since the concept of spatial dispersion appears closely related to spatial pattern and spatial cluster, it is reasonable to incorporate symbol size in the analysis.

From spatial variables, we choose the spatial arrangement of points, the number of points, and map scale. Spatial arrangement, which is sometimes called spatial pattern, indicates the locational pattern of points with respect to each other, and is invariable against rotation, translation, and scaling. This variable is closely related to the concept of spatial dispersion that have been used in geography (see King, 1969, for instance), and it appears very influential in the perception of spatial dispersion. Similar to spatial arrangement, the number of point objects is also important because it affects the visibility of spatial pattern on dot maps. It is hard to perceive and understand spatial pattern in a point distribution if the map shows too many points. Regarding the perception of spatial dispersion, it is expected that the perception is ambiguous and fuzzy when there are many points. In addition to these variables, map scale is considered to affect the perception of spatial dispersion. Let us suppose a spatial arrangement of point objects depicted in Figure 1a. When the points are displayed at a large scale (Figure 1b), the distribution appears to be fairly uniform. As the scale decreases, however, the distribution drastically changes to a tight cluster (Figure 1c).

Figure 1. A spatial arrangement of points displayed at three scales. (a) normal scale, (b) large scale, (c) small scale.

Analysis of the above variables will be done in three steps. Firstly, we perform Experiment 1 to investigate the relationship between the perception of spatial dispersion and the four variables. To represent this relationship, we build simple regression models using the results of the experiment. Secondly, we conduct Experiment 2 for further analysis of the relationship between the perceived spatial dispersion and the spatial variables, and revise the regression models obtained from Experiment 1. Thirdly, we perform the Monte-Carlo simulations based on the revised models to assist in illustrating the effect of the number of points and map scale.

Experiment 1

Method

Experiment 1 was conducted in the Department of Urban Engineering at the University of Tokyo. Thirty-three undergraduate students and nineteen graduate students served as subjects. They were naive as to the purpose of the experiment.

In the experiment, we used 36 test maps portraying the distribution of point objects (Figure 2), which composes a single map set. Each twelve of maps had 15, 30, and 45 points. The point distributions were initially determined randomly (the coordinates were generated through a C-program), and then modified by hand so that a variety of patterns appeared in the map set. Each point distribution was drawn on a paper of 6.5 cm in height and 12 cm in width. The points were indicated by black circles whose diameter was 2 mm (Figure 3a) in half of the map sets, and 5 mm (Figure 3b) in the remaining half. To avoid order-of-exposure effects in the task, we sorted the 36 maps randomly within each set. The subjects were randomly assigned to one of the map sets.

Figure 2. Maps showing point distributions used in Experiment 1. (a) 15 points, (b) 30 points, (c) 45 points. Maps are arranged in the order of the MPSD values computed in the subsequent analysis.

Figure 3. Examples of a test map. Maps where points are indicated by (a) black circles of 2 mm in diameter, (b) black circles of 5 mm in diameter.

In addition to the map sets, the subjects were given three standard maps of point distributions, each of which contained 15, 30, and 45 points (Figure 4). They were then asked to compare the spatial dispersion of the point distribution shown on a test map with that on the standard map having the same number of points, and to put a mark on the line below the test map to indicate the degree of spatial dispersion (see Figure 3).

Figure 4. Standard maps. (a) 15 points, (b) 30 points, (c) 45 points.

Results

We measured the position of marks from the left to the right with respect to the center of the line. The measured values, however, were not comparable among subjects because a variation existed in the way of marking among them: some preferred extreme evaluations while others tended to put the marks near the center of the line. We thus converted them so that the variance of the values for each subject was equal to 1.0. We call the obtained values the *Perceived Spatial Dispersions* (abbreviated as *PSD*). If a subject considered a point distribution dispersed, the PSD shows a positive value. If a subject judged a distribution to be clustered, the PSD shows a negative value. The PSD has a value close to zero, if a subject thought that the spatial dispersion of a point distribution was similar to that of the standard map.

Averaging the PSD values for each map, we obtained the *Mean Perceived Spatial Dispersion* (abbreviated as *MPSD*). We denote the MPSD value of Map i as $MPSD_i$. Figure 5 illustrates the values of MPSD.

Figure 5. The values of MPSD resulted from Experiment 1. (a) 15 points, (b) 30 points, (c) 45 points. The numbers above the figures correspond to those shown in Figure 2.

Let us first investigate whether the size of point symbols affects the perception of spatial dispersion. Comparing the values of MPSD presented in the upper and lower rows of Figure 5, we find that they are very similar in all three cases, especially in their orders. Spearman's rank correlation coefficients are 1.00, 0.92, and 0.89, respectively, all of which are larger than the critical value at a significance level of 1%. From this, we can say that symbol size does not greatly affect the average perception of spatial dispersion.

We then discuss the relationship between the spatial distribution of points and the perception of spatial dispersion. This will help us to build a mathematical model representing the perception of spatial dispersion. One of the keys to describe this relationship is spatial clusters contained in the point distributions. Examining the maps of 15 points in Figure 2, for instance, we notice that maps showing no distinct clusters such as Maps 15-1, 15-2, and 15-3, are considered to be more dispersed than those containing distinct clusters such as Maps 15-11 and 15-12. Similar results are shown for maps of 30 and 45 points. Thus we can say that spatial clusters weaken the impression of dispersion in point distributions. Moreover, cluster size affects the perception of spatial dispersion. Maps containing small clusters such as Maps 15-8, 30-8, and 45-6 have larger values of MPSD than maps of big clusters such as Maps 15-11, 30-11, and 45-12. This implies that point distributions look more dispersed as contained spatial clusters become small.

Another characteristic of spatial distribution that affects the perception of spatial dispersion is its uniformity. Among maps of 30 points, for instance, Maps 30-1 and 30-2 are very similar in the spatial arrangement of points. These maps, however, differ considerably in the MPSD value (see Figure 5b). This appears to be caused by the uniformity of point distributions: points on Map 30-1 are almost uniformly distributed while those on Map 30-2 are located rather in the lower half of the map than in the upper half. This suggests that even a slight difference of point arrangement significantly affects the perception of spatial dispersion when points are almost uniformly distributed.

The maps used in Experiment 1 include three pairs of maps that have the same arrangement of points presented in different scales: Maps 15-5 and 15-12, 30-4 and 30-12, and 45-4 and 45-10. We hence briefly take a look at the effect of map scale in the perception of spatial dispersion by examining these maps. The scale of Map 15-5 is twice as large as that of Map 15-12, and its MPSD value is somewhat larger than that of Map 15-12. The ratio of map scale of Map 30-4 to 30-12 and that of Map 45-4 to 45-10 are 2.0 and 1.4, respectively, and the MPSD values of the large-scale maps are larger than those of small-scale maps. This indicates that point distributions look more dispersed as map scale increases.

Now we turn to the variation in the perception of spatial dispersion among subjects. To measure this, we compute the variance of PSD for each map, and call it the *Variance of Perceived Spatial Dispersion* abbreviated as *VPSD*. Using this measure, we first examine the relationship between symbol size and the variation of perceived spatial dispersion. In Figure 6, we notice that symbol size clearly affects the VPSD values. For instance, maps showing large MPSD values such as Maps 15-1 and 30-1 have larger VPSD values when symbol size is 5 mm. Conversely, if a map has distinct clusters, the larger symbol gives smaller VPSD value. Consequently, to make the perception of spatial dispersion consistent among subjects, it is better to use a small symbol for dispersed distributions and a large symbol for clustered distributions.

We then examine the variation in the perception of spatial dispersion with respect to the spatial distribution of points. It can be seen from Figure 6 that maps on which small clusters are dispersedly distributed such as Maps 15-4, 15-8, and 30-7, have large VPSD values. This suggests that the spatial dispersion of such distributions is difficult for map readers to evaluate. Regarding the number of point objects, it is clear that the VPSD value decreases as points increase. From this, we can say that an increase in the number of points lowers the uncertainty in the perception of spatial dispersion.

Figure 6. The values of VPSD resulted from Experiment 1. (a) 15 points, (b) 30 points, (c) 45 points. The numbers above the figures correspond to those shown in

Figure 2.

Modelling

Keeping the above discussion in mind, we build simple regression models representing the perception of spatial dispersion in point distributions. We use aggregate models (the MPSD values) rather than disaggregate models (the PSD values), because the former are more robust against modelling error than the latter. The regression equation used for the models is written as

$$MPSD_i = \alpha SV_i + c, \qquad (1)$$

where SV_i is a measure representing the three spatial variables of the point distribution shown on Map *i*, and α and *c* are constant parameters to be estimated. For the measure SV_i, we propose two types of variables as follows.

<u>Distance variables</u>

Distance methods are very popular in spatial statistics, especially in point pattern analysis. They are based on the average point-to-point distance, which represents the degree of spatial dispersion of point distributions. Dispersed distributions show a large value, while clustered distributions have a small value.

A set of point pairs used for distance calculation can be defined in various ways. We adopt the following four methods for the definition.

1) Nearest neighbor method

This method considers every point and its nearest point as a pair for distance calculation (Figure 7b). Hence, the resultant set consists of as many point pairs as distributed points.

2) Delaunay neighbor method

This method employs the Delaunay triangulation generated by the point distribution for defining point pairs. Delaunay triangulation is often used to define "natural" neighboring points (see, for instance, Ahuja, 1982; Elliot, 1985; Okabe et al., 1992). The set consists of all the pairs of points connected by the Delaunay edges (Figure 7c).

3) Trimmed Delaunay neighbor method

Though Delaunay triangulation is useful for detecting natural neighboring points, it generates some "unnatural" point pairs around the boundary of triangulation. Such point pairs can be eliminated by using the Voronoi diagram, the dual graph of a Delaunay triangulation. Point pairs whose Voronoi regions are not adjacent within the map are regarded as "unnatural" neighboring points (Okabe et al., 1992). We thus eliminate such pairs to obtain "natural" neighboring points (Figure 7d).

4) Extended Delaunay neighbor method

In the trimmed Delaunay neighbor method, boundary points of Delaunay triangulation are connected by fewer Delaunay edges than interior points, which means that they are less influential in the perception of spatial dispersion. To correct this inequality, we additionally consider point pairs, each of which consists of a border point of Delaunay triangulation and the foot of a perpendicular from the border point to the map boundary (Figure 7e). This method is based on the assumption that the distance between a border point and map boundary affects the perception of spatial dispersion equally to the distance between neighboring points.

Five statistics are then computed from the point-to-point distances given by the above methods: mean (μ), variance (σ^2), standard deviation (σ), relative variance (σ^2/μ), and relative standard deviation (σ/μ). As a result, we obtain twenty statistics for each map, and call them *distance variables*. We successively apply them to the regression model defined by Equation (1) as the independent variable SV_i.

Figure 7. Point pairs defined for distance calculation. (a) A point distribution, (b) point pairs defined by the nearest neighbor method, (c) Delaunay neighbor method, (d) trimmed Delaunay neighbor method, (e) extended Delaunay neighbor method.

<u>Areal variables</u>

Another type of independent variable is called an *areal variable*, which represents the area of a neighborhood region of points. We hypothesize that a point distribution having a large neighborhood region will look dispersed rather than clustered. We define the neighborhood region of a point by a circle centered at the point, and call it the *neighborhood circle*. The radius of a circle is given by the following methods.

