Magazine of Civil Engineering. 2024. 17(6). Article No. 13009

Magazine of Civil Engineering

ISSN 2712-8172

journal homepage: http://engstroy.spbstu.ru/

Research article

UDC 624

DOI: 10.34910/MCE.130.9

# Evaluation of prior probability distribution of undrained cohesion for soil in Nasiriyah

***R.S. Shakir*** ✉ **,** ***Z. Abd Al-Haleem***

*University of Thi-Qar, Nasiriyah, Iraq*

✉ *rrshakir@utq.edu.iq*

**Keywords:** clay soil, prior distribution, prior knowledge, site characterization, undrained cohesion

**Abstract.** The objective of the study is to evaluate the prior probability distribution (PPD) of undrained cohesion ($C_u$) parameter for soil in Nasiriyah, southern Iraq, based on prior knowledge and observations. Estimated PPD of $C_u$ can be used in Bayesian approach to update the observed value in any project in this region using the posterior probability distribution, because it is considered as a measure of the initial belief about a random variable before considering any data. The research used five methods to express the PPD of $C_u$. Two of them are for non-informative data, i.e. uniform distribution and Jeffreys prior, and three of them – for informative data, which include maximum entropy, regression analysis and subjective probability. They were applied to data collected from different sources in Nasiriyah, based on site investigation reports. The ranges of mean, standard deviation and vertical scale of $C_u$ fluctuation were found to be 12–62 kPa, 0.5–27.6 kPa and 6–8 m, respectively. It was concluded that Jeffreys method is used well with individual models at the mean value of cohesion of 28.66 kPa and the standard deviation of 1.19 kPa. The maximum entropy can be used for the least informative data, while respecting the given constraints. The mean value of cohesion was 28.7 kPa, and the standard deviation was 1.2 kPa. Finally, for a finite number of 152 cohesion values, the subjective probability assessment approach, which takes into account expert knowledge and judgment, is the most appropriate method with the mean value of cohesion of 37 kPa and the standard deviation of 8.8 kPa.

**Acknowledgements:** The authors gratefully acknowledge the data provided by the Office of Scientific and Advisory Services (College of Engineering, University of Al-Qadisiyah), the National Company for Engineering Examination and Designs, and the Al-Sabtain Company for Engineering Tests.

**Citation:** Shakir, R.R., Al-Haleem, Z. Abd. Evaluation of prior probability distribution of undrained cohesion for soil in Nasiriyah. Magazine of Civil Engineering. 2024. 17(6). Article no. 13009. DOI: 10.34910/MCE.130.9

## *1. Introduction*

Undrained cohesion ($C_u$) is a significant and necessary property used to characterize the undrained shear strength of soil. It can be used to design and analyze different geotechnical problems, such as bearing capacity of shallow and deep (pile) foundation. Laboratory tests, such as triaxial shear or unconfined compression tests are commonly used to measure the $C_u$ value performed on undisturbed samples from field investigation. These processes are relatively time-consuming and expensive, and there are many sources of error that make the $C_u$ value uncertain, such as quality of consistency and degree of saturation.

Prior distribution is an essential and first step in the Bayesian approach, which includes knowledge of the uncertain parameters combined with the probability distribution of recent data to get the posterior distribution [1]. It is used in the Bayesian approach to update the observed value of cohesion in a project in this area. Prior knowledge includes general information about the hypothesis that may be relevant or unclear. It may be previously gained knowledge of any type of distribution that correctly reflects the state of the model or parameter under study. When the posterior distributions belong to the same family of probability distribution of prior distribution, they are known as conjugate distributions [2]. Previously gained information can be divided into two groups: informative prior knowledge and non-informative prior knowledge, according to its accuracy and quantity.

Non-informative prior knowledge can be used to probabilistically characterize a homogeneous soil layer, which requires three model parameters, e.g. mean $\mu$, standard deviation $\sigma$, vertical scale of fluctuation $\lambda$ [3]. It depends on the value of the rate, i.e. the limits of the parameter, with the most dependent uniform distribution in describing the uncertainty [2]. It led to the conclusion that prior knowledge is consistent with the rates cited in [4, 5]. Jeffreys [6] analyzed the optimal selection of non-informative prior distributions. He found that the prior probabilities should be assumed to be uniform over the parameter for variables with domains of $(-\infty, \infty)$ and uniform over the logarithm for variables with domains of $[0, \infty]$. The latter becomes inversely proportional to the soil parameter in arithmetic space [7].

Informative prior knowledge shows that certain prior estimates are preferable to others. It is used, when there is an abundance of collected information about soil properties, which is divided into three methods that help to obtain the probability distribution (e.g., subjective probability, maximum entropy, regression analysis). Subjective probability assessment framework (SPAF) is used to evaluate the plausibility of previously acquired uncertain statistical estimation [8, 9]. There are some circumstances, in which a different kind of distribution might be necessary to hold the available data without the risk of obstructing the conclusions of the analysis. The maximum entropy principle could be used to define the prior prior distribution [10, 11]. In this situation, entropy refers to randomness of the information and is relatively similar to the concept of entropy in physical systems [12]. The maximum entropy method was used to estimate the probability distribution and to develop the reliability of the slopes [13]. The probability distribution can also be obtained through regression analysis [14, 15] with the μ and σ of the data. This analysis provides a practical way to construct prior distributions from prior knowledge. It usually provides a weight similar to the prior information. In addition, it is challenging to include systematically subjective judgments in statistical analysis, which are sometimes a crucial component of prior knowledge [16].

