

Northumbria Research Link

Citation: Qi, Xiaohui, Wang, Hao, Chu, Jian and Chiam, Kiefer (2022) Effect of autocorrelation function model on spatial prediction of geological interfaces. Canadian Geotechnical Journal, 59 (4). pp. 583-600. ISSN 0008-3674

Published by: Canadian Science Publishing

URL: <https://doi.org/10.1139/cgj-2020-0644> <<https://doi.org/10.1139/cgj-2020-0644>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/47659/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

1 **Effect of autocorrelation function model on spatial prediction of geological interfaces**

2 Xiaohui Qi^{1,2}, Hao Wang², Jian Chu^{2*}, Kiefer Chiam³

3 ¹Department of Mechanical and Construction Engineering, Northumbria University, Newcastle
4 upon Tyne, UK, NE1 8ST

5 ²School of Civil and Environmental Engineering, Nanyang Technological University, 639798,
6 Singapore

7 ³Building and Construction Authority, 200 Braddell Road, 579700, Singapore

8 * Corresponding author

9 E-mail addresses: xiaohui.qi@northumbria.ac.uk (X. Qi), wang.hao@ntu.edu.sg (H. Wang),
10 cjchu@ntu.edu.sg (J. Chu), Kiefer_CHIAM@bca.gov.sg (K. Chiam)

11 **Abstract:** This study evaluated the performances of various autocorrelation function (ACF)
12 models in predicting the geological interface using a well-known conditional random field method.
13 Prediction accuracies and uncertainties were compared between a flexible Matérn model and two
14 classical ACF models: the Gaussian model and the single exponential model. The rockhead data
15 of Bukit Timah granite from boreholes at two sites in Singapore as well as simulated data were
16 used for the comparisons. The results showed that the classical models produce a reasonable
17 prediction uncertainty only when its smoothness coefficient is consistent with that of the geological
18 data. Otherwise, the classical models may produce prediction errors much larger than that of the
19 Matérn model. On the other hand, the prediction accuracy of the Matérn model is affected by the
20 spacing of the boreholes. When the borehole spacing is relatively small ($< 0.4 \times$ scale of
21 fluctuation), the Matérn model can reasonably quantify the prediction uncertainty. However, when
22 the borehole spacing is large, the prediction by the Matérn model becomes less accurate as
23 compared with the prediction using the classical models with the right value of smoothness
24 coefficient due to the large estimation error of the smoothness coefficient.

25 **Keywords:** Spatial prediction, autocorrelation function, Matérn autocorrelation model, geological
26 interface

27 **1. Introduction**

28 The interface of geological formations or layers, such as the soil-rock interface or the so-called
29 rockhead, is a major factor to be considered in tunnel constructions as the location of the interface
30 will affect the selection of construction methods or supporting measures. The location of the
31 geological interface at a construction site can vary considerably due to the complex tectonics
32 events and environmental effects experienced by geological layers and yet can only be detected
33 using a limited number of boreholes. It is thus a challenging task to identify the location of the
34 geological interface in areas between the boreholes. This is particularly the case when the borehole
35 spacing is too large. Therefore, it is necessary to establish an accurate spatial prediction method to
36 identify the location of the geological interfaces.

37 Various methods have been proposed to address the spatial prediction problems, such as the
38 coupled Markov chain method (Qi et al. 2016), the Bayesian compressive method (Wang and Zhao
39 2016, 2017; Wang et al. 2020a), multivariate adaptive regression spline and neural network
40 methods (e.g., Zhang and Goh 2013, 2016; Wang et al. 2020b), as reviewed in Qi et al. (2020a, b).
41 Among them, the geostatistical methods, such as the kriging and the random field method, have
42 acquired a wider application, as shown in Dasaka and Zhang (2012), Lloret-Cabot et al. (2012),
43 Firouziandbandpey et al. (2015), Li et al. (2016a), Li et al. (2016b), Lo and Leung (2017), Cai et al.
44 (2019), Qi and Liu (2019a) and Qi et al. (2020a). The reason may be that the geostatistical methods
45 can provide a predicted value as well as reasonably quantify the prediction uncertainty. Another
46 advantage of the geostatistical method over some aforementioned methods such as the spline
47 regression method is that the predicted curve or surface runs across the observed data points when
48 there is no measurement error. This feature is vital to engineers who need a prediction that is
49 consistent with observed data. One disadvantage of the geostatistical methods is that it requires

50 relatively large quantities of data to infer the spatial variability information. However, this problem
51 can be alleviated by adopting a reasonable borehole layout scheme and a resampling method, as
52 shown in Qi and Liu (2019b).

53 The main parameters involved in the geostatistical method for characterizing the spatial
54 variability of the underlying engineering properties include the mean, standard deviation, and scale
55 of fluctuation. The last one is a parameter of the autocorrelation function (ACF), which depicts
56 how the spatial autocorrelation decays with an increasing distance between two points. These three
57 random field parameters are commonly regarded to be the dominating factors that affect the spatial
58 prediction, and thus were the focus of parametric studies (e.g., Firouziandbandpey et al. 2015).
59 However, rare attention was paid to the role of the ACF type in the spatial prediction of geological
60 parameters using geostatistical methods. For geostatistical methods, a predicted value at a target
61 point can be regarded to be the weighted average value of neighboring observations. The weights
62 of the observed points depend primarily on the autocorrelation between the target point and
63 observed points. Different ACFs produce different autocorrelations and will induce different
64 weights and predicted values. Also, it is expected that the different ACFs lead to different
65 prediction uncertainties. This aspect, however, has rarely been addressed in existing studies. The
66 effect of the ACF model on spatial predictions should be investigated.

67 In geotechnical engineering, the commonly used ACFs are mainly the classical models, such
68 as the single exponential model, Gaussian model, and second-order Markov model as shown in
69 Ching et al. (2019). These classical models contain only one parameter, namely the scale of
70 fluctuation to represent the degree of spatial autocorrelation. Besides the classical models, another
71 group of models, the so-called non-classical ACF models, has also been used to model the
72 autocorrelation property of spatial or temporal processes (Ching et al. 2019). One example of such

73 a model is the Matérn model proposed by Matérn (1960). This model has an additional parameter,
74 the smoothness coefficient, to control the smoothness (or degree of differentiability) of random
75 fields or processes. Like other random field parameters, the smoothness parameter can be
76 estimated using the maximum likelihood method, as shown in Ching et al. (2019). Most of the
77 classical ACF models can be considered as a special case of the Matérn model with a prescribed
78 value of the smoothness coefficient. Therefore, the main advantage of the Matérn model over the
79 classical models is that it can automatically capture the smoothness of the engineering properties.
80 Also, the usage of the Matérn model avoids the difficulty of the ACF model selection. The Matérn
81 model has been used in spatial predictions in the area of soil science and mathematics (e.g., Stein
82 1999). However, it is rarely adopted in geotechnical and geological engineering, possibly because
83 in these areas the model was published much after the popular random field theory (e.g.,
84 Vanmarcke 1977, 1983). Nevertheless, as a few exceptions, the Matérn model has been used to
85 interpolate the geological properties (such as Liu et al. 2017), but its advantages over the classical
86 models are not well understood. In the area of mathematics, Stein (1999) illustrated that the Matérn
87 model produced smaller errors than the Gaussian model. But the study of Stein (1999) mainly used
88 the mean squared error to measure the prediction accuracy and thus cannot reflect the rationality
89 of the prediction uncertainty.

90 This paper studied the effect of the ACF model on the spatial prediction of the location of
91 rockhead. Borehole data from two sites in Singapore were used to evaluate the effect of the ACF
92 model on spatial predictions using the conditional random field method. These borehole data
93 reveal the rockhead of an igneous rock formation, Bukit Timah granite. The Matérn model was
94 compared with two widely used ACF models in geotechnical engineering, the Gaussian model and
95 the single exponential model. The prediction accuracies and prediction uncertainties for different

96 types of ACFs were assessed through a cross-validation procedure. Furthermore, simulated data
97 were also analyzed to consider the effect of data spacing and the smoothness of the geological
98 data.

99 **2. Conditional random field method**

100 This study used the conditional random field method (termed as CRF hereafter) to spatially predict
101 the rockhead elevations at unsampled locations. This method is a well-established geostatistical
102 method and has been widely used to predict geological or geotechnical properties (e.g., Li et al.
103 2016b; Lo and Leung 2017; Li et al. 2018; Cai et al. 2019; Qi and Liu 2019b; Qi et al. 2020). The
104 conditional random field method is a Bayesian method essentially. Sometimes a “Bayesian-based”
105 modifier is added in front of the “conditional random field method” to differentiate itself from
106 other conditional random field methods. For simplicity, the “Bayesian-based” modifier is not used
107 in this study. The basic idea of the method is to update the distributions of the unknown parameters
108 at unsampled locations using (i) the prior information regarding the probability distribution of the
109 unknown rockhead elevations and (ii) spatial autocorrelation information between rockhead
110 elevations at sampled and unsampled locations. The statistics for the prior distribution and updated
111 distribution of the unknown parameters are called prior statistics and posterior statistics,
112 respectively. The prior statistics and autocorrelation information of the unknown parameters are
113 expressed by a random field model. In this study, the random field of the rockhead elevation is
114 assumed to possess a mean trend represented by a polynomial function and a constant standard
115 deviation across the area of interest. The autocorrelation of rockhead elevations at various
116 locations is described by an autocorrelation function. The general steps of the CRF method are
117 listed as follows. More details of the conditional random field method can be found in Li et al.
118 (2016b), Lo and Leung (2017), and Qi et al. (2020b).

- 119 (1) Estimate random field parameters with borehole data using the maximum likelihood
120 estimation method, as shown in DeGroot and Baecher (1993), Fenton (1999), and Ching et
121 al. (2019). The random field parameters include regression coefficients of the mean trend
122 function, the standard deviation, and the scales of fluctuation (and smoothness coefficients
123 if the Matérn model is used). In this study, polynomial trend functions with degrees up to 3
124 are considered as higher degrees with too many unknown parameters make the estimation
125 uncertainty in the trend parameters increase significantly (Baecher and Christian 2003). Note
126 that the random field parameters can also be determined based on prior knowledge if it is
127 available. Besides, the moment method is not used to estimate the random field parameters
128 as it is not suitable for cases with irregularly spaced data.
- 129 (2) The optimal degree of the polynomial trend function is selected using a well-known model
130 selection method, Bayesian information criterion (BIC) (e.g., Yuen et al. 2010). The BIC
131 considers the fitting capacity and the complexity of various models. The model with a
132 minimum BIC value is regarded to be the optimal model and the associated random field
133 parameters are used for spatial predictions.
- 134 (3) Update the posterior statistics of the rockhead elevations at unexplored locations using the
135 prior statistics, observed rockhead elevations and autocorrelation information of the
136 rockhead elevation (e.g., Li et al. 2016b, Lo and Leung 2017). Herein the prior statistics of
137 the rockhead elevation including prior mean and standard deviation are the mean trend and
138 standard deviation for the optimal model determined in steps 1-2.

139 This study considers three commonly-used two-dimensional autocorrelation models: the
140 separable single exponential (SSE) model, the Gaussian model (also called the squared exponential
141 model), and the Matérn model, given by

$$142 \quad \text{Separable single exponential: } \rho_{\text{SSE}}(\Delta x_{ij}, \Delta y_{ij}) = \exp\left(-\frac{2|\Delta x_{ij}|}{\text{SoF}_x} - \frac{2|\Delta y_{ij}|}{\text{SoF}_y}\right) \quad (1a)$$

$$143 \quad \text{Gaussian: } \rho_{\text{G}}(\Delta x_{ij}, \Delta y_{ij}) = \exp\left(-\pi\left(\frac{\Delta x_{ij}}{\text{SoF}_x}\right)^2 - \pi\left(\frac{\Delta y_{ij}}{\text{SoF}_y}\right)^2\right) \quad (1b)$$

$$144 \quad \text{Matérn: } \rho_{\text{M}}(\Delta x_{ij}, \Delta y_{ij}) = \frac{2^{1-\nu_x}}{\Gamma(\nu_x)} \times \left(\frac{|\Delta x_{ij}|}{\xi_x}\right)^{\nu_x} K_{\nu_x}\left(\frac{|\Delta x_{ij}|}{\xi_x}\right) \times \frac{2^{1-\nu_y}}{\Gamma(\nu_y)} \times \left(\frac{|\Delta y_{ij}|}{\xi_y}\right)^{\nu_y} K_{\nu_y}\left(\frac{|\Delta y_{ij}|}{\xi_y}\right) \quad (1c)$$

145 where ρ_{SSE} , ρ_{G} , and ρ_{M} are the autocorrelation coefficients for the SSE, Gaussian, and Matérn
 146 autocorrelation functions, respectively; $|\Delta x_{ij}| = |x_i - x_j|$ and $|\Delta y_{ij}| = |y_i - y_j|$ are the distances (also
 147 called lags) between two points (x_i, y_i) and (x_j, y_j) in the x and y directions, respectively (note that
 148 x and y denote two orthogonal horizontal directions); SoF_x and SoF_y are the scales of fluctuations
 149 in the directions of x and y ; ν_x and ν_y are the smoothness coefficients; $\Gamma(\cdot)$ is the Gamma function;
 150 ξ_x and ξ_y are scale parameters, and $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind of order
 151 ν . For the Matérn model, the scale of fluctuations is related to the scale parameters through
 152 (Hristopulos and Žukovič 2011).

$$153 \quad \xi_x = \frac{\text{SoF}_x \times \Gamma(\nu_x)}{2\sqrt{\pi} \times \Gamma(\nu_x + 0.5)} \quad (1d)$$

$$154 \quad \xi_y = \frac{\text{SoF}_y \times \Gamma(\nu_y)}{2\sqrt{\pi} \times \Gamma(\nu_y + 0.5)} \quad (1e)$$

155 The Matérn model is a flexible ACF model in the sense that it has an additional parameter
 156 controlling the smoothness (degree of differentiability) of the random field. Other commonly used
 157 ACF models, such as the SSE and Gaussian model are special cases of the Matérn model. The SSE
 158 and Gaussian models correspond to the smoothness coefficient $\nu = 0.5$ and $+\infty$, respectively. To
 159 illustrate the effect of the smoothness coefficient, Fig. 1(a) plots the Matérn ACF with different

160 smoothness coefficients using the same value of the scale of fluctuation (SoF). As shown, a larger
161 smoothness coefficient exhibits a higher autocorrelation coefficient at small lags, such as the
162 separation distance $< 0.5 \times SoF$, which decays faster with the increased separation distance. Fig.
163 1(b) plots typical one-dimensional realizations of the rockhead-elevation random field with
164 different values of ν using the same SoF . As shown, the rockhead curve becomes smoother as the
165 value of the ν increases.