1) Nearest neighbor radius method

The radius of neighborhood circle of a point is given by the distance from the point to its nearest neighboring point. The neighborhood circles, therefore, have variation in size (Figure 8a).

2) Fixed radius method

The radius is arbitrarily given, and it is commonly applied to all the neighborhood circles (Figure 8b).

The area of the union of neighborhood regions inside a map, which is indicated by the shaded region in Figure 8, is used as the measure SV_i in the regression equation. Note that the neighborhood region outside the map boundary is not taken into account.

Figure 8. Neighborhood regions defined for area calculation. (a) Neighborhood circles whose radius is defined by the nearest neighbor radius method, (b) fixed radius method.

Using the above variables, we estimate the simple regression models defined by Equation (1) individually for maps of 15, 30, and 45 points. For fixed radius method, we adopt circles of 5 mm, 10 mm, 15 mm, and 20 mm in radius. The results of model estimation are shown in Table 1.

Table 1. Coefficients of determination R^2 and the F-values of the regression models. Star marks indicate that the model is significant at the 5% level. (a) Distance variables, (b) areal variables.

Table 1 indicates that the mean (μ) and relative standard deviation (σ/μ) of point-to-point distances defined by the trimmed Delaunay neighbor method give good results among distance variables. All of their F-values are larger than the critical value at a significance level of 5%. Among areal variables, the area of neighborhood region defined by circles of 10 mm in radius shows the largest R^2, whose F-values are also significant at the level of 5%.

We then consider the appropriateness of the regression analysis examining whether the assumptions required for regression hold. Figure 9 illustrates the residuals of the regression functions mentioned above. The distribution of residuals does not reveal obvious patterns such as curvatures or funnels (Montgomery and Peck, 1992) which violate the independency and normality assumptions. Hence it seems reasonable to apply the regression models to the obtained data.

Figure 9. The residual of the regression models. The horizontal axes indicate (a) the mean (μ) of point-to-point distances defined by the trimmed Delaunay neighbor method, (b) the relative standard deviation (σ/μ) of point-to-point distances defined by the trimmed Delaunay neighbor method, (c) the area of neighborhood region defined by circles of 10 mm in radius.

Now we have three variables that appear to be useful for representing the perception of spatial dispersion. We hereafter call these variables *Mean point-to-point Distance* (abbreviated as *MD*), *Relative Standard Deviation of point-to-point Distance* (*RSDD*), and *Area of Neighborhood Region* (*ANR*). The values of these variables for Map *i* are denoted by MD_i, $RSDD_i$, and $ANR_i(r)$, respectively, where r is the radius of neighborhood circles. We use them in a further modelling of the perceived spatial dispersion in the following section.

Experiment 2

Experiment 2 was conducted to test if the regression models obtained from Experiment 1 are applicable for a wider range of the number of points, and to examine which model is the best for representing the perception of spatial dispersion.

Method

Similar to Experiment 1, Experiment 2 was conducted in the Department of Urban Engineering at the University of Tokyo. Eighteen undergraduate students and two graduate students served as subjects. They were naive as to the purpose of the experiment.

We produced 30 test maps for the experiment. Each map showed the distribution of points whose number ranged from 5 to 50 (Figure 10). The point objects were indicated by black circles of diameter 2 mm on a paper of the same size as that used in Experiment 1 (Figure 11). We arranged 30 maps randomly for each map set, and randomly assigned the subjects to one of the map sets.

Figure 10. Maps displaying point distributions used in Experiment 2. Maps are arranged in the order of the MPSD values computed in the subsequent analysis. The number of points is indicated in parentheses.

Figure 11. An example of a test map.

Before the experiment we showed two example maps, representing completely clustered (Figure 12a) and completely dispersed distributions (Figure 12b). We then asked the subjects to put a mark on the line below a test map to indicate the degree of spatial dispersion. Unlike Experiment 1, Experiment 2 requires the subjects to compare maps containing different number of points on the same scale. This enables us to analyze explicitly how the number of points affects the perception of spatial dispersion.

Figure 12. Example maps. (a) Completely clustered distribution, (b) completely clustered distribution.

Results

We measured the PSD values in the same way as that used in Experiment 1. We should note, however, that the zero value of PSD in Experiment 2 implies that the subject could not judge the point distribution to be clustered or dispersed, which is different from that in Experiment 1. From the PSD values we computed the MPSD and VPSD values for each map (Figures 13 and 14).

The results are very similar to those of Experiment 1. Maps having no distinct clusters have large MPSD values, while maps containing big clusters show small MPSD values. The maps numbered from 1 to 6 on which points are almost uniformly distributed have fairly larger values of MPSD than other maps. Maps 17, 22, and 24, which show dispersedly distributed small clusters, have large VPSD values.

Figure 13. The values of MPSD in Experiment 2. The numbers above the figures correspond to those shown in Figure 10.

Figure 14. The values of VPSD in Experiment 2. The numbers above the figures correspond to those shown in Figure 10.

Modelling

Having obtained the results of Experiment 2, we revise the simple regression models obtained from Experiment 1 to build multiple regression models. We define the multiple regression equation by

$$MPSD_i = \alpha SV_i + \beta f(n_i) + c, \tag{2}$$

where n_i is the number of point objects shown on Map *i*, and α, β, and *c* are parameters to be estimated.

In Experiment 1, the variables MD, RSDD, and ANR fit the experiment results well. These variables, however, are not applicable to Equation (2) as the measure SV_i because they are dependent on the number of points. Hence, to standardize these variables, we divide them by their expectations when points are randomly distributed on the map. Mathematically, this standardization is written as

$$SV_i = \frac{MD_i}{\mathrm{E}\left[MD_i\right]}, \tag{3}$$

$$SV_i = \frac{RSDD_i}{\mathrm{E}\left[RSDD_i\right]}, \tag{4}$$

and

$$SV_i = \frac{ANR(r)_i}{\mathrm{E}\left[ANR(r)_i\right]}, \tag{5}$$

where E[MD_i], E[$RSDD_i$], and E[$ANR(r)_i$] are the expectations of MD, RSDD, and ANR for Map *i*, respectively. The expectations of MD and RSDD are computed by a Monte-Carlo simulation using the VORONOI 2 program (Sugihara and Iri, 1983). The expectation of ANR is numerically calculated by the method shown in the Appendix. We call the regression models by the name of the variable used for measuring SV_i: the MD models, the RSDD models, and the ANR models.

The second term of Equation (2) represents the effect of the number of points on the perception of spatial dispersion that is not fully taken into account by the above standardization. We adopt the following definitions of $f(n_i)$:

$$f\left(n_i\right) = 0, \tag{6}$$

$$f\left(n_i\right) = n_i, \tag{7}$$

and

$$f\left(n_i\right) = \log\left(n_i\right). \tag{8}$$

If the perception of spatial dispersion is fully dependent on the measure SV_i, Equation (6) will give a good result. However, if the MPSD value increases or decreases linearly with both SV_i and the number of points, it is better to use Equation (7). If the MPSD value increases slowly with the number of points, Equation (8) will be a good choice. We tried all the above equations to find a better model for representing the perception of spatial dispersion.

For the radius of neighborhood circles used in the ANR models, we try values from 1 to 20 mm incremented by 1 mm. The results of model estimation are shown in Figure 15 and Table 2.

Figure 15. Relationship between the coefficient of determination R^2 and the radius of neighborhood circles used in the ANR models.

Table 2. Coefficients of determination R^2, the *F*-values, and the *t*-values of the independent variables. Star marks indicate that the model or the variable is significant at the 5% level.

Figure 15 demonstrates the relationship between the coefficient of determination R^2 and the size of neighborhood circles employed in the ANR models. This figure indicates that the models of radius 8 mm give the largest R^2 among the ANR models, irrespective of the form of $f(n_i)$. This result is consistent with that of Experiment 1.

Table 2 shows the coefficients of determination R^2, the F-values, the t-values of the estimated models. The RSDD and ANR models show fairly large R^2, while the MD models have smaller values. All the independent variables employed in the RSDD and ANR models are significant at the level of 5%. When examining the residuals of these models, we did not find any obvious pattern in their distribution and the correlation between the residuals and the independent variables. Regarding the form of $f(n_i)$, we find that Equations (7) and (8) show similar results that are better than those given by Equation (6).

From the results of Experiments 1 and 2, we can conclude that the ANR models are the best among the models we proposed, because they showed the best fitness for the MPSD values in both experiments. The ANR models give better results if we add the term $f(n_i)$ which represents the effect of the number of points on the perception of spatial dispersion.

Simulating Perception of Spatial Dispersion

In the preceding section, we obtained the ANR models to represent the perceived spatial dispersion. The models implicitly describe the relationship between the perceived spatial dispersion and spatial variables, that is, the spatial arrangement of points, the number of points, and map scale. It is, however, difficult to understand this relationship intuitively by the regression equations. Does a point distribution usually look more dispersed if the number of points increases? How does the map scale affect the perceived spatial dispersion? To answer these questions, we employ the Monte-Carlo simulation to examine the effects of the number of points and map scale.

The simulation is performed as follows. We first determine the number of point objects and map scale. The points are then randomly distributed on a map. Given the point distribution, we compute the MPSD value by

$$MPSD_i = \alpha \frac{ANR(8\text{mm})_i}{\mathrm{E}\left[ANR(8\text{mm})_i\right]} + \beta \log\left(N_i\right) + c, \tag{9}$$

where α, β, and c are parameters estimated in the preceding section. For a given number of points and map scale, this procedure is repeated 10,000 times so that the probability distribution of MPSD is obtained.

Effect of Number of Point Objects

Fixing map scale to 1.0, we performed the Monte-Carlo simulations for various numbers of points ranging from 3 to 100. For each probability distribution of MPSD, we computed its expectation and variance (Figure 16). The expectation of MPSD decreases as the number of points increases. This implies that, in the random configuration, an increase of point objects raises the probability of point distributions considered to be clustered. Similar to the expectation, the variance of MPSD decreases as points increase. This indicates that an increase of points makes it difficult to find a difference in spatial dispersion between point distributions. This result is consistent with what we expected earlier.

Figure 16. The expectations and variances of MPSD in relation to the number of point objects.

Effect of Map Scale

The effect of map scale was investigated as follows. Given the number of points (5, 10, 20, 50, and 100), we performed the Monte-Carlo simulation and obtained the probability distribution of MPSD at a scale of 1.0 (Figure 17a). We then shrank map scale to 0.9 (Figure 17b), and performed the simulation. Map scale was successively decremented at 0.1 step until it reached 0.1 (Figures 17c and 17d). The expectations and variances of MPSD are depicted in Figures 18 and 19, respectively.

Figure 17. Changing map scale. Broken lines indicate the boundary of original map. (a) Scale at 1.0, (b) 0.9, (c) 0.5, (d) 0.1.

Figure 18. The expectations of MPSD in relation to map scale.

Figure 19. The variances of MPSD in relation to map scale.

As shown in Figure 18, an increase of map scale generally raises the expectation of MPSD, in other words, it makes point distributions look more dispersed. This effect is drastic in maps displaying many points. One might think it is strange that the expectations show their peaks at a scale of 0.9 when points are less than 50. This result is caused by edge effects. If the ANR model used in the simulations perfectly fits the perception of spatial dispersion, the maximum of the expectations would actually be given at a scale of 0.9. The model, however, contains estimation residuals, hence the map scale maximizing the expectations of MPSD still remains unknown.

