Study on autocorrelation model for spatial distribution of soil properties using Gaussian process regression

Y. Tomizawa*1, I. Yoshida2, and Y. Otake3

1 School of Integrative Science and Engineering, Tokyo City University, Japan
2 Department of Urban and Civil Engineering, Tokyo City University, Japan
3 Department of Civil Environmental Engineering, Tohoku University, Japan

Abstract

The spatial distribution of soil properties is often modeled by dividing it into trend and random components. Gaussian process regression with multiple stochastic fields is used to separate and model them in this study. The random component is important for understanding the characteristics of local variability in the spatial distribution. Autocorrelation models for the random components are examined based on 1D and 3D cone penetration test data. We compared various autocorrelation functions such as Gaussian, Markovian, Binary Noise, Whittle-Matérn (WM), and Gaussian-Markovian geometric mean (GMGM) model in terms of Akaike and Bayesian information criterion, AIC and BIC. As the result of comparison of the information criteria, GMGM and WM model are selected for the model site. GMGM is a very simple model and requires only exponential function, while WM requires special functions such as gamma and Bessel functions. From the viewpoint of computational cost, GMGM has advantage compared with WM. Spatial distribution of trend and random components in 1D and 3D are estimated by using GMGM model for the autocorrelation function of the random component.

Keywords: soil properties, Gaussian process regression, random components, autocorrelation function

1. Introduction

Since the safety of a geotechnical structure is highly dependent on the surrounding soil properties, it is very important to understand the spatial distribution of the geotechnical properties. In order to understand the spatial distribution of soil properties, they are often separated into trend and random components [1]. The uncertainties of spatial distribution of the soil properties should also be properly evaluated. The general regression theory, Gaussian Process Regression (GPR) [2], can be interpreted as a special case of the linear inverse problem by Bayes’ theorem, which allows the evaluation of not only the representative values but also the uncertainty of them.

When performing spatial distribution estimation using GPR, the model selection for the autocorrelation function is very important. Ching et al. [3], [4] investigated the use of a two-parameter autocorrelation function, Whittle-Matérn, for more flexible estimation, and compared to various other autocorrelation functions. Yoshida et al. [5] proposed a method to simultaneously estimate the trend and random components of geotechnical properties at arbitrary locations using GPR, which uses a superposition of multiple random fields, and examined the autocorrelation function focusing on the trend component.

In this study, we use the method of Yoshida et al. [5] to compare various existing autocorrelation function models [6] and the proposed Gaussian-Markovian geometric mean (GMGM) model for the random component, which is important for understanding the characteristics of local variations in spatial distributions.

2. Formulation of estimation using GPR

The GPR formulation can be induced as a special case of a linear inverse problem based on Bayesian estimation [7], and its main points are described below. We define the prior information of the unknown quantity vector \( x \) to be estimated as Eq. 1.

\[
x = \bar{x} + w
\]

where \( \bar{x} \) and \( w \) are the mean and the probabilistic component of the prior information, respectively. The observation equation, which is the probabilistic forward model for predicting observation \( z \) from \( x \), is generally expressed as a nonlinear function of \( x \) contaminated with Gaussian noise:

\[
z = H(x) + v
\]

where \( w \) in Eq. 1 and \( v \) in Eq. 2 are Gaussian random variable vectors with zero mean whose covariance matrices are \( M \) and \( R \), respectively. The problem to obtain maximum a posteriori estimate of \( x \) is equivalent to the optimization problem of minimizing the objective function \( J \) in Eq. 3, which is the negative logarithm of the numerator of \( p(x)p(z|x) \).

\[
J = \frac{1}{2}(x - \bar{x})^T M^{-1} (x - \bar{x}) + \frac{1}{2}(z - H(x))^T R^{-1} (z - H(x))
\]

(3)

If the nonlinear function \( H(x) \) in Eq. 2 is replaced by the linear function \( Hx \), then the minimization of Eq. 3 can be analytically performed by taking the derivative of \( J \) with respect to \( x \):

\[
\frac{dJ}{dx} = M^{-1} (x - \bar{x}) - H^T R^{-1} (z - Hx) = 0
\]

