SIMULATION STUDIES OF GEOMETRIC ANISOTROPY DETECTION METHODS FOR GEOSTATISTICAL DATA

SEPTEMBER 2012

Takafumi KUBOTA

Contents

1 Introduction **4**

2 Geostatistical Data **8**

2.1 What Is Geostatistical Data? 8

2.2 R Packages and Their Data Sets 9

2.3 Environmental Data 13

2.4 Omni-directional Variogram 15

2.5 Directional Variogram and Exploratory GA detection 16

2.6 Visualizing Application 18

3 Parameter Estimation of Variogram **20**

3.1 Cutoff Value 20

3.2 Variogram Cloud 21

3.3 Empirical Variogram 23

3.4 Theoretical Variogram 24

3.5 Geometric Anisotropy 26

3.6 Visualizing Application 27

4 Detection and Correction of Geometric Anisotropy **31**

4.1 Geometric Anisotropy Detection Method 31

4.1.1 Directional Variogram and Least-squares Method for Fitting to An Ellipse 31

4.1.2 Likelihood-based Estimation 34

4.2 Geometric Anisotropy Correction Method 35

5 Simulation Study **37**

5.1 Calculations 37

5.2 Parallel computing 41
5.3 Parameters 42
5.4 Results 43

6 Concluding Remarks **50**

A Theoretical Variograms **52**

B CRAN Task Views **55**

List of Figures

1 Plots of the meuse data 10
2 Plots of the scallop data 12
3 Plots of the Okayama water quality data 14
4 Omni-directional variogram with theoretical variogram of meuse data. 15
5 Four directional variograms of the meuse data. 16
6 Exploratory GA detection 17
7 Observation points on the Google Maps 19
8 Typical variogram model 22
9 Variogram models of the areal environmental data 25
10 The example of rose diagram 26
11 Visualizing application (variogram cloud) 28
12 Visualizing application (conection between observed points) . 29
13 Visualizing application (empirical and theoretical variogram) . 30
14 An example of fitting to an ellipse 33
15 An example of correction method 36
16 Random point (left) and constricted square (right). 39

17 Typical theoretical variograms 54

List of Tables

1 $CV^{*(t)}$ values in $d = 0.5$ 44
2 $CV^{*(t)}$ values in $d = 1.0$ 45
3 $CV^{*(t)}$ values in $d = 2.0$ 46
4 Frequencies of minimum of $CV_s^{(t)}$ values in $d = 0.5$ 49
5 Frequencies of minimum of $CV_s^{(t)}$ values in $d = 1.0$ 49
6 Frequencies of minimum of $CV_s^{(t)}$ values in $d = 2.0$ 49

1 Introduction

A common theme in geostatistical analysis is to detect a spatial structure of observed data. Using the estimated spatial structure, data at unobserved points can be predicted. A well-known prediction methods is kriging. To apply kriging to geostatistical data, second-order stationarity is required. Second-order stationarity shows the directional homogeneity of spatial autocorrelation, including mean and variance. Therefore, it is very important to determine whether the spatial autocorrelation for each direction is homogeneous.

One of the criteria for spatial autocorrelation is the variogram, a model that ordinarily has three parameters: nugget effect, sill and range values. Zimmerman (1993) introduced three kinds of anisotropy: nugget effect anisotropy, sill anisotropy and range anisotropy. Geostatistical analysis is often applied in fields such as environmental science, soil science and water quality science. Cheng et al. (2000) applied an anisotropic spatial modeling approach of image rectification in the field of remote sensing.

However, almost all applications assume isotropy, not anisotropy. Geometric anisotropy (GA) is another way of expressing the range anisotropy, because GA data have different range values in different directions. The derivation and the properties of GA were given by Chilès and Delfiner (1999) and Wackernagel (1995). It was derived from isotropy to GA by axis rotation and extension of one axis.

In previous studies, Journel and Froidevaux (1982) presented a case study in which anisotropic data were applied to the hole-effect model. Guan et al. (2006) proposed a formal nonparametric approach to test for isotropy.

Since GA is important for geostatistics, this paper focused on only GA, not nugget effect anisotropy nor zonal anisotropy (sill value anisotropy). Be-

cause nugget effect is jump value of zero distance, nugget effect anisotropy is not theoretically existent. The zonal anisotropy is combined of isotropic model plus GA. The zonal anisotropy is sometimes resolved by decreasing directional trend.

Kubota and Tarumi (2007) applied environmental data in Okayama prefecture to detect GA interactively by using the web system. The system was applied GoogleMapsAPI and SVG to visualize GA.

In another previous study that focused on the estimation of the parameters of GA, Kubota and Tarumi (2008a) proposed fitting an ellipse to points whose coordinates are calculated by the range values of the directional variograms in all directions. In that paper, the ellipse parameters were estimated as follows: calculation of directional variograms, parameter estimation of each directional variogram and parameter estimation of the ellipse model.

There are many kinds of parameters which define geostatistical data through a process of generating data and estimating GA: the density of data, the ratio of anisotropy and the number of directions for directional variograms. Kubota and Tarumi (2008b) produced data from a simulated random point pattern, which included several kinds of densities in random fields, in order to discuss the relation between these densities and the number of directions. In that paper, we used the *RandomFields* R package to simulate the anisotropic data. Four parameters were used for estimating range values: two kinds of anisotropic parameters (the ratio of the semi-major and semi-minor axes, and the angle of the semi-major axis), data densities (the number of points per unit area), and the number of directions.

Kubota and Tarumi (2009) focused on two parameters: data densities and anisotropic ratio. We also introduced a statistical test to clarify the geostatistical predictions of isotropic assumptions and the correction of GA,

detected by Kubota and Tarumi (2008a). Only in the case of large sample size and strong anisotropic ratio, the correction of anisotropy showed good results.

On the other hand, another detection method uses likelihood in estimating anisotropic parameters. Kubota and Tarumi (2010) compared this likelihood-based detection method with that of Kubota and Tarumi (2008a), which used the environmental data of water quality in Okayama. Cross validation was used to compare the two detection methods. The prediction criteria were as follows: mean square error of cross validation and final prediction error. The number of directions, 2, 4, 8 and 16, was used to compare the results of the cross validations. The result of the paper showed that the prediction method of Kubota and Tarumi (2008a) was better than the likelihood-based model.

In this paper, we apply three kinds of detection methods: the four-directional restriction of Kubota and Tarumi (2008a), the maximum likelihood method and the method that assumes isotropy, after generating several kinds of geostatistical data having GA. The objective of this simulation study is to confirm whether the proposed method of Kubota and Tarumi (2008a) is better than the maximum likelihood method or the method assuming isotropy. We compensated for unconfirmed data areas other than those in the practical case study of Okayama water quality data (Kubota and Tarumi, 2010).

We describe the geostatistical data that were used in this study in Section 2. Then, we describe the estimation of variogram parameters with the assumption of isotropy in Section 3. Also, visualizing applications are also explained in both Section 2 and 3. Meanwhile, we describe the estimation and the correction of the GA parameters in Section 4 that is the proposed

method in this paper. Next, we describe the simulation study, including generating data from a random point pattern which has GA, and the method of cross validation in Section 5 that is the main simulation part of this study. Finally, we offer our concluding remarks in Section 6.

2 Geostatistical Data

In spatial data, there are three kinds of data type: geostatistical data, lattice data and point pattern data. In this paper, geostatistical data are used to detect and correct geometric anisotropy.

In this section, we explain geostatistical data. First, we explain what is geostatistical data in Section 2.1. Next, R packages and related data sets with geostatistical data are briefly introduced in Section 2.2, and environmental data of geostatistical data is also introduced to be applied to R system in Section 2.3. Because we focused on the GA, we explain what are two kinds of variogram; omni-directional variogram and directional variogram in Section 2.4 and 2.5 respectively. By using real geostatistical data, the exploratory GA detection is also explained in Section 2.5. Finally, visualizing application of geostatistical data is explained in Section 2.6.

2.1 What Is Geostatistical Data?

Geostatistical data consist of two kinds of information; spatial reference and attribute. Spatial reference has coordinate values and system of reference for these coordinates, for example pairs of longitude and latitude degree values and pairs of horizontal and vertical distances from determined central point.

Attribute is the characteristic values at the observed point, for example the number of scallops on a coast, pH values in a watershed area, and the densities of CO_2 (carbon dioxide) or $SO_{\mathbf{x}}$ (sulphur dioxide or sulphur trioxide) near large roads

2.2 R Packages and Their Data Sets

The *R* system is a free software environment for statistical computing and graphics. There are a lot of expanded packages that analyze geostatistical data, for example *gstat*, *geoR* and so on. They include geostatistical data. For example, meuse of gstat is the data set that gives locations and topsoil heavy metal concentrations of Meuse river (collected by Ruud van Rijn and Mathieu Rikken; data were compiled for R by Edzer Pebesma (Burrough and McDonnell, 1998)).

Some *R* packages have special functions to analyze geostatistical data. The *gstat* provides a wide range of functions for univariate and multivariate geostatistics, while the *geoR* contains functions for model-based geostatistics.

By using plot function of the *geoR*, figure 1 shows the four plots of the attribute; logarithm of topsoil zinc concentration (mg/kg soil); the top left figure shows the observation points with four classes, the top right figure shows the attribute versus Y coordinate with lowess line, the bottom left figure shows the attribute versus X coordinate with lowess line and the bottom right figure shows the histogram of the attribute.

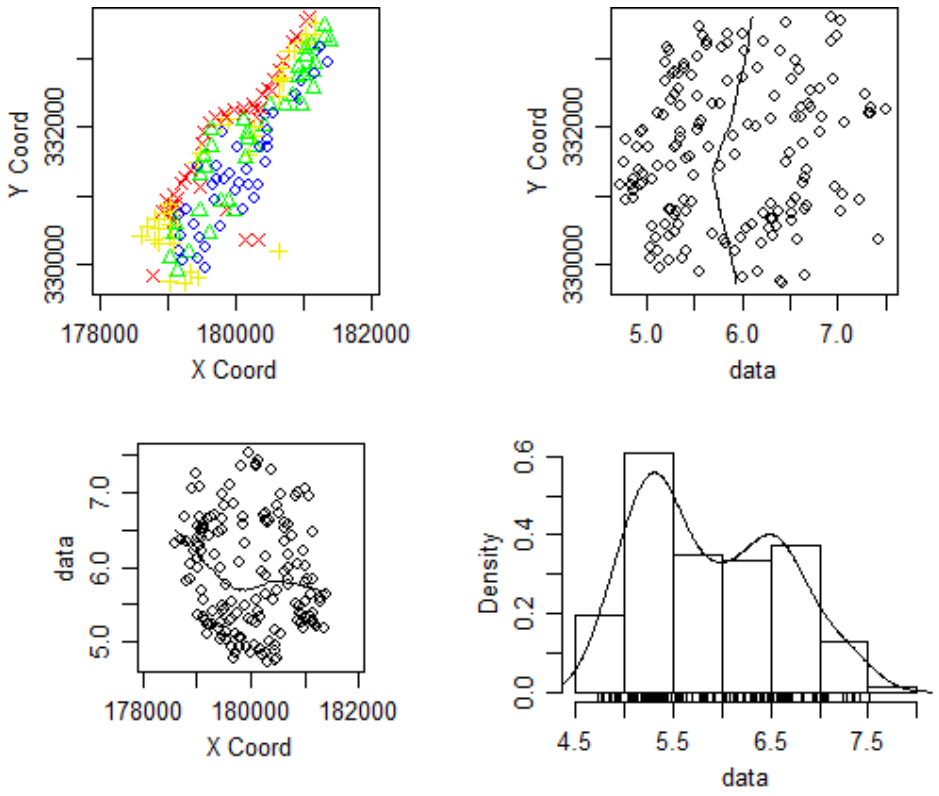

Figure 1: Plots of the meuse data; observation points(top left), latitude v.s. attribute (top right), longitude v.s. attribute (bottom left) and histogram (bottom right) of meuse data.