Similar to the expectation of MPSD, its variance increases with map scale. This is mainly because any point arrangement looks clustered if it is presented in a small scale. In maps showing less than 50 points, however, the variances of MPSD have their maximum at a scale below 0.9. This arises from the nature of the ANR model rather than by edge effects.

Conclusions

In the present paper, we have analyzed the perception of spatial dispersion in the distribution of point objects. Conducting two experiments, we obtained some useful findings about the relationship between the perception and map characteristics. On the basis of the experiment results, we built regression models representing the average degree of spatial dispersion perceived by map readers. The regression equations enable us to predict the degree of perceived spatial dispersion and provides design guidelines for map authors to communicate the concept of spatial dispersion. Using these models, we finally performed the Monte-Carlo simulations to investigate how the number of points and map scale affect the perceived spatial dispersion.

The major results obtained are summarized as follows.

1) Spatial arrangement of point objects greatly affects the perception of spatial dispersion. The results of Experiment 1 indicate that the size of spatial clusters and the uniformity of point distributions are influential in the perceived spatial dispersion, which is also supported by the high R^2 values of the regression models.
2) The number of point objects affects the perception of spatial dispersion. An increase of point objects raises the probability of point distributions considered to be clustered in the random configuration. It also makes point distributions look very similar in spatial dispersion.
3) Map scale greatly affects the perception of spatial dispersion. An increase of map scale makes point distributions look more dispersed, and this effect is drastic in maps displaying many points.
4) The size of point symbols are not influential, at least in the average perception of spatial dispersion.

The above results are suggestive for the communication of spatial dispersion. To convey the concept of spatial dispersion, we should (i) simplify spatial clusters, (ii) reduce point objects by map generalization, and (iii) shrink map scale. If we do the opposite, map readers will receive the concept of spatial clustering.

Finally we should note some limitations of the present study. In the experiments we used maps containing not more than 50 points. Maps of many points, say, 1000 points, may differ in the structure of perceived spatial dispersion. In addition to this, we did not employ distinctive distributions like uniform distributions, perfectly clustered distributions, and distributions where points follow a constant direction group (good continuation). As seen in the experiments, it is possible that particular characteristics of such distributions affect the perception of spatial dispersion. Subsequent research should consider a greater variety of maps to deepen our understanding of the perception of spatial dispersion and to improve the models obtained from the experiments.

REFERENCES

Ahuja, N. 1982. Dot Pattern Processing Using Voronoi Neighborhoods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-4**: 336-343.

Cressie, N. 1991. *Statistics for Spatial Data*. New York: John Wiley & Sons.

de Lepper, M. J. C, Scholten, H. J. and Stern, R. M. 1995. *The Added Value of Geographical Information Systems in Public and Environmental Health.* Dordrecht: Kluwer.

Elliot, H. M. 1985. Cardinal Place Geometry. *Geographical Analysis* **17 (1)**: 16-35.

Fotheringham, S. and P. Rogerson. 1994. *Spatial Analysis and GIS*. London: Taylor and Francis.

Jenks, G. F. 1973. Visual Integration in Thematic Mapping: Fact or Fiction? *International Yearbook of Cartography* **13**: 27-35.

King, L. J. 1969. *Statistical Analysis in Geography*. Englewood Cliffs: Prentice-Hall.

McCleary, G. F. 1975. In Pursuit of the Map User. *Proceedings. Auto Carto II* (Washington DC, U. S. Bureau of the Census and ACSM): 238-250.

Montgomery, D. C. and E. A. Peck. 1992. *Introduction to Linear Regression Analysis*. New York: John Wiley & Sons.

Okabe, A., B. Boots, and K. Sugihara. 1992. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. New York: John Wiley & Sons.

Pielou, E. C. 1977. *Mathematical Ecology.* New York: John Wiley & Sons.

Ripley, B. D. 1981. *Spatial Statistics.* New York: John Wiley & Sons.

Robinson, A. H., J. L. Morrison, P. C. Muehrcke, A. J. Kimerling, and S. C. Guptill. 1995. *Elements of Cartography*. New York: John Wiley & Sons.

Sadahiro, Y. 1997. Cluster Perception in the Distribution of Point Objects. *Cartographica* **34 (1)**, to appear.

Sugihara, K. and M. Iri. 1983. VORONOI2 Reference Manual - Topology-Oriented Version for the Incremental Method for Constructing Voronoi Diagrams. *Research Memorandum* **RMI 89-04**. Department of Mathematical Engineering and Information Physics, University of Tokyo.

APPENDIX

We consider a rectangular map R whose width and height are X and Y, respectively. Assume that we randomly drop n points $P_1, P_2, ..., P_n$ inside R and draw neighborhood circles C_i (i=1, ..., n) of radius r ($r<X/2, Y/2$) centered at P_i. Let S be the area of $\bigcup_i C_i \cap R$. The expectation of S, denoted by E[S], is calculated as follows.

We define a coordinate system in such a way that the origin is located at the lower-left vertex of R and the axes are parallel to the sides of R, and then divide R into nine subregions $R_1, R_2, ..., R_9$ as shown in Figure A1. We consider a point Q located at (x, y) and denote the probability that Q is covered by the neighborhood circles as $P(x, y)$. The expectation E[S] is then written as

$$\mathrm{E}[S] = \int_0^X \int_0^Y P(x,y)\mathrm{d}y\mathrm{d}x. \quad \text{(A 1)}$$

When Q is located in subregion R_1, the probability that Q is covered by a circle C_i is given by

$$\Pr[Q \in C_i] = 1 - \frac{\pi r^2}{XY}. \quad \text{(A 2)}$$

Hence, the probability $P(x, y)$ is

$$\begin{aligned} P(x,y) &= 1 - \{\Pr[Q \in C_i]\}^n \\ &= 1 - \left(1 - \frac{\pi r^2}{XY}\right)^n \end{aligned}. \quad \text{(A 3)}$$

Similarly, the probability that a point Q in subregion R_2 is covered by a circle C_i is given by

$$\Pr[Q \in C_i] = \frac{1}{XY}\left(\frac{\pi r^2}{2} + x\sqrt{r^2 - x^2} + r^2 \sin^{-1}\frac{x}{r}\right). \quad \text{(A 4)}$$

Using this equation, we have

$$\begin{aligned} P(x,y) &= 1 - \{\Pr[Q \in C_i]\}^n \\ &= 1 - \left\{1 - \frac{1}{XY}\left(\frac{\pi r^2}{2} + x\sqrt{r^2 - x^2} + r^2 \sin^{-1}\frac{x}{r}\right)\right\}^n \end{aligned}. \quad \text{(A 5)}$$

When Q is located in subregion R_3, $P(x, y)$ is

$$P(x,y) = 1 - \left\{1 - \frac{1}{XY}\left(\frac{\pi r^2}{2} + y\sqrt{r^2 - y^2} + r^2 \sin^{-1}\frac{y}{r}\right)\right\}^n. \quad \text{(A 6)}$$

Replacing (x, y) by (X-x, Y-y) in Equations (A 4) and (A 5), we obtain $P(x, y)$'s when Q is located in subregions R_4 and R_5, respectively.

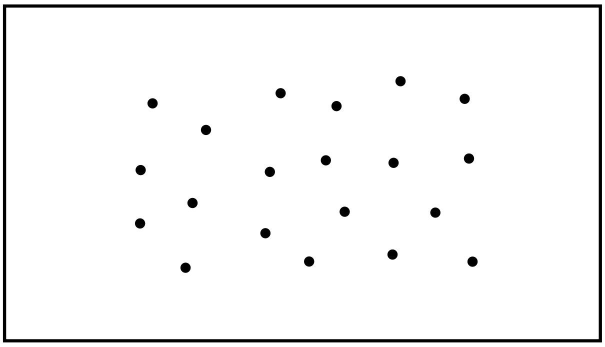
For a point Q in subregion R_6, $P(x, y)$ is given by

$$P(x,y) = 1 - \left\{1 - \frac{1}{XY}\left(\frac{\pi r^2}{4} + x\sqrt{r^2 - x^2} + r^2 \sin^{-1}\frac{x}{r} + y\sqrt{r^2 - y^2} + r^2 \sin^{-1}\frac{y}{r}\right)\right\}^n. \quad \text{(A 7)}$$

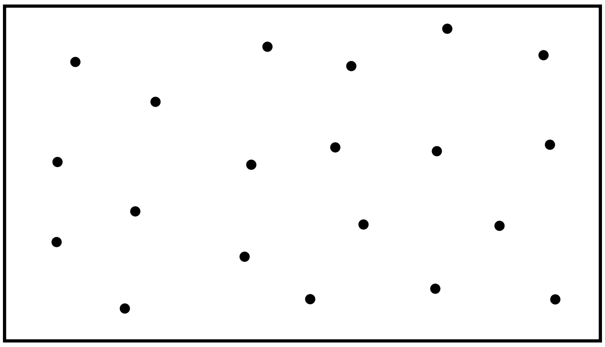
The substitution of Equations (A 3), (A 5), (A 6) and (A 7) into (A 1) yields

$$\begin{aligned} \mathrm{E}[S] &= (X-2r)(Y-2r)\left\{1-\left(1-\frac{\pi r^2}{XY}\right)^n\right\} \\ &+2(X+Y-4r)\left[r-\int_0^r\left\{1-\frac{1}{XY}\left(\frac{\pi r^2}{2}+x\sqrt{r^2-x^2}+r^2\sin^{-1}\frac{x}{r}\right)\right\}^n \mathrm{d}x\right] \\ &+4\left[r^2-\int_0^r\int_0^r\left\{1-\frac{1}{XY}\left(\frac{\pi r^2}{4}+\frac{x}{2}\sqrt{r^2-x^2}+\frac{r^2}{2}\sin^{-1}\frac{x}{r}+\frac{y}{2}\sqrt{r^2-y^2}+\frac{r^2}{2}\sin^{-1}\frac{y}{r}+xy\right)\right\}^n \mathrm{d}x\mathrm{d}y\right] \end{aligned} \quad . \quad \text{(A 8)}$$

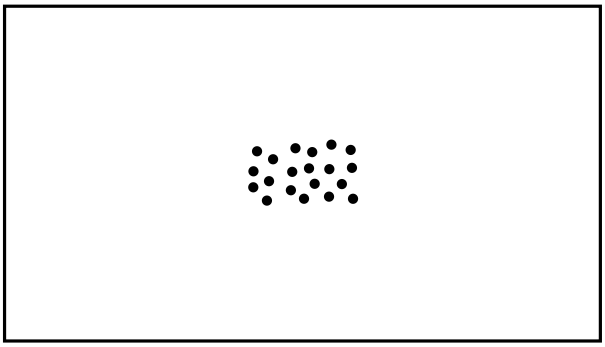
Using numerical integration, we can compute E[S] by Equation (A 8).