(4)

The maximum a posteriori estimate can be obtained analytically:

\[
\hat{x} = \bar{x} + PH^T R^{-1} (z - H\bar{x})
\]

(5)

\[
P = (H^T R^{-1} H + M^{-1})^{-1}
\]

(6)

Eqs. 5 and 6 can be rewritten in the form of the Kalman gain \( K_0 \) in the Kalman filter algorithm:

\[
\hat{x} = \bar{x} + K_0 (z - \bar{x})
\]

(7)

\[
P = M - K_0 H M
\]

(8)

\[
K_0 = MH^T (MH^T + R)^{-1}
\]

(9)

The GPR can also be derived from the minimization of the above objective function.

Consider the case for which the observation vector \( z \) is sampled from a random field. We assume linear function

* E-Mail: ktskkmn@gmail.com
fluctuation (SOF) \[8\]. The horizontal distance and vertical directions, respectively. Points of the coordinate values obtained from Eq. 19 for the distance between the two separability in previous studies \[9\], \[10\]. Autocorrelation functions, which have been adopted as represented by the product of the horizontal and vertical autocorrelation functions and is defined as

\[ M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} \tag{12} \]

where \( M_j = E[x_j x_j^T] \). Substituting Eqs. 11 and 12 into Eqs. 7, 8 and 9 yields

\[ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} M_{11} \\ M_{12}^T \end{bmatrix} \begin{bmatrix} M_{11} + R \\ M_{22} \end{bmatrix}^{-1} z \tag{13} \]

\[ P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} \tag{14} \]

where \( P = M_j - M_i(M_{i+j} + R)^{-1} M_j \),

The formulation can be extended to the superposition of \( n \) random fields \[5\]. In this study, we examine two cases, the first case uses two random fields, the trend component and the random component, and the other case uses three probability fields, the trend component, the random component, and the white noise. Estimated vector \( w_1 \) is obtained by,

\[ w_1 = M_{1,12}^T[M_{11} + M_{21} + R]^{-1} z \tag{15} \]

where, \( R \) is the covariance matrix of the white noise, the subscript \( k=1, 2 \) indicates trend and random component respectively. The estimate \( x_2 \) is the sum of the trend and random components, i.e. \( x_2 = w_1 + w_2 \). Eqs. 16 and 17 indicate covariance of one-dimensional and three-dimensional data respectively.

\[ \text{cov}_j(s, s') = \sigma_j^2 \times \rho_j(d_j | \delta_{ij}) \tag{16} \]

\[ \text{cov}_j(s, s') = \sigma_j^2 \times \rho_j(d_j | \delta_{ij}) \rho_j(d_j | \delta_{ij}) \tag{17} \]

where, \( \sigma \) is the standard deviation of the random field, \( \rho(d \mid \delta) \) is the autocorrelation function, \( \delta \) is the scale of fluctuation (SOF) \[8\]. The horizontal distance \( d_h \) is obtained from Eq. 18 and the vertical distance \( d_v \) is obtained from Eq. 19 for the distance between the two points of the coordinate values \( s_j \) and \( s_j' \) at two points.

\[ d_h = \sqrt{(s_j - s_j')^2 + (s_s - s_{s'})^2} \tag{18} \]

\[ d_v = |s_j - s_j' | \tag{19} \]

where the subscript \( j \) in the coordinates \( s_j \) and \( s_j' \) is 1, 2 for the horizontal direction and 3 for the vertical direction. We assume that the autocorrelation function in 3D space is represented by the product of the horizontal and vertical autocorrelation functions, which have been adopted as separability in previous studies \[9\], \[10\].

We use the ML method for the estimation of parameters such as SOF and standard deviation in the covariance function. The log-likelihood function \( \ln L \) can be written as

\[ \ln L = -\frac{1}{2} \frac{1}{2} \ln |M_{ij}| + \frac{1}{2} \ln |M_{21} + M_{21} + R| + \frac{m}{2} \ln (2\pi) \tag{20} \]

3. Characteristics of each autocorrelation function and their comparison

Cami et al. \[6\] investigated the autocorrelation functions used in previous literatures in the field of geotechnical engineering and summarized their frequency of use. In this study, the autocorrelation functions of the following Eqs. 21 to 25, which are frequently used, are selected and compared.

