Applying a Python script to predict the geotechnical properties of the Nasiriyah soil

Haneen Mohammed Ali1, Ressol R Shakir2

1Civil Engineering Department, College of Engineering, University of Thi-Qar, Iraq
2Assistant Professor of Geotechnical Engineering, Civil Engineering Department, College of Engineering, University of Thi-Qar, Iraq

haneen1994@utq.edu.iq, haneen2111996@gmail.com, rrshakir@eng.utq.edu.iq, rrshakir@yahoo.com

Abstract. Soil is a natural material that suffers from intrinsic spatial variability resulting from natural factors and their influence on the soil. It became controversial and debated how to estimate the characteristic value of soils to obtain a reliable geotechnical design with low cost and less effort. Usually, foundations are not built on the same site as the screening; investigations are carried out to excavate a little at essential sites. In this paper (423), test wells were collected in the study area to be obtained and tabulated in Excel. The kriging statistics is applied using a python script to predict the values of geotechnical site properties and reliability of the method in estimating spatially varying soil properties values based on measurement data and prior knowledge. The program implements probabilistic kriging statistics and predicts the desired value by entering the coordinates of the locations whose properties you want to know and based on the previously prepared Excel file of known points, coordinates, and property values. The program will be used in two soil sites in the city of Nasiriyah to predict its properties. These points were selected from the examination of soil investigation reports to determine the reliability and accuracy of the program in predicting values. To get more reliable probability values using the kriging method and python scripts. A huge database of prepared and analyzed engineering soil properties has been created based on field investigation reports for projects in Nasiriyah.

Keywords: python script; kriging statistics; Reliability; Characteristic values; Nasiriyah city.

1. Introduction
Soil is a natural material whose properties change with spatial variance. Therefore, geotechnical site characterization is essential in civil engineering, as it aims to demonstrate the spatial distribution of geotechnical properties through laboratory tests. An accurate description of site requests requires deep study. Reporting measurements are usually limited and scattered due to many limitations that are difficult to take advantage of in characterizing the site as needed[1]. Experienced engineers tend to much less than the mean, perhaps less than one standard deviation[2]. Accurate estimation of corresponding to cohesion and soil friction angle plays an essential role in integrity decision-making. Conservatively reliable values are obtained when the combined behaviour of soil shear parameters follows a normal distribution[3][4]. Geotechnical engineers usually seek to describe the soils' engineering properties to be analyzed and designed[5].

The world has recently witnessed a wide trend towards reliability in construction, which directly impacts the development of buildings. The availability of data and requisition requirements for the python programming language leads to many instances of machine learning, automation, and visualization that
enhance building productivity. Geotechnical engineering is the most damaging sector of civil engineering. Using Python forces geotechnical engineers to use a data-driven decision-making tool, leading to reliable and economical geotechnical design[6]. The application of statistical models has become a significant concern for studying complex engineering problems using computers in the past era. Interpolation methods are often used to solve complex engineering problems repetitively because they create an effective tool for modelling and optimizing multiple engineering objectives[7][8]. Probability theory predictors determine the uncertainty associated with distorted values and have a statistical basis. Statistical models depend on the data used, and the data is often too complex to describe mathematically. The kriging model that can be represented in a python script is readily available[9]. Although ARC GIS contains interpolation tools such as kriging methods and others when implementing different values are assigned to parameters until the need arises to get more reliability of geostatistical interpolation methods a python script[11][12]. Python package can simulate soil with a single python script, saving you effort and money and providing reliable results[13][14]. The statistical model is implemented using a highly accurate, reliable, and easy-to-use python script for soil simulation[10].