(a)



(b)



(c)

Figure 1

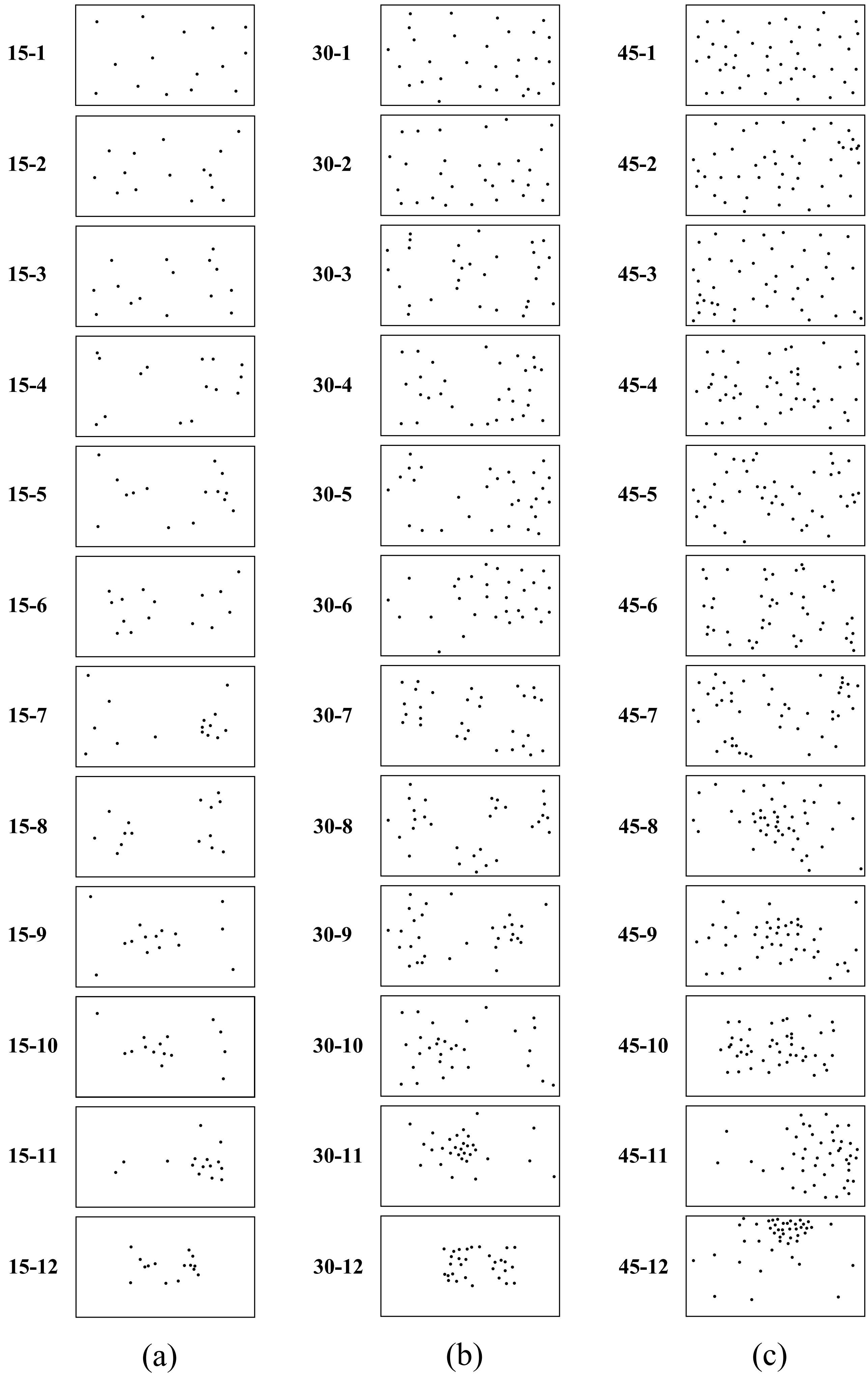
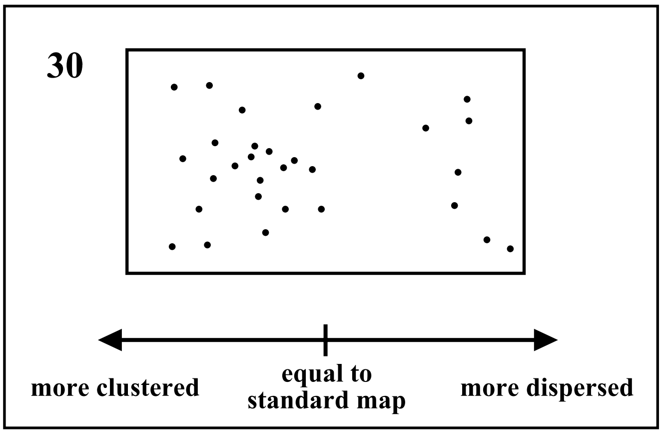
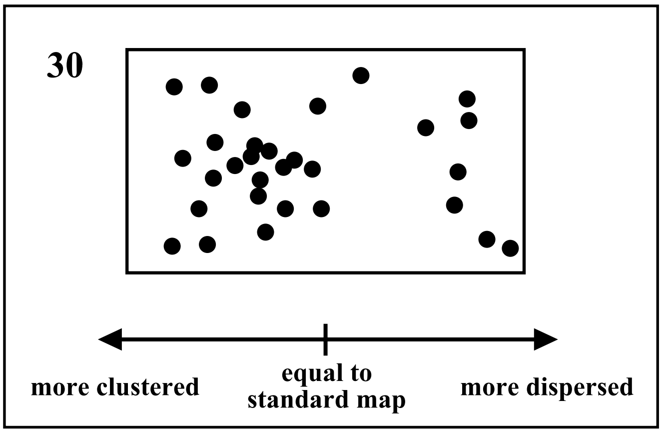

15-1
15-2
15-3
15-4
15-5
15-6
15-7
15-8
15-9
15-10
15-11
15-12
30-1
30-2
30-3
30-4
30-5
30-6
30-7
30-8
30-9
30-10
30-11
30-12
45-1
45-2
45-3
45-4
45-5
45-6
45-7
45-8
45-9
45-10
45-11
45-12
(a)
(b)
(c)

Figure 2

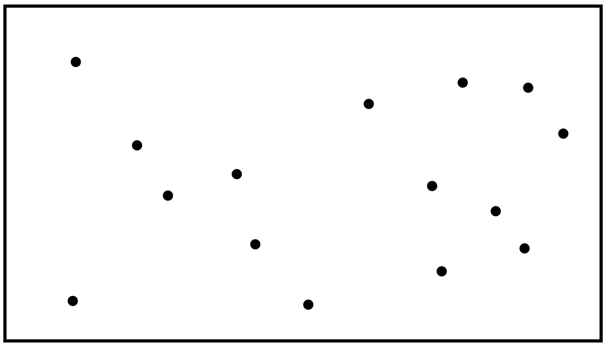

30
more clustered
equal to
standard map
more dispersed

(a)

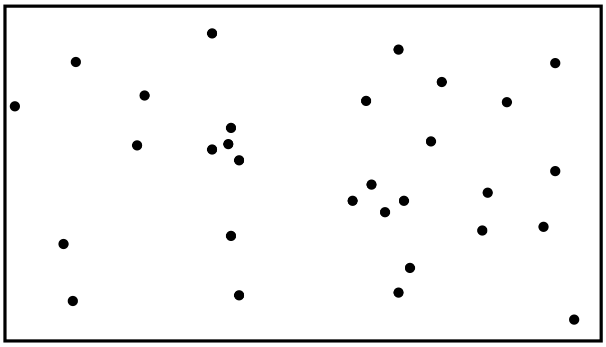

30
more clustered
equal to
standard map
more dispersed

(b)

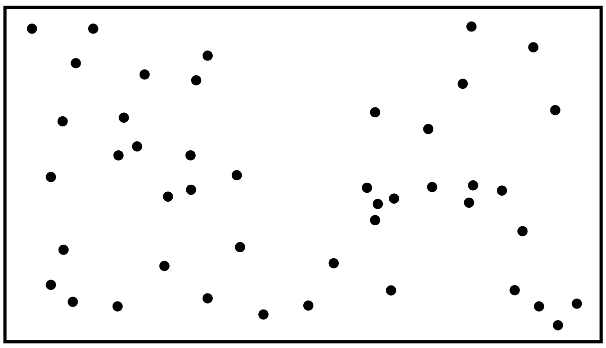
Figure 3



(a)



(b)



(c)