Markovian (single exponential)

\[ \rho(d \mid \delta) = \exp \left(-\frac{d}{\delta} \right) \tag{21} \]

Gaussian (squared exponential)

\[ \rho(d \mid \delta) = \exp \left(-\pi \frac{d^2}{\delta^2} \right) \tag{22} \]

Binary noise

\[ \rho(d \mid \delta) = \begin{cases} 1 - \frac{d}{\delta} & \text{if } d < \delta \\ 0, & \text{otherwise} \end{cases} \tag{23} \]

Cosine exponential

\[ \rho(d \mid \delta) = \exp \left(-\frac{d}{\delta} \right) \cos \left(\frac{d}{\delta} \right) \tag{24} \]

Whittle-Matérn

\[ \rho(d \mid \delta) = \frac{2}{\Gamma(\nu)(\nu + 0.5)d^\nu} \left(\frac{2\sqrt{\pi(\nu + 0.5)d}}{\Gamma(\nu)\delta} \right)^{\nu} \tag{25} \]

where, for Eq. 25, \( \Gamma \) is the gamma function, \( K_\nu \) is the modified Bessel function of quadratic \( \nu \), and \( \nu \) is the smoothness parameter (SP). In this study, in addition to these models, we propose the Gaussian-Markovian geometric mean (GMGM) model shown in Eq. 26.

\[ \rho(d \mid \delta) = \left(\exp \left(-\pi \left(\frac{d}{\delta} \right)^2 \right) \right)^\alpha \left(\exp \left(-\frac{d}{\delta} \right) \right)^{1-\alpha} \tag{26} \]

where, \( \alpha \) represents the ratio of Gaussian and Markovian model, which is equivalent to Markovian model when \( \alpha = 0 \), and Gaussian model when \( \alpha = 1 \).

Fig. 1 shows the relationship between the distance and the correlation coefficient when SOF is 1 using these autocorrelation functions. Whittle-Matérn and GMGM have one more parameter than other autocorrelation functions and include Markovian and Gaussian. The binary noise shown in Fig. 1 (2) and the cosine exponential shown in Fig. 1 (3) have correlation characteristics that cannot be expressed by Whittle-Matérn or GMGM.

The correlation coefficient at the small distance has a large effect on the estimation of spatial distribution by GPR. As shown in Fig. 2, the shapes of Whittle-Matérn and GMGM do not match, which means that they have different characteristics.
4. An example of estimation by GPR and evaluation of autocorrelation function

The measured data used for the study are the cone penetration test conducted on the river embankment [11]. Sleeve friction and converted N values are used for the numerical examples. Hereafter, property $P_1$ and $P_2$ denote the sleeve friction and the converted N below. In this study, the methodology for estimation of spatial distribution is focused, and the geotechnical characteristics of the data are not discussed.

Fig. 3 shows the plan view of locations for observation and estimation. Figs. 4 and 5 show spatial distribution of $P_1$ and $P_2$ at each observation location. Observation data are obtained at intervals of 0.05m in the vertical direction. For $P_1$ and $P_2$, model selection of autocorrelation functions is performed for one-dimensional data at point A, and then a three-dimensional data at all four points.

The Gaussian autocorrelation function is used for the trend component [5], and the models of Eqs. 21 to 26 with or without white noise are compared for the modelling of random component. The model selection of the autocorrelation function is performed by the information criterion AIC or BIC.

$$AIC = -2 \ln L + 2n_h$$

$$BIC = -2 \ln L + \ln(m)n_h$$

where, $n_h$ is the number of hyperparameters. Information Criterion AIC and BIC have two terms, and the first term represents the fitness between the model and the data, and the second term represents the complexity of the model. The smaller the value of AIC and BIC, the better the model is.