The objective of geotechnical site description is to identify the soil layers and to evaluate the properties of the soil and rocks for the analysis of geological and geotechnical systems [9, 17]. The correct site characterization requires comprehensive measures at several places since soils are common geomaterials with spatial heterogeneity [18].The process of geotechnical site characterization involves several steps, including desk research, site surveys, laboratory tests [3], analysis or comprehension of site data, and inference of soil and rock properties [19]. Any method used in the site characterization that solely relies on measured data is called “data-driven site characterization”, and this includes both site-specific data collected for the current project and existing data of any kind collected from previous stages of the same project or previous projects at the same site, adjacent sites, or elsewhere [20, 21]. Geotechnical engineering frequently deals with uncertainty, and engineering design must take it into account [22]. There are three main categories of uncertainties: test errors, existing model uncertainty, and inherent variability [23, 24]. The evaluation of soil properties and their statistics based on past information are not clearly known results rather than clear-cut conclusions because of uncertainties in the currently studied information and engineers' qualification. As a result, such not clearly known evaluations are referred to as uncertain evaluation [3].

This article investigated approaches to enhance and estimate prior knowledge. It included data obtained from the geotechnical reports of projects implemented in Nasiriyah in southern Iraq. The purpose of collecting this data, which was considered as prior knowledge, is to quantify it and calculate its prior distributions to achieve the most appropriate geometric judgment sense and insert it into a Bayesian framework. The non-informative knowledge (e.g. uniform distribution) was based on the average maximum and minimum value of $\mu$, $\sigma$, and $\lambda$ of $C_u$, in addition to the maximum entropy method and the regression analysis method, besides the SPAF, which in turn was based on many stages to achieve the prior distribution appropriate to the soil property.

# 2. Methods

## 2.1. Study Area

Thi-Qar is a province in the south of Iraq that borders the provinces of Basra, Wassit, Muthanna, Missan, and Qadissiya (Fig. 1) [25, 26]. The province is located about 370 km southeast of Baghdad between latitude 31°14' N and longitude 46°19' E. It has a total area of 13552 km$^2$ [27]. The province features a hot desert climate with very hot and dry summers and mild winters. The mean daily maximum in the summer exceeds 40 °C [28]. It is located in the Mesopotamian plain. There are deposits of alluvial silt from the Euphrates and Tigris rivers. The soil in this region is a floodplain formation consisting of clay, silt, and sand, where the silt forms 60 % of the deposits. The silt soil and sand settle in the swamps, the mud runs below the Shatt al-Arab, and one million tons of sediments are dried up annually (12000 years) in these rivers that flow from the northwest to the southeast of Turkey through this basin [29, 30]. Since the area is free from surface erosion of ancient rocks, the city of Nasiriyah belongs to the floodplain zone and represents the last formation of surface of Iraqi geology [31, 32].


**Figure 1. Location of Thi-Qar Province on the map of Iraq.**

## *2.2. Data Collection*

In this study, the data was gathered from a variety of places within Thi-Qar Province. The province is divided into many sections (for example, Nasiriyah, Al-Chibayish, Suq Al-Shuyukh, Al-Rifai, Al-Shatrah, Qalat Sukkar and Al-Nasr), as shown in Fig. 2. The available soil investigations for these areas were included in the data. To analyze the geotechnical properties, as well as to measure soil consistency at various depths, the standard penetration test of approximately 200 boreholes was used.

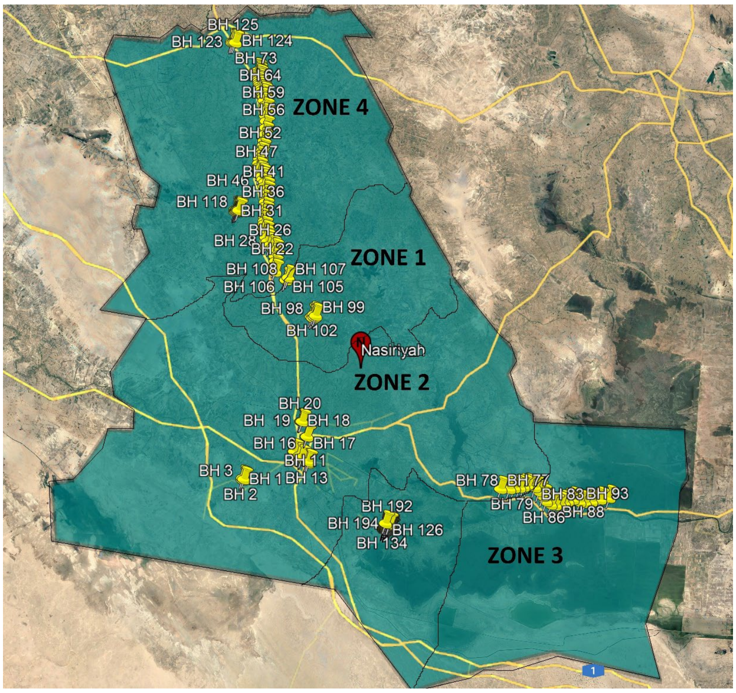



**Figure 2. Location of the study area.**

## 2.3. Description of Thi-Qar Soil

The data collected from multiple projects in Nasiriyah included zones 1–4 (Fig. 2). Zone 1 represents the description of soils of Qalat Sukkar and Al-Nasr. Zone 2 represents the soils of Al-Shatrah. Zone 3 and 4 represent the soils of Nasiriyah, Al-Chibayish, and Suq Al-Shuyukh, respectively.

The soil layers consist of silty clayey or silty clay soils of little to medium plasticity ML, that varies in color from light gray to light brown and have a fairly firm texture. It is also characterized by patches of salt and gypsum crystals that extend in Zone 1 at a depth of 14 m and 20–24 m, and in Zone 2 at a depth of 12–14 m and in Zone 3 at a depth of 11–13 m.