166 When the Matérn model is used, an upper limit value such as 50 is set for the smoothness
167 coefficient parameter, ν , in the maximum likelihood optimization. The reason is that sometimes
168 the value of ν keeps increasing during the optimization, indicating the optimized value of ν is $+\infty$
169 (corresponding to the Gaussian model). When the ν reaches a relatively large value such as 50, the
170 corresponding ACF is already very close to a Gaussian ACF, and a further increase of ν hardly
171 changes the shape of the ACF. Therefore, $\nu = 50$ is taken as a sign that the estimated ACF model
172 is a Gaussian model. By setting this upper limit value of ν , the computational time can be
173 significantly reduced. Matlab codes regarding the maximum likelihood estimation of random field
174 parameters are provided in appendix A.

175 **3. Borehole data**

176 Borehole data from two sites in Singapore are used to perform the spatial predictions. Site 1 is
177 around Yishun Park and Site 2 is at the Canberra Link (Sembawang). Geological cross-sections
178 from the two sites are plotted in Fig. 2(a) and 2(b), respectively. As shown, the main geological
179 formation present at the two sites is the igneous rock formation, Bukit Timah granite (BTG). The
180 BTG covers around one-third of Singapore Island and is considered to be the base bedrock of
181 Singapore Island. The BTG was formed from a large body of acid mass, which intruded from
182 greater depths (Pitts 1984; Sharma et al. 1999). The fresh material of the BTG has good

183 engineering properties (Pitts 1984; Sharma et al. 1999). Thus, the BTG-occupied area provides an
184 ideal space for underground constructions. However, due to the humid tropical climate and the
185 past tectonic events in Singapore, the BTG has undergone an intensive weathering. As such, the
186 location of the interface of soil and rock layers, or the so-called rockhead, in the BTG may vary
187 significantly even within a short distance. Herein the soil and rock layer in a rock formation are
188 distinguished mainly according to the weathering degree of the rock masses. In the engineering
189 practice of Singapore, the weathering degrees of the BTG are classified into six grades: fresh
190 (grade I), slightly weathered (grade II), moderately weathered (grade III), highly weathered (grade
191 IV), completely weathered (grade V), and residual soil (grade VI) (e.g., Zhao et al. 1994). Grades
192 I, II, and III are regarded as rock, whereas Grades IV, V, and VI as soil-like materials. Since the
193 soil layers have weaker engineering properties than the rock layers, it is essential to identify the
194 location of the rockhead in order to determine a suitable excavation method or supporting scheme
195 in underground constructions. It is worth noting that the geological conditions can be quite variable
196 and sometimes a layer having a weathering grade of I, II or III overlie a soil layer (Qi et al. 2020a,
197 b). In this case, the rock layer may likely be boulders or rock intrusion.

198 The borehole data for the two sites are plotted in Fig. 2(c) and 2(d), respectively. In Figs. 2(c)
199 and 2(d), the rockhead data are plotted together with the rockhead surface linearly interpolated
200 using a MATLAB function, *fit*. As shown, 188 and 135 boreholes were distributed in the two sites
201 with an area of 1600 m × 600 m and 600 m × 400 m, respectively. The rockhead elevation for
202 boreholes at Site 1 ranges from – 48.0 m to 3.0 m while that at Site 2 ranges from – 50.8 m to 3.6
203 m. The elevation herein refers to the height relative to the average sea level measured from 1935
204 to 1937 at Victoria Dock located in the south of Singapore. The standard deviation of the rockhead
205 elevation data at Site 1 is 9.8 m while that at Site 2 is 11.6 m. Evidently, the rockhead elevation

206 has a large variability in these two areas. The coordinates of the boreholes and associated rockhead
 207 elevations at Site 1 are listed in Table A1 in appendix B while those for the other site can be found
 208 in Qi et al. (2020b). The rockhead elevations were obtained from site investigation reports.

209 **4. Spatial prediction of rockhead elevation using actual borehole data**

210 This section presents comparisons of the performances of three selected ACF models using the
 211 borehole data at the two sites mentioned above. The comparisons are carried out using a cross-
 212 validation procedure composed of the following steps:

- 213 1) All the rockhead data are randomly divided into two groups, training group and testing
 214 group.
- 215 2) The training data are used to estimate random field parameters and perform spatial
 216 prediction using the CRF method. The CRF provides a predicted value as well as a posterior
 217 standard deviation (or confidence interval) for the rockhead elevation at any testing point.
- 218 3) The prediction errors are evaluated by comparing estimated rockhead elevations at the
 219 locations of testing data with the corresponding actual values. The rationality of the
 220 prediction uncertainties is also assessed.
- 221 4) The above three steps are repeated a few times to evaluate the prediction performance for
 222 various ACF models under various scenarios.

223 The prediction error is quantified by the two indexes, root mean squared error, *RMSE*, and root
 224 mean squared relative error, *RMSRE*, while the rationality of the prediction uncertainty is
 225 quantified by the normalized prediction variance, *NPV* (Cressie 1993), as shown by.

$$226 \quad RMSE = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (\hat{z}_i - z_i)^2} \quad (2a)$$

$$227 \quad RMSRE = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} \left(\frac{\hat{z}_i - z_i}{d_i} \right)^2} \quad (2b)$$

$$NPV = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\frac{\hat{z}_i - z_i}{\sigma_{z_i}} \right)^2 \quad (2c)$$

where N_p is the number of data points in the testing group; \hat{z}_i is the predicted value of the rockhead elevation for the i th testing point, Z_i , which is treated as a random variable; z_i is the observed value of the rockhead elevation Z_i , d_i is the actual depth of the rockhead for the i th testing point, and σ_{z_i} is the predicted standard deviation for the rockhead elevation Z_i . The former two indexes quantify the prediction error, and the last one quantifies the rationality of the prediction uncertainty. Note that the depth rather than actual elevation is used in the denominator of Eq. 2(b) because the elevation may have a value close to 0. Elevation rather than depth is used as the geological parameter investigated in this study as the former is a relatively constant quantity that is immune to erosion and human activities, such as excavation and backfilling.

It is worth noting that the optimal value of the NPV is 1. The meaning of NPV can be well appreciated by considering a random variable N with a mean of μ and a variance of σ^2 . If one generates many samples (such as m) of the variable N , n_1, n_2, \dots, n_m , based on the definition, the variance of the random variable can be estimated from the samples, i.e., $\sigma^2 \approx \hat{\sigma}^2 =$

$\frac{1}{m} \sum_{i=1}^m (n_i - \mu)^2$, making the statistic $\frac{1}{m} \sum_{i=1}^m \left(\frac{n_i - \mu}{\hat{\sigma}} \right)^2$ close to 1. However, if another estimation

of σ^2 , σ'^2 , which is estimated from very limited samples, is smaller than the actual value, the

statistic $\frac{1}{m} \sum_{i=1}^m \left(\frac{n_i - \mu}{\sigma'} \right)^2$ would be larger than 1. Hence, an NPV with a value much larger than 1 is

an indication that the uncertainty of a rockhead elevation is underestimated. On the contrary, an

NPV with a value much smaller than 1 means that the uncertainty of the rockhead elevation is

overestimated.

248 The spatial prediction was first performed using all the rockhead data to obtain an overall
249 picture of the rockhead trend in the area. Then the prediction errors and prediction uncertainties
250 for various ACF models are compared using the cross-validation procedure. The rockhead data at
251 the two sites were analyzed in sections 4.1 and 4.2, respectively.

252 **4.1 Bukit Timah granite rockhead at Site 1**

253 **4.1.1 Rockhead surface predicted using all the data**

254 The random field parameters of the rockhead elevation at Site 1 were estimated by the maximum
255 likelihood estimation method using all the rockhead data. The BIC values for the selection of the
256 suitable trend functions were summarized in Table 1 (a) while the estimated values for the most
257 suitable model are summarized in Table 1 (b). Correspondingly, the autocorrelation functions for
258 various models are plotted in Fig. 3(a, b). As shown in Table 1(a), 10 polynomial functions with
259 degrees up to 3 are considered. The constant mean model has the minimum value of BIC and is
260 the optimal form of trend function. As shown in Table 1(b), the estimated values of the smoothness
261 coefficients of the Matérn model, ν_x and ν_y , are close to 0.5, indicating the actual ACF model is
262 likely to be the SSE model.

263 Another observation in Table 1(a) is that the scale of fluctuation (SoF) for the Gaussian
264 model is much smaller than those for the other two models. To explore the reason, Fig. 3(c) plots
265 the borehole pairs with small lags in the x -direction (i.e., $|\Delta x| < 10$ m) while Fig. 3(d) plots the
266 frequencies of y -lags for the borehole pairs shown in Fig. 3(c). In Fig. 3(d), the height of each bin
267 denotes the number of data pairs with the y -lag, $|\Delta y|$ falling within a certain range, such as [0 m,
268 10 m] and [10 m, 20 m]. As shown, the $|\Delta y|$ interval of [20 m, 30 m] has the most (i.e., 27) data
269 pairs, indicating the autocorrelation coefficients for $|\Delta y| = [20 \text{ m}, 30 \text{ m}]$ is likely to dominate the
270 maximum likelihood estimations. Additionally, Fig. 3(e) plots the rockhead elevation associated

271 with the 27 data pairs. As shown, the 27 data pairs are highly correlated and have a relatively high
272 autocorrelation coefficient, 0.64. Fig. 3(f) plots two autocorrelation functions with an
273 autocorrelation coefficient of 0.64 at $|\Delta y| = 25$ m. As shown, to achieve such a high
274 autocorrelation coefficient for $|\Delta y| = 25$ m, the SSE model has to adopt a relatively high value of
275 SoF_y , i.e., 112 m. By contrast, the Gaussian model just needs a SoF_y of 66 m to achieve the same
276 autocorrelation coefficient for $|\Delta y| = 25$ m. Note that the 112 m and 66 m are not the same as the
277 corresponding values in Table 1(b) because only the data pairs with a data spacing range of 20 m
278 ~ 30 m are considered herein to make the explanation more understandable. The horizontal scale
279 of fluctuation can be explained in the same way. Furthermore, a similar phenomenon is observed
280 by Qi and Liu (2019), which found if the Gaussian autocorrelation function is used to estimated
281 random field parameters for soil properties with a single exponential autocorrelation structure, the
282 estimated autocorrelation distance is smaller than the actual value when the data spacing is
283 relatively small. It must be noted that the analysis here may not be a good quantitative
284 representation because the borehole data are not sufficient and regularly spaced, making it difficult
285 to evaluate accurately the autocorrelation coefficient for a certain lag. For example, the
286 autocorrelation coefficient evaluated in Fig. 3(e) is just an approximate solution of the
287 autocorrelation coefficient for $\Delta x = 0$ m and $\Delta y = 25$ m because all the data pairs in Fig. 3(e) have
288 slightly different x coordinates (i.e., $\Delta x \neq 0$) and the actual Δy varies from 20 m to 30 m.

289 Figs. 4(a, b, c) plot the predicted surface of the rockhead for various models at Site 1 while
290 Fig.4(d) and 4(e) plot the predicted curves and the 95% confidence intervals of the rockhead
291 elevation along $y = 260$ m, respectively. The 95% confidence intervals (CIs) of the rockhead
292 elevation in Fig. 4(e) are evaluated by assuming the rockhead elevation has a normal distribution.
293 Hence, the 95% CI is bounded by the mean value $\pm 1.96 \times$ posterior standard deviations. It can be

294 seen from Fig. 4(d) that when a point deviates away from the observed data, the corresponding
295 predicted elevation gradually approaches the prior mean value for all the models, i.e., -18.1 m, -
296 18.5 m, and -18.4 m for the SSE, Gaussian, and Matérn model respectively, as given in Table 1(b).
297 Besides, the known points have a relatively small influential range when the Gaussian model is
298 used. This means that the predicted elevation quickly converges to the prior mean trend when a
299 target point gradually deviates from a known point, as shown by the green line segment with 150
300 $m < x < 550$ m in Fig. 4(d). By contrast, for the other two models, convergences of the predicted
301 curves of rockhead are not achieved in the range, $150 \text{ m} < x < 550$ m. Similarly, as a target point
302 gradually deviates from a known point, the 95% CI for the Gaussian model converges to a steady
303 level faster than those for the other two models, as shown in Fig. 4(e). As a result, the average
304 width of the 95% CI for the Gaussian model, 29.2 m, is also larger than those for the other two
305 models, i.e., 26.7 m for both. The main reason for these phenomena is that the Gaussian model
306 produced smaller values of estimated *SoF* (see Table 1(b)) and the autocorrelation decays much
307 faster than the other two models, as shown in Fig. 3(b).