Even in the other *R* packages, which focus on other functions, there are also many data sets of geostatistical data. For example, the *SemiPar* package provides the *scallop* data frame that has triplets concerning scallop abundance (from Ecker and Heltshe (1994)).

Figure 2 also shows the four plots of the scallop; the top left figure shows the observation points with four classes, the top right figure shows the attribute versus Y coordinate with lowess line, the bottom left figure shows the attribute versus X coordinate with lowess line and the bottom right figure shows the histogram of the attribute.

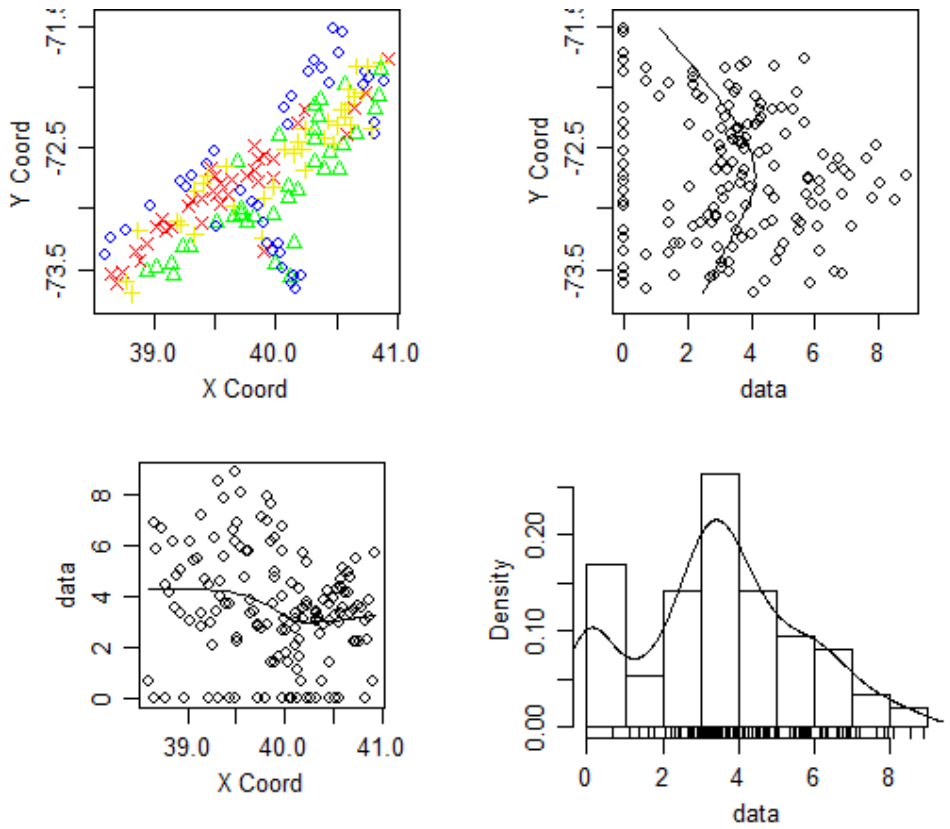

Figure 2: Plots of the scallop data; observation points(top left), latitude v.s. attribute (top right), longitude v.s. attribute (bottom left) and histogram (bottom right) of scallop data.

2.3 Environmental Data

Environmental numerical database (National Institute for Environmental Studies, 2012) also provides geostatistical data that is discussed on environment. They can be downloaded from the web site and can be loaded into *R* to use geostatistical functions. There are the data of areal environment and water quality data of public water area.

Figure 3 also shows the four plots of the water quality data at Okayama prefecture (the Okayama water data); the top left figure shows the observation points with four classes, the top right figure shows the attribute versus Y coordinate with lowess line, the bottom left figure shows the attribute versus X coordinate with lowess line and the bottom right figure shows the histogram of the attribute.

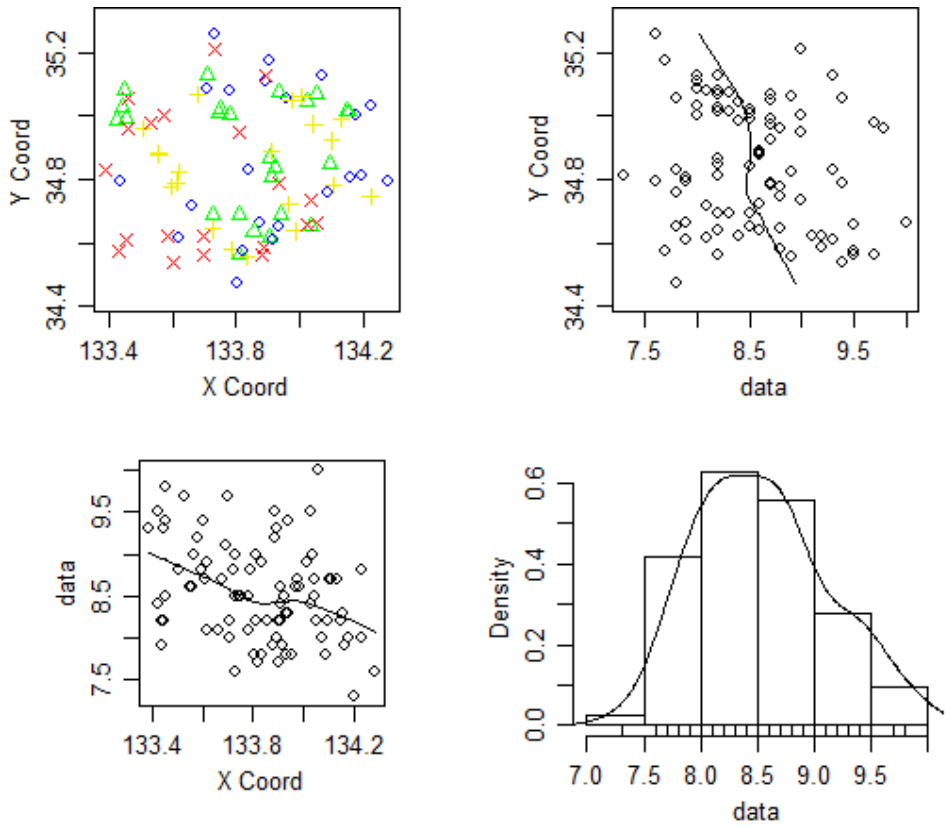

Figure 3: Plots of the Okayama water data; observation points(top left), latitude v.s. attribute (top right), longitude v.s. attribute (bottom left) and histogram (bottom right) of Okayama water quality data.

2.4 Omni-directional Variogram

Variogram is the graph of variance versus distances of pairs of observation points, and it explains the spatial structures of dependences of data. Omni-directional variogram is calculated by using attributes of all observed points.

Figure 4 shows empirical variogram and theoretical variogram model of the meuse data. The details of calculating empirical variogram and fitting theoretical variogram model are explained in Section 3.

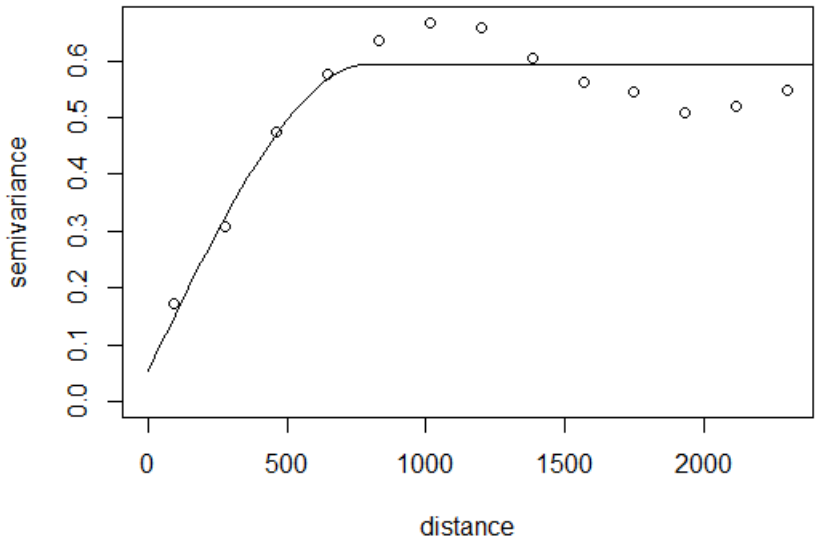

Figure 4: Omni-directional variogram with theoretical variogram of meuse data.

2.5 Directional Variogram and Exploratory GA detection

On the other hand, directional variogram is restricted for directions with some tolerances. For example figure 6 shows four directional variograms of the meuse data; 0, 45, 90 and 135° by clockwise rotation from the northern direction.

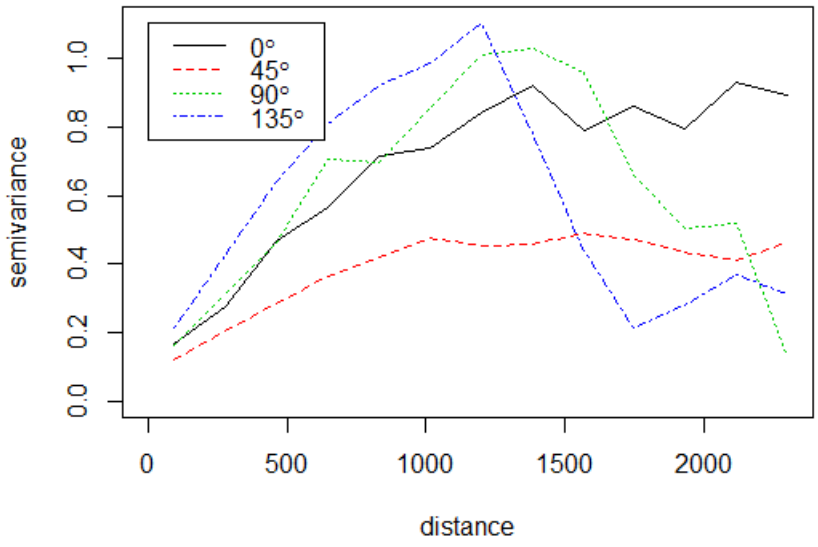

Figure 5: Four directional variograms of the meuse data.

From figure 6, we can perform exploratory detection of the GA. We can detect different range values as follows and we pointed them in figure 6.