Zone 3 consists of a hard to very hard layer and medium plasticity. Layer of light green clay appears in the upper layer at a depth of 4 m, at 8–12 m, and also at 14–24 m. Zone 4 extends at a depth of 2–12 m and at 15–24 m. It classified by type CL.

Soil layers of little plasticity or non-plasticity are located in Zone 1, as it consists of medium to dense gray soil to soft brown of soft sandy grains (SM), at a depth of 14–20 m. The same applies to Zone 2 at a depth of 15–18 m and at a depth of 4–8 m and at a depth of 12–14 m in Zones 3 and 4, respectively. Fig. 3 shows the profile of each of these layers of Nasiriyah soils in the Thi-Qar Province.

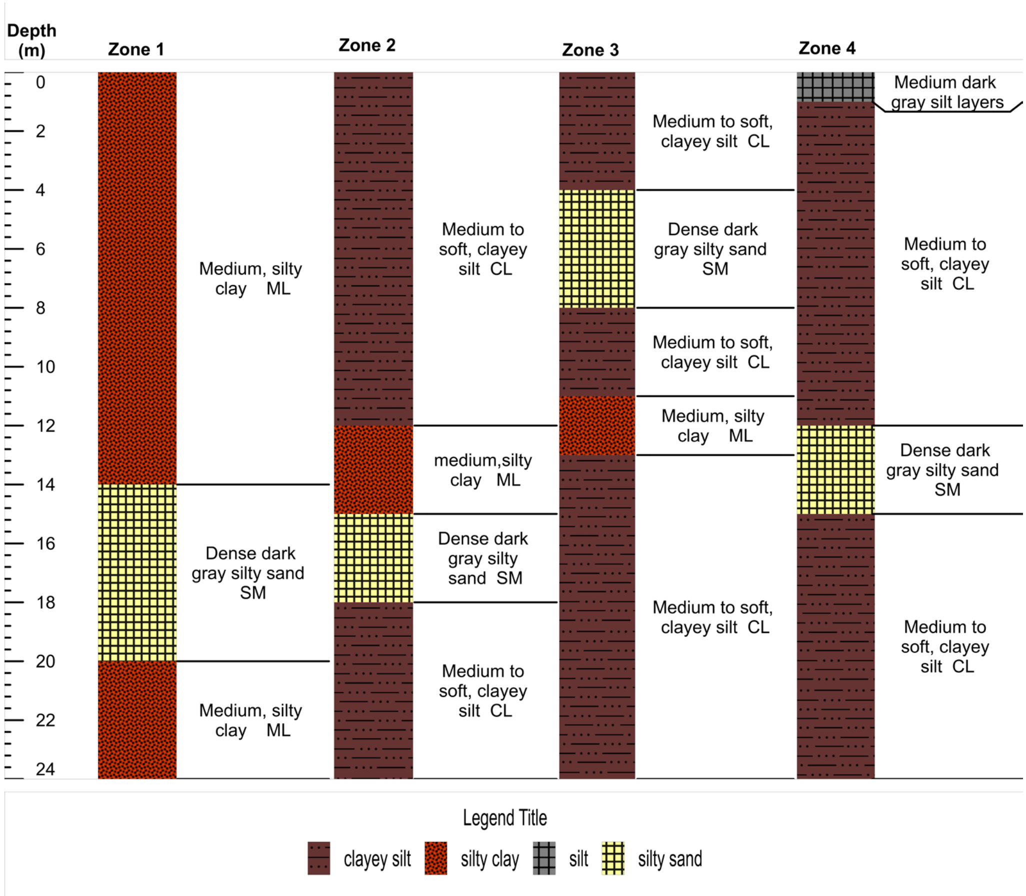



**Figure 3. Soil profile of Nasiriyah based on borehole samples considered in the current study.**

## 2.4. Prior knowledge

During geotechnical site characterization, both previously acquired knowledge and observation data that are site-specific are used to determine soil parameters, that can be expressed as $\theta_i$ (e.g., $\mu$, $\sigma$, $\lambda$). From a Bayesian perspective, such a procedure may be viewed as an update of previously acquired knowledge using observational data from a project site [19]. The previously acquired distribution is an essential part of Bayesian inference because it resembles information about an uncertain parameter that is accompanied by the probability distribution of new data to produce the posterior distribution. The prior distribution can be obtained by applying the methods shown in Fig. 4.