308 **4.1.2 Comparison of predictions using different ACF models**

309 The prediction accuracies and uncertainties for various ACF models are compared using the cross-
310 validation procedure described at the beginning of section 4. 100 rounds of cross-validations were
311 performed as the associated results are quite close to those for 200 rounds. In each round, 70% of
312 data were randomly drawn as training data and the remainder as testing data. Each round of cross-
313 validation produces one set of indexes, i.e., *RMSE*, *RMSRE* and *NPV*. The means and standard
314 deviations of the 100 sets of indexes are given in Table 2. As shown, the mean values of *RMSE*
315 and *RMSRE* associated with the Gaussian model are 7.67 m and 0.34 which are larger than the
316 values for the SSE model (6.66 m and 0.27) and the values for the Matérn model (6.73 m and

317 0.29). The higher mean *RMSE* value for the Gaussian model indicates a lower accuracy of this
318 model. Moreover, the mean value of *NPV* for the Gaussian model is 1.48 which is larger than 1,
319 indicating the uncertainties of the rockhead elevations at testing points are significantly
320 underestimated. To explore the reasons, one typical cross-validation example is plotted in Fig. 5.
321 The plan views of the training and testing boreholes and the autocorrelation coefficients for various
322 models are plotted in Figs. 5(a, b). The predicted curves and 95% CIs of the rockhead along a
323 cross-section are plotted in Figs. 5(c-e). The following has been observed from Figs. 5.

324 (1) The Gaussian model has larger prediction errors than the other two models. The *RMSE* for
325 the former is 6.92 m which the *RMSE* for the *SSE* and Matérn models are 6.29 m and 6.35 m,
326 respectively. In particular, when the Gaussian model is used, the prediction errors for the
327 testing points at $(x, y) = (815 \text{ m}, 261 \text{ m})$ and $(1145 \text{ m}, 261 \text{ m})$ are -15.5 m and -9.6 m (see
328 Fig. 5(d)), the absolute values of which are much larger than those for the *SSE* Model (- 5.8
329 m and - 4.1 m as shown Fig. 5(c)) or those for the Matérn model (-7.9 m and - 5.3 m as shown
330 Fig. 5(e)). The main reason is that the Gaussian model produces smaller estimations of the
331 *SoF* than the other two models (e.g., 51 m vs 114 m and 81 m as shown in Fig. 5(b)). Since
332 the spatial prediction using the conditional random field method is mainly dominated by
333 borehole data within an *SoF* of a target point, the Gaussian model uses data in a smaller
334 neighborhood than the other two models. As a result, the predicted rockhead elevations for
335 the Gaussian model are close to those at the several nearest boreholes (see the solid curve
336 around $x = 835 \text{ m}$ and 1150 m in Fig. 5(d)). By contrast, the predicted rockhead elevations
337 for the *SSE* and Matérn models are not so close to those for the nearest boreholes and the
338 prediction error is consequently smaller.

339 (2) Another interesting finding is that the Gaussian model underestimates the prediction
340 uncertainty although it produces a wider prediction interval than the other two models. On
341 the contrary, the SSE and Matérn models generally can reasonably quantify the prediction
342 uncertainty. Figs. 5(c, d, e) also provide the average widths of the 95% CIs of the rockhead
343 elevations along the investigated cross-sections. As shown, the average width of the 95% CI
344 for the Gaussian model is wider than those for the other two models. This phenomenon is
345 expected because the Gaussian model produces a smaller estimation of the SoF and reduces
346 the uncertainty of unknown parameters in a smaller neighborhood of the training data than
347 those for the other two models. Nevertheless, the Gaussian model always covers the least
348 testing points. As shown in Figs. 5(c, d, e), three testing points cannot be covered by the 95%
349 CI for the Gaussian model while only one testing point cannot be covered by the 95% CI for
350 the other two models. This phenomenon agrees with the observation that the NPV value for
351 the Gaussian model (i.e., 1.85) is much larger than 1 while those for the other two models are
352 close to 1 (i.e., 0.98 for the SSE model and 1.59 for the Matérn model). The main reason is
353 that the Gaussian model excessively reduces the prediction uncertainty of the testing data
354 close to the training data points. As shown in Fig. 5(b), for the Gaussian model, the
355 autocorrelation coefficient corresponding to small lags is larger than those for the other
356 models. In other words, when the actual value of ν does not reach $+\infty$, the Gaussian model
357 is likely to overestimate the autocorrelation between an observation and a closely located
358 unknown point, making the associated confidence interval too narrow to cover the actual
359 value.

360 **4.2 Bukit Timah granite rockhead at Site 2**

361 **4.2.1 Rockhead surface predicted using all the data**

362 The BIC values for the selection of the trend function for site 2 were summarized in Table 1(c).
363 As shown, the optimal trend function for the single exponential and the Matérn model is the
364 constant trend while that for the Gaussian model is a linear function of the x coordinate. However,
365 it is subsequently found that a constant trend and a linear function model produce nearly the same
366 prediction accuracy. For an easy comparison of various autocorrelation models, the constant trend
367 is used for all autocorrelation models. The random field parameters of the rockhead elevation
368 estimated from all the data using the maximum likelihood estimation method were summarized in
369 Table 1(d). With this information, the autocorrelation coefficients for various ACF models were
370 plotted in Figs. 6(a, b), respectively. As shown, the estimated ν_x and ν_y for the Matérn model are
371 0.38 and 1.62, respectively, indicating the actual ACF in the x -direction is close to an SSE ACF
372 while the actual ACF in the y -direction is close to a second-order Markov ACF model (e.g., Ching
373 et al. 2019). Besides, the estimated values of the SoF for the Gaussian model are still significantly
374 smaller than those for the other two models. The reasons are still that the dominant data spacings
375 are relatively small, such as 40 m \sim 50 m in the y -direction and the dominant data pairs give rise
376 to a large autocorrelation coefficient, as shown in Figs. 6(c, d, e).

377 Figs. 7(a-c) plotted the predicted surfaces of rockhead for various models using all the data
378 points at Site 2 while Fig. 7(d) and 7(e) plot the predicted curves and 95% CI of the rockhead
379 elevation along $y = 170$ m, respectively. As shown, the differences among the predicted surfaces
380 and curves for various ACF models are not as distinct as those for Site 1 because the borehole data
381 at Site 2 is denser than those at Site 1. However, one can still find the effect of ACF from the
382 average width of the 95% CI of the rockhead elevation along one cross-section with $y = 170$ m,
383 which is 21.5 m, 23.9 m and 21.3 m for the SSE, Gaussian, and Matérn model, respectively. Similar

384 to the result for Site 1, the Gaussian model still produces the widest 95% CI, indicating that on the
385 whole, the Gaussian model cannot effectively reduce the prediction uncertainty.

386 **4.2.2 Comparison of the predictions for different ACF models**

387 The performances of various ACF models are also compared using the cross-validation procedure.
388 Similar to Site 1, 100 rounds of cross-validations were performed and in each round, 70% of data
389 were randomly drawn as training data while the reminder as testing data. The means and standard
390 deviations of the 100 sets of indexes were summarized in the last two rows of Table 2. As shown,
391 the averaged value of *RMSE* and *RMSRE* for the Gaussian model is the largest and the averaged
392 value of *NPV* for the Gaussian model is the farthest from 1. These results imply that the Gaussian
393 model still produces the largest prediction errors and the least reasonable prediction uncertainties.

394 To show the reason, Fig. 8 plots a typical example of the cross-validation, including the plan
395 view of the training and testing data in Fig. 8(a), the autocorrelation coefficient for various models
396 in Fig. 8(b), the predicted curve of rockhead along $y = 33$ m in Fig. 8(c), and the 95% CI of the
397 rockhead elevation along $y = 33$ m in Fig. 8(d). As shown in Fig. 8(c), the Gaussian model
398 generally produces larger errors than the SSE and Matérn models, especially for the 1st, 3rd and 4th
399 testing data along the cross-section of $y = 33$ m. Besides, a large part of the predicted curve for the
400 Gaussian model is close to the prior mean trend, - 20.6 m, indicating the known data points do not
401 contribute too much to the updating in these areas. The main reason is that the Gaussian model
402 produces a relatively small value of the estimated scale of fluctuations (see Fig. 8(b)). Based on
403 this Gaussian model, the data points on the cross-section have very weak autocorrelation with the
404 nearest training data points. As a result, the prediction error and prediction uncertainty are also
405 large.

406 Moreover, similar to the example of Site 1, the Gaussian model still underestimates the
407 uncertainty of rockhead elevation at testing points although in general, it produces the widest 95%
408 CI. Actually, the *NPV* value of the Gaussian model is much larger than 1 (i.e., 1.86) while those
409 for the other two models are close to 1 (i.e., 0.76 for the SSE model and 0.92 for the Matérn model).
410 Correspondingly, the 95% CI of the Gaussian model covers the least testing data (see Fig. 8(d)).
411 For instance, the testing point at $(x, y) = (79 \text{ m}, 48 \text{ m})$ cannot be covered by the 95% CI of the
412 Gaussian model but can be covered by those for the other two models. This is caused by the large
413 error of the predicted value, or more specifically, the prediction error larger than $1.96 \times$ posterior
414 standard deviation.

415 **5. Spatial prediction using simulated data**

416 Section 4 just investigates two cases of borehole data and the performance of various ACFs in
417 situations with a different data density or a different value of the smoothness coefficient are
418 unknown. Hence, simulated data were generated and analyzed in this section to evaluate the effect
419 of borehole spacing and the smoothness of the actual data using the following procedure:

- 420 1) One set of random field parameters is assumed to be the “actual” random field parameters
421 of the rockhead elevation in an imaginary site;
- 422 2) 50 realizations of the rockhead-elevation random field are generated based on the assumed
423 random field parameters using the covariance matrix decomposition method (e.g., Li et al.
424 2019). Each artificially generated realization can be considered to be the ‘real’ rockhead
425 elevation at the site;
- 426 3) One cross-validation is performed for each realization to compute the prediction error.

427 Similar to Section 4, the three ACF models, SSE, Gaussian, and Matérn are used for spatial
428 predictions of the rockhead for the testing data points. For each set of the random field parameters

429 and each ACF model, 50 sets of $RMSE$ and NPV are calculated for the 50 realizations. The mean
430 value of the two indexes, $RMSE$ and NPV are evaluated. Note that the realization number, 50 is
431 chosen because the corresponding mean values of $RMSE$ and NPV are similar to those for 100
432 realizations. In other words, 50 realizations are adequate for an accurate evaluation of the
433 prediction error.

434 The data selection criterion for the training and testing data is designed as follows. Firstly,
435 the random field realizations composed of relatively dense grid points are simulated. The spacing
436 of any two adjacent data points in each direction is set as 10 m. This spacing is considered to be
437 adequately small as a smaller spacing such as 5 m yields similar statistics of $RMSE$ and NPV in
438 cross-validations. Secondly, $15 \times 15 = 225$ sparse grid points are withdrawn from the dense grid
439 points as training data. Several spacings of neighboring training data are selected to ensure the
440 normalized spacing (defined as $\Delta/SoF =$ the spacing of the training data / the actual value of the
441 scale of fluctuation) has a value of 0.2, 0.3, 0.4, 0.5, 1 or 1.5. The remaining data points in the
442 area covered by the training data are considered to be testing data. One example of the plan view
443 of the training and testing data is plotted in Fig. 9.

444 The Matérn ACF model is used to generate the random field realization of the rockhead
445 elevation at the imaginary site. The ‘actual’ standard deviation and mean value of the rockhead
446 elevation are set as 10 m and -50 m, respectively. The ‘actual’ scale of fluctuation is set to be
447 $SoF_x = SoF_y = 40$ m or 100 m. Finally, the value of the smoothness parameter is set to be $\nu_x = \nu_y$
448 $= 0.5, 1.5,$ and 10. These value generally covers the typical range of ν because when ν reaches a
449 relatively large value such as 10, the shape of the ACF hardly change with an increasing ν . The
450 statistics of the $RMSE$ and NPV for various cross-validation schemes are listed in Table 3(a). The
451 following discussion can be made based on the results in Table 3(a).

452 (1) As a special case of the Matérn model, a classical model produces a reasonable prediction
453 uncertainty when the prescribed value of the smoothness coefficient is consistent with the
454 actual ν of the rockhead data. Otherwise, the classical model produces an unreasonable
455 prediction uncertainty and a prediction error that is generally larger than the Matérn model.
456 For instance, the Gaussian model usually underestimates while the SSE model overestimates
457 the prediction uncertainty when the actual value of ν is larger than 0.5 and smaller than $+\infty$. As
458 shown in Table 3(a), all values of μ_{NPV} for the Gaussian model are larger than 1. The reason
459 is that the Gaussian model overestimates the autocorrelation at small lags, making the
460 uncertainties of unknown parameters at locations close to known data to be excessively
461 reduced.

462 (2) The Matérn ACF model normally produces a reasonable prediction uncertainty (as evidenced
463 by a mean value of NPV close to 1) and an error similar to the classical model with a suitable
464 value of ν . However, since this model has additional smoothness parameters to estimate, it
465 may be subjected to estimation errors of the ν parameters. This side-effect can be well shown
466 in Table 3(b), which summarizes the number of experiments producing an estimated ν of $+\infty$
467 when the actual ν is 0.5. As shown, the number of experiments with estimated $\nu = +\infty$
468 gradually increases as the normalized spacing increases. This side-effect makes the prediction
469 accuracy of the Matérn ACF model slightly lower than the classical model with a proper value
470 of ν and also makes the former underestimate the prediction uncertainty. As shown in Table
471 3(a), the effect of the side-effect on the performance of the Matérn model is the most prominent
472 when the actual value of ν is small and the data spacing (e.g., $> 0.4 \times SoF$) is large. Hence, it
473 is crucial to drill some closely spaced boreholes in site investigation to ensure an accurate

474 estimation of the smoothness parameter.