4.1 1-D spatial distribution

Tabs. 1 and 2 shows the maximum likelihood estimates and AIC and BIC for the data of $P_1$, $P_2$ at point A. The estimated information criterion is shown in order of the number of parameters in the autocorrelation function of the random component, starting from the model with the smallest number of parameters. The Gaussian model is
used for the autocorrelation function of the trend component in all cases as described above. In the table, \( \nu \) and \( \alpha \) are parameters that control smoothness of distribution, which only GMGM and Whittle-Matérn have. If autocorrelation function does not have such parameter, "-" is shown in the table. When white noise is taken into consideration, the standard deviation obtained by the maximum likelihood method is shown.

Figs. 6 and 7 show the AIC and BIC values of autocorrelation models. Information criteria of

| Table 1. MLE, AIC and BIC for each autocorrelation function for \( P_1 \) at point A. |
|-------------------------------|--------------------|-----------------|----------------|----------------|----------------|----------------|
| Random component              | SD                 | Vertical SOF    | \( \nu \) or \( \alpha \) | SD              | Vertical SOF    | AIC             | BIC             | Number of parameters |
| 1: Binary noise               | 0.32               | 0.21            | -              | 0.87            | 10.2            | -843            | -824            | 4                  |
| 2: Markovian                  | 0.28               | 0.40            | -              | 0.97            | 13.3            | -976            | -957            | 4                  |
| 3: Cosine exponential         | 0.25               | 0.18            | -              | 0.91            | 8.85            | -989            | -961            | 4                  |
| 4: Gaussian + white noise     | 0.23               | 0.16            | 0.080          | 0.93            | 8.57            | -946            | -922            | 5                  |
| 5: Binary + white noise       | 0.24               | 0.14            | 0.032          | 1.01            | 7.99            | -921            | -898            | 5                  |
| 6: Markovian + white noise    | 0.28               | 0.40            | \( 2 \times 10^{-5} \) | 0.97            | 13.3            | -974            | -950            | 5                  |
| 7: Cosine exponential + white noise | 0.25         | 0.17            | \( 7.7 \times 10^{-7} \) | 0.91            | 8.65            | -978            | -954            | 5                  |
| 8: Whittle-Matérn            | 0.27               | 0.28            | 0.63           | 0.89            | 10.0            | -982            | -958            | 5                  |
| 9: GMGM                      | 0.26               | 0.21            | 0.46           | 0.90            | 8.90            | -982            | -959            | 5                  |
| 10: Whittle-Matérn + white noise | 0.26         | 0.26            | 0.85           | 0.947           | 9.55            | -981            | -953            | 6                  |
| 11: GMGM + white noise       | 0.26               | 0.21            | 0.46           | \( 5.0 \times 10^{-7} \) | 0.90          | 8.90            | -989            | -952            | 6                  |

| Table 2. MLE, AIC and BIC for each autocorrelation function for \( P_2 \) at point A. |
|-------------------------------|--------------------|-----------------|----------------|----------------|----------------|----------------|
| Random component              | SD                 | Vertical SOF    | \( \nu \) or \( \alpha \) | SD              | Vertical SOF    | AIC             | BIC             | Number of parameters |
| 1: Binary noise               | 10.0               | 2.65            | -              | 22.5            | 28.5            | 2692            | 2711            | 4                  |
| 2: Markovian                  | 3.96               | 1.39            | -              | 22.0            | 29.4            | 2389            | 2408            | 4                  |
| 3: Cosine exponential         | 2.96               | 0.40            | -              | 21.4            | 21.1            | 2377            | 2396            | 4                  |
| 4: Gaussian + white noise     | 2.55               | 0.20            | 0.59           | 20.3            | 3.62            | 2438            | 2461            | 5                  |
| 5: Binary + white noise       | 10.5               | 3.05            | \( 1.5 \times 10^{4} \) | 22.2            | 33.4            | 2664            | 2688            | 5                  |
| 6: Markovian + white noise    | 3.94               | 1.37            | \( 1.4 \times 10^{4} \) | 20.0            | 20.2            | 2391            | 2414            | 5                  |
| 7: Cosine exponential + white noise | 2.96         | 0.40            | \( 1.3 \times 10^{4} \) | 21.4            | 21.1            | 2379            | 2402            | 5                  |
| 8: Whittle-Matérn            | 3.66               | 0.60            | 0.72           | 21.0            | 26.5            | 2338            | 2361            | 5                  |
| 9: GMGM                      | 3.35               | 0.33            | 0.74           | 21.1            | 21.3            | 2332            | 2356            | 5                  |
| 10: Whittle-Matérn + white noise | 3.55         | 0.48            | 0.99           | 0.37            | 19.9            | 23.4            | 2334            | 2362            | 6                  |
| 11: GMGM + white noise       | 3.35               | 0.33            | 0.74           | \( 5.0 \times 10^{-7} \) | 21.1          | 21.3            | 2334            | 2362            | 6                  |