Figure 4

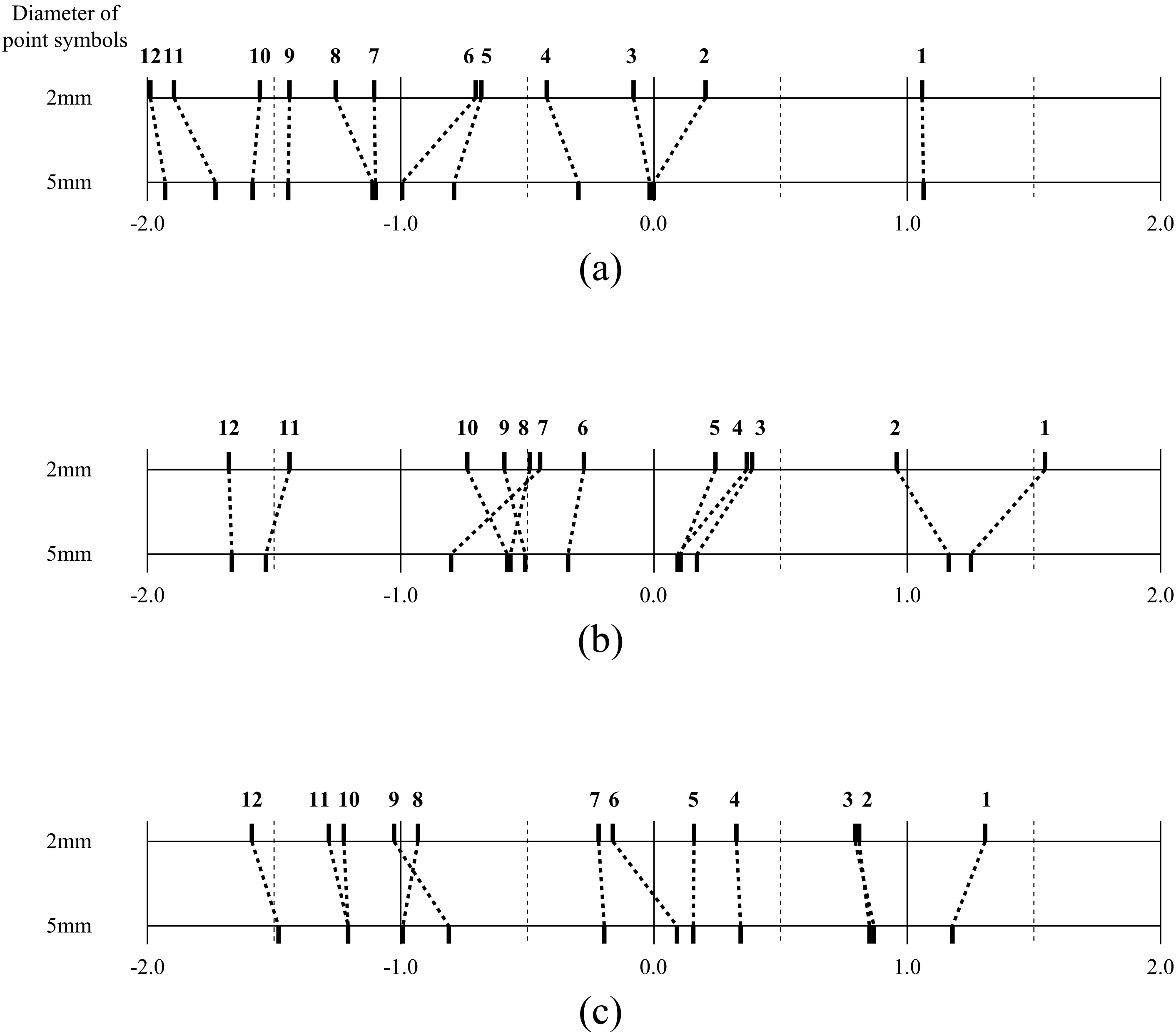
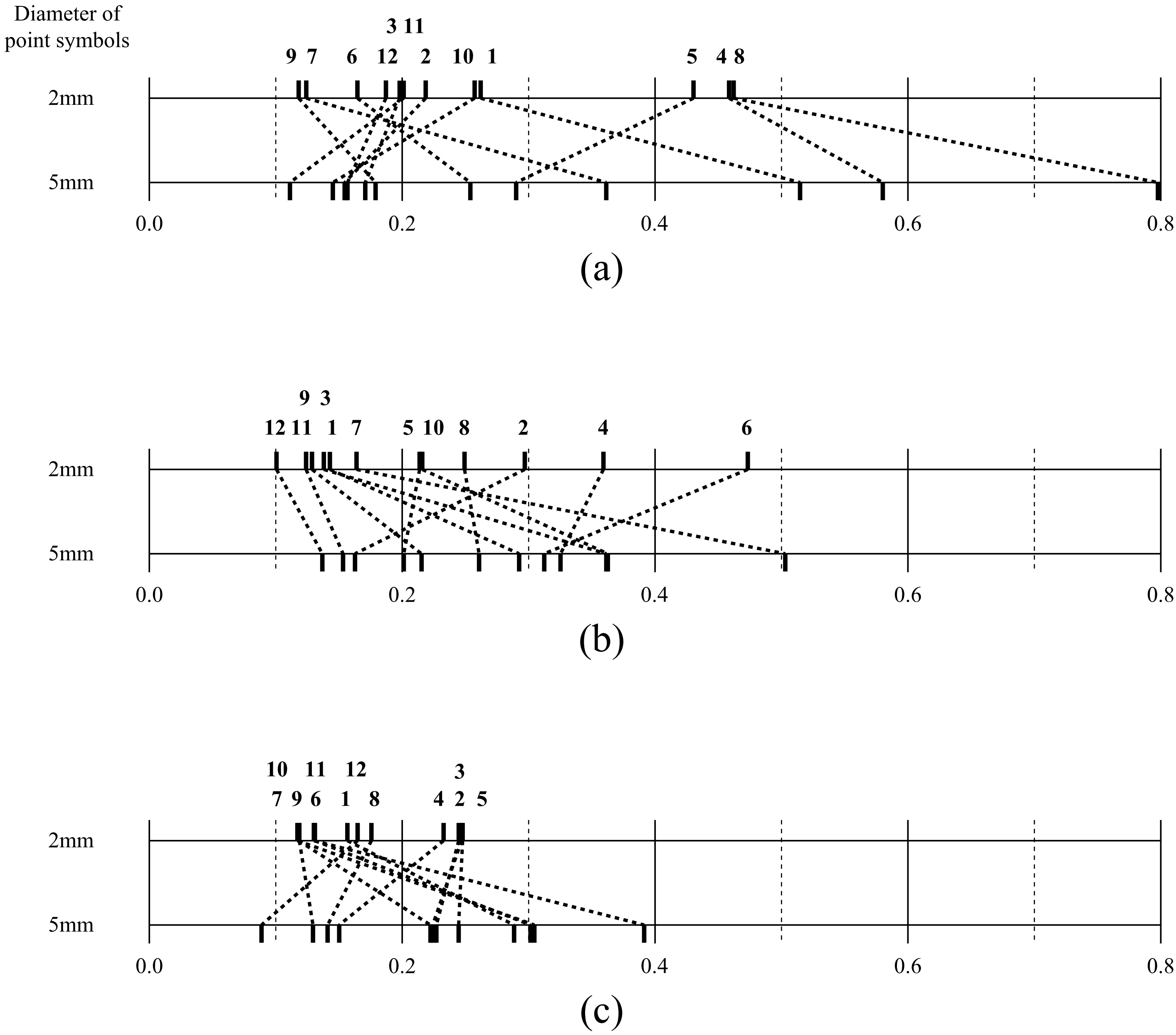

Diameter of
point symbols
2mm
5mm
12 11 10 9 8 7 6 5 4 3 2 1
-2.0 -1.0 0.0 1.0 2.0
(a)
12 11 10 9 8 7 6 5 4 3 2 1
(b)
12 11 10 9 8 7 6 5 4 3 2 1
(c)

Figure 5



Diameter of point symbols
3 11
9 7 6 12 2 10 1 5 4 8
2mm
5mm
0.0 0.2 0.4 0.6 0.8
(a)
9 3
12 11 1 7 5 10 8 2 4 6
2mm
5mm
0.0 0.2 0.4 0.6 0.8
(b)
10 11 12 3
7 9 6 1 8 4 2 5
2mm
5mm
0.0 0.2 0.4 0.6 0.8
(c)

Figure 6

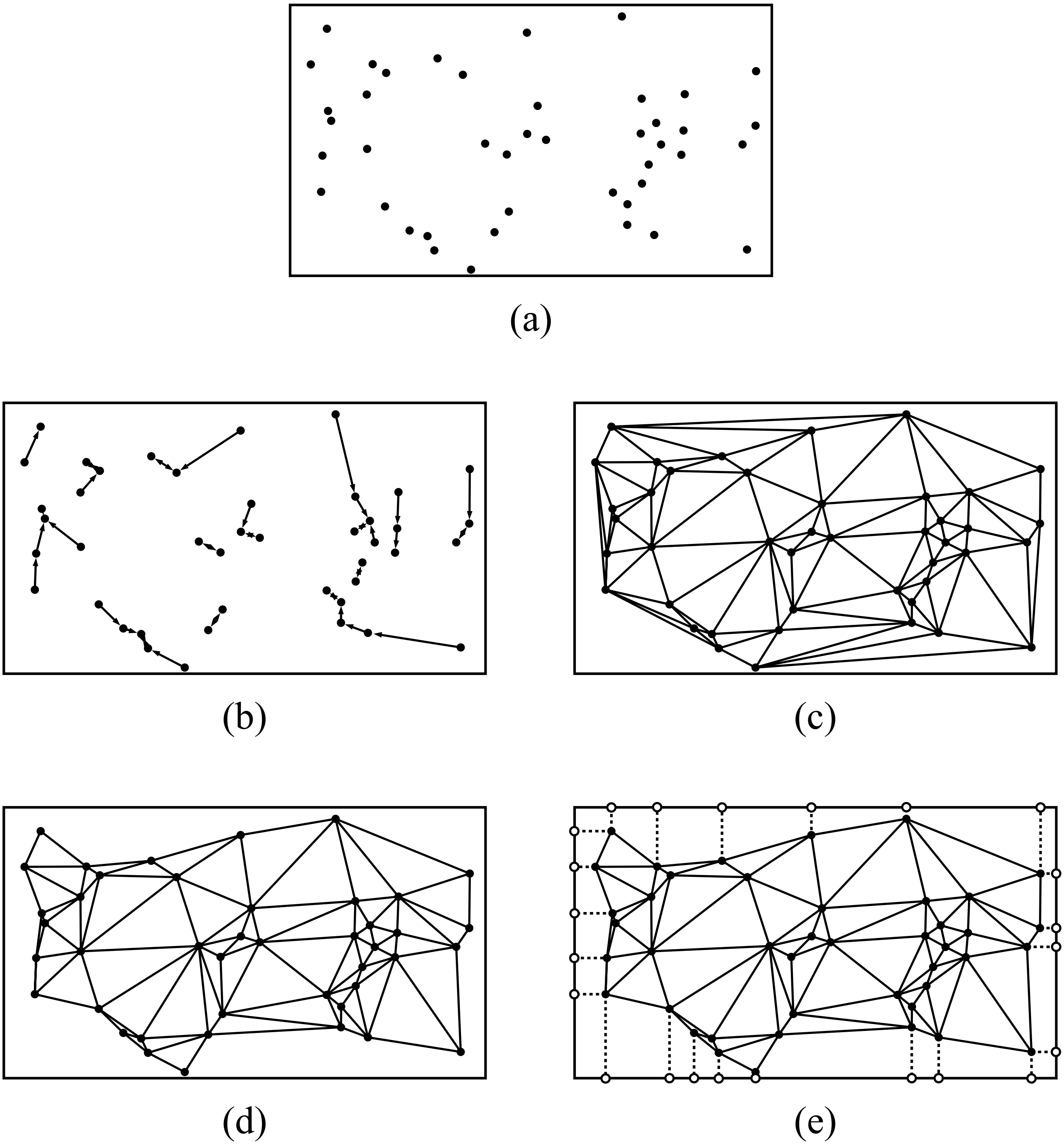
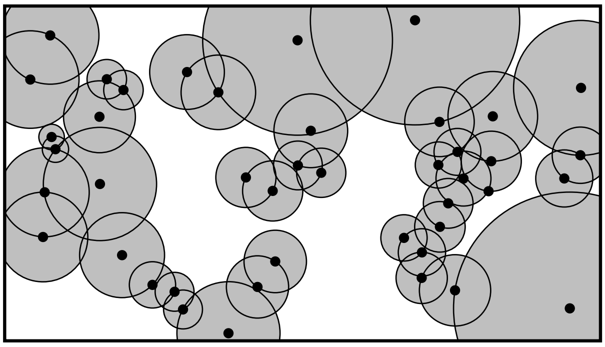
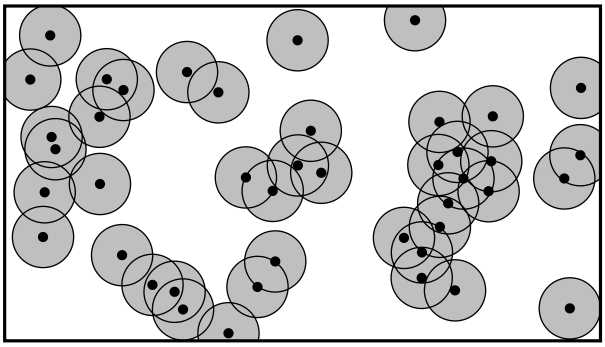


Figure 7



(a)



(b)