475 (3) For the Gaussian model, the μ_{NPV} first decreases and then increases with an increasing δ , as
476 shown in the 7th row under the table header of Table 3(a). This trend is attributed to two
477 primary factors, the data density, and the estimated SoF . On one hand, a large data spacing
478 makes the borehole data sparse, resulting in large prediction uncertainty (denominator of the
479 NPV). On the other hand, an increasing data spacing also induces a growing estimated SoF .
480 This trend can be found in Table 3(c), which summarizes the mean of the estimated values of
481 SoF as the actual $\nu = 1.5$. This increased value of SoF in turn reduces the prediction
482 uncertainty. The former factor plays a dominant role as spacing is relatively small and is
483 responsible for the decreasing trend of NPV while the latter factor is responsible for the
484 increasing trend of NPV .

485 As for the SSE mode, the NPV exhibits an increasing trend with an increasing borehole
486 spacing. This main reason is that as the borehole spacing is large, the prediction error (i.e., the
487 numerator of the NPV) is also relatively large. Note that a growing borehole spacing also leads
488 to a reduced value of the estimated SoF (see the last row of Table 3(c)) and sparser data,
489 thereby resulting in large prediction uncertainty (denominator of NPV). However, the effect
490 of these two factors on the NPV is not as large as the prediction error. Thus the NPV for the
491 SSE model generally increases with an increasing data spacing.

492 In summary, the classical ACF model produces relatively small prediction error and
493 satisfactory uncertainty only when the underlying ν value is consistent with the actual ν value of
494 the geological data. Hence, it is preferable to use the classical ACF model only when reliable prior
495 information of the ν parameter is available. Besides, the Matérn ACF generally produces an
496 accuracy and prediction uncertainty similar to the classical model with a suitable value of ν .

497 However, its performance relative to the classical model with a proper with of ν deteriorates when
498 the spacing of the borehole becomes large. Therefore, it is vital to drill some closely spaced
499 boreholes, which is also emphasized by Qi and Liu (2019a).

500 **6. Conclusions**

501 This study investigated the effect of the autocorrelation function model on the spatial prediction
502 of the interface of geological layers using the conditional random field method. The prediction
503 accuracies and rationalities of the computed prediction uncertainties for the single exponential,
504 Gaussian, and Matérn models were compared via a cross-validation procedure. Borehole data from
505 two sites of Singapore and revealing rockhead of an igneous rock formation, Bukit Timah granite
506 were used for the comparison. Simulated data of rockhead elevation were also analyzed to
507 investigate the effect of the borehole spacing and smoothness of the data. The following
508 conclusions were tentatively drawn from the analyses.

509 (1) A classical model produces a reasonable prediction uncertainty when its prescribed value
510 of smoothness coefficient is consistent with the actual smoothness coefficient of the geological
511 data. Otherwise, the error is normally larger than that of the Matérn model and the prediction
512 uncertainty from the classical model is not reasonable. To be specific, the Gaussian model may
513 underestimate, or the single exponential model overestimate the uncertainty in the predicted
514 rockhead elevation when the actual smoothness coefficient of the geological data has a value
515 between 0.5 and $+\infty$. These conclusions are supported by analyses using both the actual borehole
516 data and simulated data. For example, the study using the actual borehole data shows that the
517 Gaussian model on average has an error of 1 m larger than the other two models and can
518 underestimate the standard deviation of the prediction by almost 40% (i.e., $1-\sqrt{1/2.70}=0.39$).

519 (2) The prediction accuracy of the Matérn model is affected by the spacing of the boreholes.
520 When the borehole spacing is relatively small ($< 0.4 \times$ actual scale of fluctuation), the Matérn
521 model can reasonably quantify the prediction uncertainty and has an accuracy similar to that of the
522 classical model with the right value of smoothness coefficient. However, when the borehole
523 spacing is large, the prediction by the Matérn model becomes less accurate as compared with the
524 prediction using the classical models with the right value of smoothness coefficient because of the
525 large estimation error of smoothness coefficient. Besides, the Matérn model may underestimate
526 the prediction uncertainty under this circumstance. The deterioration of the performance of the
527 Matérn model is the most prominent when the smoothness parameter is small (such as 0.5).

528 Based on these observations, the following suggestions are provided to guide the selection of
529 the autocorrelation models in spatial predictions using the conditional random field method. (i)
530 When the borehole spacing is relatively small ($< 0.4 \times$ actual scale of fluctuation), the Matérn
531 model should be used in spatial predictions. (ii) When the borehole spacing is relatively large and
532 prior information regarding the smoothness coefficient (e.g., smoothness coefficient for
533 neighboring sites as reported in the literature) is available, the corresponding classical model could
534 be used. For example, the single exponential model can be used if the users are confident that the
535 value of the smoothness coefficient is close to 0.5. (iii) When the borehole spacing is relatively
536 large and there is no prior information regarding the smoothness coefficient, additional site
537 investigations need to be performed to acquire closely spaced data for the use of the Matérn model.

538 The conclusions of this study can also be applied to spatial predictions of geotechnical
539 properties, such as the N values in standard penetration tests or the tip resistance in cone
540 penetration tests. Furthermore, it is beneficial to use the predicted geotechnical and geological

541 parameters in reliability analyses of geotechnical structures when there is a lack of real
542 geotechnical data.

543 **Acknowledgments**

544 This research is supported by the Singapore Ministry of National Development and the National
545 Research Foundation, Prime Minister's Office under the Land and Liveability National Innovation
546 Challenge (L2 NIC) Research Programme (Award No. L2NICCFP2-2015-1). Any opinions,
547 findings, and conclusions or recommendations expressed in this material are those of the authors
548 and do not reflect the views of the Singapore Ministry of National Development and National
549 Research Foundation, Prime Minister's Office, Singapore.

550 **References**

- 551 Baecher, G.B., and Christian, J.T. 2003. Reliability and statistics in geotechnical engineering. John
552 Wiley & Sons.
- 553 Cai, Y., Li, J., Li, X., Li, D., and Zhang, L. 2019. Estimating soil resistance at unsampled locations
554 based on limited CPT data. *Bulletin of Engineering Geology and the Environment*, **78**(5):
555 3637-3648.
- 556 Ching, J., Wu, T.-J., Stuedlein, A.W., and Bong, T. 2018. Estimating horizontal scale of fluctuation
557 with limited CPT soundings. *Geoscience Frontiers*, **9**(6): 1597-1608.
- 558 Ching, J., and Phoon, K.-K. 2019. Impact of autocorrelation function model on the probability of
559 failure. *Journal of Engineering Mechanics*, **145**(1): 04018123.
- 560 Ching, J., Phoon, K.-K., Stuedlein, A.W., and Jaksa, M. 2019. Identification of sample path
561 smoothness in soil spatial variability. *Structural Safety*, **81**: 101870.
- 562 Cressie, N.A. 1993. *Statistics for spatial data*. John Wiley & Sons, Inc, 1993. p. 102-102.

- 563 DeGroot, D.J., and Baecher, G.B. 1993. Estimating autocovariance of in - situ soil properties.
564 Journal of Geotechnical Engineering, **119**(1): 147-166.
- 565 Dasaka, S., and Zhang, L. 2012. Spatial variability of in situ weathered soil. Géotechnique, **62**(5):
566 375-384.
- 567 Firouziandbandpey, S., Ibsen, L.B., Griffiths, D.V., Vahdatirad, M.J., Andersen, L.V., and Sørensen,
568 J.D. 2015. Effect of spatial correlation length on the interpretation of normalized CPT data
569 using a kriging approach. Journal of Geotechnical and Geoenvironmental Engineering,
570 **141**(12): 04015052.
- 571 Fenton, G.A. 1999. Random field modeling of CPT Data. Journal of Geotechnical and
572 Geoenvironmental Engineering, **125**(6): 486-498.
- 573 Hristopulos, D.T., and Žukovič, M. 2011. Relationships between correlation lengths and integral
574 scales for covariance models with more than two parameters. Stochastic Environmental
575 Research and Risk Assessment, **25**(1): 11-19.
- 576 Lloret-Cabot, M., Hicks, M.A., and Eijnden, A.P.v.d. 2012. Investigation of the reduction in
577 uncertainty due to soil variability when conditioning a random field using Kriging.
578 Géotechnique Letters, **2**(3): 123-127.
- 579 Li, D.-Q., Xiao, T., Zhang, L.-M., and Cao, Z.-J. 2019. Stepwise covariance matrix decomposition
580 for efficient simulation of multivariate large-scale three-dimensional random fields. Applied
581 Mathematical Modelling, **68**: 169-181.
- 582 Li, J., Cassidy, M.J., Huang, J., Zhang, L., and Kelly, R. 2016a. Probabilistic identification of soil
583 stratification. Géotechnique, **66**(1): 16-26.

- 584 Li, X., Zhang, L.M., and Li, J. 2016b. Using conditioned random field to characterize the
585 variability of geologic profiles. *Journal of Geotechnical and Geoenvironmental Engineering*,
586 **142**(4): 04015096.
- 587 Li, X.Y., Zhang, L.M., Zhu, H., and Li, J.H. 2018. Modeling geologic profiles incorporating
588 interlayer and intralayer variabilities. *Journal of Geotechnical and Geoenvironmental*
589 *Engineering*, **144**(8): 04018047.
- 590 Liu, W., Leung, Y.F., and Lo, M.K. 2017. Integrated framework for characterization of spatial
591 variability of geological profiles. *Canadian Geotechnical Journal*, **54**(1): 47-58.
- 592 Lo, M.K., and Leung, Y.F. 2017. Probabilistic analyses of slopes and footings with spatially
593 variable soils considering cross-correlation and conditioned random field. *Journal of*
594 *Geotechnical and Geoenvironmental Engineering*, **143**(9): 04017044.
- 595 Matern, B. 1960. Spatial variation. Stochastic models and their application to some problems in
596 forest surveys and other sampling investigations. *Meddelanden fran Statens*
597 *Skogsforskningsinstitut*, **49**(5): 1-144.
- 598 Pitts, J. 1984. A review of geology and engineering geology in Singapore. *Quarterly Journal of*
599 *Engineering Geology and Hydrogeology*, **17**(2): 93-101.
- 600 Qi, X.-H., and Liu, H.-X. 2019a. An improved global zonation method for geotechnical parameters.
601 *Engineering Geology*, **248**: 185-196.
- 602 Qi, X.-H., and Liu, H.-X. 2019b. Estimation of autocorrelation distances for in-situ geotechnical
603 properties using limited data. *Structural Safety*, **79**: 26-38.
- 604 Qi, X.-H., Li, D.-Q., Phoon, K.-K., Cao, Z.-J., and Tang, X.-S. 2016. Simulation of geologic
605 uncertainty using coupled Markov chain. *Engineering Geology*, **207**: 129-140.

- 606 Qi, X., Pan, X., Chiam, K., Lim, Y.S., and Lau, S.G. 2020a. Comparative spatial predictions of
607 the locations of soil-rock interface. *Engineering Geology*, **272**: 105651.
- 608 Qi, X., Wang, H., Pan, X., Chu, J., and Chiam, K. 2020b. Prediction of interfaces of geological
609 formations using the multivariate adaptive regression spline method. *Underground Space*. doi:
610 <https://doi.org/10.1016/j.undsp.2020.02.006>.
- 611 Sharma, J.S., Chu, J., and Zhao, J. 1999. Geological and geotechnical features of Singapore: an
612 overview. *Tunnelling and Underground Space Technology*, **14**(4): 419-431.
- 613 Stein, M.L. 1999. *Interpolation of spatial data: some theory for Kriging*. Springer Science &
614 Business Media.
- 615 Vanmarcke, E.H. 1977. Probabilistic modeling of soil profiles. *Journal of the geotechnical*
616 *engineering division*, **103**(11): 1227-1246.
- 617 Vanmarcke, E. 1983. *Random fields: Analysis and synthesis*. MIT Press, Cambridge, MA.
- 618 Wang, Y., and Zhao, T. 2016. Interpretation of soil property profile from limited measurement
619 data: a compressive sampling perspective. *Canadian Geotechnical Journal*, **53**(9): 1547-1559.
- 620 Wang, Y., and Zhao, T. 2017. Statistical interpretation of soil property profiles from sparse data
621 using Bayesian compressive sampling. *Géotechnique*, **67**(6): 523-536.
- 622 Wang, Y., Hu, Y., and Zhao, T. 2020a. Cone penetration test (CPT)-based subsurface soil
623 classification and zonation in two-dimensional vertical cross section using Bayesian
624 compressive sampling. *Canadian Geotechnical Journal*, **57**(7): 947-958.
- 625 Wang, L., Wu, C., Gu, X., Liu, H., Mei, G., and Zhang, W. 2020b. Probabilistic stability analysis
626 of earth dam slope under transient seepage using multivariate adaptive regression splines.
627 *Bulletin of Engineering Geology and the Environment*, **79**(6): 2763-2775.