Figure 6. Model selection of autocorrelation for \( P_1 \) at point A.  
Figure 7. Model selection of autocorrelation for \( P_2 \) at point A.
autocorrelation number 8 (Whittle-Matérn) and 9 (GMGM) are smaller than the others. They have one more parameter than other autocorrelation functions. There are few models whose evaluation is improved by adding white noise. According to Tabs.1 and 2, BIC is minimal when 3 (Cosine exponential) is used for P1 at point A, and 9 (GMGM) is used for P2 at point A.

Fig. 8 shows the maximum likelihood estimates and the contour maps of the negative log-likelihoods around them when the autocorrelation of the smallest BIC is used. Fig.8 (1) and (2) are contour maps of the SD and SOF, and α and SOF of random components. The maximum likelihood estimates matches the minimum points of the contour maps naturally. The trend and random components are separately shown using the autocorrelation function of the smallest BIC in Fig. 9.

### 4.2 3-D spatial distribution

Using the boring data at four points A, B, C, and D, the three-dimensional spatial distribution is estimated. Tabs. 3 and 4 shows the maximum likelihood estimates, AIC and BIC for P1 and P2 respectively. The 4 types of autocorrelation functions are selected from the one-dimensional study. The cases with and without white noise are considered for each autocorrelation functions. Gaussian autocorrelation is used for the trend components as in the one-dimensional study. Since the number of boring data is small, the horizontal SOF of the trend component is assumed to be 80 (m) referring Cami et al. [6].

#### Table 3. MLE for each autocorrelation function and AIC, BIC for P1 in 3D space.

<table>
<thead>
<tr>
<th>Autocorrelation</th>
<th>Random component</th>
<th>White noise</th>
<th>Trend component (Gaussian)</th>
<th>AIC</th>
<th>BIC</th>
<th>Number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>v or α</td>
<td>SD Vertical SOF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1: Markovian</td>
<td></td>
<td>4.11</td>
<td>5.99 1.51</td>
<td>10.3</td>
<td>15.3</td>
<td>9469 9100 5</td>
</tr>
<tr>
<td>2: Cosine exponential</td>
<td></td>
<td>3.19</td>
<td>6.05 0.47</td>
<td>16.7</td>
<td>17.6</td>
<td>9414 9445 5</td>
</tr>
<tr>
<td>3: Markovian + white noise</td>
<td></td>
<td>4.10</td>
<td>6.00 1.51</td>
<td>15.2</td>
<td>18.1</td>
<td>9468 9594 5</td>
</tr>
<tr>
<td>4: Cosine exponential + white noise</td>
<td></td>
<td>3.19</td>
<td>6.05 0.48</td>
<td>16.7</td>
<td>17.5</td>
<td>9416 9453 6</td>
</tr>
<tr>
<td>5: Whittle-Matérn</td>
<td></td>
<td>3.97</td>
<td>5.39 0.58</td>
<td>14.9</td>
<td>17.6</td>
<td>9024 9060 6</td>
</tr>
<tr>
<td>6: GMGM</td>
<td></td>
<td>3.63</td>
<td>4.26 0.30</td>
<td>15.7</td>
<td>16.4</td>
<td>9013 9049 6</td>
</tr>
<tr>
<td>7: Whittle-Matérn + white noise</td>
<td></td>
<td>3.57</td>
<td>6.28 0.41</td>
<td>15.7</td>
<td>17.5</td>
<td>8979 9021 7</td>
</tr>
<tr>
<td>8: GMGM + white noise</td>
<td></td>
<td>3.63</td>
<td>4.00 0.30</td>
<td>15.0</td>
<td>17.0</td>
<td>9015 9058 7</td>
</tr>
</tbody>
</table>