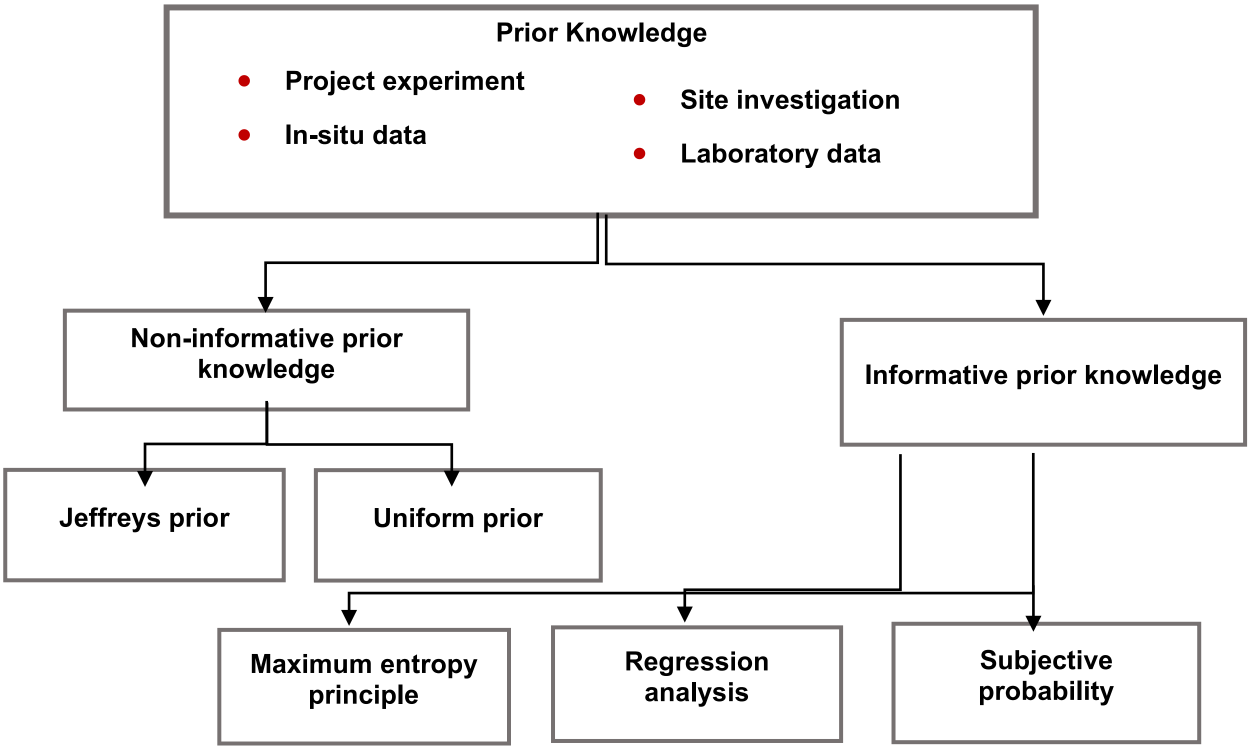



**Figure 4. Methods of expressing the prior knowledge.**

### *2.4.1. Non-informative prior knowledge*

Since the early studies of Bayes and Laplace [33], there has been some consensus that assigning a uniform distribution to a scalar parameter $\theta_i$ is one of the easiest methods to depict a state of ignorance about its value. Each value of $\theta_i$ has the same prior probability with a uniform prior (or probability density). It expresses vague data regarding the specified variable. In the lack of specific prior knowledge, vague priors are frequently used as the default prior option, however, it is crucial to emphasize that ambiguous priors are not always non-informative to the analysis, which is a common mistake in the literature with vague priors being incorrectly referred to as non-informative priors. The use of a non-informative antecedent in parameter estimation issues yields findings that are not materially different from those of conventional statistical analysis. Based on prior distribution, there is no predisposition for any value that falls within the range of potential values $\theta_i$. The prior distribution of $\theta_i$ can be written as follows.

$$P\left(\theta_p\right)=\prod_{i=1}^{n_m} P\left(\theta_i\right); \quad (1)$$

$$\left(\theta_i\right)=\begin{cases}\dfrac{1}{\left(\theta_{i,\max}-\theta_{i,\min}\right)} & \text{for } \theta_i \in\left[\theta_{i,\min},\ \theta_{i,\max}\right],\ i=1,2,\ldots,n_m \\ 0 & \text{otherwise}\end{cases}. \quad (2)$$

Consider a model parameter $\theta_i$ (e.g., $\mu,\ \sigma,\ \lambda$).

$$P(\mu)=\begin{cases}\dfrac{1}{\mu_{\max}-\mu_{\min}} & \text{for } \mu \in\left[\mu_{\min},\ \mu_{\max}\right] \\ 0 & \text{otherwise}\end{cases}; \quad (3)$$

$$P(\sigma)=\begin{cases}\dfrac{1}{\sigma_{\max}-\sigma_{\min}} & \text{for } \sigma \in\left[\sigma_{\min},\ \sigma_{\max}\right] \\ 0 & \text{otherwise}\end{cases}; \quad (4)$$

$$P(\theta_P) = \begin{cases} \dfrac{1}{\mu_{\max} - \mu_{\min}} \times \dfrac{1}{\sigma_{\max} - \sigma_{\min}} \times \dfrac{1}{\lambda_{\max} - \lambda_{\min}} & \text{for } \mu \in [\mu_{\min}, \mu_{\max}], \sigma \in [\sigma_{\min}, \sigma_{\max}], \text{ and } \lambda \in [\lambda_{\min}, \lambda_{\max}] \\ 0 & \text{others} \end{cases}. \quad (5)$$

$P(\theta_P)$ refers to the prior probability distribution, $\mu$ is the mean, $\sigma$ is the standard deviation, $\lambda$ is the vertical scale of fluctuation.

This form of prior knowledge of the soil property $C_u$ in Nasiriyah city was investigated. During the site characterization, the maximum and minimum values of the mean, standard deviation and vertical fluctuation scale were determined and presented as shown in Tables 1 and 2.

***Table 1. Ranges of the mean and standard deviation of soil properties of Thi-Qar.***

| Soil property | Range of prior evaluation of mean | Range of the prior evaluation of standard deviation | Range of COV% |
|---|---|---|---|
| Cohesion $C_u$ (kPa) | 12–62 | 0.5–27.58 | 1.6–84.85 |

***Table 2. Ranges of the vertical scale of fluctuation of soil properties of Thi-Qar.***

| Soil property | Range of the prior evaluation of vertical scale of fluctuation(m) |
|---|---|
| $C_u$ (kPa) | 6–8 |

Jeffreys proposed prior, non-informative method used to estimate parameters when an appropriate prior distribution is not obtainable.

Jeffreys noticed that a non-informative prior proportionate to the square root of the information matrix's determinant is non-informative. Then, Jeffreys prior is:

$$P(\theta) \propto I(\theta)^{\frac{1}{2}}. \quad (6)$$

Regarding the non-invariance of prior distributions based on Fisher's information matrix, Jeffreys shows up a larger finding. Fisher [34] proposed that the data contained in a set of observations $x = x_1, ..., x_n$, with regard to a parameter $\theta$, are to be as follows:

$$I(\theta) = E_\theta\left[\left(\frac{d}{d\theta}\ln\left(p\left(x/\theta\right)\right)\right)^2\right]. \quad (7)$$

$I(\theta)$ is the Fisher information.