This paper adds the kriging method to show the usefulness of using Python in everyday geotechnical applications. In some areas of the soil layer, the average spatial characteristics of the soil represent the relevant soil parameters in geotechnical judgment. It is estimated that the spatial mean of soil properties includes statistical uncertainty, inherent spatial uncertainty, and regular uncertainty. The resulting variance can be evaluated utilizing a random field[15]. This work presents a challenge in geotechnical design codes because a small amount of data does not produce meaningful statistics, making choosing an appropriate property value very difficult. One of the most critical problems facing a geotechnical engineer is soil layers’ spatial variation and characteristics. Despite the geographical location and the economic and cultural importance of Nasiriyah city, located in the southeastern part of Mesopotamia, it lacks the study of geotechnical characteristics of this type. Furthermore, there have been few studies available by some researchers on the geotechnical characteristics of some of the southern provinces of Mesopotamia. Therefore, site investigations cannot be considered a valuable tool for a geotechnical engineer unless thoroughly studied. Engineering practice is clearly described in comparison with traditional methods. In this paper, the geotechnical properties of Nasiriyah soil were collected. Then it was tabulated in an Excel file to create a reliable database from which soils could be simulated to produce probability values using the Kriging Stats script in Python. The previous data were added to the investigations examination to estimate the Nasiriyah soil's design values and establish a reliable database, which saves effort and money. It has been concluded that the kriging script effectively determines the characteristic value city of Nasiriyah, and a reliable database of soil properties can be made in any area with minimal cost and effort.

2. Methodology

2.1. The Study area

Nasiriyah is the capital of Thi Qar Governorate and is located in southwestern Iraq, on the banks of the Euphrates River, 225 miles southwest of the capital Baghdad (Figure 1). Desert plains dominate the landscape of the province. This study covers Nasiriyah Governorate, located at coordinates 46.266667E and 31.053888N, with approximately 12,900 square kilometres (5,000 mi), or roughly 3.1% of the area of Iraq Source: Google Earth. It is located in the Mesopotamian plain near the Euphrates River[16]. The tables of Mesopotamia dominate the tetrahedral river. The reactive sediments of the Tigris and Euphrates merge into the northern Basra swamp and the Shatt al-Arab delta plain between Basra and the Persian Gulf. About (60%) of the sediments, silt, and sand, are deposited in the sewers, and clay flows into the Shatt al-Arab[17][18]. A dry continental desert characterizes Nasiriyah with hot summers and cool, humid winters with significant temperatures during the day. The general direction of the wind is
northwest, and wind speeds are characterized by their high, especially in the summer, where the average annual value of wind speed is 5 or more due to the flatness of the earth's surface [31].

![Map of Iraq](image)

**Figure 1.** The location of the study area is on the map of Iraq.

2.2. Data Collection

This study obtained geotechnical data for 85 soil investigation reports from (Construction Consulting Laboratory, Thi Qar University Multidisciplinary Laboratory), and the University of Technology Laboratory. Each lot contains (3-4) test wells except for one of the sites studied, including 225 test wells. The total number of wells is 500 test wells. Some wells were neglected due to the inability to obtain accurate coordinates, and 423 test wells were carefully studied to get comprehensive coordinates. The geolocation of some wells was determined using the Global Positioning System (GPS). Figure 2 shows this evaluation of the soil on which the foundation will be built and the basic resettlement requirements for the design. Exploration of the subsoil conditions of the proposed sites is essential to support the geotechnical design and availability of the construction program as per the needs of the structural engineer.
2.3. Software
In this paper, a Python script is designed to store, retrieve, display, manage and analyze all kinds of geospatial data. The excel file prepared as a database contains the values of the measured properties of the location and the coordinates. It also includes all the tables that have the data that was used in the analysis process. This file was then used within the Python program to analyze and predict the geotechnical site values using the kriging method and produce the maps. The pykriging library in Python was used for this purpose. Finally, the site values were predicted by entering the geotechnical site coordinates. The Python script displays the characteristic value and produces the geotechnical map of soil properties in any area of Nasiriyah city.

2.4. Kriging
It is a probabilistic predictor, which is the name given to a particular class of statistical techniques. It has been used in different countries and areas where a statistical data model is assumed to determine the uncertainty associated with expected values. Match the actual values with the predictor kriging. Kriging is a typeface that, depending on the measurement model, can be precise or homogeneous. It is very flexible and enables you to study spatial graphs, spatial and cross-relationship. Kriging uses statistical models for various output areas, including forecasts, standard prediction errors, probabilities, and quantities. Kriging's flexibility can require a large amount of decision-making. Kriging assumes that the data comes from a static random process, and specific methods assume that the data is usually distributed[20]. The purpose of studying kriging, also known as the best linear unbiased estimate (BLUE), is to obtain the best estimate of any soil characteristic at a given point based on known data. The basic idea of this theory is to estimate X(x) at any point using a linear set computed from the values
of \( X \) at each observation point. Suppose \( X_1, X_2, \ldots X_n \) is the random observations of the field, \( X(x) \) at the point \( x_1, x_2, \ldots x_n \). Then the BLUE at \( X(x) \) at \( x \) is expressed by the equation.