1. north direction (0° from the north direction axis), range: 1400
2. northeast direction (45° from the north direction axis), range: 1100
3. east direction (90° from the north direction axis), range: 1000
4. southeast direction (135° from the north direction axis), range: 800

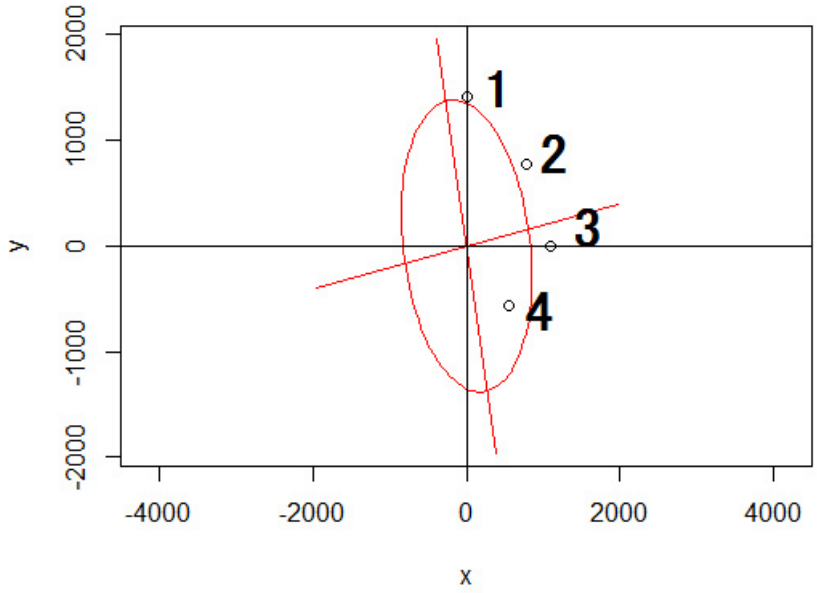

Figure 6: Exploratory GA detection; plotted range values and the detected ellipse.

Then, we detect the GA parameter; the ratio of the semi-major and semi-minor axes was 1.7, and the angle of the semi-major axis was 168.75° (15 π / 16) by trial-and-error method. These parameters means that the effect toward south of south-southeast is 70% stronger than that toward east of east-northeast.

2.6 Visualizing Application

In order to find the spatial structure of geostatistical data, to visualize in effective for application, because spatial reference relates to maps with the coordinates. Kubota and Tarumi (2007) developed the system to use Google Maps API to visualize geostatistical data. In that paper, they overlaid the attribute on the Google Maps by using colored squares at the coordinate points. Figure 7 shows the example of that system that shows the data of areal environment at Okayama prefecture.

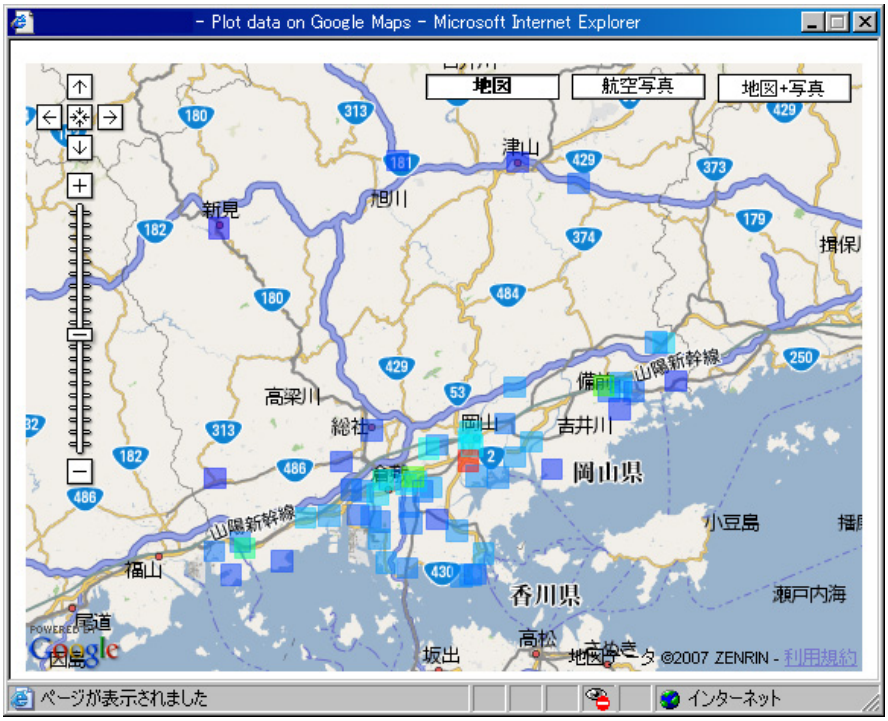

Figure 7: Visualizing application; Observation points on the Google Maps (colors are based on the values of attributes)

3 Parameter Estimation of Variogram

In this paper, the range value, which is one of the estimated variogram parameters of directional theoretical variogram, was used for proposed method. In this section, we explain the details of how to detect the range values. First, we explain how to select the appropriate cutoff value that is limitation of calculating relations of observation points, in Section 3.1. Next, the calculation of variogram cloud is explained in Section 3.2, then the calculation of empirical variogram is explained in Section 3.3. Then, some theoretical variograms are explained and parameter estimation of theoretical variogram from the emperical variogram by least squared method is explained in Section 3.4 Finally, visualizing application of variogram for goeostatistical data is explained in Section 3.5.

3.1 Cutoff Value

The cutoff value is the length of distance to user for calculation and model fitting of variogram. Variogram models are depended on the cutoff value because in the case of small cutoff value only the small distance effects are modeled while in the case of large cutoff value large distance which only uses edge observation of observed area are sensitive.

In order to detect the appropriate cutoff value, Kubota and Tarumi (2005b) simulated geostatistical data in the area of 100x100 by using Random Fields packages. Then, they calculated the differences between given variogram model and estimated variogram model by changing various cutoff values. Furthermore, Kubota and Tarumi (2006) used the data of areal environment at Tokyo and Osaka prefectures (the areal environmental data) to estimate the characteristic values at the same point of observation points

and discuss the differences between the estimated value and observed value by changing various cutoff values. Meanwhile, Kubota, et al. (2005a) simulated geostatistical data in different observation area sizes; 20x20, 40x20 and 40x40. They calculated the differences between given variogram model and estimated variogram model by changing various cutoff values in the same way of Kubota and Tarumi (2006).

From these previous studies, we assumed the cutoff value as the half of maximum distance in all pairs of observation and used it in this paper.

3.2 Variogram Cloud

If the observed data are assumed isotropic, we can develop the variogram model which express the spatial dependence. Figure 8 shows typical variogram model with parameters; the nugget effect is 0.1, the silll value is 0.9 and the range value is 1.

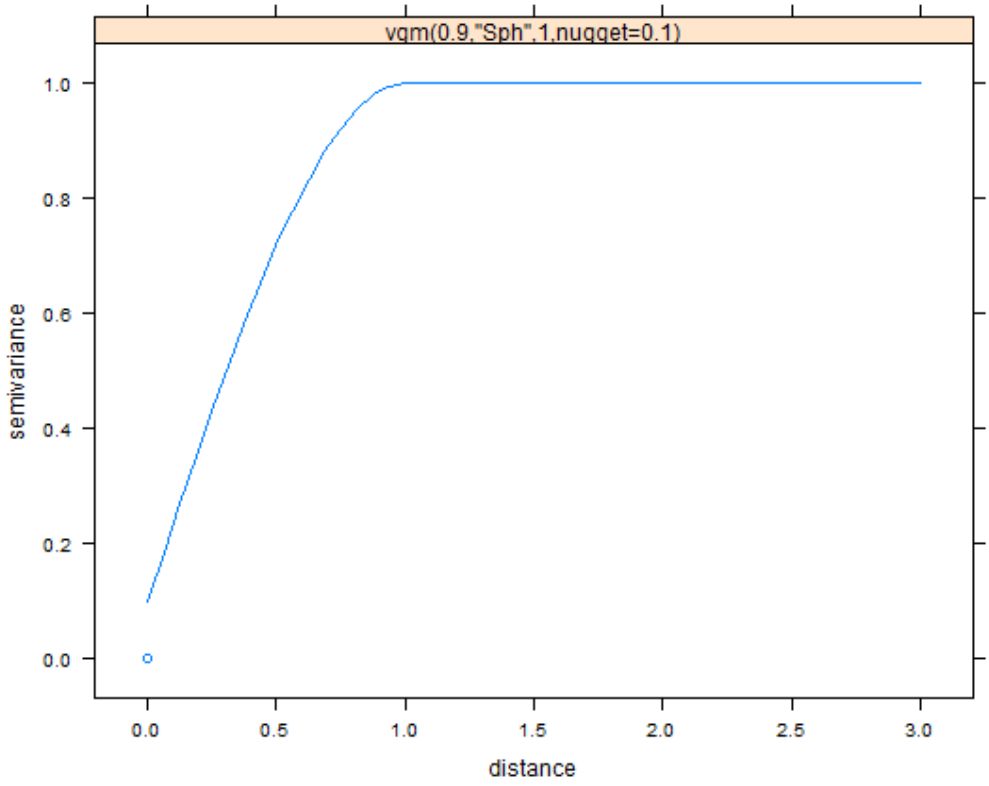

Figure 8: Typical variogram model.

The variogram model is predicted in the following order: calculate a variogram cloud, calculate an empirical variogram, and estimate the parameters of the theoretical variogram. To obtain a variogram cloud, we measure the difference between pairs of characteristic values, $z(\mathbf{x}_{p_1})$ and $z(\mathbf{x}_{p_2})$, which are located at the points $\mathbf{x}_{p_1}$ and $\mathbf{x}_{p_2}$. Some examples of characteristic values are the number of scallops on a coast, pH values in a watershed area, and the densities of CO_2 (carbon dioxide) or $SO_{\mathbf{x}}$ (sulphur dioxide or sulphur trioxide) near large roads. The dissimilarity of squared difference ($z(\mathbf{x}_{p_1})$ and $z(\mathbf{x}_{p_2})$) at points $\mathbf{x}_{p_1}$ and $\mathbf{x}_{p_2}$ is given by

$$\gamma(\mathbf{h})^* = \frac{1}{2}(z(\mathbf{x}_{p_1} + \mathbf{h}) - z(\mathbf{x}_{p_1}))^2, \mathbf{h} = \mathbf{x}_{p_2} - \mathbf{x}_{p_1}. \tag{1}$$

A variogram cloud is a graph with plots of dissimilarities γ_l^* which is calculated by (1) versus distances $d_{p_1,p_2} = d_l = |\mathbf{x}_{p_1} - \mathbf{x}_{p_2}|$ in all pairs of observation $((p_1, p_2) \in \{1, 2, \ldots, n\}, l = 1, 2, \ldots, n(n+1)/2$, where n is the number of observations).

3.3 Empirical Variogram

An empirical variogram is calculated as follows:

- Separation and classification

  The distance axis is separated

  $I_1 = (0, \alpha], I_2 = (\alpha, 2\alpha], \ldots, I_K = ((K-1)\alpha, K\alpha]$

  where K is the number of classes and α is the distance on each class.

- Average distances and dissimilarities

  The points of a variogram cloud are classified and the calculated mean values of distance and dissimilarities are

$h_k = \frac{1}{|N_k|} \sum_{N_k} d_{p_i p_j}$
$\hat{\gamma}_k = \frac{1}{|N_k|} \sum_{N_k} \frac{(z(\mathbf{x}_{p_i}) - z(\mathbf{x}_{p_j}))^2}{2}$
where N_k is the class in which the points include the distance $d_{p_i p_j} = |\mathbf{x}_{p_i} - \mathbf{x}_{p_j}| \in I_k$, and $|N_k|$ is the number of elements in N_k.

An empirical variogram is used to fit the theoretical variogram.