Please note that $p(x/\theta)$ is the probability of $\theta$.

## *2.4.2. Informative prior knowledge*

Informative prior distributions are those based on knowledge other than the immediate measured data at hand (e.g., prior data or engineering opinion). Prior information about the value of a parameter exists, the prior distribution might be informative rather than non-informative. Any probability distribution may be used to depict prior information. Increase in prior knowledge or data is more beneficial. The right informative distribution of prior $\theta_i$ information must be estimated by appropriate evaluating of the plausibility of various prior $\theta_i$ evaluation in light of previous experience.

### *2.4.2.1. Maximum entropy principle and statistical analysis method*

This method is the least biased method and by selecting a distribution with entropy, the most appropriate probability distribution of $C_u$ can be determined, which increases the uncertainty measure. Prior knowledge regarding the model parameters may also be used to evaluate the informative prior distribution, which resembles the degree of belief (or certainty level) of prior information according to the model parameters. When there is insufficient prior information on $\theta_i$, the maximum entropy method can also be used to infer the prior distribution from the established data [35, 11]. Using the maximum entropy principle, prior probability distributions for Bayesian inference are frequently generated. Jaynes [36] was an ardent supporter of this method, believing that the distribution with the maximum entropy was the least informative. Within the maximal entropy technique, the information entropy $H_I$ is used as a measure of the uncertainty of the prior probability density function (PDF) $P(\theta_i)$ of $\theta_i$ and is defined as [35]:

$$H_I = -\int P(\theta_i)\ln\left[P(\theta_i)\right]d\theta, \tag{8}$$

where $\theta_i$ is a random variable with a continuous distribution and $P(\theta_i)$ is the PDF. A function is constructed to assess the entropy difference between $P_1(\theta_i)$ and $P_2(\theta_i)$ probability assignments [37].

$$H\left[P_1(\theta_i), P_2(\theta_i)\right] = \int P_1(\theta_i)\ln\left[\frac{P_1(\theta_i)}{P_2(\theta_i)}\right]d\theta. \tag{9}$$

According to Jaynes, the minimum biased assignment of probabilities is one that reduces entropy while satisfying the restrictions given by the available information.

$$H\left[P(\theta_i), P_0(\theta_i)\right] = \int P(\theta_i)\ln\left[\frac{P(\theta_i)}{P_0(\theta_i)}\right]d\theta. \tag{10}$$

The maximum entropy principle implies that, given specified restrictions on the prior, the prior should be the distribution with the maximum entropy that follows these requirements. The most fundamental condition is that $P$ must reside in the probability simplex, i.e.

$$\begin{gathered} P(\theta_i) = 1 \text{ and } P(\theta_i) \geq 0. \\ \int \theta_i P(\theta_i) d\theta = \mu_i,\ i = 1, 2, \ldots, n_m, \end{gathered} \tag{11}$$

where $P_0(\theta_i)$ is the prior distribution function, $P(\theta_i)$ is the actual distribution function of the random variable $\theta_i$, and $n_m$ is the greatest level of moments examined for the random variable.

The maximum entropy of the normal distribution gives the following equation:

$$E(x) = \mu,\ E(x-\mu)^2 = \sigma^2. \tag{12}$$

The statistical analysis method can also be used to produce the prior distribution. The following equation represents the probability density function of normal distribution.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \text{ for } -\infty < x < \infty. \tag{13}$$

The first step in this technique is to obtain the maximum likelihood equation applied to normal distribution equation. The following statistical quantities were obtained for average and variance, respectively.

$$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}; \tag{14}$$

$$\bar{\sigma}^2 \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{\mu}\right)^2; \tag{15a}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{\mu}\right)^2. \tag{15b}$$

*2.4.2.2. Subjective probability*

The SPAF method was established in response to cognitive biases. Each action taken by the engineer demonstrates the effective application of prior information and the reduction of unfavorable effects and complications [3]. It is developed using the cognitive process, which is broken down into a number of cognitive models, including the following.

**Specification of assessment objectives**

Evaluation objectives are an essential and important step during information gathering because they can cause cognitive biases and the evaluation objective must be clearly understood. In the start of the SPAF, it helps engineers to know and understand the objective of the assessment clearly by providing many solutions:

a) Record the soil property $C_u$ of interest and establish an overall evaluation target (for example, "probabilistic characterization of the soil attribute $C_u$").

b) Break the overall objective into multiple sub-goals. Each sub-goal corresponds to the statistic $\theta_i,\ i=\left[1,2,\ldots,n_m\right]$ of $C_u$. The relevant statistics $\theta_i=\left[\theta_1,\theta_2,\ldots,\theta_{n_m}\right]$, which depends on the theory of probability, used to explain the underlying variability of $C_u$ within a Bayesian framework. The important statistics are the model parameters of the random field, namely $\mu$, $\sigma$, and $\lambda$ of $C_u$. Thus, $\theta$ is composed of three random variables.

c) Identify unfamiliar probability terminology (including $C_u$ statistics) for engineers.

**Collecting relevant data and making a preliminary estimate**

The second stage is to compile the essential information on evaluation objectives from the prior knowledge (i.e. the acquired existing data and the engineers' ability). Using known correlations (e.g., real regressions or theoretical correlations) or intuitive reasoning, a key part of knowledge may resulted in several questionable estimates of the soil property $C_u$ and/or its statistics in the past. It then provides several examples of assessment goals. Two sorts of evidence exist: disconfirming evidence and supporting evidence [8]. These evidences provide a set of information that is consistent with prior information and engineering experience, which includes soil property $C_u$ or its statistics. The serious attempts to uncover information and evidence that are related to prior information, which was obtained from a number of projects of geotechnical exploration implemented in the province. The cohesion value were obtained for Al-Shatrah, Nasiriyah, Suq Al-Shuyukh and Al-Rifai, depending on the data collected from many projects for each region with the depth as shown in Fig. 5.