\[
\hat{X}(x) = \sum_{i=1}^{n} \beta_i X_i
\]  

(1)

Where the \( n \) unknown weights \( \beta_i \) are to be determined.

The mean can be expressed in regression analysis, equation

\[
\mu_x(x) = \sum_{k=1}^{m} a_k g_k(x)
\]

(2)

Where \( a_k \) is the unknown coefficients and \( g_1(x) = 1, g_2(x) = x, g_3(x) = x^2 \) and \( m = 2 \), and so on in a one dimension. In two dimensions, the corresponding mean function is linear, with \( g_1(x) = 1, g_2(x) = x_1, g_3(x) = x_2 \) and \( g_4(x) = x_1x_2 \). A similar approach is used in higher dimensions [21].

Then the unknown weights can be obtained from the matrix, equation

\[
K\beta = M
\]

(3)

Where \( K \) and \( M \) depend on the covariance structure,

\[
K = \begin{bmatrix}
C_{11} & C_{12} & \cdots & C_{1n} & g_1(x_1) & g_2(x_1) & \cdots & g_m(x_1) \\
C_{21} & C_{22} & \cdots & C_{2n} & g_1(x_2) & g_2(x_2) & \cdots & g_m(x_2) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
C_{n1} & C_{n2} & \cdots & C_{nn} & g_1(x_n) & g_2(x_n) & \cdots & g_m(x_n) \\
g_1(x_1) & g_2(x_1) & \cdots & g_m(x_1) & 0 & 0 & \cdots & 0 \\
g_2(x_1) & g_2(x_2) & \cdots & g_m(x_2) & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
g_m(x_1) & g_m(x_2) & \cdots & g_m(x_n) & 0 & 0 & \cdots & 0 \\
\end{bmatrix}
\]

In which \( C_{ij} \) is the covariance between \( X_i \) and \( X_j \)

\[
\beta = \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_n \\
-\eta_1 \\
-\eta_2 \\
\vdots \\
-\eta_m
\end{bmatrix}, \quad M = \begin{bmatrix}
C_{1x} \\
C_{2x} \\
\vdots \\
C_{nx} \\
g_1(x) \\
g_2(x) \\
\vdots \\
g_m(x)
\end{bmatrix}
\]

(4)

The quantities \( \eta_i \) are Lagrangian parameters used to solve the variance reduction problem subject to bias conditions. That the \( k \)-matrix is just a function of matching control points and common variations, as it can be flipped once. Then equations (3) and (4) were used iteratively at different spatial points to get the best estimates for each spatial point where the \( M \) vector changes with the weight vector. This
method depends on knowing how the mean differs moderately with the position, which means it must be determined \( g_1, g_2, \ldots \) and knowledge of the covariance structure of the field is usually an assumption of an \( M=1, g_1=1, a_1=\mu_x \) constant, or a suitable linear variable\[22\] \[23\].

To define \( K \) and \( M \) (with, perhaps some interpolation of common heterogeneities is not directly estimated). When there is insufficient data, the coefficient of the covariance function is often executed. A typical model is one in which the exponential covariance diminishes the separation distance.

\[
\tau_{ij} = |x_1 - x_2| \tag{5}
\]

\[
C_{ij} = \sigma_x^2 \exp\left\{-\frac{2|\tau_{ij}|}{\theta}\right\} \tag{6}
\]

The parameter \( \theta \) is called the scale of fluctuation. This model requires the estimation of only two and \( \sigma_x^2 \) Parameters assume that the field is statistically homogeneous and is often suitable for soils displaying strata.