- 628 Yuen, K.-V. 2010. Bayesian methods for structural dynamics and civil engineering. John Wiley &
629 Sons.
- 630 Zhang, W.G., and Goh, A.T.C. 2013. Multivariate adaptive regression splines for analysis of
631 geotechnical engineering systems. *Computers and Geotechnics*, **48**: 82-95.
- 632 Zhang, W., and Goh, A.T.C. 2016. Multivariate adaptive regression splines and neural network
633 models for prediction of pile drivability. *Geoscience Frontiers*, **7**(1): 45-52.
- 634 Zhao, J., Broms, B.B., Zhou, Y., and Choa, V. 1994. A study of the weathering of the Bukit Timah
635 granite part a: Review, field observations and geophysical survey. *Bulletin of the*
636 *International Association of Engineering Geology*, **49**(1): 97-106.
- 637

Table 1 Random field parametric values of the rockhead elevation evaluated from the maximum likelihood estimation method

(a) BIC values for various polynomial trends of rockhead elevation for site 1 (Yishun Park)

Model	Constant	Linear 1 ^a	Linear 2 ^b	Linear 3 ^c	Quadratic 1 ^d	Quadratic 2 ^e	Quadratic 3 ^f	Cubic 1 ^g	Cubic 2 ^h	Cubic 3 ⁱ
SSE	1293	1296	1298	1302	1300	1307	1304	1309	1309	1314
Gaussian	1333	1334	1338	1339	1335	1343	1337	1342	1342	1347
Matérn	1302	1305	1307	1310	1309	1316	1313	1318	1318	1324

Note: ^aLinear 1: Trend function = $\beta_0 + \beta_1 \times x$; ^bLinear 2: Trend function = $\beta_0 + \beta_1 \times y$;
^cLinear 3: Trend function = $\beta_0 + \beta_1 \times x + \beta_2 \times y$; ^dQuadratic 1: Trend function = $\beta_0 + \beta_1 \times x + \beta_2 \times y + \beta_3 \times x^2$;
^eQuadratic 2: Trend function = $\beta_0 + \beta_1 \times x + \beta_2 \times y + \beta_3 \times y^2$;
^fQuadratic 3: Trend function = $\beta_0 + \beta_1 \times x + \beta_2 \times y + \beta_3 \times x^2 + \beta_4 \times y^2$;
^gCubic 1: Trend function = $\beta_0 + \beta_1 \times x + \beta_2 \times y + \beta_3 \times x^2 + \beta_4 \times y^2 + \beta_5 \times x^3$;
^hCubic 2: Trend function = $\beta_0 + \beta_1 \times x + \beta_2 \times y + \beta_3 \times x^2 + \beta_4 \times y^2 + \beta_5 \times y^3$;
ⁱCubic 3: Trend function = $\beta_0 + \beta_1 \times x + \beta_2 \times y + \beta_3 \times x^2 + \beta_4 \times y^2 + \beta_5 \times x^3 + \beta_6 \times y^3$, where β_0, \dots, β_6 are regression coefficients.

(b) Random field parametric values for rockhead elevation at site 1 (Yishun Park)

Model	SoF_x (m)	SoF_y (m)	ν_x	ν_y	σ (m)	μ (m)
SSE	177.8	138.7	—	—	9.8	-18.1
Gaussian	72.5	37.4	—	—	9.5	-18.5
Matérn	166.4	112.6	0.5	0.7	9.9	-18.4

(c) BIC values for various polynomial trends of rockhead elevation at site 2 (Canberra Link)

Model	Constant	Linear 1 ^a	Linear 2 ^b	Linear 3 ^c	Quadratic 1 ^d	Quadratic 2 ^e	Quadratic 3 ^f	Cubic 1 ^g	Cubic 2 ^h	Cubic 3 ⁱ
SSE	956	959	960	964	968	968	972	976	977	981
Gaussian	1002	999	1006	1003	1004	1004	1005	1009	1010	1014
Matérn	962	966	967	970	975	974	978	983	983	988

Note: the expression for all the polynomial functions are provided in Table 1(a).

(d) Random field parametric values for rockhead elevation at site 2 (Canberra Link)

Model	SoF_x (m)	SoF_y (m)	ν_x	ν_y	σ (m)	μ (m)
SSE	84.3	164.7	—	—	10.2	-21.2
Gaussian	26.3	57.2	—	—	10.6	-20.6
Matérn	83.3	133.2	0.4	1.6	10.4	-21.3

Table 2 Prediction accuracies for the spatial prediction of rockhead elevation

		Single exponential model			Gaussian model			Matérn model		
		<i>RMSE</i> (m)	<i>RMSRE</i>	<i>NPV</i>	<i>RMSE</i> (m)	<i>RMSRE</i>	<i>NPV</i>	<i>RMSE</i> (m)	<i>RMSRE</i>	<i>NPV</i>
Site 1	mean	6.66	0.27	1.01	7.67	0.34	1.48	6.73	0.29	1.21
	^a SD	0.72	0.05	0.27	0.73	0.08	0.61	0.75	0.06	0.38
Site 2	mean	7.27	0.89	1.02	8.66	1.38	2.70	7.38	0.91	1.22
	^a SD	0.98	0.39	0.33	1.19	0.81	2.13	0.95	0.45	0.41

Note: ^aSD = standard deviation

Table 3 Prediction accuracies for the spatial prediction of rockhead elevation using simulated data(a) Statistics of *RMSE* and *NPV* for cross-validations

¹ ν	² δ	0.2	0.3	0.4	0.5	1	1.5						
		³ μ_{RMSE}	⁴ μ_{NPV}	μ_{RMSE}	μ_{NPV}								
0.5	Gau	5.26	12.78	6.42	8.35	7.09	8.20	7.80	5.53	9.56	3.15	10.15	2.35
	SSE	4.94	1.07	5.91	0.99	6.61	1.04	7.25	1.00	8.97	0.99	9.70	1.04
	Matérn	4.94	1.07	5.95	1.14	6.80	1.34	7.41	1.43	9.32	1.49	9.87	1.49
1.5	Gau	1.57	11.09	2.66	6.47	3.77	4.16	4.97	2.97	8.40	1.75	9.69	1.49
	SSE	1.80	0.24	2.98	0.34	4.06	0.43	5.23	0.54	8.56	0.85	9.70	0.98
	Matérn	1.52	1.10	2.71	1.24	3.83	1.28	4.99	1.34	8.87	1.32	9.85	1.20
10	Gau	0.08	5.07	0.50	2.14	1.34	1.54	2.55	1.33	7.92	1.19	9.45	1.42
	SSE	0.76	0.06	1.64	0.13	2.64	0.21	3.81	0.32	8.63	0.81	9.53	0.99
	Matérn	0.07	1.07	0.50	1.06	1.34	1.05	2.60	0.94	8.95	0.97	9.64	1.17

Note: ¹ ν = input (actual) value of the smoothness coefficient for the simulated data;

² δ = normalized borehole spacing = borehole spacing / actual scale of fluctuation, e.g., 0.2 = 20 m / 100 m;

³ μ_{RMSE} = mean of the root mean squared error for 50 rounds of cross-validations;

⁴ μ_{NPV} = mean of the normalized prediction variance for 50 rounds of cross-validations.

(b) Number of cross-validations which produce an estimated ν of $+\infty$ (input $\nu = 0.5$, Matérn model used)

δ	0.2	0.3	0.4	0.5	1	1.5
Number	0	0	2	6	22	21
Percentage	0%	0%	4%	12%	44%	42%

(c) Normalized estimated scale of fluctuation for the Gaussian and SSE model when the actual $\nu = 1.5$

δ	0.2	0.3	0.4	0.5	1	1.5
	SoF_x, SoF_y					
Gaussian	0.53, 0.53	0.69, 0.69	0.80, 0.81	0.88, 0.89	1.02, 1.07	1.07, 1.04
SSE	1.88, 1.84	1.47, 1.44	1.26, 1.28	1.13, 1.13	0.80, 0.82	0.66, 0.60

Note: normalized SoF = mean of the estimated SoF / input SoF

Appendix A

The Matlab codes used to estimate the random field parameter and evaluate the BIC values are provided as follows.

(1) The main program

```
%estimate the random field parameters, assume geological parameter normally distributed
clc; clear
format short g
%[X,Y]: coordinate of borehole data; GeoPara: rockhead elevation; ACFT:
%type of autocorrelation function; deltaX and deltaY are matrixes denoting
%the lags in the x and y directions, respectively
global X Y GeoPara ACFT FunNum deltaX deltaY
num = xlsread('UsedBHdata.xlsx',1); %user input
X = num(:,1);
Y = num(:,2);
GeoPara = num(:,3);
NumP = length(X);
%considered polynomial functions
f1=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2);
f2=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2;
f3=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,2).^2;
f4=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2+p(5)*x(:,2).^2;
f5=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2+p(5)*x(:,2).^2+p(6)*x(:,1).^3;
f6=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2+p(5)*x(:,2).^2+p(6)*x(:,2).^3;
f7=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2+p(5)*x(:,2).^2+p(6)*x(:,1).^3+p(7)*x(:,2).^3;
f8=@(p,x) p(1)+p(2)*x(:,1);
f9=@(p,x) p(1)+p(2)*x(:,2);
opt=optimset('TolFun',1e-4,'TolX',1e-5);
%%
%estimation using the maximum likelihood method
ACFT = 'WM'; %select from 'SinExp','Gau','WM','Exp'
k = [3 4 4 5 6 6 7 2 2];
deltaX = X - X';
deltaY = Y - Y';
options=optimset('LargeScale','off','display','iter');
for iTrend = 0:0
    FunNum = iTrend;
    %generate initial value for regression coefficient
    switch iTrend
        case 0
            p0 = mean(GeoPara);
        case 1
            [p0,~,r] = regress(GeoPara,[ones(size(X)),X,Y]);
        case 2
            [p0,~,r] = regress(GeoPara,[ones(size(X)),X,Y,X.^2]);
        case 3
            [p0,~,r] = regress(GeoPara,[ones(size(X)),X,Y,Y.^2]);
        case 4
            [p0,~,r] = regress(GeoPara,[ones(size(X)),X,Y,X.^2,Y.^2]);
        case 5
            [p0,~,r] = regress(GeoPara,[ones(size(X)),X,Y,X.^2,Y.^2,X.^3]);
        case 6
            [p0,~,r] = regress(GeoPara,[ones(size(X)),X,Y,X.^2,Y.^2,Y.^3]);
        case 7
            [p0,~,r] = regress(GeoPara,[ones(size(X)),X,Y,X.^2,Y.^2,X.^3,Y.^3]);
        case 8
```

```

    [p0,~,r] = regress(GeoPara,[ones(size(X)),X]);
case 9
    [p0,~,r] = regress(GeoPara,[ones(size(X)),Y]);
end
%initial value for standard deviation
switch iTrend
case 0
    Nvalfit = p0;
case 1
    Nvalfit=f1(p0,[X,Y]);
case 2
    Nvalfit=f2(p0,[X,Y]);
case 3
    Nvalfit=f3(p0,[X,Y]);
case 4
    Nvalfit=f4(p0,[X,Y]);
case 5
    Nvalfit=f5(p0,[X,Y]);
case 6
    Nvalfit=f6(p0,[X,Y]);
case 7
    Nvalfit=f7(p0,[X,Y]);
case 8
    Nvalfit=f8(p0,[X,Y]);
case 9
    Nvalfit=f9(p0,[X,Y]);
end
SD0 = std(Nvalfit-GeoPara);
if size(p0,1)>1
    p0 = p0';
end
%the initial value of all random field parameters
if strcmpi(ACFt,'WM')
    iniParas = [50,50,0.5,0.5,SD0,p0; 50,50,10,10,SD0,p0; ...
        200,100,0.5,0.5,SD0,p0;200,100,10,10,SD0,p0];
else
    iniParas = [40,20,SD0,p0;200,50,SD0,p0;10,2,SD0,p0];
end
minNLL = 10000000;
for j = 1:size(iniParas,1)
    [x,fval] = fminsearch(@NLL2dNorm,iniParas(j,:));
    if fval < minNLL
        minNLL = fval;
        MPVx = x;
    end
end
EstPara = MPVx;
if strcmpi(ACFt,'WM')
    if iTrend == 0
        BIC = log(NumP)*(1+5)+2*minNLL;
    else
        BIC = log(NumP)*(k(iTrend)+5)+2*minNLL;
    end
else
    if iTrend == 0
        BIC = log(NumP)*(1+3)+2*minNLL;
    else
        BIC = log(NumP)*(k(iTrend)+3)+2*minNLL;
    end
end
startRow = 1;
xlswrite('Output.xlsx',[EstPara,minNLL,BIC],1,['A',num2str(iTrend+startRow)],1,...

```

```

    'A'+length(EstPara)+1,num2str(iTrend+startRow]));
end
(2) Subroutine

```

```

function nLL = NLL2dNorm(input)
%normally distributed random field and measurement noise not considered
%nLL: -log(likelihood); input: random field parameters; lx, ly: correlation distance in two horizontal directions
global X Y GeoPara ACFt FunNum deltaX deltaY nLLsteady
n= length(X);
if strcmpi(ACFt,'WM')
    %scale of fluctuation
    SOFx = input(1);    SOFy = input(2);
    nux = input(3);    nuy = input(4);
    sigmaSV = input(5);
    p = input(6:end);
    if SOFx<0 || nux<0 || SOFy<0 || nuy<0 || sigmaSV<0
        nLL=10^7;
        return
    end
else
    lx=input(1);    ly=input(2);
    sigmaSV = input(3);    p = input(4:end);
    if lx<0 || ly<0 || sigmaSV<0
        nLL=10^7;
        return
    end
end
end
f1=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2);
f2=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2;
f3=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,2).^2;
f4=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2+p(5)*x(:,2).^2;
f5=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2+p(5)*x(:,2).^2+p(6)*x(:,1).^3;
f6=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2+p(5)*x(:,2).^2+p(6)*x(:,2).^3;
f7=@(p,x) p(1)+p(2)*x(:,1)+p(3)*x(:,2)+p(4)*x(:,1).^2+p(5)*x(:,2).^2+p(6)*x(:,1).^3+p(7)*x(:,2).^3;
f8=@(p,x) p(1)+p(2)*x(:,1);
f9=@(p,x) p(1)+p(2)*x(:,2);
switch FunNum
case 0
    Nvalfit = p;
case 1
    Nvalfit = f1(p,[X,Y]);
case 2
    Nvalfit = f2(p,[X,Y]);
case 3
    Nvalfit = f3(p,[X,Y]);
case 4
    Nvalfit = f4(p,[X,Y]);
case 5
    Nvalfit = f5(p,[X,Y]);
case 6
    Nvalfit = f6(p,[X,Y]);
case 7
    Nvalfit = f7(p,[X,Y]);
case 8
    Nvalfit = f8(p,[X,Y]);
case 9
    Nvalfit = f9(p,[X,Y]);
end
end
%the autocorrelation coefficient
RhoRF=zeros(n);
if strcmpi(ACFt,'Exp')