Figure 8

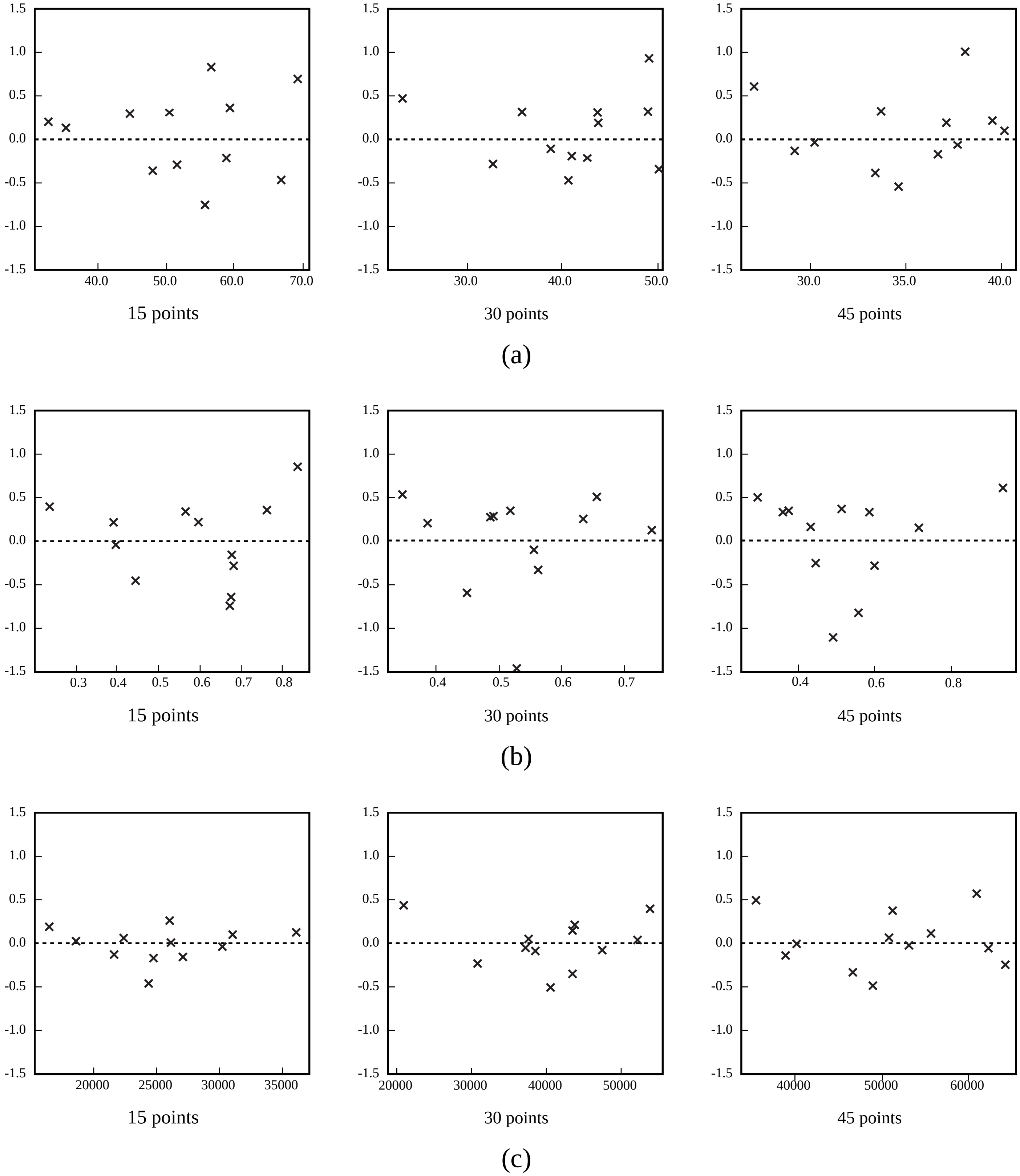

15 points
30 points
45 points
(a)
15 points
30 points
45 points
(b)
15 points
30 points
45 points
(c)

Figure 9

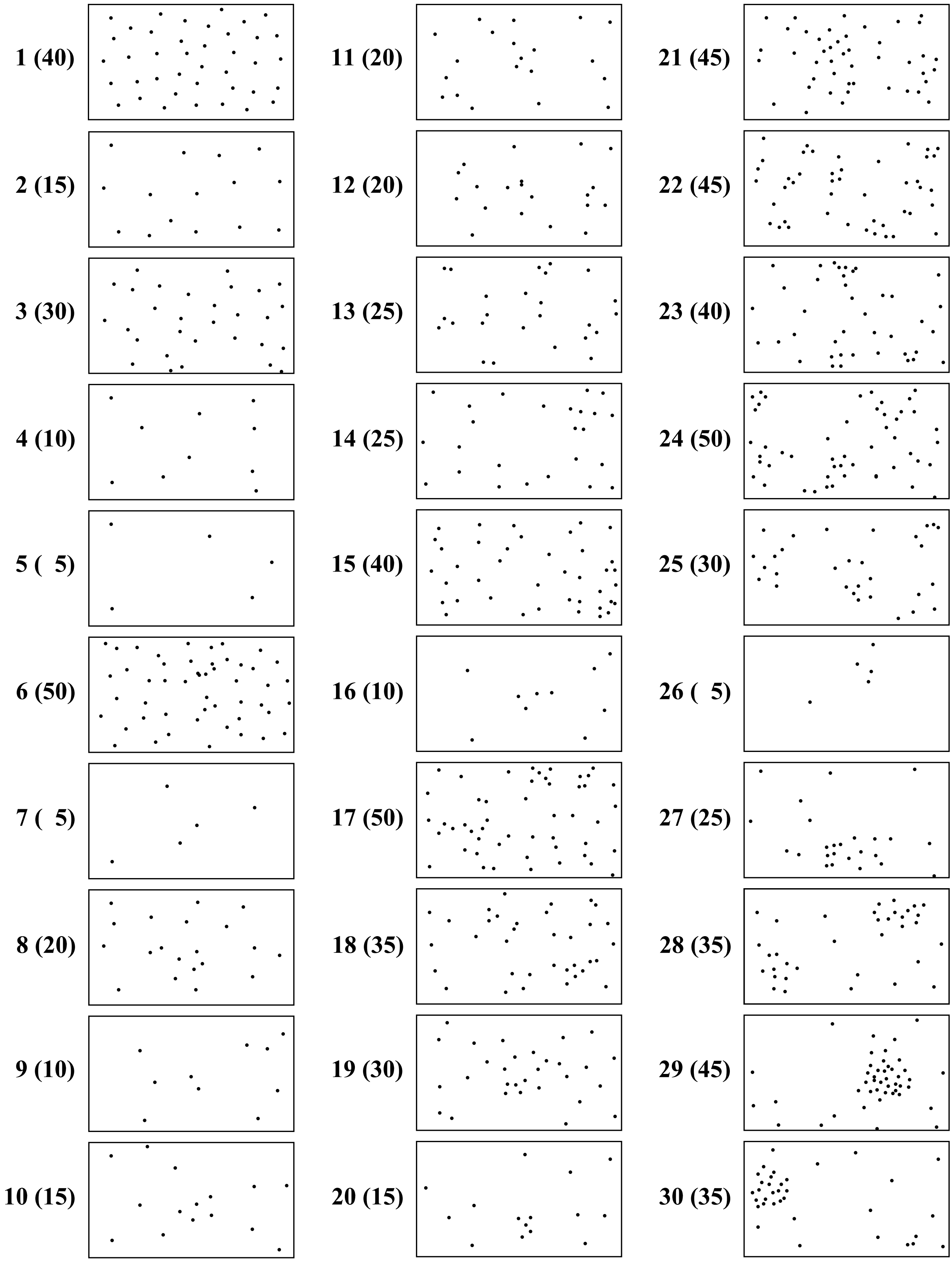

1 (40)
2 (15)
3 (30)
4 (10)
5 (5)
6 (50)
7 (5)
8 (20)
9 (10)
10 (15)
11 (20)
12 (20)
13 (25)
14 (25)
15 (40)
16 (10)
17 (50)
18 (35)
19 (30)
20 (15)
21 (45)
22 (45)
23 (40)
24 (50)
25 (30)
26 (5)
27 (25)
28 (35)
29 (45)
30 (35)

Figure 10

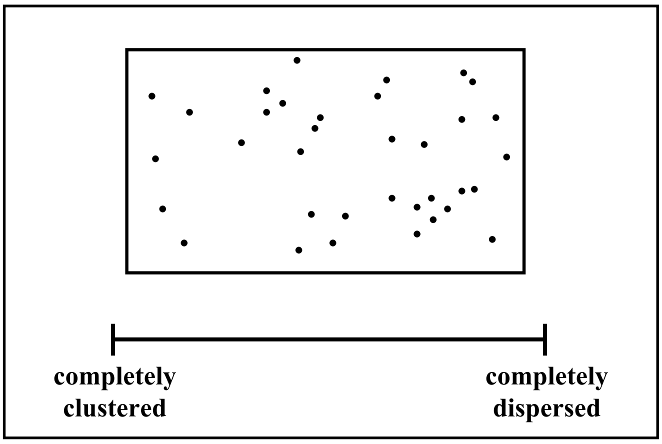
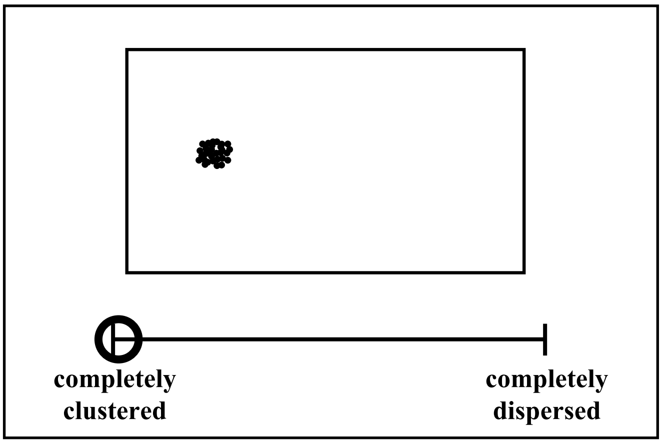
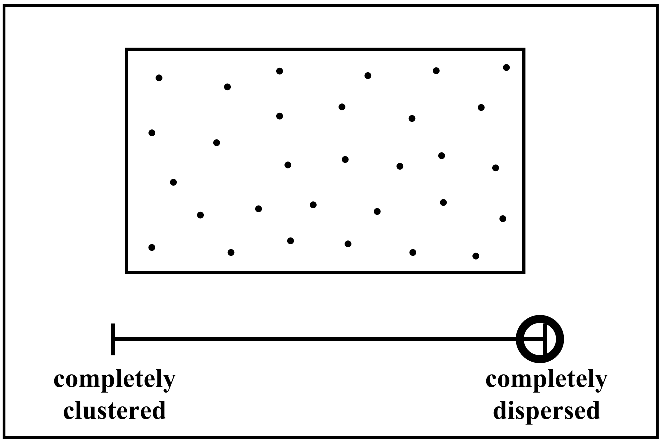


Figure 11

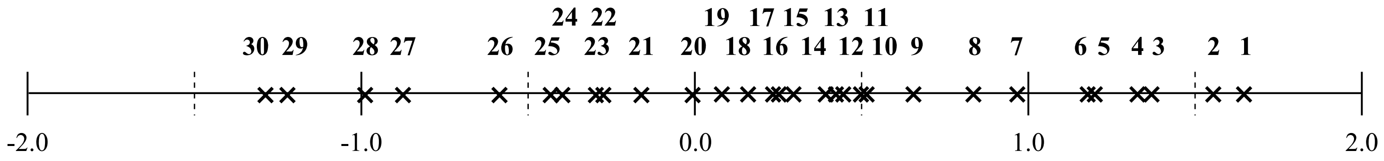

completely
clustered
completely
dispersed

(a)


completely
clustered
completely
dispersed

(b)

Figure 12



24 22
19 17 15 13 11
30 29 28 27 26 25 23 21 20 18 16 14 12 10 9 8 7 6 5 4 3 2 1
-2.0
-1.0
0.0
1.0
2.0