3. Results and Discussion

Researchers usually suggest empirical relationships to relate soil parameters to one another and guess properties that cannot be tested. Usually, the correlations are not of sufficient strength to determine the desired properties. Therefore, a database of investigational tests should be created for neighbouring regions. Because soils differ spatially, kriging statistics are applied, which estimates values based on past data and regression distance between parameters to estimate values to be known. Figure 3 shows two sites. The first site is located within coordinates 591067.03E, 3484104.26N and the second site is situated in coordinates 590716.00E, 3485034.00N. These sites are known for their geotechnical characteristics in the soil investigation reports and were not entered into the excel file prepared as previous data. In this study, this location was randomly selected to check the efficiency of the Python script in applying the kriging method and the accuracy in displaying results based on location data and prior knowledge.

![Figure 3. The Location of the tested points.](image-url)
3.1. Expected values of geotechnical properties

In site 1, Figure 4 shows the value of the cohesion rate in this area, 26.0 kN/m². The value was compared with the investigation report of the soil investigation. The values were verified at the geotechnical site in Table 1. Based on the kriging method for statistical analysis, the Python script provides reliable and accurate geotechnical descriptions. A Python script was applied to evaluate the importance of a standard penetration test at various depths. In Figure 5, the program shows when you enter the coordinates 591067.03E 3484104.26N, the SPT-N value of 4 appears at the depth (0-2) m in this region. It also shows a chart of points with known values adopted in the computation and guides the behaviour of the measured values with the computed value in kriging. This value gives the impression that the soil in this area and at a depth of (0-2) m is weak soil with a consistency of medium stiff sandy clay soil, and its bearing capacity does not exceed 100 kN/m², and the soil friction angle(ø°) ranges between (23-30)°. Based on the relationship between (SPT) and (ø°) of sand, according to Kelhawi and Main 1990.

Table 1. On-site verification results.

| coordinates | E | N | E | N |
|-------------|-----------------|-----------------|-----------------|-----------------|
| characteristic | Site 1 | Site 2 | Site 1 | Site 2 |
| SPT-N (0-2) m | 5 | 4 | 7 | 6 |
| SPT (2-4) m | 3 | 4 | 3 | 3 |
| SPT (14-16) m | 6 | 7 | 6 | 7 |
| SPT (18-20) | 24 | 23 | 24 | 25 |
| Cc (0-2) m | 0.19 | 0.2 | 0.19 | 0.18 |
| Cc (8-12) | 0.24 | 0.235 | 0.25 | 0.26 |
| Cohesion | 26 | 27.5 | 26 | 25.4 |

Figure 6 shows the SPT value at a depth of (2-4) m at site 2, estimated at 3 boreholes. A decrease in the number of boreholes was observed, which indicates that the soil of this layer is less solid than the soil that preceded it, and its bearing rate does not exceed 50 kN / m². In Figure 7, the value of SPT-N at a depth of (14-16) m is shown as 6 boreholes. A noticeable increase in the number of boreholes was observed, meaning that the greater the depth, the greater the number of boreholes. There is a direct relationship between the number of wells and depth, and there is a relationship between the number of wells and the bearing capacity of the soil. The limit in this category does not exceed 100 kN/m². In addition, it can be seen that the values of the probability graph and the limitations they take are an estimate of the most appropriate value. The results of the SPT-N examination program at a depth of (18-20)m also showed the about 24 boreholes, and this value is very logical since an increase in the value of the number of boreholes with depth was observed. The maximum soil tolerance based on SPT-N values range between (200-400)kN/m².
Figure 4. Apply kriging stats to cohesion value using Python script.

Figure 5. Estimation of SPT-N value at depth (0-2) in Site 1 and Site 2.
Figure 6. Estimation of SPT-N value at depth (2-4) in Site 1 and site 2.

Figure 7. Estimation of SPT-N value at depth (14-16) in Site 1 and Site 2.
3.2. Soil compressibility

The study of soil compressibility is critical, as it is an essential characteristic in determining the ability of the soil to swell when constructing engineering structures. Since it is an expensive and complex test, it was chosen in this study to create its database and apply it in a Python script to test the program's prediction efficiency. The compression index is used to find settlement in naturally compacted clay soils when the applied stress is greater than the field stress to which the soil has been subjected in the past. Figure 9 shows the program result for the compression index value at a depth of (0-2) m at site 2; its coordinates give 591067.03E 3484104.26N. The results showed that the compression index value is 0.19. Figure 9 shows a geotechnical map that provides a clear description of the distribution. The unknown values are determined based on known points in the text file for site 2 and neighbouring sites and at different depths. Figure 10 shows the value of the compression index for the same point at a depth of about (10-12) m equal to 0.24. The value of the compression coefficient varies according to the type of soil. From here, we notice an increase in the compression index at the depth (10-12) m.