```

```

    equL=sqrt((deltaX./lx).^2+(deltaY./ly).^2);
    RhoRF=exp(-equL);
elseif strcmpi(ACFt,'SinExp')
    RhoRF=exp(-abs(deltaX./lx)-abs(deltaY./ly));
elseif strcmpi(ACFt,'Gau')
    RhoRF=exp(-(deltaX./lx).^2-(deltaY./ly).^2);
elseif strcmpi(ACFt,'WM')
    equLx= abs(deltaX./SOFx)*2*sqrt(pi)*gamma(nux+0.5)/gamma(nux);
    RhoRFx= 2^(1-nux).*(equLx.^nux).*besselk(nux,equLx)/gamma(nux);
    flag = isnan(RhoRFx);
    RhoRFx(flag) = 1;
    equLy= abs(deltaY./SOFy)*2*sqrt(pi)*gamma(nuy+0.5)/gamma(nuy);
    RhoRFy= 2^(1-nuy).*(equLy.^nuy).*besselk(nuy,equLy)/gamma(nuy);
    flag = isnan(RhoRFy);
    RhoRFy(flag) = 1;
    RhoRF = RhoRFx.*RhoRFy;
end
COVRF = corr2cov(sigmaSV*ones(size(X)),RhoRF);
detCOV = det(COVRF);
invCOV = COVRF^(-1);
vec = GeoPara - Nvalfit;
if strcmpi(ACFt,'Gau')
    if detCOV<0 || det(invCOV)<0 || 0.5*vec'*invCOV*vec<0 %|| detCOV == Inf || det(invCOV) == Inf
        nLL=10^7;
        return
    end
else
    if detCOV<0 || det(invCOV)<0 || 0.5*vec'*invCOV*vec<0
        nLL=10^7;
        return
    end
end
%{
%method 1
nLL= n/2*log(2*pi) + 0.5*log(detCOV) + 0.5*vec'*invCOV*vec
%method 2
nLL = -log(mvnpdf(GeoPara,Nvalfit,COVRF)) %modified on 180612
%}
%method 3
if strcmpi(ACFt,'WM')
    if nux > 60 || nuy > 60
        nLL = nLLsteady;
        return
    end
end
L = chol(COVRF,'lower');
LogdetRho2 = 2*sum(diag(log(L)));
nLL = n/2*log(2*pi) + 0.5*LogdetRho2 + 0.5*vec'*invCOV*vec;
if strcmpi(ACFt,'WM')
    if nux > 50 || nuy > 50
        nLLsteady = nLL;
    end
end
end

```

Appendix B

The coordinates (x , y) and rockhead elevation (Ele) of the Bukit Timah granite at Site 1 are summarized in Table A1.

Table A1 Borehole data used for spatial predictions at Site 1

x (m)	y (m)	Ele (m)	x (m)	y (m)	Ele (m)	x (m)	y (m)	Ele (m)	x (m)	y (m)	Ele (m)
0.0	287.3	3.0	958.0	557.2	-22.7	1127.8	448.3	-25.3	1158.5	249.4	-12.4
103.4	330.4	-4.5	979.2	561.0	-23.7	1248.0	538.7	-27.7	1142.0	231.8	-17.4
118.9	321.5	-5.5	646.5	279.6	-7.9	1179.4	480.3	-4.2	1043.9	144.4	-8.5
169.2	342.4	-5.4	759.0	360.4	-13.4	824.4	181.8	-9.4	1132.0	207.8	-27.3
73.8	247.3	-4.5	469.9	112.3	-28.6	1157.7	454.5	-19.3	941.3	45.0	-4.4
206.0	343.1	-5.8	646.6	257.6	-10.4	1234.4	517.6	-20.4	1170.7	234.8	-10.9
179.7	301.3	-5.3	796.6	380.0	-14.3	1175.6	463.8	-9.1	1030.8	114.7	-6.4
91.8	215.0	-2.4	690.7	281.3	-13.4	863.9	203.1	-7.5	1154.2	217.2	-15.4
213.0	310.8	-7.3	645.3	235.7	-15.4	1143.9	431.8	-24.4	963.7	56.1	-9.7
543.1	565.0	-19.4	459.3	74.8	-27.0	1138.3	421.6	-21.4	986.5	62.6	1.7
335.1	349.4	-18.6	690.0	258.7	-13.1	921.3	236.8	-26.9	1083.1	140.4	-27.0
545.8	506.4	-14.9	439.0	39.6	-35.0	1198.9	462.0	-27.0	1031.6	91.2	-4.3
563.7	505.8	-10.9	745.5	290.5	-26.5	1171.2	438.6	-13.9	994.0	58.3	0.0
352.9	318.6	-21.3	644.0	204.6	-19.5	1234.0	488.0	-25.5	1056.9	91.0	0.2
606.1	529.5	-10.2	478.7	66.4	-24.4	943.1	226.0	-28.9	1292.7	285.6	-38.4
555.7	481.3	-24.1	688.8	236.8	-20.6	1246.1	474.6	-35.6	1001.2	39.3	-3.1
381.2	334.1	-20.6	642.8	184.1	-18.5	936.3	211.7	-26.4	1026.8	60.4	-4.7
426.2	370.2	-17.3	785.8	296.8	-29.4	1238.6	459.6	-41.0	1091.8	106.6	-18.4
603.1	498.1	-17.0	755.2	264.8	-26.5	941.0	205.7	-27.0	1338.6	291.9	-26.9
564.3	461.1	-23.9	687.2	204.2	-21.6	1218.0	427.3	-15.8	1247.3	215.6	-16.1
360.3	284.2	-30.4	859.6	344.1	-24.5	1186.0	400.3	-37.1	1126.3	108.5	-33.6
450.4	338.5	-15.1	641.5	162.2	-17.4	1162.0	380.1	-25.0	1075.6	57.3	-4.8
572.1	433.3	-18.1	686.8	184.4	-24.0	945.7	199.8	-27.6	1244.6	196.2	-14.5
562.6	408.1	-11.8	800.5	279.1	-27.1	959.6	209.2	-29.7	1264.4	211.4	-15.9
549.0	390.8	-10.3	771.9	244.8	-18.6	1250.0	449.5	-31.2	1270.7	214.0	-12.6
541.5	371.4	-13.0	685.7	163.9	-18.9	1140.5	348.4	-27.6	1113.9	79.8	-21.4
631.9	444.7	-26.6	1131.4	533.3	-2.8	969.7	192.8	-27.6	1272.4	197.8	-16.1
676.9	465.3	-20.7	1122.9	520.8	-8.5	998.0	208.2	-31.9	1267.5	190.5	-9.8
606.7	393.6	-9.9	721.7	184.8	-32.7	1087.3	264.7	-31.6	1373.2	276.0	-19.6
673.1	445.3	-11.9	815.3	261.5	-27.0	1101.9	275.2	-35.4	1362.1	265.9	-26.9
639.8	416.5	-20.6	1156.6	536.8	-11.7	1135.2	297.7	-35.6	1094.2	32.8	-12.4
582.9	357.8	-15.2	520.3	6.2	-23.9	942.2	136.6	-18.2	1330.7	229.3	-19.5
872.3	592.2	-17.9	1119.6	503.6	-14.1	961.6	145.2	-23.4	1319.2	218.9	-24.9
720.2	463.0	-13.7	1129.7	504.5	-3.0	1228.7	361.7	-26.5	1123.2	49.5	-20.2
680.9	420.1	-7.2	880.3	292.7	-47.0	1114.1	260.6	-24.0	1321.3	196.3	-17.5
650.4	390.2	-1.1	794.3	217.9	-20.6	1134.8	277.9	-28.9	1396.3	243.0	-23.8
347.4	137.2	-18.5	544.6	2.4	-20.9	988.7	151.5	-16.1	1256.3	48.6	-30.5
613.6	355.1	-20.6	895.8	293.6	-48.0	957.6	112.5	-5.6	1573.4	296.3	-15.6
734.8	442.1	-10.3	780.6	192.2	-27.0	1063.5	197.8	-42.3	1553.0	277.9	-3.9
655.0	366.9	-24.0	835.8	236.9	-9.1	1100.3	227.9	-21.3	1532.6	259.5	-2.5
740.2	415.7	-16.9	1221.2	546.8	-23.5	1144.7	261.3	-25.3	1512.1	241.1	-18.5
659.6	347.6	-13.5	1204.8	532.2	-29.9	1130.5	245.6	-14.4	1321.7	79.8	-9.4
694.0	374.4	-25.9	809.6	199.5	-13.2	1252.9	342.8	-25.9	1299.5	47.6	-28.0
960.2	586.9	-16.7	570.6	0.0	-27.0	1240.5	331.6	-23.9	1493.4	142.4	-14.2

733.2	385.3	-28.4	1159.8	486.4	-1.3	1115.7	225.0	-18.7	1440.0	55.4	-16.4
702.8	348.4	-22.4	1143.7	470.6	-11.6	1005.0	128.4	-15.7	1403.3	19.8	-6.3
786.5	414.9	-12.0	849.8	220.0	-23.0	946.3	74.1	-7.0	1433.2	29.9	-18.0

Caption of figures

Fig. 1 Matérn autocorrelation function with different smoothness coefficients and typical random field realizations of rockhead elevation

Fig. 2 Geological profiles of two cross-sections and borehole data at the two sites

Fig. 3 Autocorrelation coefficients of the rockhead elevation at Site 1 and reason for the estimated smaller SoF for the Gaussian model

Fig. 4 Rockhead elevation predicted using all the data at Site 1

Fig. 5 Rockhead elevation predicted by different autocorrelation models at Site 1

Fig. 6 Autocorrelation coefficients of the rockhead elevation at Site 2 and reason for the estimated smaller SoF for the Gaussian model

Fig. 7 Rockhead elevation predicted using all the data at Site 2

Fig. 8 Rockhead elevation predicted by different autocorrelation models at Site 2

Fig. 9 Plan view of training and testing data when borehole spacing of training data = 20 m

Caption of figures

Fig. 1 Matérn autocorrelation function with different smoothness coefficients and typical random field realizations of rockhead elevation

Fig. 2 Geological profiles of two cross-sections and borehole data at the two sites

Fig. 3 Autocorrelation coefficients of the rockhead elevation at Site 1 and reason for the estimated smaller SoF for the Gaussian model

Fig. 4 Rockhead elevation predicted using all the data at Site 1

Fig. 5 Rockhead elevation predicted by different autocorrelation models at Site 1

Fig. 6 Autocorrelation coefficients of the rockhead elevation at Site 2 and reason for the estimated smaller SoF for the Gaussian model

Fig. 7 Rockhead elevation predicted using all the data at Site 2

Fig. 8 Rockhead elevation predicted by different autocorrelation models at Site 2

Fig. 9 Plan view of training and testing data when borehole spacing of training data = 20 m

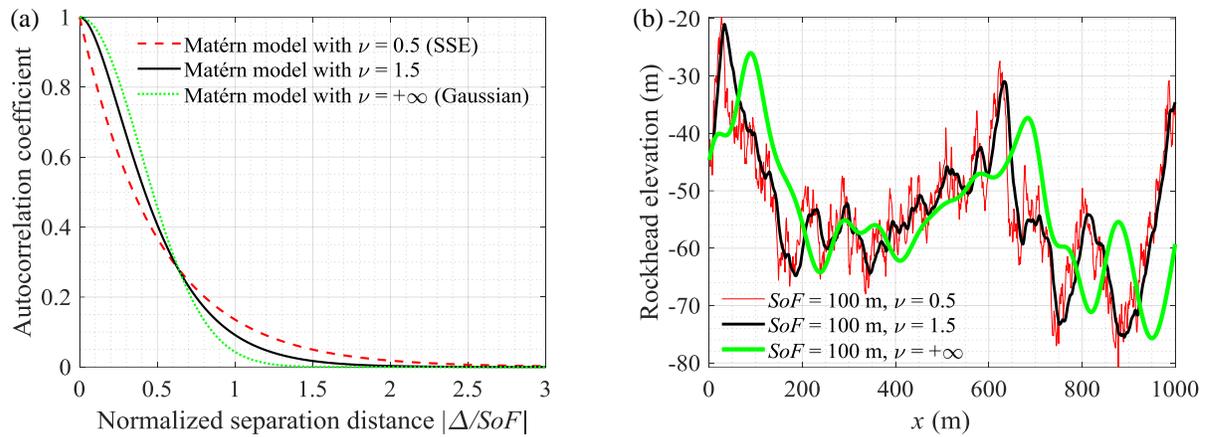


Fig. 1 Matérn autocorrelation function with different smoothness coefficients and typical random field realizations of rockhead elevation: (a) Autocorrelation functions; (b) Typical realizations of rockhead elevation (mean = - 50 m, standard deviation = 10 m).