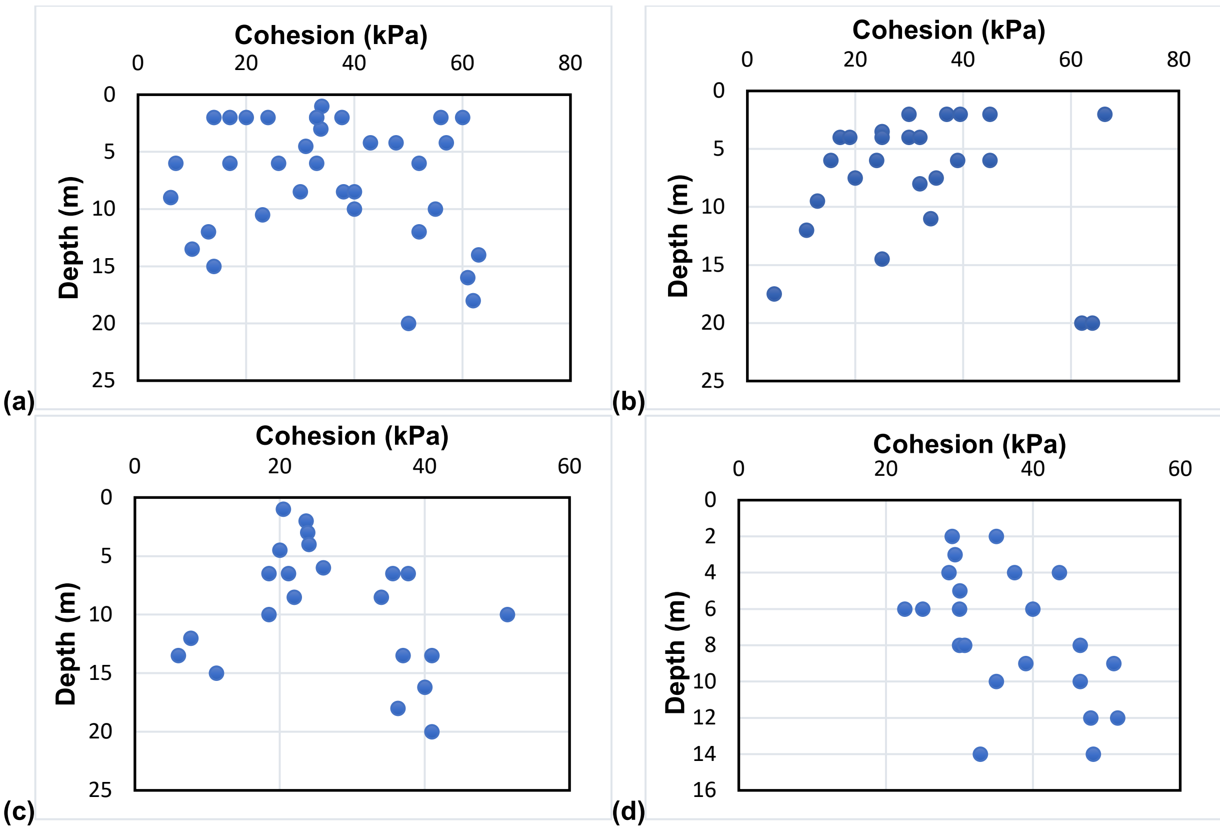



**Figure 5. Mean value of cohesion with depth for (a) AL-Shatrah; (b) AL-Nasiriyah; (c) Suq AL-Shuyukh, Al-Chibayish; (d) AL-Rifai.**

### Synthesis of the evidence

Engineers make use of the collected data to develop internal engineering judgments, for example property of soil $C_u$ and its statistics. The evidence shows two crucial cognitive attributes: weight and strength [8]. Evidence synthesis is a research process that helps researchers to gather all relevant information about the research subject. This occasionally leads to engineers being overconfident in powerful but untrustworthy evidence while underemphasizing (or ignoring) fairly weak proof with relatively high weight (e.g., huge quantity and good quality), which leads to overconfidence bias, representativeness bias, and inadequate correction. There is a requirement to correctly balance the impacts of evidence strength and weight, as well as further synthesis the evidence for subjective probability evaluation. Table 3 shows the strength and weight of the cohesive soil property $C_u$ obtained from in-situ SPTs N and the calculation of the correlation function, which were evaluated according to synthesis of the evidence. Strength in group (I) is represented as weak, (II) – moderate, (III) – weak, while in regions (IV) it is characterized as strong, (V) – moderate, (VI) – weak, (VII) – weak, (VIII) – moderate, and moderate weight.