Estimating the value of the compression index helps the geotechnical engineer in the appropriate planning for building structures on the soil without the need to conduct all tests to determine the properties of the soil. Other aspects such as time, cost and effort are invested based on previous data to determine the soil and statistics kriging the required property of the soil. The results obtained from this study indicate that the pressure index values are typical for the soil of the city of Nasiriyah, as the soil is clay with silt. When uniformity testing is not possible, kriging with a Python script is an excellent way to add the previous data for guesswork and get the average value for predicting soil C parameter.
4. Conclusion
In this work, a Python script was created to know the characteristic value of soil properties directly. Using the statistical kriging method to obtain a safe structure when designing is checked and compared with the measured values, it is very close to the site and can be adopted. In a simple Python script, it is
possible to determine the characteristic value of any property in any part of Nasiriyah city based on the measured values of the location. This work poses a challenge in the world of geotechnical engineering design. Python has proven to be a powerful tool in probabilistic characterization and addressing spatial uncertainty in geotechnical engineering, as indicated by the sample site for the studied projects. It can process data efficiently and automate repetitive data easily. Python can help geotechnical engineers make a decision based on historical data. Neighbouring projects can get economical and reliable geotechnical design. Python scripts have many advantages for future work by adding additional data to an Excel file or applying it to another study area. There is no problem with changing the property parameters of the interpolation method in Python script files.

ACKNOWLEDGEMENTS
Thanks and appreciation to the Deanship of the College of Engineering and its laboratories for their assistance in preparing the official books to obtain the reports formally and legally. Special thanks to the Engineering Construction Laboratory and the Multidisciplinary Laboratory for submitting investigation reports.

python script
from pyKriging.krige import kriging
import numpy as np
import pandas as pd
import os
fixed_path = os.path.join(os.path.expanduser ('~'),"desktop")
target_path = input('Enter folder name then the file name ex:folder\data.xlsx:
')
full_path = fixed_path +target_path
# We generate our observed values based on our sampling plan and the test function
data = pd.read_excel(full_path)
x = data.iloc[:, :-1].values
y = data.iloc[:, -1].values
y = np.array(y).flatten()
print('Setting up the Kriging Model')
# Now that we have our initial data, we can create an instance of a kriging model
k = kriging(x, y, name='simple')
k.train(optimizer='ga')
# #And plot the model
print('Now plotting final results...')
k.plot()
print(k.predict())

References
[1] J. Xu, L. Zhang, Y. Wang, C. Wang, J. Zheng, and Y. Yu, "Probabilistic estimation of cross-variogram based on Bayesian inference," Eng. Geol., vol. 277, no. July, p. 105813, 2020, doi: 10.1016/j.enggeo.2020.105813.
[2] S. Lacasse, F. Nadim, I. Centre, N. Geotechnical, and I. Ngi, "OTC 19117 Statistical Description of Characteristic Soil Properties," no. January, 2007, doi: 10.4043/19117-MS.
[3] X. Z. Wu, "Modelling dependence structures of soil shear strength data with bivariate copulas and applications to geotechnical reliability analysis," Soils Found., vol. 55, no. 5, pp. 1243–1258, 2015, doi: 10.1016/j.sandf.2015.09.023.
[4] L. Di, M. Daniela, and V. Remo, "Laboratory shear strength parameters of cohesive soils: variability and potential effects on slope stability," pp. 101–106, 2013, doi: 10.1007/s10064-013-0459-6.
[5] T. T. Bakala and S. Thanappan, "Study on the shear strength and some of index properties of
cohesive soil," vol. 2, no. 2.

[6] B. A. Tirta and J. Assessment, "Python Application in Geotechnical Engineering Practices," no. February, 2021.