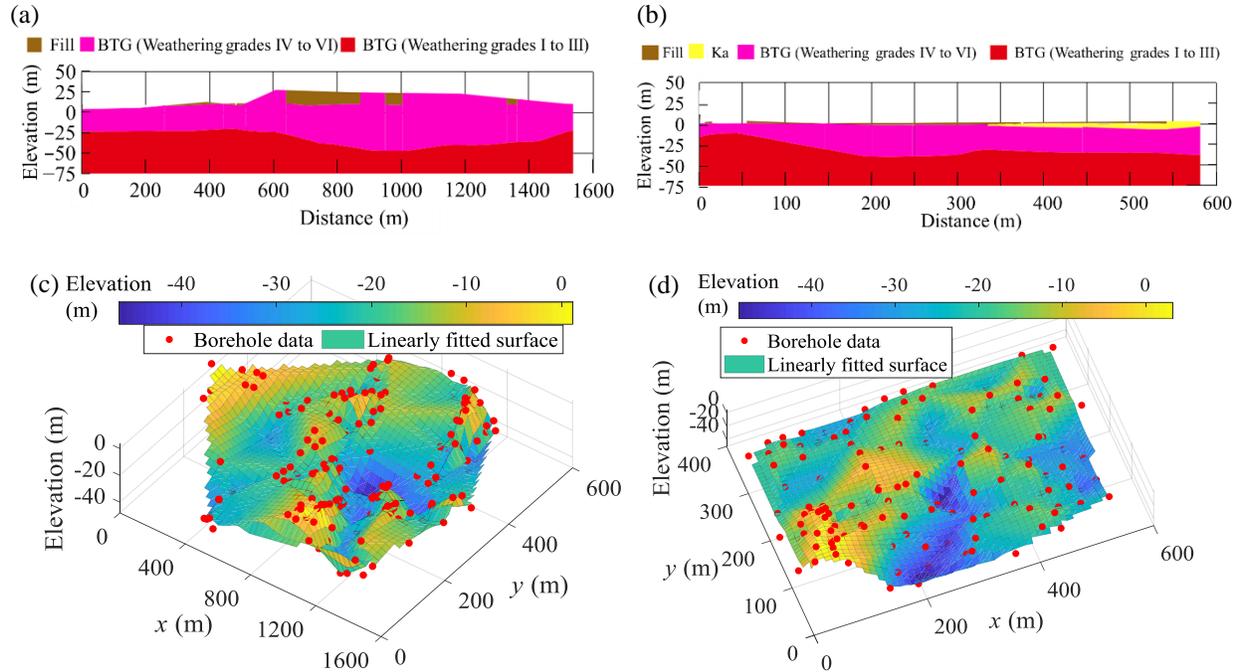


Fig. 2 Geological profiles of two cross-sections and borehole data at the two sites: (a) Geological profile for a cross-section at Site 1; (b) Geological profile for a cross-section at Site 2; (c) Borehole data at Site 1; (d) Borehole data at Site 2.

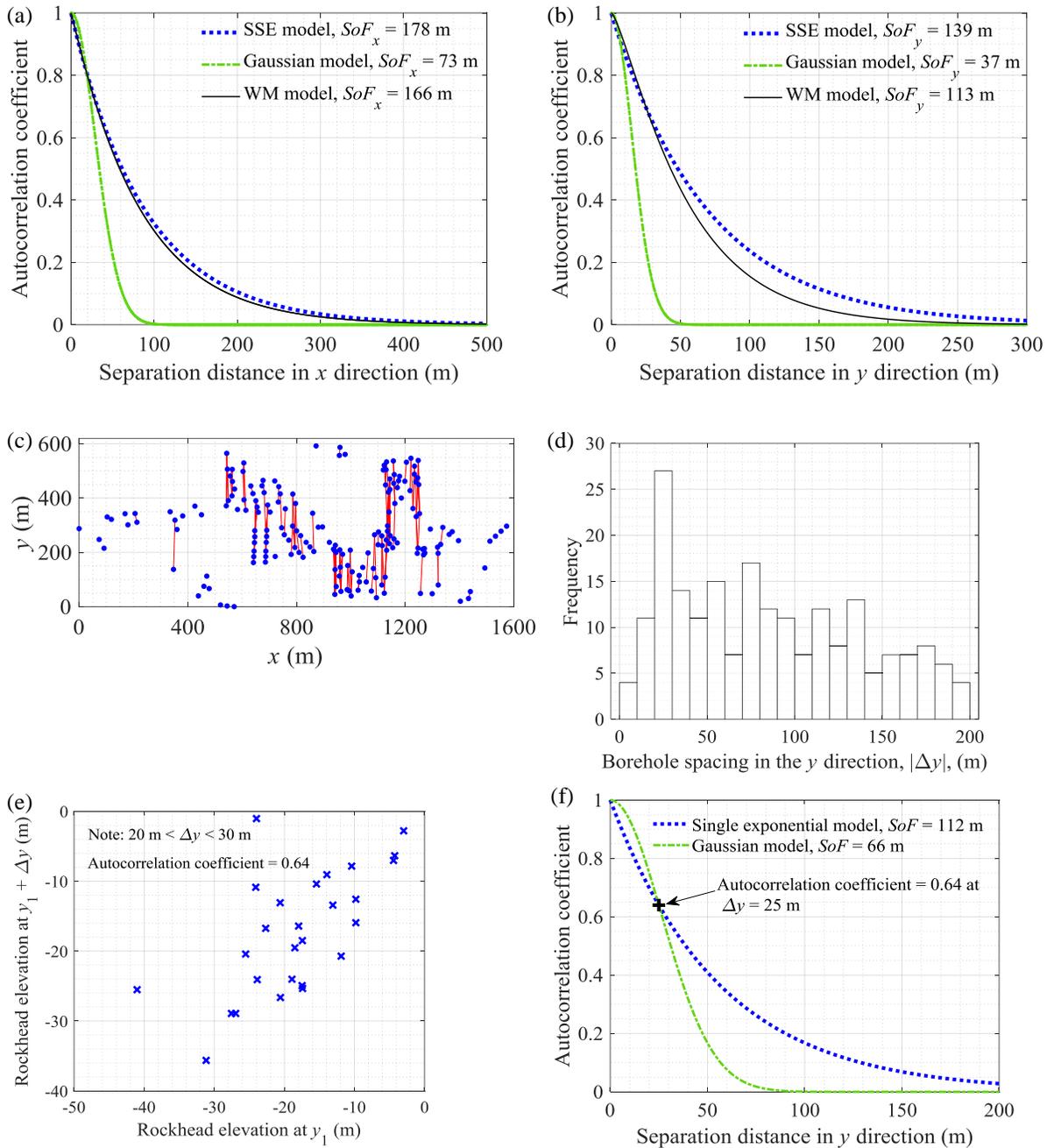


Fig. 3 Autocorrelation coefficients of the rockhead elevation at Site 1 and reason for the estimated smaller SoF for the Gaussian model; (a) x -direction; (b) y -direction; (c) Borehole pairs with $|\Delta x| < 10$ m; (d) Frequency of $|\Delta y|$ for data pairs with $|\Delta x| < 10$ m; (e) Rockhead elevation for the data pairs with $|\Delta x| < 10$ m and 20 m $< \Delta y < 30$ m; (f) Two autocorrelation functions running through the point $(25$ m, 0.64).

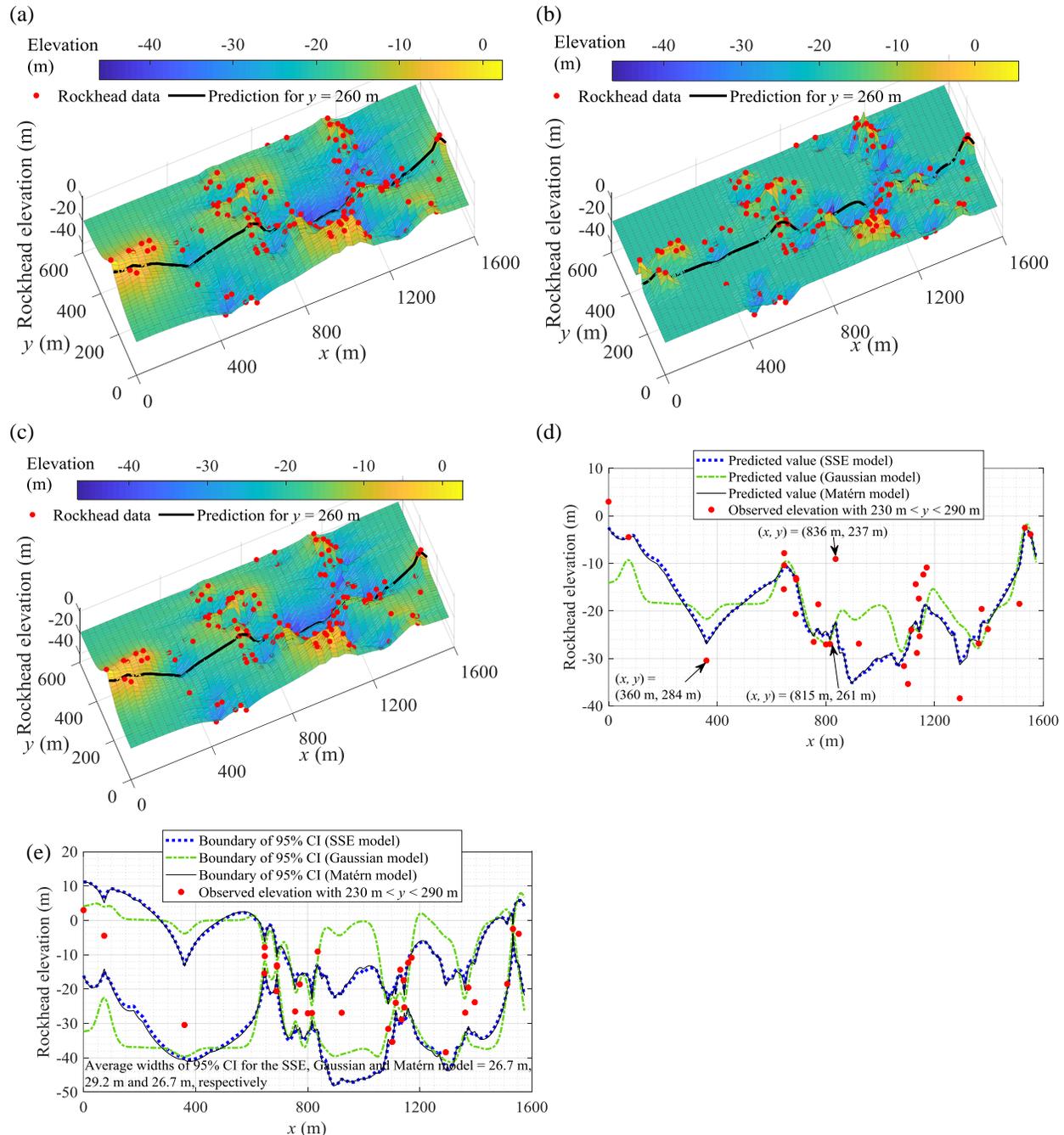


Fig. 4 Rockhead elevation predicted using all the data at Site 1: (a) Posterior mean of rockhead (SSE); (b) Posterior mean of rockhead (Gaussian); (c) Posterior mean of rockhead (Matérn); (d) Predicted rockhead elevation along $y = 260$ m; (e) 95% confidence interval of rockhead elevation along $y = 260$ m.

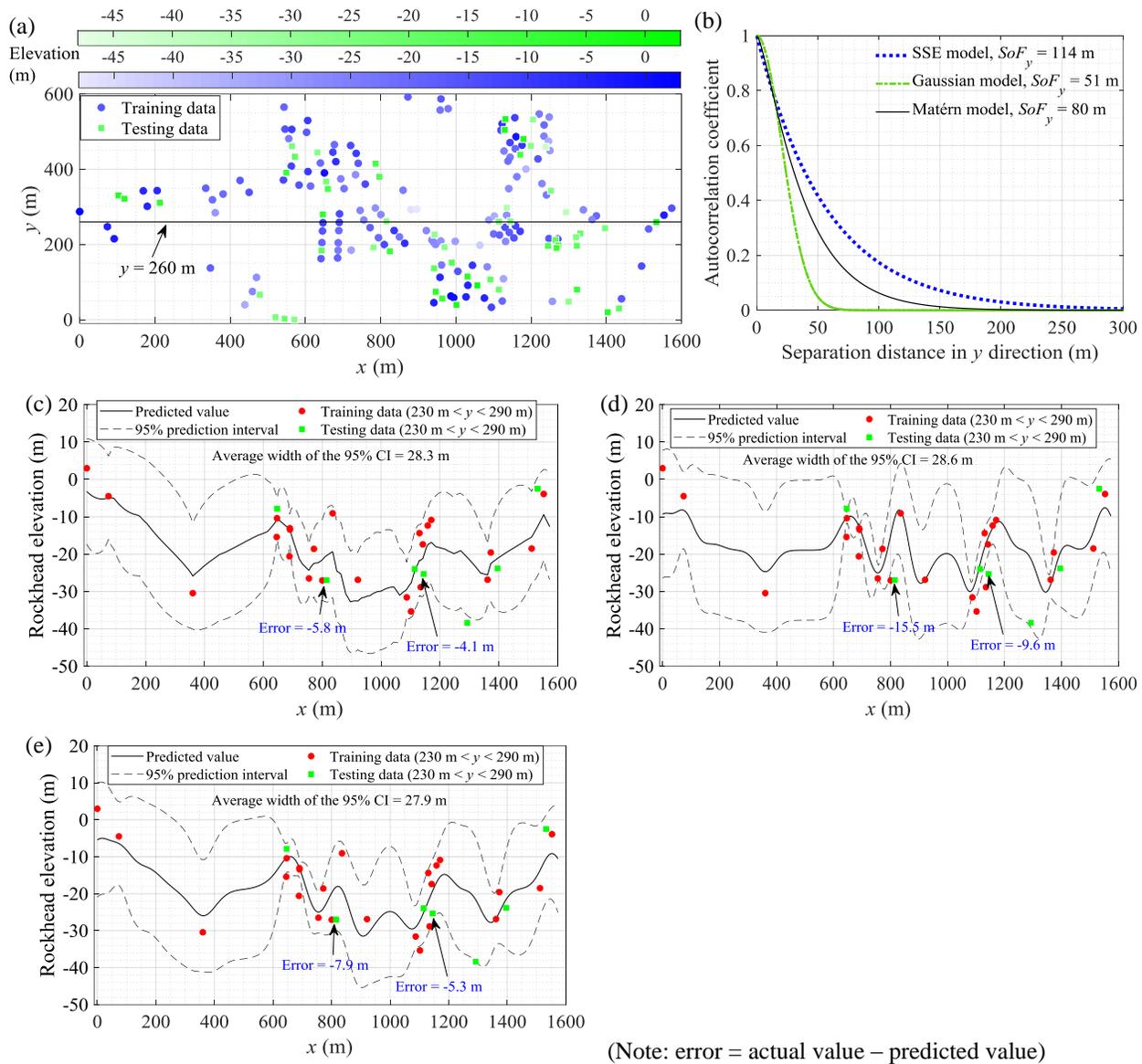


Fig. 5 Rockhead elevation predicted by different autocorrelation models at Site 1: (a) Plan view of training and testing data; (b) Autocorrelation coefficient in the y-direction; (c) Prediction along $y = 260$ m (SSE); (d) Prediction along $y = 260$ m (Gaussian); (e) Prediction along $y = 260$ m (Matérn)