3.4 Theoretical Variogram

Many kinds of models of theoretical variograms are proposed in various fields. We used the spherical model

$$\gamma(\mathbf{h}; \xi_0, \xi_1, \xi_2) = \begin{cases} \xi_0 + \xi_1(\frac{3}{2}|\mathbf{h}|/\xi_2 - \frac{1}{2}[|\mathbf{h}|/\xi_2]^3), & 0 < |\mathbf{h}| \leq \xi_2 \\ \xi_0 + \xi_1, & |\mathbf{h}| > \xi_2 \\ 0, & |\mathbf{h}| = 0 \end{cases} \quad (2)$$

where ξ_0 is the nugget effect value, ξ_1 is the sill value and ξ_2 is the range value. We used the least-squares method to estimate the parameters of the theoretical variogram. Figure 9 shows calculated variogram cloud, empirical variogram (plots) and fitted theoretical variogram (solid line) of the areal environmental data. In this case the cutoff value is set to 0.45, and the distance from zeto to 0.45 is divided to fifteen classes to calculate the empirical variogram.

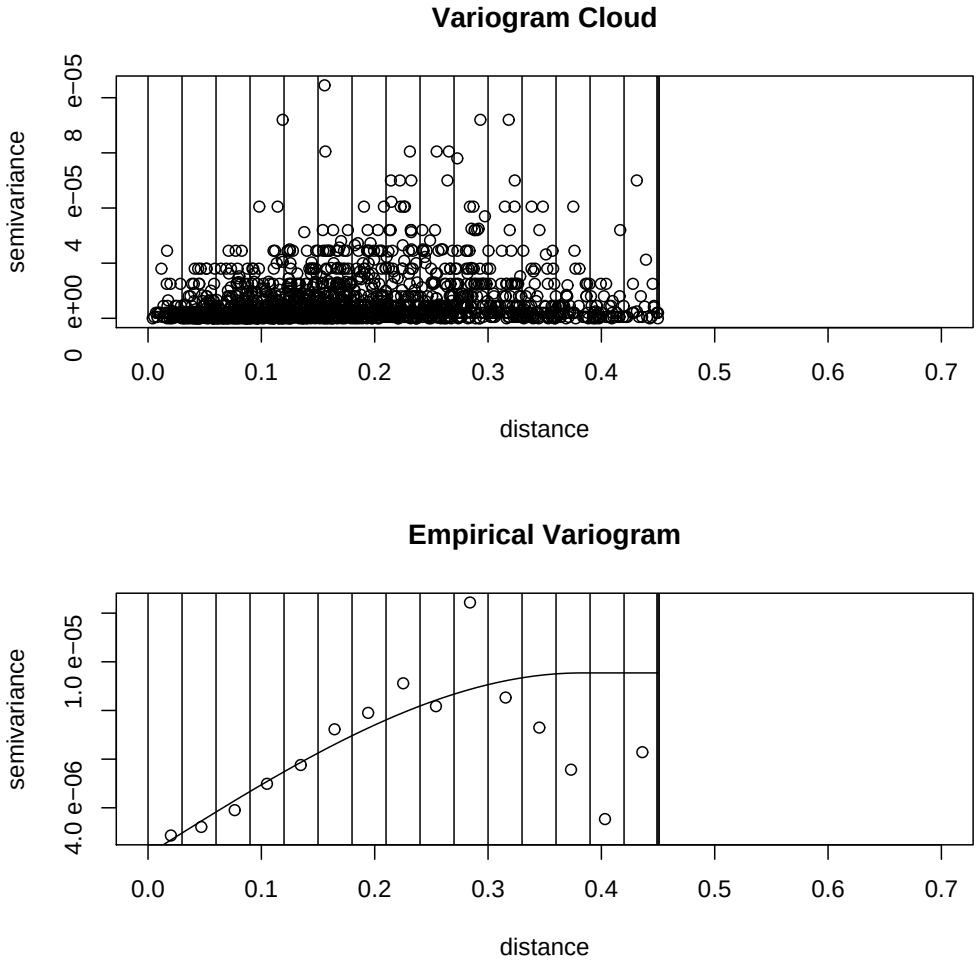

Figure 9: Variogram models of the areal environmental data; variogram cloud (top) and empirical variogram with fitted theoretical variogram.

3.5 Geometric Anisotropy

The theoretical variogram model (2) with the isotropic assumption has the same parameters (ξ_0, ξ_1, ξ_2) in different directions. Practically, these parameters sometimes have different values in different directions, because the geostatistical data exhibit anisotropy. A case with different ξ_1 values in different directions is called zonal anisotropy. A case with different ξ_2 values in different directions is called geometric anisotropy (GA). In the former problem, one solution for zonal anisotropy is to remove the directional trend. On the other hand, in the latter problem, little attention has been given to solving it. For example, using exploratory data analysis to draw a rose diagram. Figure 10 shows the example of rose diagram of uniform data which uses the *CircStats* package of *R* (Jammalamadaka and SenGupta, 2001).

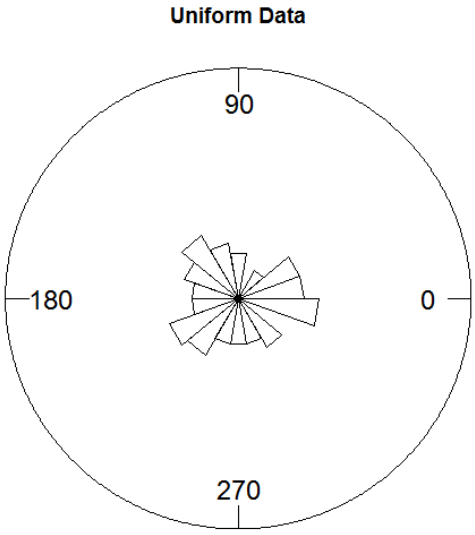

Figure 10: The example of rose diagram of uniform data.

The other and parametric way for detection geometric analysis was fitting ellipse to points which coordinates were calculated by range values of directional variogram in all directions (Kubota and Tarumi, 2008a).

3.6 Visualizing Application

In order to find the relation between the observed points and the variogram cloud, the interactive handling system, which used SVG and HTML, was also developed by Kubota and Tarumi (2007). SVG is the abbreviation of Scalable Vector Graphics, and is an XML-based file format for two-dimensional vector graphics. Therefore, the user can easily zoom in the selected point of the variogram cloud.

In that system, the point of variogram cloud in figure 11 is connected to the pair of the observed points in figure 12. For example, the #1 point of the variogram cloud in figure 11, which is the pair with long distance and large difference, corresponds to the #1 pair (the both edges of the red line) of the observed points in figure 12, while the #2 point of the variogram cloud in figure 11, which is the pair with short distance and small difference, corresponds to the #2 pair (the both edges of the red line) of the observed points in figure 12.

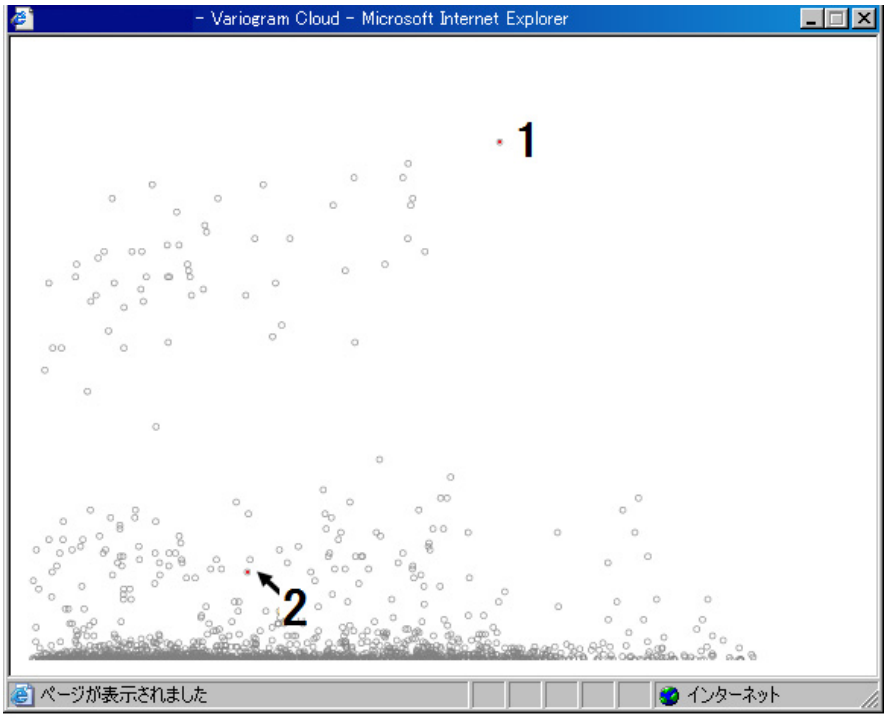

Figure 11: The example of the visualizing application (variogram cloud)

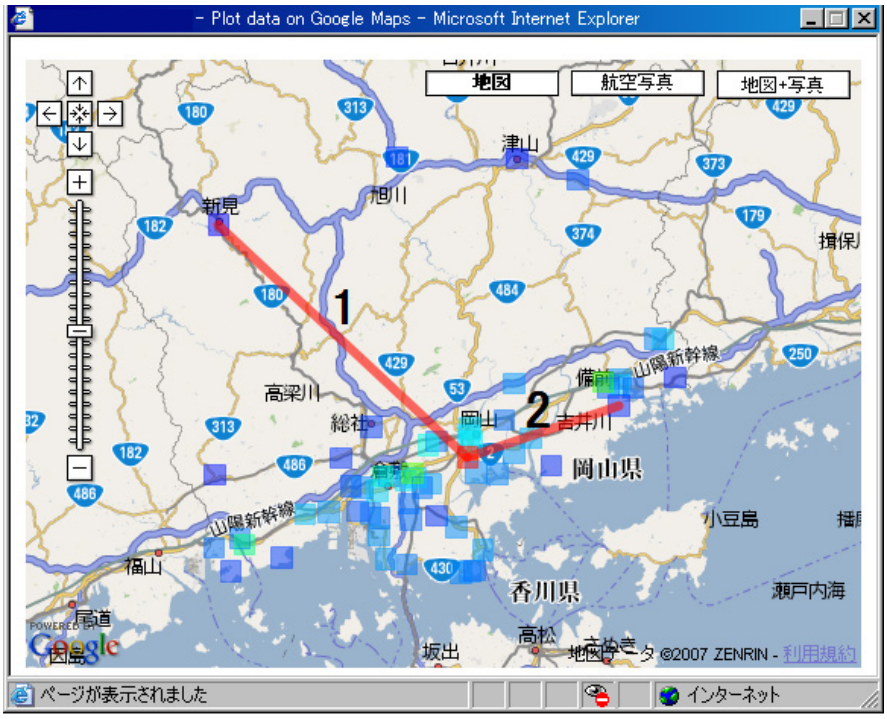

Figure 12: The example of the visualizing application (conection between observed points; solid lines are the pairs of observations which selected in variogram cloud)

Also, Kubota and Tarumi (2007) developed interactive changing of cutoff values and model selection of theoretical variogram. Figure 13 shows the example of that system which includes the empirical variogram, cutoff values and theoretical variogram.