[7] G. Arcidiacono et al., "Kriging models for payload distribution optimization of freight trains," *Int. J. Prod. Res.*, vol. 7543, no. February, pp. 1–13, 2017, doi: 10.1080/00207543.2016.1268275.

[8] M. A. Bouhlel, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier, and J. R. R. A. Martins, "Advances in Engineering Software A Python surrogate modeling framework with derivatives," no. September 2018, pp. 1–13, 2019.

[9] "PyKriging Documentation," 2021.

[10] H. Jürgens and S. Henke, "The design of geotechnical structures using numerical methods," 2021, doi: 10.1088/1755-1315/727/1/012021.

[11] C. Republic, "AUTOMATIC GENERATION OF DIGITAL ELEVATION MODELS USING PYTHON SCRIPTS Ing. Zdena Dobesova, Ph.D.," vol. 2, pp. 599–604, 2011, doi: 10.5593/SGEM2011/S11.102.

[12] B. K. Krivoruchko and S. D. Team, "Empirical Bayesian Kriging Implemented in ArcGIS Geostatistical Analyst," pp. 6–10, 2012.

[13] W. Zhang, M. J. Cassidy, and Y. Tian, "Practical Large-Deformation Finite-Element Method for 3D Geotechnical Problems Involving Free Surface Deformations," vol. 20, no. 2014, pp. 1–13, 2020, doi: 10.1061/(ASCE)GM.1943-5622.0001648.

[14] M. Zhu, F. Mckenna, and M. H. Scott, "SoftwareX OpenSeesPy: Python library for the OpenSees finite element framework," *SoftwareX*, vol. 7, pp. 6–11, 2018, doi: 10.1016/j.softx.2017.10.009.

[15] "Principles of Probabilistic Characterization of Soil Properties," p. 1984, 1984.

[16] E. A. Mandhour, "Assessment of the liquefaction susceptibility of Al-Nasiriya city soil based on plasticity properties," vol. 7, no. 1, pp. 542–551, 2020.

[17] A. A. M. AQRAWI and G. EVANS, "Sedimentation in the lakes and marshes (Ahwar) of the Tigris-Euphrates Delta, southern Mesopotamia," *Sedimentology*, vol. 41, no. 4, pp. 755–776, 1994, doi: 10.1111/j.1365-3091.1994.tb01422.x.

[18] R. R. Shakir, "Probabilistic-based analysis of a shallow square footing using Monte Carlo simulation," *Eng. Sci. Technol. an Int. J.*, vol. 22, no. 1, pp. 313–333, 2019, doi: 10.1016/j.jestch.2018.08.011.

[19] E. Mandhour, "SWELLING POTENTIAL AND MINERALOGY OF THI – QAR," no. December 2011, 2020.

[20] B. Nas and A. Berktay, "Groundwater quality mapping in urban groundwater using GIS," *Environ. Monit. Assess.*, vol. 160, no. 1–4, pp. 215–227, 2010, doi: 10.1007/s10661-008-0689-4.

[21] G. A. Fenton and D. V. Griffiths, "Settlement of Shallow Foundations," *Risk Assess. Geotech. Eng.*, pp. 311–346, 2008, doi: 10.1002/9780470284704.ch10.

[22] S. Firouzanbandpey, L. B. Ibsen, D. V. Griffiths, M. J. Vaahdatirad, L. V. Andersen, and J. D. Sørensen, "Effect of Spatial Correlation Length on the Interpretation of Normalized CPT Data Using a Kriging Approach," *J. Geotech. Geoenvironmental Eng.*, vol. 141, no. 12, p. 04015052, 2015, doi: 10.1061/(asce)gt.1943-5606.0001358.

[23] S. H. Marques, A. T. Gomes, and A. A. Henriques, "Reliability assessment of Eurocode 7 spread foundations design methodology," 2011, doi: 10.1061/41183(418)63.

[24] S. Saray, T. Satir, and N. Dogan-Saglamtimur, “Proficiency of Maritime English course: An investigation in Istanbul, Turkey”, Heritage and Sustainable Development, vol. 3, no. 1, pp. 6–15, Jan. 2021.