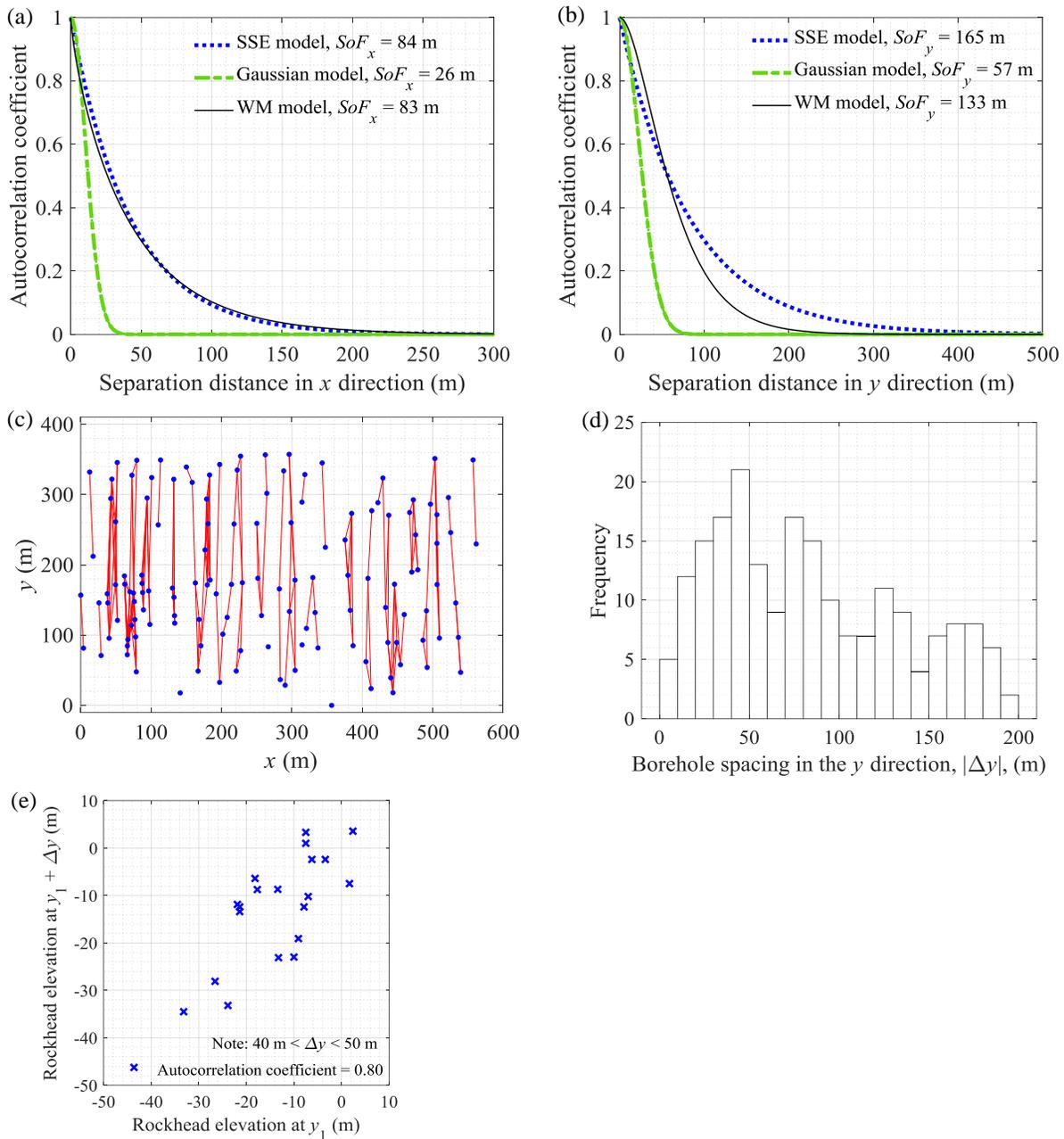


Fig. 6 Autocorrelation coefficients of the rockhead elevation at Site 2 and reason for the estimated smaller SoF for the Gaussian model: (a) x -direction; (b) y -direction; (c) Borehole pairs with $|\Delta x| < 10$ m; (d) Frequency of $|\Delta y|$ for data pairs with $|\Delta x| < 10$ m; (e) Rockhead elevation for the data pairs with $|\Delta x| < 10$ m and $40 \text{ m} < |\Delta y| < 50$ m.

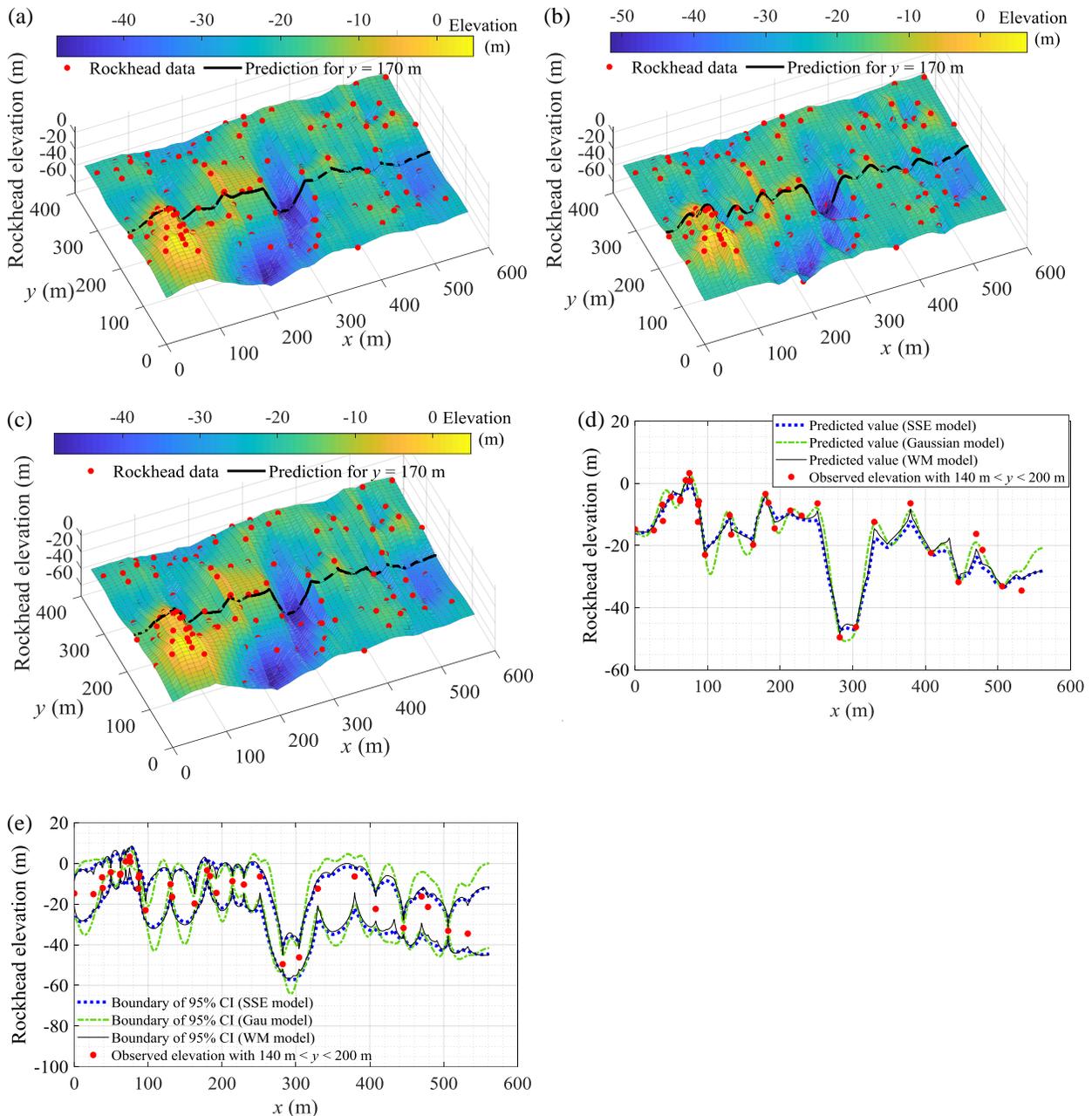


Fig. 7 Rockhead elevation predicted using all the data at Site 2: (a) Posterior mean of rockhead (SSE); (b) Posterior mean of rockhead (Gaussian); (c) Posterior mean of rockhead (Matérn); (d) Predicted rockhead elevation along $y = 170$ m; (e) 95% confidence interval of rockhead elevation along $y = 170$ m.

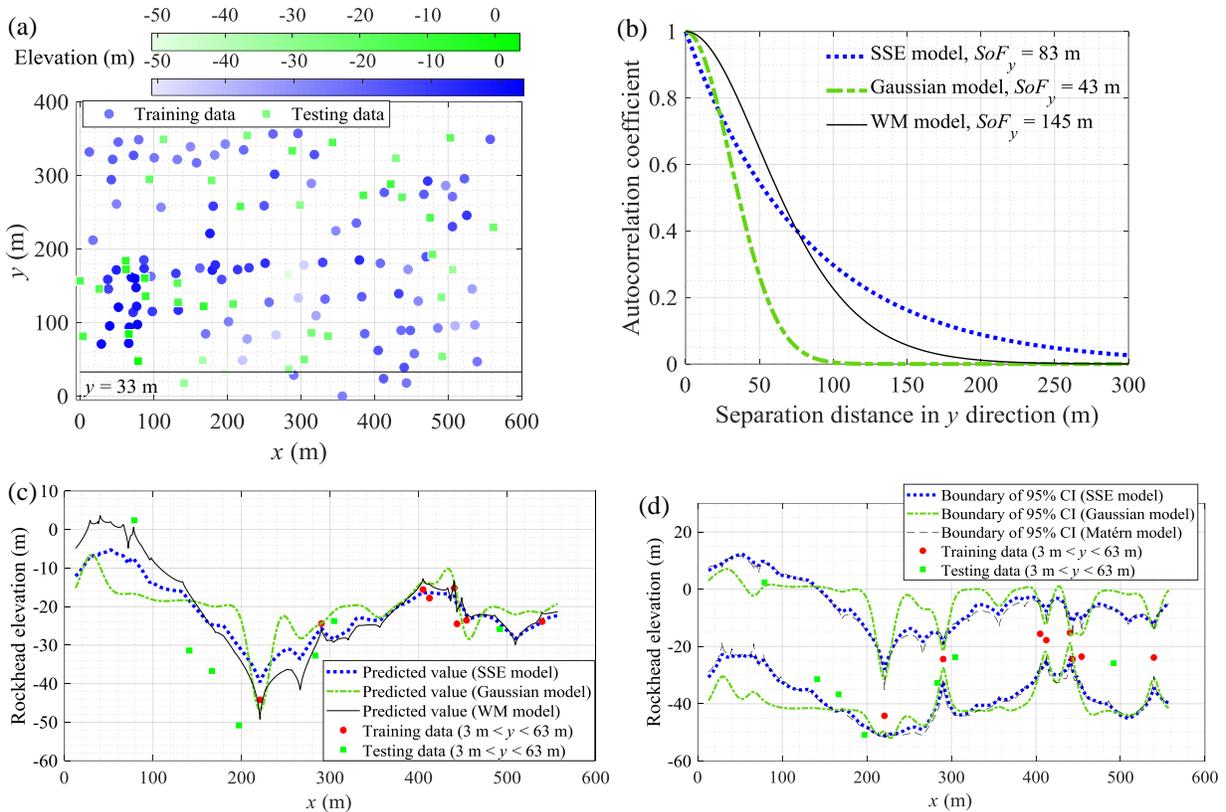


Fig. 8 Rockhead elevation predicted by different autocorrelation models at Site 2: (a) Plan view of training and testing data; (b) Autocorrelation coefficient in the y-direction; (c) Predicted rockhead elevation along $y = 33$ m; (d) 95% CI of the rockhead elevation along $y = 33$ m
(Note: error = actual value – predicted value)

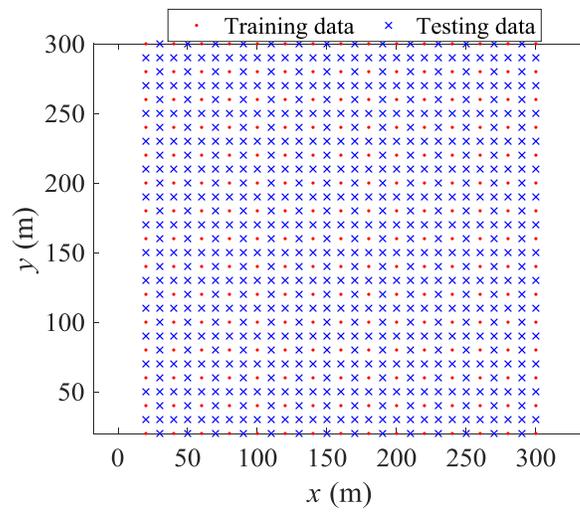


Fig. 9 Plan view of training and testing data when borehole spacing of training data = 20 m