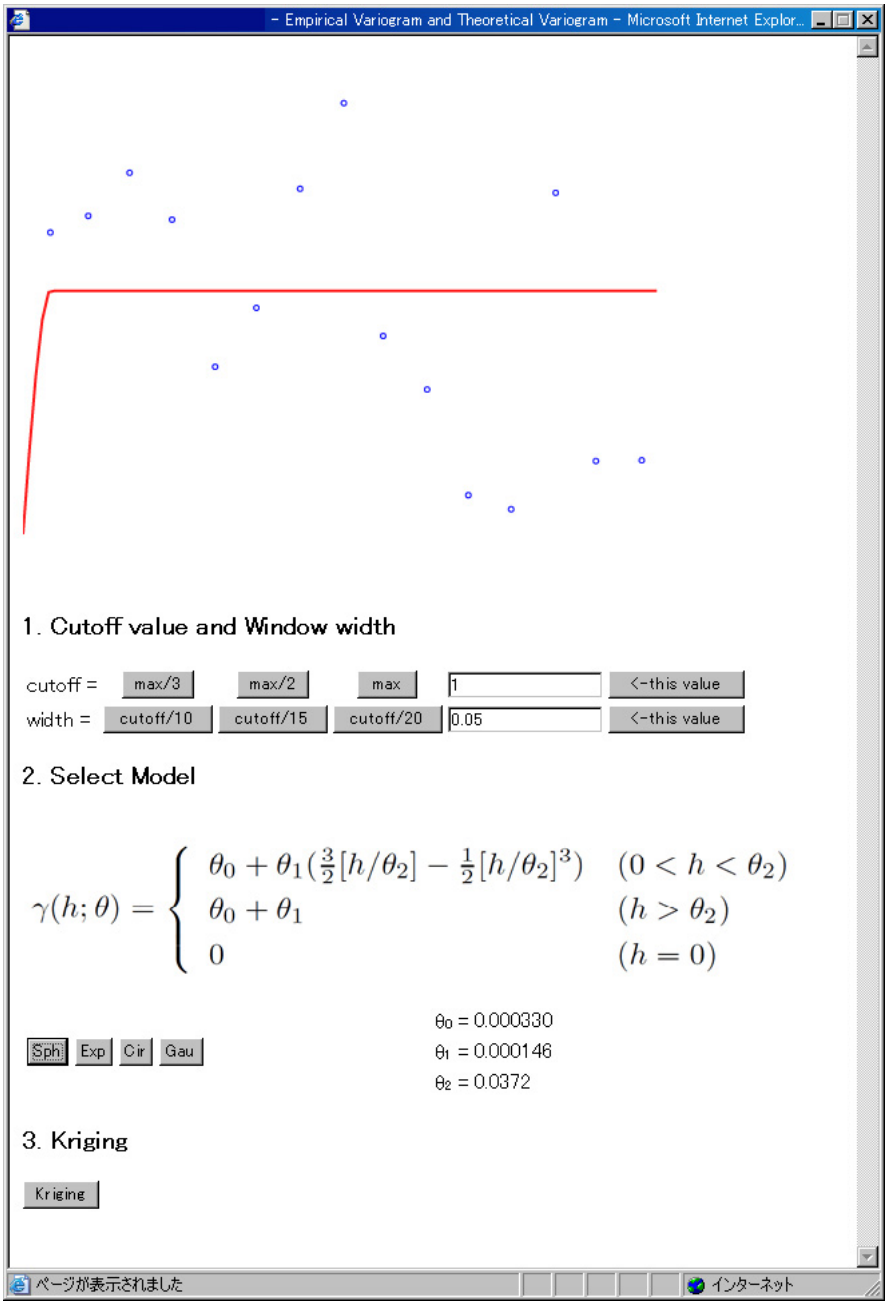

Figure 13: The example of the visualizing application (empirical variogram, cutoff values and theoretical variogram)

4 Detection and Correction of Geometric Anisotropy

In this section, we explain how to detect and correct GA. First, both least-squares method of parameter estimation by the method of Kubota and Tarumi (2008a) and likelihood-based parameter estimation are explained in section 4.1. Next, Geometric anisotropy correction method, which is common method in two estimated parameters, is explained in section 4.2.

4.1 Geometric Anisotropy Detection Method

4.1.1 Directional Variogram and Least-squares Method for Fitting to An Ellipse

In this paper, for the simulation, we applied only the case in which the number of directions (n_d) was four. Fewer n_ds (such as two or three) were too unstable for fitting to an ellipse and greater n_ds caused small number of pairs in each directional (empirical) variogram, which creates instability for fitting to a theoretical variogram. Therefore, we chose four directions for fitting to the ellipse model.

At the beginning, we set the directional start angle and tolerance values (tol). Because a small change in the directional start angle has no effect on fitting to an ellipse, we set the four directions as 0, 45, 90 and 135° by clockwise rotation from the northern direction. In order to cover all observation data, we used 22.5° of tolerance on both sides of each direction.

To obtain a variogram cloud, we measured the difference between the pairs of characteristic values, $z(\mathbf{x}_\alpha)$ and $z(\mathbf{x}_\beta)$, located at points $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$ in each direction. The difference is given by (1) in $\mathbf{h}$ of $\mathbf{h} = \mathbf{x}_\beta - \mathbf{x}_\alpha$. We had to consider not only the distance between each pair of the observed points but also the corresponding direction. In the calculation of a variogram cloud, it

is important to choose a cutoff value that is the limit of the distance between observation pairs. We adopted half the maximum value of the distance in all pairs as the cutoff value (Kubota et al., 2005). We used the spherical model in (2). For the j-th direction, we applied the above method and got the j-th range values ξ_2. Then, we calculated the coordinates of point $\mathbf{P}_{0j}$ in the j-th direction. The coordinates are expressed by the direction and the corresponding range value, which are parameters of the fitted variogram model. $\mathbf{P}_{0j}$ is as follows:

$$\mathbf{P}_{0j} = \begin{pmatrix} d_j \cos\theta_{0j} \\ d_j \sin\theta_{0j} \end{pmatrix}, (j = 1, 2, 3, 4). \tag{3}$$

$\mathbf{P}_{1j}$, the point on the ellipse line corresponding to $\mathbf{P}_{0j}$, is defined as follows:

$$\mathbf{p}_{1j} = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} a\cos\theta_{1j} \\ b\sin\theta_{1j} \end{pmatrix}, (j = 1, 2, 3, 4). \tag{4}$$

The parameter a is the length of the semi-major axis, b is the length of the semi-minor axis, and ψ is the position angle of the semi-major axis, where $\tan\theta_{1j} = \frac{a}{b}\tan\theta_{0j}$. We calculated the weighted least-squares criterion, the weighted regression sum of the squared error of the ellipse (WRSSE), by using $\mathbf{OP}_{0j}$ and $\mathbf{OP}_{1j}$ that are the distance between $\mathbf{O}$ and $\mathbf{P}_{0j}$, and the distance between $\mathbf{O}$ and $\mathbf{P}_{1j}$, respectively:

$$\mathbf{WRSSE} = \sum_{j=1}^{n_d} \frac{\nu_j}{N} (\mathbf{OP}_{0j} - \mathbf{OP}_{1j})^2 \tag{5}$$

where N is the total number of all pairs and ν_j is the number of pairs in the j-th direction. We fitted the parameters a, b and ψ by optimizing the WRSSE with the restrictions $a > 0$ and $b > 0$.

Thus, the estimated anisotropy parameters are as follows: the semi-major axis' angle of the fitted ellipse ($\psi_A^{(EL)} = \psi$), and the ratio of the semi-major and semi-minor axes ($\psi_R^{(EL)} = \frac{a}{b}$). By using this detection method, we used the *scallop* data from *SemiPar* (Ecker and Heltshe (1994)) and *meuse* data from *gstat* (Burrough and McDonnell, 1998). In the former data, the logarithm of the number of scallops, there was a geometric anisotropy (GA) of $(\psi_A^{(EL)}, \psi_R^{(EL)}) = (47, 4.1)$. In the latter data, the logarithm of zinc, there was a GA of $(\psi_A^{(EL)}, \psi_R^{(EL)}) = (120, 1.0)$. The former results showed that the effects of the northeastern direction were four times larger than that of the southeastern direction. The latter results showed there was no GA effect because the ratio was equal to 1. In other example, figure 14 shows the four fitted theoretical variograms and the fitted ellipse with the GA parameter of $(\psi_A^{(ML)}, \psi_R^{(ML)}) = (11, 2.64)$ in the Okayama water data.

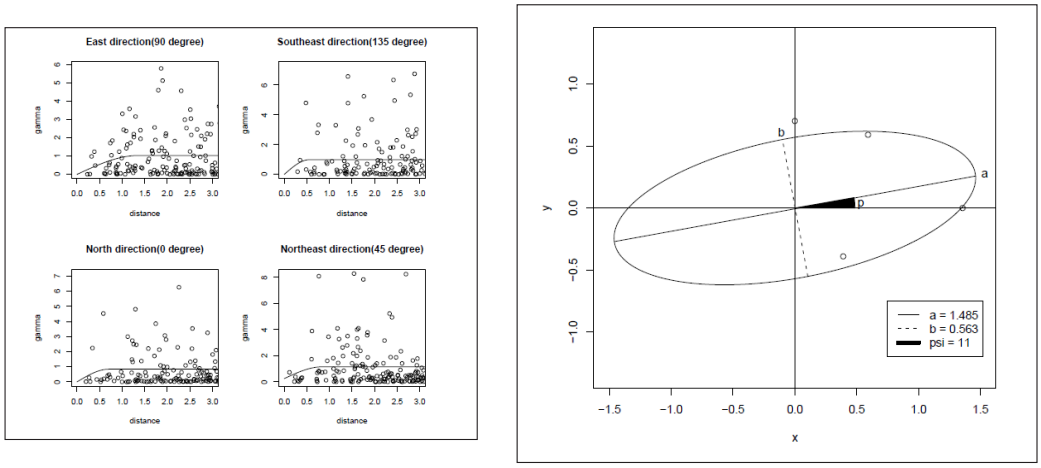

Figure 14: An example of fitting to an ellipse; the four fitted theoretical variograms (left) and the fitted ellipse (right).

4.1.2 Likelihood-based Estimation

Another method involves the likelihood-based estimation for a Gaussian random field. Assuming that the characteristic values $Z(\mathbf{y})$ of the observed points $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ are n dimensional normally distributed with mean $\mu(\mathbf{y})$ and covariance $S(\mathbf{y})$, the probability density function f is as follows:

$$f(Z(\mathbf{y})) = \frac{1}{(2\pi)^{n/2}|S(\mathbf{y})|^{1/2}} \exp(-\frac{1}{2}\left(Z(\mathbf{y}) - \mu(\mathbf{y}))'S(\mathbf{y})^{-1}(Z(\mathbf{y}) - \mu(\mathbf{y}))\right). \tag{6}$$

If we assumed GA, the observed point $\mathbf{x}$ with GA is written using $\mathbf{y}$ without the GA and GA parameters, ψ_A and ψ_R as follows:

$$\mathbf{x} = \mathbf{y} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\psi_R} \end{pmatrix} \begin{pmatrix} \cos(\psi_A) & \sin(\psi_A) \\ -\sin(\psi_A) & \cos(\psi_A) \end{pmatrix}. \tag{7}$$

We denote (7) as $\mathbf{x} = \tau(\mathbf{y}; \psi_A, \psi_R)$; in other words, $\mathbf{y}$ can be transformed by using the inverse function (τ^{-1}) as $\mathbf{y} = \tau^{-1}(\mathbf{x}; \psi_A, \psi_R)$. The probability density function in this case is as follows:

$$g(Z(\mathbf{x}); \psi_A, \psi_R) = f(Z(\tau(\mathbf{y}; \psi_A, \psi_R))) \tag{8}$$

We estimated the GA parameters (ψ_A, ψ_R) as ($\psi_A^{(ML)}$, $\psi_R^{(ML)}$) as well as other parameters.