Figure 13

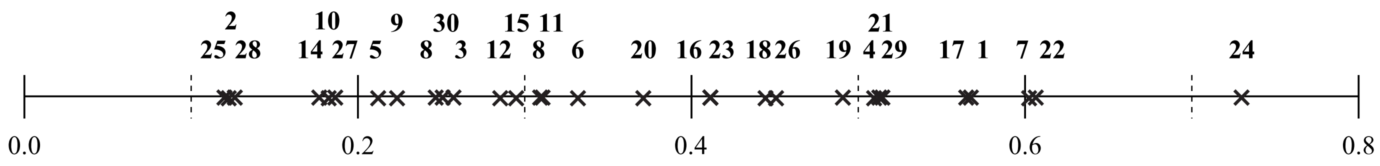
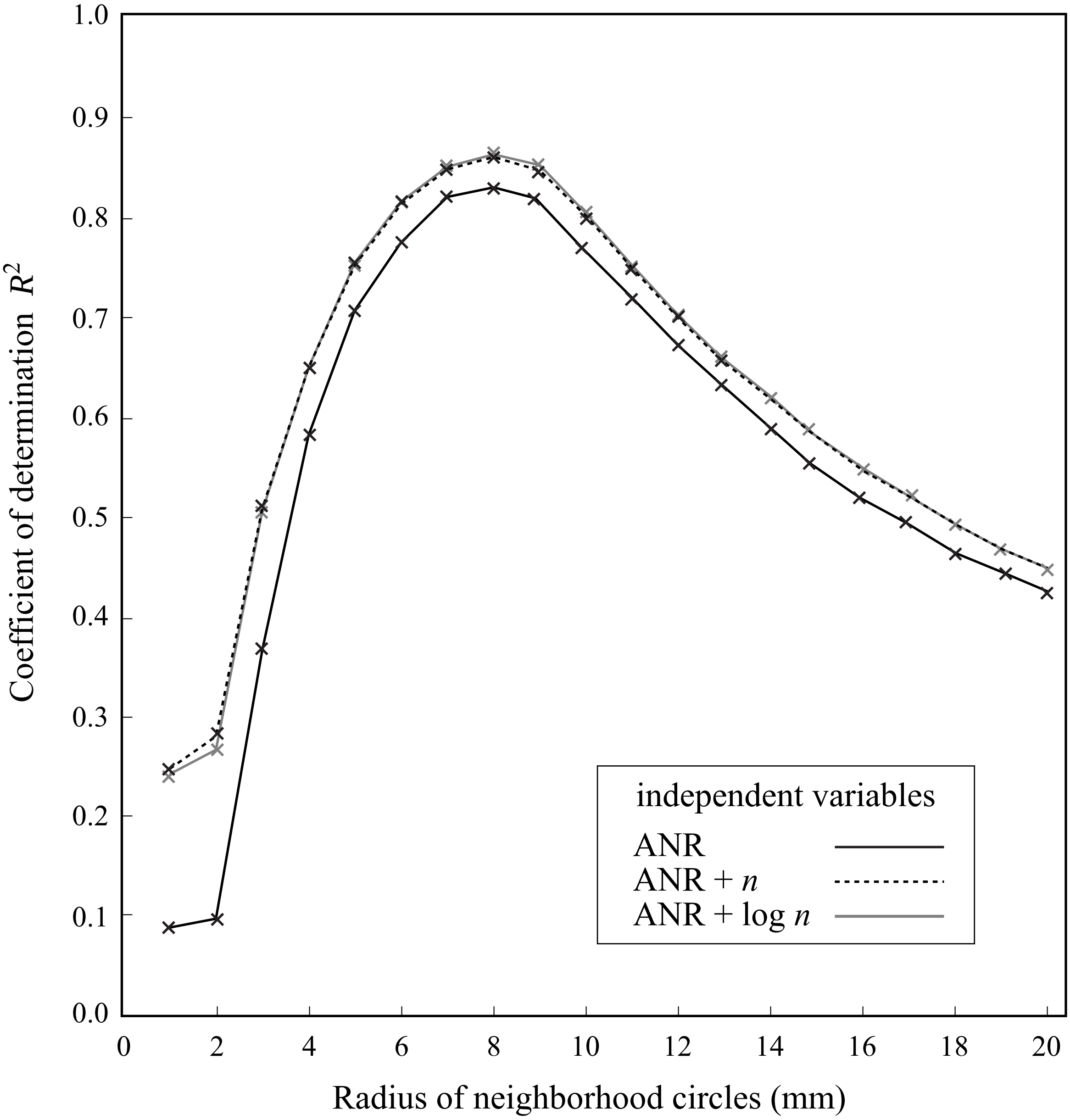

2
25 28
10
14 27 5
9
30
8 3
15 11
12 8 6
20
16 23
18 26
19
21
4 29
17 1
7 22
24
0.0
0.2
0.4
0.6
0.8

Figure 14



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Coefficient of determination R^2
0
2
4
6
8
10
12
14
16
18
20
Radius of neighborhood circles (mm)
independent variables
ANR
ANR + n
ANR + log n

Figure 15

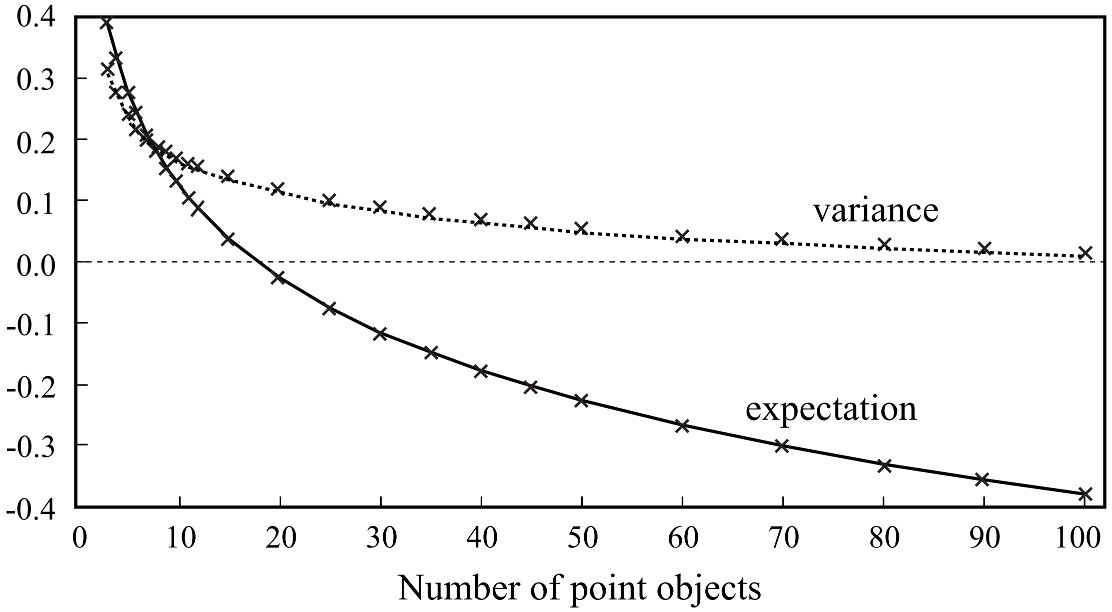

0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
0
10
20
30
40
50
60
70
80
90
100
variance
expectation
Number of point objects