***Table 3. Summary of the evidence's strength and weight.***

| No. of evidence | Source of the information | Type of correlation | Strength | Weight |
|---|---|---|---|---|
| 1 | A formal report on the location of the clay | Empirical correlation | Strong | Strong |
| 2 | A formal report on the location of the clay | Intuitive inference | Weak | Intermediate |
| 3 | A formal report on the location of the clay | Intuitive inference | Weak | Intermediate |
| 4 | A formal report on the location of the clay | Empirical correlation | Weak | Intermediate |
| 5 | A formal report on the location of the clay | Empirical correlation and logical inference | Intermediate | Intermediate |
| 6 | A formal report on the location of the clay | Empirical correlation | Strong | Intermediate |
| 7 | A formal report on the location of the clay | Empirical correlation and logical inference | Intermediate | Intermediate |
| 8 | A formal report on the location of the clay | Empirical correlation | Strong | Intermediate |
| 9 | A formal report on the location of the clay | Intuitive inference | Weak | Intermediate |
| 10 | A formal report on the location of the clay | Empirical correlation | Strong | Intermediate |
| 11 | A formal report on the location of the clay | Intuitive inference | Weak | Intermediate |
| 12 | A formal report on the location of the clay | Intuitive inference | Weak | Intermediate |
| 13 | A formal report on the location of the clay | Empirical correlation | Strong | Intermediate |
| 14 | A formal report on the location of the clay | Intuitive inference | Weak | Intermediate |
| 15 | A formal report on the location of the clay | Intuitive inference | weak | Intermediate |
| 16 | A formal report on the location of the clay | Intuitive inference | Weak | Intermediate |
| 17 | A formal report on the location of the clay | Intuitive inference | weak | Intermediate |

### Numerical assignment

The numerical assignment is a method for selecting simple and compound random probabilities that consider how prior distributions of non-random parameters are determined and derived. It involves applying many probabilities with repeated significance to randomness, and Bayesian analysis avoids it. The median value of $\theta_i$ for a given range from 0.5 to 1, 2, …, $n_m$ is 0.5. The similar "equivalent" lottery technique requires a range of $\theta_i$, namely $\theta_{i,\min}$, $\theta_{i,\max}$ to estimate the average value of the statistics. Using various ranges in a similar lottery approach results in different median values.

**Final confirmation**

The SPAF methods mentioned above are repeated to produce the relevant PDFs of the model parameters $\theta_i$, $i = 1, 2, ..., n_m$, based on prior information. The prior information about $\theta_i$ is represented in the probability distributions (e.g., CDF) and PDF, which provides means of assessing the impact of uncertainty on characterizing conditions, in which uncertainty is high or low. This theory requires a large amount of data, and the latter requires a sufficient time and effort, which provides means of assessing the impact of uncertainty on characterizing conditions, in which uncertainty is high or low.

# *3. Results and Discussion*

## *3.1. Uniform distribution*

This method relied on the results shown in Tables 1, 2 for the characteristic of undrained cohesion and based on (5). The details of uniform distribution of the $C_u$ soil parameters for the three statistical quantities are as follows:

$$P(\mu) = \begin{cases} \dfrac{1}{62-12} = \dfrac{1}{50} & \text{for } \mu \in [12,\ 62] \\ 0 & \text{otherwise} \end{cases};$$

$$P(\sigma) = \begin{cases} \dfrac{1}{27.6-0.5} = \dfrac{10}{271} & \text{for } \sigma \in [0.5,\ 27.6] \\ 0 & \text{otherwise} \end{cases};$$

$$P(\lambda) = \begin{cases} \dfrac{1}{8-6} = \dfrac{1}{2} & \text{for } \lambda \in [6,\ 8] \\ 0 & \text{otherwise} \end{cases};$$

$$P(\theta_p) = \begin{cases} \dfrac{1}{50} \times \dfrac{10}{271} \times \dfrac{1}{2} = \dfrac{1}{2710} & \text{for } \mu \in [12,\ 62],\ \sigma \in [0.5,\ 27.6], \text{ and } \lambda \in [6,\ 8] \\ 0 & \text{others} \end{cases}.$$

It implies that all combinations of $\mu$, $\sigma$, $\lambda$, and within their respective potential ranges have the same probability 1/2710. Only the ranges of soil parameters are given, they can be used in the Bayesian framework to describe $C_u$ at the clay location according to the probability distribution. Fig. 6 shows the uniform probability distributions for $C_u$ based on the data set depended on in the current study.

Fig. 6a shows the CDF of $\mu$ of $C_u$ as a solid line. It was calculated using the simpler procedure. The relation between CDF and the mean value of $C_u$ shows a linear relationship. The CDF increases linearly from 0.01 to 0.99 at mean values of 12 to 62. The PDF can be presented using a histogram with a single bin (i.e. a uniform distribution with a range of 12–62), and the PDF value of $\mu$ is around 0.019. The uniform PDF of $\mu$ (Fig. 6b) may serve as the prior distribution of $\mu$ in the Bayesian framework.

Fig. 6c displays the CDF of $\sigma$ as a solid line calculated using the simpler procedure. The CDF increases linearly from 0.01 to 0.99 as $\sigma$ rises from 0.5 to 27.6. The PDF of $\sigma$ determined using the simpler approach was represented a histogram with a single bin (i.e. a uniform distribution with a range of 0.5–27.6), and the PDF value of $\sigma$ is around 0.037 (Fig. 6d). The uniform PDF of $\sigma$ (Fig. 6d) may serve as the prior distribution of $\sigma$ in the Bayesian framework.

The solid line shown in Fig. 6e represents the CDF of $\lambda$. The value of $\lambda$ increases from 6 to 8; the CDF increases linearly from 0.01 to 0.99. Fig. 6e shows the PDF of $\lambda$ using a histogram with a single bin (i.e. a uniform distribution with a range of 6–8), and the PDF value of $\lambda$ is around 0.5. The uniform PDF of $\lambda$ (Fig. 6f) may serve as the prior distribution of $\lambda$ in the Bayesian framework.