We used this detection method and the *scallop* and *meuse* data. From the *scallop* data, the logarithm of the number of scallops, there was a GA of ($\psi_A^{(ML)}$, $\psi_R^{(ML)}$) = (40, 11). From the *meuse* data, the logarithm of zinc, there was a GA of ($\psi_A^{(ML)}$, $\psi_R^{(ML)}$) = (120, 1.0). The former showed that the effects of the northeastern direction were 11 times larger than that of the southeastern direction. The latter showed there was no GA effect.

4.2 Geometric Anisotropy Correction Method

Once parameters (ψ_A, ψ_R) are estimated, the original position is inversely transformed from an ellipse to a circle. Kubota and Tarumi (2008a) also proposed a correction method for geometric anisotropy (GA). We applied this method to the results of both the ellipse and the likelihood method. It was assumed that the observed data exhibited GA, specified as ψ_A and ψ_R. The estimated parameters, ψ_A and ψ_R, were used in the correction process: reverse rotation of the observed points by $(-\psi_A)$ and extension of the vertical direction by ψ_R. The original coordinates of the observed data, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ were transformed to $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$ as follows:

$$\mathbf{Y} = \mathbf{X} \begin{pmatrix} \cos(-\psi_A) & \sin(-\psi_A) \\ -\sin(-\psi_A) & \cos(-\psi_A) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \psi_R \end{pmatrix}. \tag{9}$$

On the other hand, the characteristics of the observed data were fixed. In this process, the GA data were transformed to isotropic data.

For example, figure 15 shows the original position and transformed position which corresponds to the ellipse and the circle in the Okayama water data.

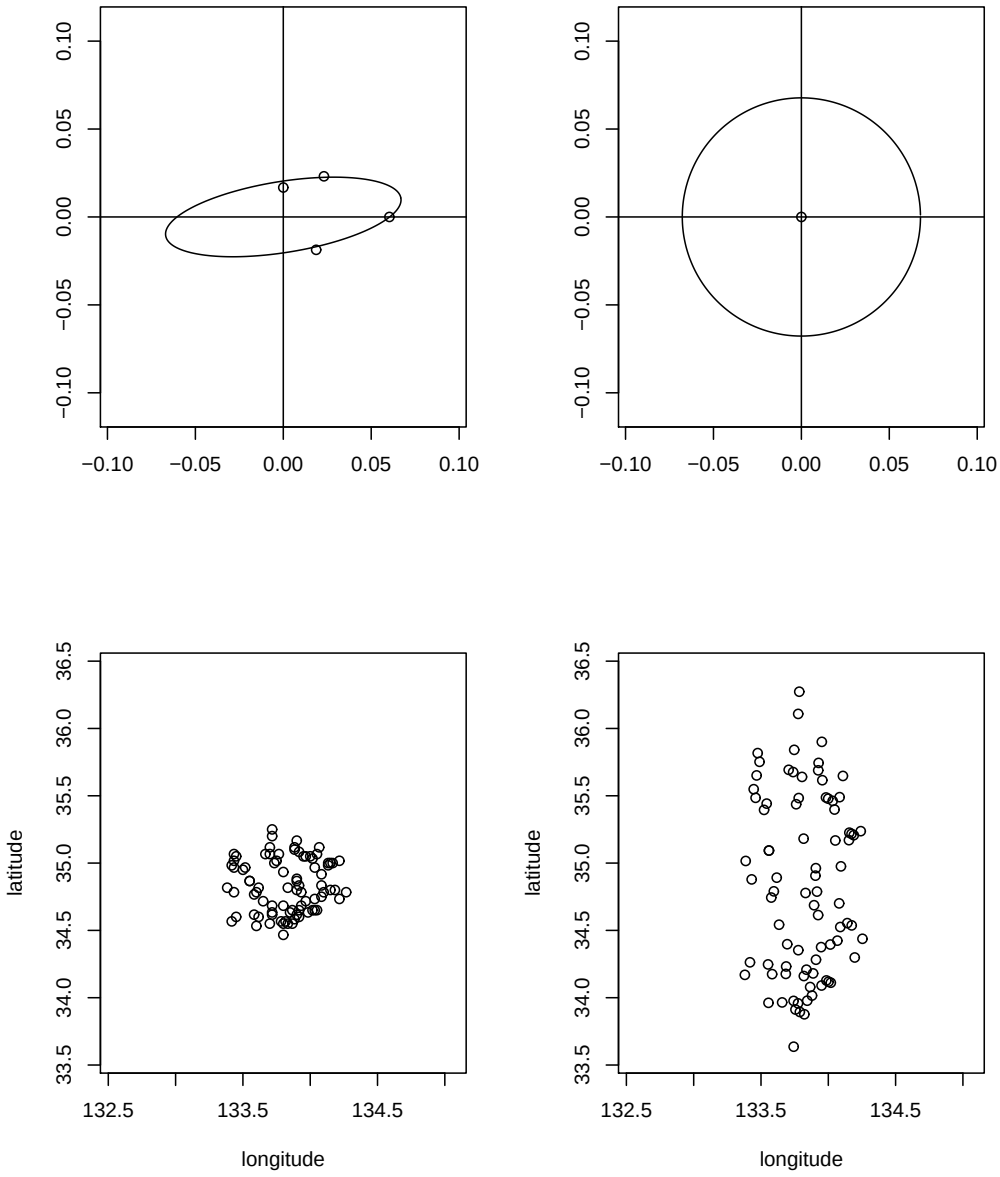

Figure 15: An example of correction method; the fitted ellipse (top left), the original position (bottom left), transformed position(bottom right) and the corresponding circle (top right).

5 Simulation Study

In this section, we explain how the anisotropy data were generated as well as the method and the results of cross validation. First, we explain the calculation steps; to simulate data, detect a GA, correct the GA, do Kriging and validate the results to compare among three types in Section 5.1. Second, we explain the application of parallel computing in Section 5.2. Third, we explain the parameters that we used in the simulation study in Section 5.3. Finally, we show the results of the simulation study in Section 5.4.

5.1 Calculations

We did a simulation study to compare the three types of calculations as follows:

Type 1 (IS) Assumption of isotropy

Type 2 (EL) Fitting to an ellipse

Type 3 (ML) Likelihood-based method

The order of simulation was as follows:

Step 1 Determine the parameters

We used the parameters mean = 0, variance = 1, a spherical variogram model of $(\xi_0, \xi_1, \xi_2) = (0.2,\ 0.8,\ r)$ with density d and the anisotropy parameter (angle of ellipse, ratio of ellipse) = $(0,\ \psi_R)$, in area (longitude, latitude)= $([0, 10], [0, 10])$, with changing parameters $r = 5$ and 10, $d = 0.5, 1$ and 2, and $\psi_R = 2, 3$ and 5. Using these parameters, the data were simulated by distribution in random fields. On the simulated random points in a rectangular area (longitude, latitude) = $([0, 10], [0, 10\psi_R])$ by geometric anisotropy (GA) the ψ_R, vari-

able characteristic values corresponding to these points were simulated. Then, only the latitudes of the data were transformed $(1/\psi_R)$, fixing the other values (longitude and characteristic values) as a rectangular area (longitude, latitude) $= ([0, 10], [0, 10\psi_R])$ which transformed it to a square area $([0, 10], [0, 10])$.

Step 2 Data simulation

The simulated data were $(z, \mathbf{x})$, where $\mathbf{x}$ were simulated points, z were the characteristic values on $\mathbf{x}$ and the size of the data was n. At first, we simulated the homogeneous Poisson point process, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)'$, for the observed points, by using the determined parameters of ψ_R. Then, we simulated the characteristic values on $\mathbf{Y}$ with the determined mean, variance and variogram parameters. In order to add GA, the scale of the latitude was shrunk from [0, $10\psi_R$] to [0, 10]. We used two packages for simulating the data: *spatstat* for simulating the Poisson point process, and *RandomFields* for simulating characteristic values. Figure 16 shows the rectangle of the points that was simulated in the Poisson point process (left), and a shrunken square of points where the scale of the latitude direction was changed with the parameters $d = 1$, $r = 10$ and $\psi_R = 3$ (right).

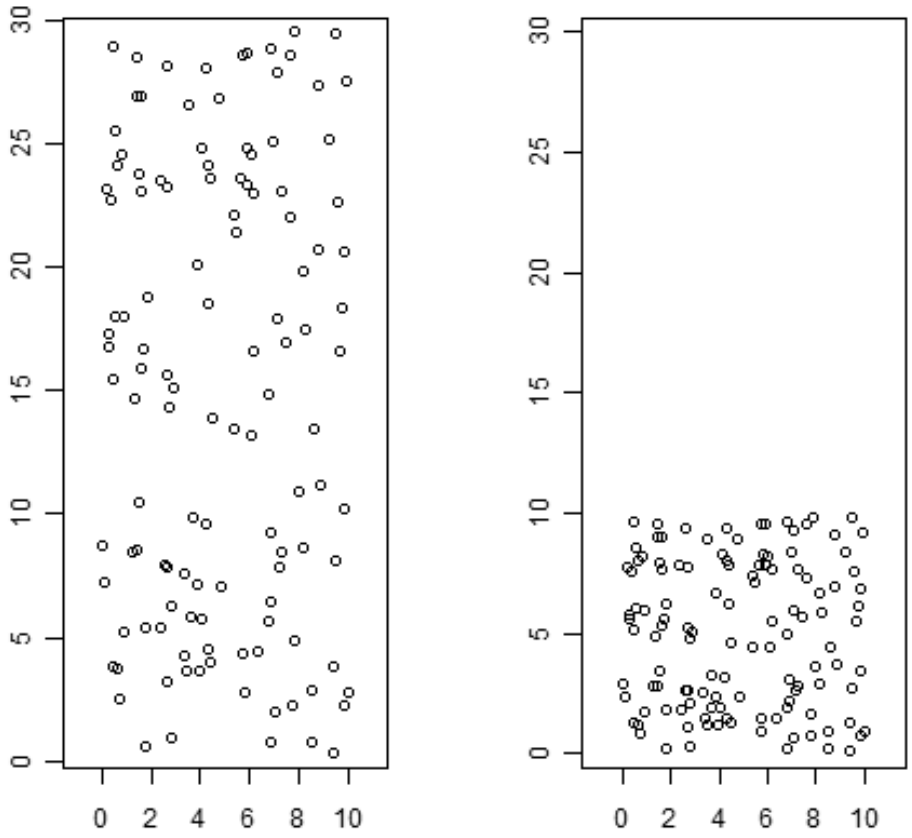

Figure 16: Random point (left) and constricted square (right).

Step 3 Detect and correct anisotropy

We used $(n-1)$ data $(z_{(i)}, \mathbf{x}_{(i)})$, where the i-th data point $(z_i, \mathbf{x}_i)$ was removed as testing data. Then, GA was detected by the three methods mentioned above:

Type 1 (IS) We assumed that the data had isotropy, therefore no transformation to correct anisotropy was done to the original simulated points $(\psi_A^{(IS)} = 0, \psi_R^{(IS)} = 1)$.

Type 2 (EL) We detected the GA parameters $(\psi_A^{(EL)}, \psi_R^{(EL)})$ by way of fitting to an ellipse, as denoted in Section 4.2.1. Then, points $\mathbf{x}$ were transformed to $\mathbf{y}^{(EL)}$ in order to correct the GA denoted in Section 4.3.