Figure 16

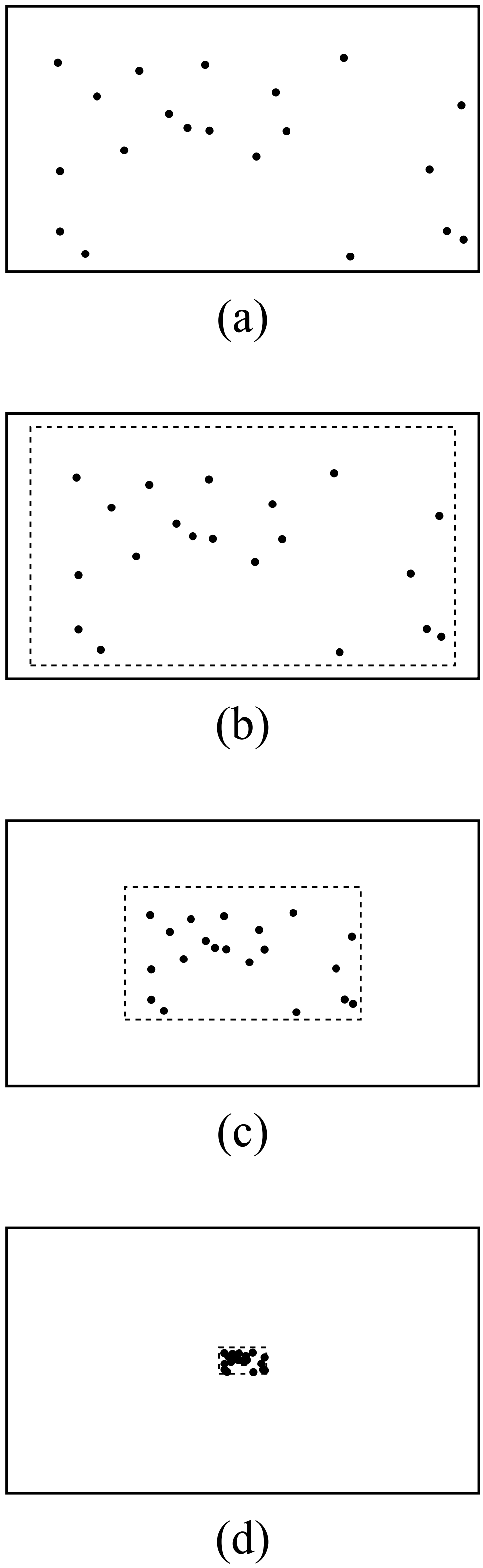


Figure 17

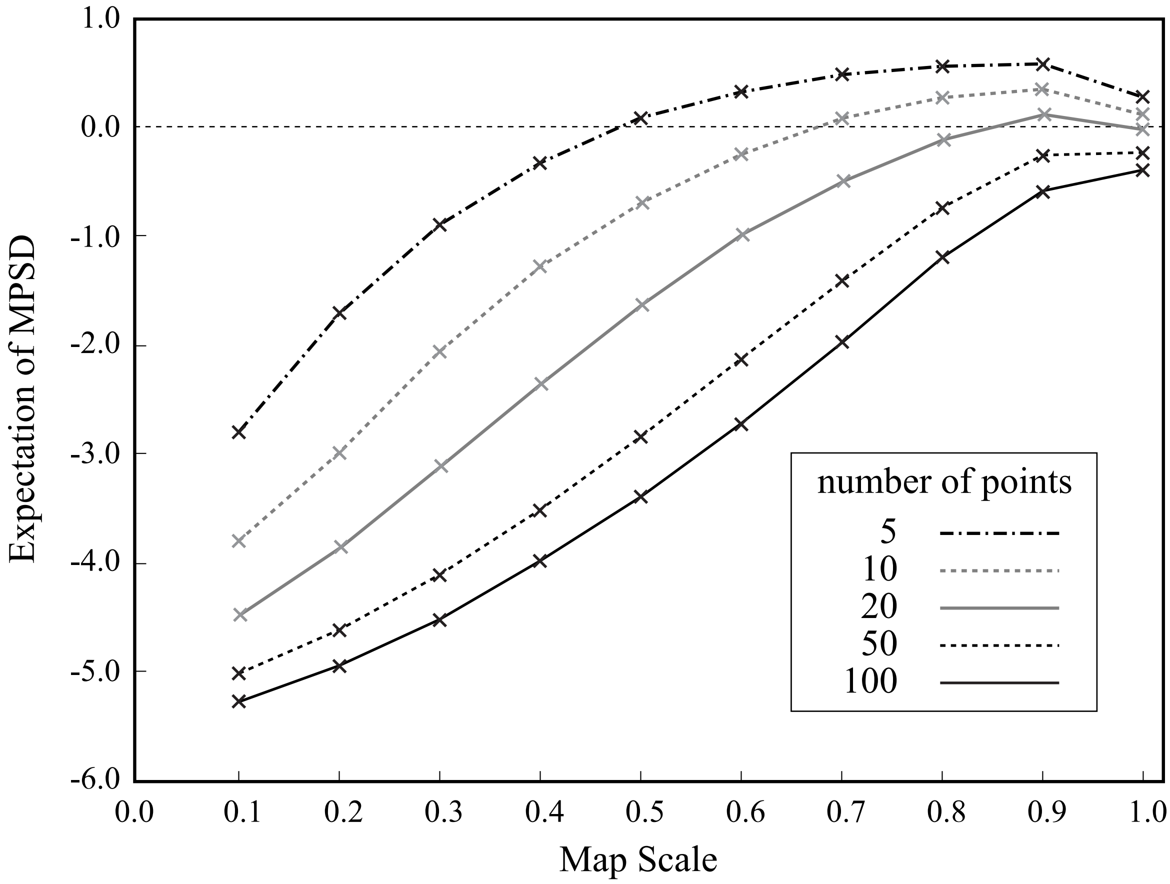
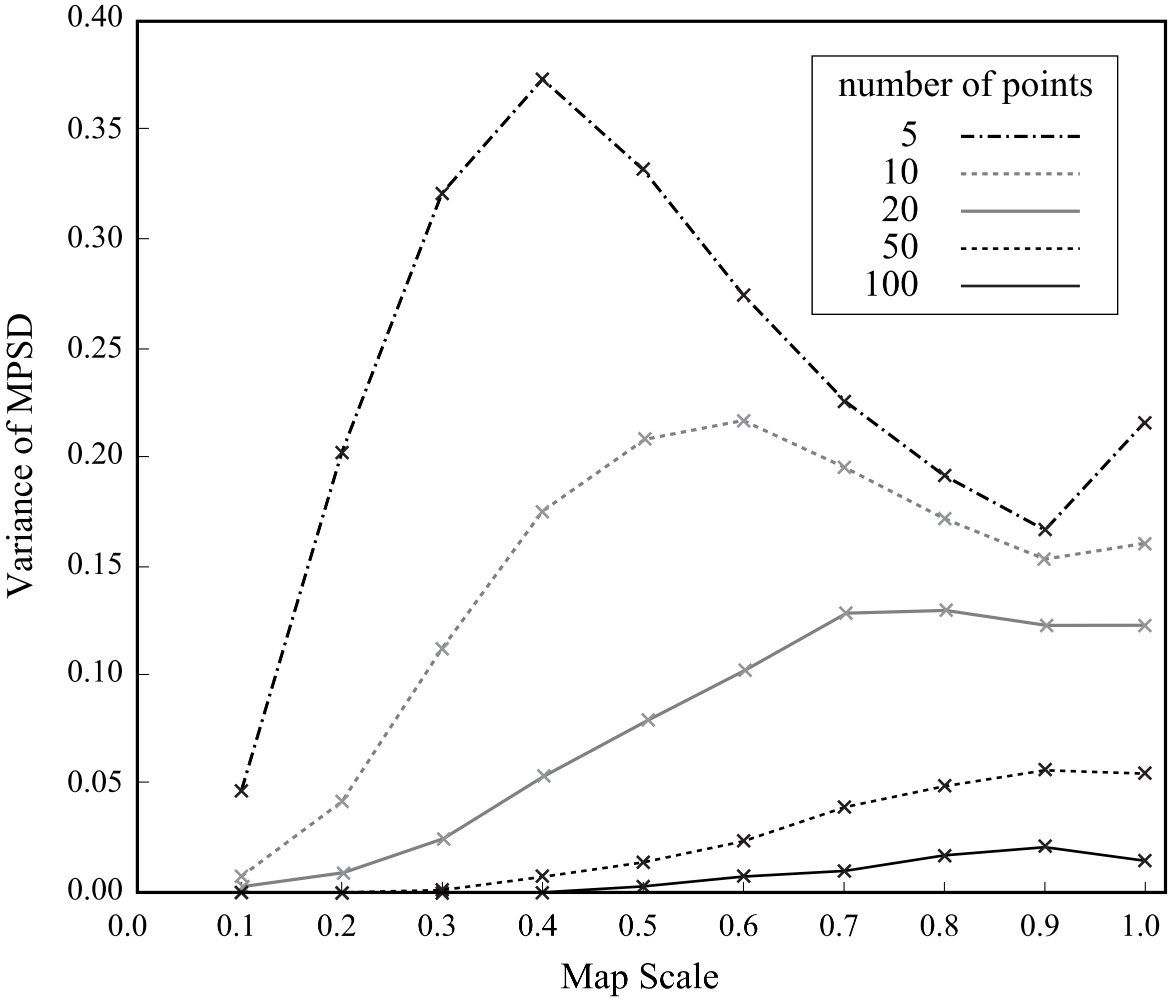

1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
Expectation of MPSD
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Map Scale
number of points
5
10
20
50
100

Figure 18



Variance of MPSD
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Map Scale
number of points
5
10
20
50
100

Figure 19

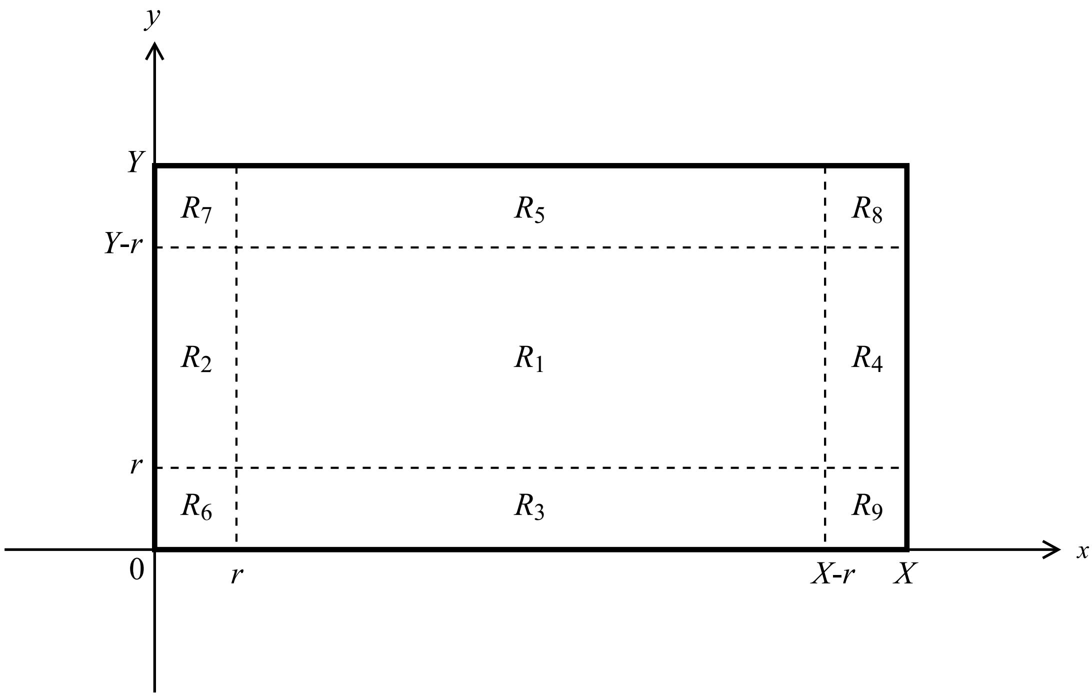


Figure A1

nearest neighbor method

number of points		μ	σ^2	σ	σ^2/μ	σ/μ
15	R^2	0.676	0.104	0.077	0.003	0.162
	F-value	20.86*	1.16	0.83	0.03	1.94
30	R^2	0.650	0.004	0.027	0.019	0.081
	F-value	15.54*	0.04	0.24	0.17	0.78
45	R^2	0.091	0.301	0.231	0.176	0.071
	F-value	0.89	3.79	2.65	1.90	0.68

Delaunay neighbor method

number of points		μ	σ^2	σ	σ^2/μ	σ/μ
15	R^2	0.416	0.129	0.106	0.340	0.486
	F-value	7.12*	1.48	1.19	5.14*	7.63*
30	R^2	0.564	0.000	0.007	0.042	0.364
	F-value	12.90*	0.00	0.07	0.44	5.70*
45	R^2	0.412	0.282	0.030	0.209	0.523
	F-value	6.93*	0.29	0.30	2.64	10.94*

trimmed Delaunay neighbor method

number of points		μ	σ^2	σ	σ^2/μ	σ/μ
15	R^2	0.650	0.099	0.138	0.285	0.719
	F-value	18.49*	1.10	1.61	3.99	25.61*
30	R^2	0.773	0.015	0.003	0.209	0.605
	F-value	33.94*	0.15	0.03	2.63	15.31*
45	R^2	0.764	0.323	0.335	0.446	0.640
	F-value	32.44*	4.77	5.04*	8.06*	17.80*

extended Delaunay neighbor method

number of points		μ	σ^2	σ	σ^2/μ	σ/μ
15	R^2	0.094	0.229	0.180	0.164	0.078
	F-value	1.04	2.97	2.20	1.96	0.85
30	R^2	0.035	0.260	0.278	0.251	0.224
	F-value	0.36	3.51	3.85	3.36	2.88
45	R^2	0.006	0.052	0.059	0.045	0.036
	F-value	0.06	0.55	0.62	0.47	0.37

Table 1a

number of points		nearest neighbor radius method	fixed radius method 5mm	10mm	15mm	20mm
15	R^2	0.108	0.706	0.949	0.869	0.741
	F-value	1.22	23.97*	95.51*	66.12*	28.77*
30	R^2	0.184	0.766	0.897	0.824	0.743
	F-value	2.25	26.59*	67.06*	37.63*	23.71*
45	R^2	0.003	0.806	0.914	0.839	0.776
	F-value	0.03	33.59*	81.96*	41.56*	28.20*

Table 1b

variables	R^2	F-value	t-value
MD_i	0.237	8.709*	2.951*
MD_i n_i	0.320	6.365*	2.895* -1.816
MD_i $\log n_i$	0.324	6.485*	2.970* -1.866
$RSDD_i$	0.841	147.43*	-12.138*
$RSDD_i$ n_i	0.876	96.25*	-12.978* -2.833
$RSDD_i$ $\log n_i$	0.891	111.43*	-14.027* -3.586
ANR_i(8mm)	0.865	179.02*	13.384*
ANR_i(8mm) n_i	0.895	114.27*	14.176* -2.755
ANR_i(8mm) $\log n_i$	0.903	123.60*	14.804* -3.181

Table 2