#### Table 4. MLE for each autocorrelation function and AIC, BIC for P2 in 3D space.

<table>
<thead>
<tr>
<th>Autocorrelation</th>
<th>Random component</th>
<th>White noise</th>
<th>Trend component (Gaussian)</th>
<th>AIC</th>
<th>BIC</th>
<th>Number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>v or α</td>
<td>SD Vertical SOF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1: Markovian</td>
<td></td>
<td>0.29</td>
<td>5.17 0.47</td>
<td>0.90</td>
<td>12.8</td>
<td>-4144 -4114 5</td>
</tr>
<tr>
<td>2: Cosine exponential</td>
<td></td>
<td>0.26</td>
<td>5.47 0.29</td>
<td>1.14</td>
<td>10.1</td>
<td>-4190 -4160 5</td>
</tr>
<tr>
<td>3: Markovian + white noise</td>
<td></td>
<td>0.29</td>
<td>5.16 0.47</td>
<td>5.3 x 10^4</td>
<td>0.90</td>
<td>12.8</td>
</tr>
<tr>
<td>4: Cosine exponential + white noise</td>
<td></td>
<td>0.26</td>
<td>1.21 0.29</td>
<td>1.6 x 10^3</td>
<td>1.09</td>
<td>10.33</td>
</tr>
<tr>
<td>5: Whittle-Matérn</td>
<td></td>
<td>0.29</td>
<td>5.26 0.31</td>
<td>1.18</td>
<td>13.7</td>
<td>-4221 -4184 6</td>
</tr>
<tr>
<td>6: GMGM</td>
<td></td>
<td>0.26</td>
<td>3.93 0.23</td>
<td>1.05</td>
<td>11.5</td>
<td>-4236 -4196 6</td>
</tr>
<tr>
<td>7: Whittle-Matérn + white noise</td>
<td></td>
<td>0.28</td>
<td>6.02 0.28</td>
<td>0.049</td>
<td>1.05</td>
<td>13.2</td>
</tr>
<tr>
<td>8: GMGM + white noise</td>
<td></td>
<td>0.29</td>
<td>3.22 0.27</td>
<td>0.91</td>
<td>11.6</td>
<td>-4231 -4189 7</td>
</tr>
</tbody>
</table>
As shown in Figs. 10 and 11, information criteria of autocorrelation function 5 (Whittle-Matérn) and 6 (GMGM), which has one more parameter than other autocorrelation functions, are smaller than the others. In the three-dimensional spatial distribution, few models improve the information criteria by adding white noise. When GMGM and Whittle-Matérn are used, the information criteria is low in all 1D and 3D cases. From Tabs. 3 and 4, BIC is minimal when GMGM is used for P1 and Whittle-Matérn + white noise is used for P2.

Figs. 12 and 13 show the maximum likelihood estimates and the contour map of the negative log-likelihood around them when the autocorrelation function with the smallest BIC is used for each of P1 and P2. These are contour map on the parameters of random components. The sensitivity of SOF in the horizontal direction is low.

Using these autocorrelation functions, the spatial distributions of the trend and random component of the a-a’ cross sections are estimated and shown in Figs. 14 and 15 for P1, P2. The trend component shows the distribution that gradually increases in the vertical direction, and the random component shows small fluctuation depending on the observed data.

5. Conclusion

Gaussian process regression with multiple random fields is one of useful tool for estimation of spatial distribution of soil properties. In this study, we investigate a suitable autocorrelation model of random component.
general, Whittle-Matérn and GMGM give good results in term of information criteria. Since these models have more parameters than other autocorrelation functions, it is possible to control the local smoothness in addition to the fluctuation of the overall distribution. We also examine usefulness of the addition of white noise, but there is almost no improvement in the information criteria.

GMGM proposed in this study has almost the same performance as Whittle-Matérn. GMGM does not require a special function such as the Bessel function. It is considered to be useful from the viewpoint of simplicity and computation cost. In the future study, we plan to examine the applicability of GMGM to various types of actual measured data.

References