Type 3 (ML) We detected the GA parameters $(\psi_A^{(ML)}, \psi_R^{(ML)})$ using the likelihood method denoted in Section 4.2.2. Then, points $\mathbf{x}$ were transformed to $\mathbf{y}^{(ML)}$ in order to correct the GA, as denoted in Section 4.3.

Step 4 Estimate the variogram parameters

The omni-directional theoretical variogram $\gamma^{*(t)}$ was estimated using the least-squares method, fitting to the corresponding empirical variogram calculated from $(z_{(i)}, \mathbf{y}_{(i)}^{(t)})$, where t was the type of method, and $t = IS, EL, ML$.

Step 5 Do kriging and calculate the prediction square error

The characteristic value at the transformed testing point $\mathbf{y}_i^{(t)}$ was predicted using the parameters of gamma and $(z_i, \mathbf{y}_i)$ as $\hat{z}_i^{(t)}$.

Step 6 Calculate the predicted mean square error of cross validation

For all n points, the calculations from Steps 3 to 5 were iterated to

predict the mean square error of cross validation (CV)
$CV^{(t)} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \hat{z}_i^{(t)})^2$.

Step 7 Iterate 100 times in different seeds

We denote $CV_s^{(t)}$ as the calculated $CV^{(t)}$ of seed s in Step 6. We performed 100 iterations from Steps 2 to 6 in different seeds to calculate the mean and standard deviation (s.d.) of $CV_s^{(t)}$

$$CV^{*(t)} = \frac{1}{100}\sum_{s=1}^{100} CV_s^{(t)}, \tag{10}$$

$$sd^{(t)} = \sqrt{\frac{1}{99}\sum_{s=1}^{100}\left(CV_s^{(t)} - CV^{*(t)}\right)^2}. \tag{11}$$

Because (10) is just the average of $CV_s^{(t)}$, $CV^{*(t)}$ is strongly affected by outliers. Furthermore, we have to compare the detection methods to determine which method is most appropriate for GA. Therefore, we also calculated the frequency of the minimum case among the three types to check for goodness of fit among the three types and to remove the effect of extremely high values of $CV_s^{(t)}$.

5.2 Parallel computing

In order to shorten the calculation time, we applied the *multicore* package (Urbanek, 2012) of *R* to these simulations. The package provides a way of running parallel computations with multiple cores or CPUs and it provides methods for results collection. Especially, the Step 7 of the simulation in Section 5.1 used the function mclapply of the package.

5.3 Parameters

In the simulation process of Steps 1 and 2 in Section 5.1, we used three kinds of changing parameters (d, r, and ψ_R). The first parameter, d, is the density of the simulated points (d points per unit area). We could control the number of observation points by the observation area, (longitude, latitude) $= ([0, 10], [0, 10])$, and d. The expected number of observation points in parameter $d = 0.5, 1.0$ and 2.0 corresponded, respectively, to 50, 100 and 200 in a 100-unit area of (longitude, latitude) $= ([0, 10], [0, 10])$. We assumed that $d = 2.0$ corresponded to the case of scallops where geometric anisotropy (GA) was detected by Kubota and Tarumi (2010) and $d = 0.5$ corresponded to the case of Okayama water quality data (Kubota and Tarumi, 2008a).

The second changing parameter, r, was the range of the theoretical variogram. We assumed two types of range values: $r = 5$ and 10. The former ($r = 5$) corresponded to the case in which points near the center (5,5) were affected by almost all of the points. In contrast, points near (0,0) were only affected by the other quarter points. This situation might have two or more groups of points. The latter ($r = 10$) corresponded to the case in which any point was affected by almost all points. In Step 4, we used a cutoff value of half the maximum value of the distance in all pairs (around 7). We assumed two kinds of range values: small range ($r = 5$) and large range ($r = 10$). If we assumed r larger than 10, it was difficult to estimate the range values and if we assumed r smaller than 5, more than two groups might be included.

The third changing parameter was ψ_R, the GA ratio. We assumed $\psi_R = 2, 3$ and 5. Because a weaker GA ratio does not make sense, we assumed $\psi_R = 2$ as the weakest ratio. Because a stronger ratio produces only one-dimensional spatial data, we assumed $\psi_R = 5$ as the strongest ratio. On the other hand, in practical data analysis, $\psi_R = 3$ is sometimes estimated as the

stronger ratio. Therefore, we also used $\psi_R = 3$ as the strongest ratio for practical data analysis.

If we use other units of interaction times such as 10, 50, 200 and 1000, there might be small differences, but EL and ML might sometimes be in error because of the extremely large GA ratio.

5.4 Results

Table 1 shows the values of the mean and standard deviation of $CV_s^{(t)}$ for Step 7, in the density parameter $d = 0.5$ by two kinds of changing parameters (ratio of anisotropy and range of theoretical variogram). The table compares the three types of correction methods (IS, EL, ML) that we mentioned in Section 5.1. Table 2 is similar to Table 1, but d is 1.0. Table 3 is also similar to Table 1, but d is 2.0.

Table 1: $CV^{*(t)}$ values in $d = 0.5$

range	5						10					
$CV^{*(t)}$	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
method / ratio	IS		EL		ML		IS		EL		ML	
2	0.626	0.172	0.626	0.163	0.634	0.175	0.438	0.104	0.435	0.104	0.443	0.108
3	0.739	0.203	0.735	0.195	0.756	0.207	0.502	0.130	0.503	0.129	0.511	0.135
5	0.900	0.254	0.892	0.247	0.915	0.260	0.646	0.193	0.646	0.188	0.653	0.196

Table 2: $CV^{*(t)}$ values in $d = 1.0$

range	5						10					
$CV^{*(t)}$	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
method / ratio	IS		EL		ML		IS		EL		ML	
2	0.507	0.100	0.522	0.106	0.510	0.103	0.368	0.071	0.399	0.094	0.370	0.075
3	0.610	0.118	0.619	0.125	0.611	0.118	0.420	0.081	0.427	0.083	0.419	0.082
5	0.765	0.150	0.766	0.161	0.767	0.150	0.536	0.114	0.520	0.116	0.536	0.113

Table 3: $CV^{*(t)}$ values in $d = 2.0$

range	5						10					
$CV^{*(t)}$	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
method / ratio	IS		EL		ML		IS		EL		ML	
2	0.424	0.052	0.468	0.117	0.424	0.052	0.327	0.039	0.349	0.060	0.328	0.039
3	0.498	0.067	0.526	0.130	0.498	0.067	0.364	0.043	0.394	0.098	0.365	0.043
5	0.646	0.097	0.644	0.138	0.647	0.097	0.442	0.059	0.438	0.108	0.442	0.059

Tables 4, 5 and 6 show the frequency of the minimum case of $d = 0.5, 1.0$ and 2.0 (corresponding respectively to Tables 1, 2 and 3). The sums of IS, EL and ML for each ratio and range are not equal to 100 because there are some equally predicted values with anisotropic parameters (ψ_A, ψ_R) of (0, 1).

If $(\psi_A, \psi_R) = (0, 1)$, the data exhibit isotropy. If the ML case was the minimum and its parameter was (0, 1), the values of IS and ML were added up.

In Table 1, EL showed good results in the range of both 5 and 10 for almost all ratios(ψ_R) of 2, 3 and 5 (except $\psi_R = 3$ of $r = 10$, which was larger) although the differences in all d, ratios and ranges were not always statistically significant (5%). In Table 2, there were differences between the small range ($r = 5$) and the large range ($r = 10$). In the three left-hand columns of Table 2, IS showed good results. In the three right-hand columns of Table 2, the goodness of fit for the results depended on the ratio; the IS results were good for $\psi_R = 2$; the ML results were good for $\psi_R = 3$, and EL had a good result in $\psi_R = 5$. The trend in Table 3 was similar to that in Table 2.

The terms of Tables 4, 5 and 6 are the same as those in Table 1. However, the figures in each cell are not the average of $CV_s^{(t)}$ but rather frequencies of minimum cases; the bigger the counts, the better the results. Tables 4, 5 and 6 correspond to Tables 1, 2 and 3, respectively. In Table 5, the results depended on the range and ratio; IS showed good results for the weak ratio of $r = 5$; ML showed good results for the weak ratio of $r = 10$, and EL had good results in a strong ratio of both $r = 5$ and 10. The trend in Table 6 was the same as that in Table 5. Upon comparison of Tables 5 and 2 (or 6 and 3), ML improved. Because the figures of Tables 4, 5 and 6 are the frequencies of the minimum cases of $CV_s^{(t)}$, the choice of ML is robust.

Table 4: Frequencies of minimum of $CV_s^{(t)}$ values in $d = 0.5$

range	5			10		
method / ratio	IS	EL	ML	IS	EL	ML
2	42	46	31	30	46	35
3	45	47	34	41	48	33
5	37	45	35	47	46	37

Table 5: Frequencies of minimum of $CV_s^{(t)}$ values in $d = 1.0$

range	5			10		
method / ratio	IS	EL	ML	IS	EL	ML
2	51	39	47	38	25	44
3	46	48	42	38	41	45
5	44	47	43	34	60	34

Table 6: Frequencies of minimum of $CV_s^{(t)}$ values in $d = 2.0$

range	5			10		
method / ratio	IS	EL	ML	IS	EL	ML
2	54	36	49	43	29	46
3	46	48	42	38	41	45
5	44	47	43	34	60	34

6 Concluding Remarks

In this paper, we applied a detection method of geometric anisotropy (GA) using four theoretical directional variograms that produced four sets of parameters. This method was one of the proposed methods by Kubota and Tarumi (2008a). The objective of this simulation study is to identify which detection method is effective in reducing prediction errors. The detection methods used in this study were the assumption of isotropy (IS), fitting to an ellipse (EL) and the likelihood-based method (ML), as denoted in Section 3.1.

On the basis of our results, we make three suggestions regarding the detection and correction of GA: (1) if the number of observation points is low (50 or less; d=0.5 or less), EL should be chosen because of minimum prediction error; (2) if the GA ratio is strong (ψ_R is 5 or more), EL should be chosen; otherwise, (3) IS produces sensitive results that the predicted values depend on the attributes or the spatial structures of observed data because $CV^{*(IS)}$ is minimum but the frequencies are not maximum, while (4) ML produces robust results because $CV^{*(ML)}$ is not minimum but the frequencies are maximum. Suggestions (1) and (2) were anticipated, but suggestion (3) and (4) were not. Suggestion (1) confirmed the case study of Okayama water quality data (Kubota and Tarumi, 2010) that corresponded to $\psi_R = 2$ or 3 of $r = 5$ in Table 1.

Practically, GA is sometimes ignored even though it is one of the parameters that express the background character in a study area. From the results of the simulations, we propose the correction of GA in the case of (1) and (2) using the detected parameter of ψ_A and ψ_R that was proposed by Kubota and Tarumi (2008a). To find more details and relations among the parameters, in the future, we have to simulate other models of theoretical

variograms, such as linear, exponential, Gaussian and so on; other types of anisotropy (zonal anisotropy); and smaller or larger number of data. Furthermore, the likelihood method sometimes produces a very strong GA ratio and isotropy (practical application of scallop and meuse data; (ψ_A, ψ_R) = (4, 11) and (120, 1.0) in 2.2.2). Therefore, determining the properties of the likelihood method will the subject of a future study.

Appendix

A Theoretical Variograms

We used the spherical model in Section 3.4. In this section we list the other typical theoretical variogram models, and figure 17 shows the corresponding graphs of the models.

1. Nugget effect model

$$\gamma(\mathbf{h};\xi_0,\xi_1) = \begin{cases} \xi_0+\xi_1, & |\mathbf{h}|>0 \\ 0, & |\mathbf{h}|=0 \end{cases} \tag{12}$$

2. Linear model

$$\gamma(\mathbf{h};\xi_0,\xi_1) = \begin{cases} \xi_0+\xi_1|\mathbf{h}|, & |\mathbf{h}|>0 \\ 0, & |\mathbf{h}|=0 \end{cases} \tag{13}$$

3. Circular model

$$\gamma(\mathbf{h};\xi_0,\xi_1,\xi_2) = \begin{cases} \xi_0+\xi_1(\frac{2|\mathbf{h}|}{\pi\xi_2}\sqrt{1-(|\mathbf{h}|/\xi_2)^2}+\frac{2}{\pi}\arcsin|\mathbf{h}|/\xi_2), & 0<|\mathbf{h}|\le\xi_2 \\ \xi_0+\xi_1, & |\mathbf{h}|>\xi_2 \\ 0, & |\mathbf{h}|=0 \end{cases} \tag{14}$$

4. Exponential model

$$\gamma(\mathbf{h};\xi_0,\xi_1,\xi_2) = \begin{cases} \xi_0+\xi_1(1-\exp(-|\mathbf{h}|/\xi_2)), & |\mathbf{h}|>0 \\ 0, & |\mathbf{h}|=0 \end{cases} \tag{15}$$

5. Logarithmic model

$$\gamma(\mathbf{h};\xi_0,\xi_1,\xi_2) = \begin{cases} \xi_0 + \xi_1(\log(|\mathbf{h}| + \xi_2)), & |\mathbf{h}| > 0 \\ 0, & |\mathbf{h}| = 0 \end{cases} \tag{16}$$

6. Gaussian model

$$\gamma(\mathbf{h};\xi_0,\xi_1,\xi_2) = \begin{cases} \xi_0 + \xi_1(1 - \exp(-|\mathbf{h}|/\xi_2)^2), & |\mathbf{h}| > 0 \\ 0, & |\mathbf{h}| = 0 \end{cases} \tag{17}$$

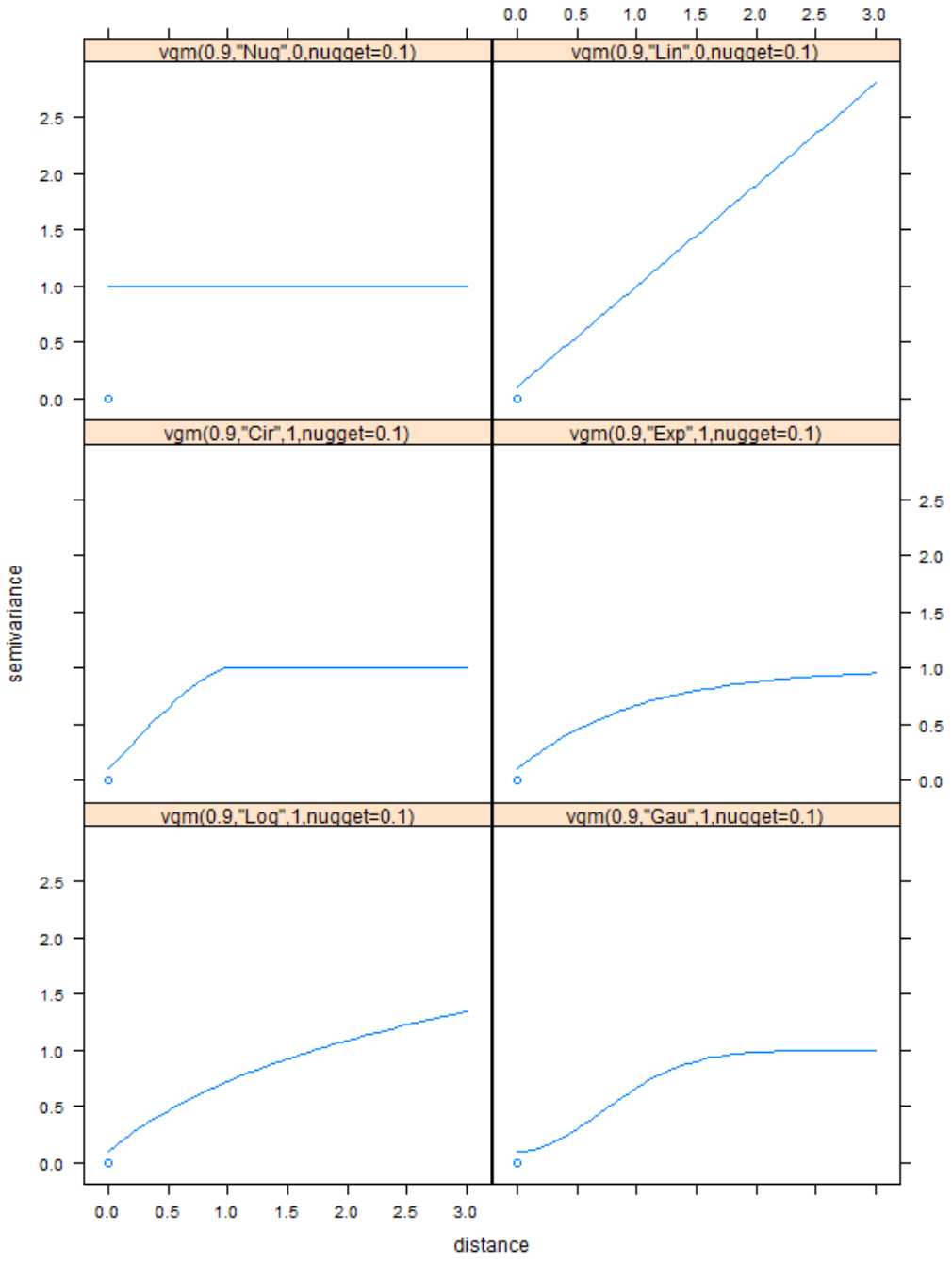

Figure 17: Typical theoretical variograms; .

B CRAN Task Views

In this paper, we used a lot of packages in *R*, which related CRAN Task Views (CTV). CTV provides the packages and functions for specific research areas. In order to use CTV, we used the *ctv* packages in *R*. We list the CTV and URL that we used in this paper.

- Analysis of Spatial Data
  http://cran.r-project.org/web/views/Spatial.html
- Analysis of Ecological and Environmental Data
  http://cran.r-project.org/web/views/Environmetrics.html
- High-Performance and Parallel Computing with R
  http://cran.r-project.org/web/views/HighPerformanceComputing.html

References

[1] Burrough, P.A. and McDonnell, R.A. (1998). *Principles of Geographical Information Systems*. Oxford University Press.

[2] Cheng, K. S., Yeh, H. C. and Tsai, C. H. (2000). An Anisotropic Spatial Modeling Approach for Remote Sensing Image Rectification. *Remote Sensing of Environment*, Vol. **73**, Issue 1, pp. 46-54.

[3] Chilès, J. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics.

[4] Ecker, M.D. and Heltshe, J.F. (1994). *Geostatistical estimates of scallop abundance*. Case Studies in Biometry, John Wiley and Sons, pp. 107-124

[5] Guan, Y., Sherman, M. and Galvin, J. A. (2006). Assessing Isotropy for Spatial Point Processes. *Biometrics*, Vol. **62**, Issue 1, pp. 119-125.

[6] Jammalamadaka, S. R. and SenGupta, A. (2001) *Topics in circular Statistics*. World Scientific.

[7] Journel, A. G. and Froidevaux, R. (1982). Anisotropic Hole-Effect Modeling. *Mathematical Geology*, Vol. **14**, No. 3, pp. 217-239.

[8] Kubota, T., Iizuka, M., Fueda, K. and Tarumi, T. (2005a). The Selection of the Cutoff in Estimating Variogram Model. *The 5th IASC Asian Conference on Statistical Computing*, pp. 97-100.

[9] Kubota, T. and Tarumi, T. (2005b). The Sellection of the Cutoff for Variogram Estimation. *Proceedings of the 19th Annual Meeting of Japanese Society of Computational Statistics*, pp. 115-118. (in Japanese)

[10] Kubota, T. and Tarumi, T. (2006). The Sellection of the Cutoff for Variogram Estimation and Its Validation. *Proceedings of the 19th Symposium Meeting of Japanese Society of Computational Statistics*, pp. 119-122. (in Japanese)

[11] Kubota, T. and Tarumi, T. (2008a). Using Geometric Anisotropy in Variogram. *COMPSTAT2008 Proceedings in Computational Statistics*, Physica-Verlag HD, pp. 793-801.

[12] Kubota, T. and Tarumi, T. (2008b). A study on detection and correction of anisotropy for geostatistical data. *Proceedings of IASC2008; Joint Meeting of The 4th World Conference of the IASC and The 6th Conference of the Asian Regional Section of the IASC on Computational Statistics and Data Analysis*, pp. 933-942.

[13] Kubota, T. and Tarumi, T. (2010). Comparison of detecting geometric anisotropy between geometric method and likelihood based method. *Proceedings of Joint Meeting of Japan-Korea Special Conference of Statistics and The 2nd Japan-Korea Statistics Conference of Young Researchers*, pp. 173-180.

[14] National Institute for Environmental Studies (2012). Environmental numerical database (in Japanese). Retrieved July 30, 2012, from http://www.nies.go.jp/igreen/index.html

[15] Urbanek (2012). multicore - Parallel processing in R on machines with multiple cores or CPUs. Retrieved July 30, 2012, from http://www.rforge.net/multicore/index.html

[16] Wackernagel, H. (1995). *Multivariate Geostatistics: An Introduction With Applications*. Springer.

[17] Zimmerman, D. L. (1993). Another Look at Anisotropy in Geostatistics. *Mathematical Geology*, Vol. **25**, No. 4, pp. 453-470.

Acknowledgments

I have countless people to thank for having contributed to the completion of this thesis.

First of all, I would like to express my deep gratitude to Professor Tomoyuki Tarumi of the Graduate School of Environmental and Life Science, Okayama University, for giving me the opportunity to study, for his passionate teaching, guidance, helpful comments, valuable suggestions and continuous encouragement in my study.

I would also like to express my gratitude to Professor Koji Kurihara, Dr. Kaoru Fueda and Dr. Masaya Iizuka of the Graduate School of Environmental and Life Science, Okayama University, for their valuable pieces of advice and helpful suggestions in my study.

I would also like to thank Professor Satomi Tani, Professor Hiroshi Sano, Professor Naozumi Kurokami, Dr. Fumio Ishioka, Dr. Myung Jin Na and all the members of the Faculty of Law and the Graduate School of Humanities and Social Sciences, Okayama University, for their continuous encouragement and support.

I would also like to thank Professor Toshiharu Fujita, Professor Hiroe Tsubaki, Professor Yutaka Tanaka and all the members of the Institute of Statistical Mathematics, for their assistances and various pieces of advice.

I would also like to thank Dr. Makoto Tomita of Tokyo Medical and Dental University, Dr. Tomokazu Fujino of Fukuoka Woman's University, Professor Yoshiro Yamamoto of Tokai University, Professor Yuichi Mori of Okayama University of Science, Dr. Kikuo Yanagi of Okayama University of Science, all the members of MIKAWAYA and all the members of Okayama Statistical Association.

Finally, I am thankful to my family, for their understanding and love.