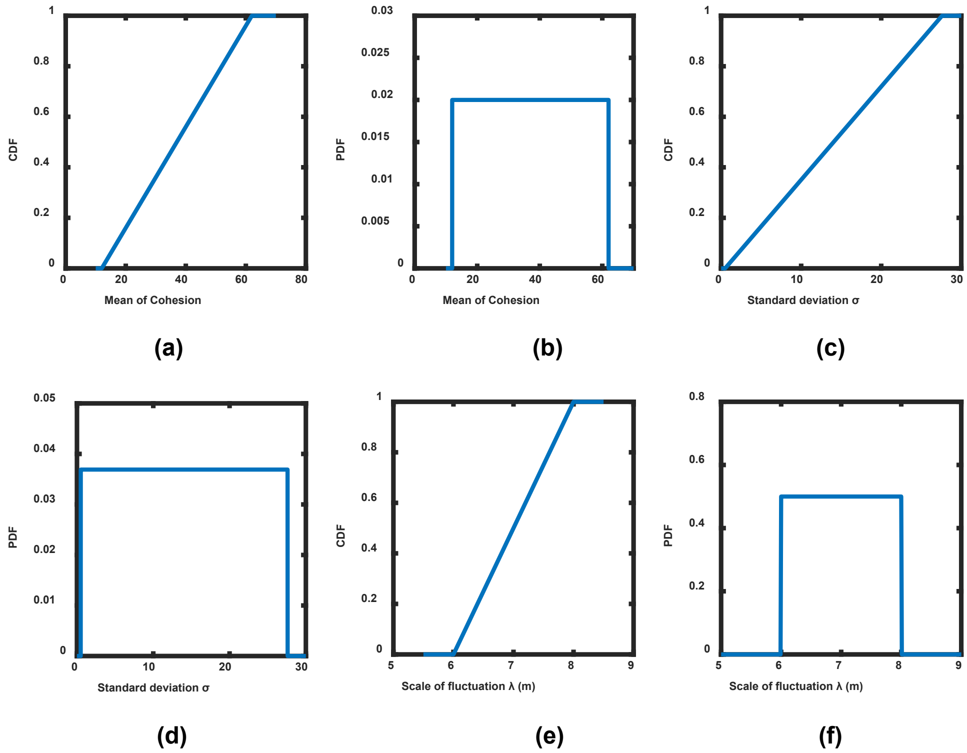



**Figure 6. Non-informative prior distribution of μ, σ, λ.**

### 3.2. *Jeffreys method*

Jeffreys prior has a limited effect on the posterior distribution. Application of this method shows that the mean was 28.7 kPa, with a standard deviation of 1.19 kPa, based on the data included in the MATLAB code developed in this paper. Fig. 7 shows the probability and cumulative distributions. It is noteworthy that the maximum probability distribution reaches 0.33.

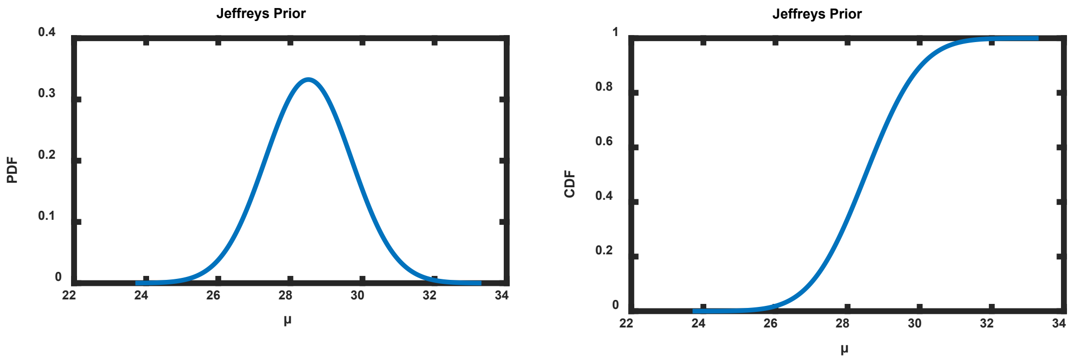



**Figure 7. Non-informative prior distribution (Jeffreys prior) of $C_u$.**

### 3.3. *Regression analysis*

In this method, the probability distribution of prior knowledge can be found based on the data of undrained cohesion. Producing the probability distribution implies computing the average and standard deviation of the data and the maximum likelihood function. The data can be fitted with a suitable probability distribution using regression analysis. In this research, two types of probability distribution were introduced for the data: normal and lognormal distribution. The disadvantage of this method was that the data collected from different sources had similar weight and no strength or weakness for the evidence was taken into account. Fig. 8 shows the normal and lognormal probability distribution fitted to the cohesion data. The results of this statistical analysis do not reflect the reality of data because some data may be very significant and some of them are not significant based on the source of data that has not been taken into account in the analysis. Engineering judgment in this analysis will be difficult to make since the data is not recognized according to its quality.

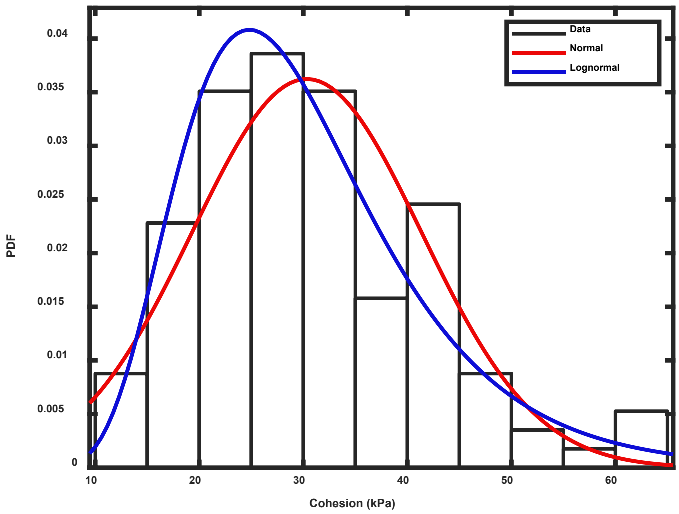
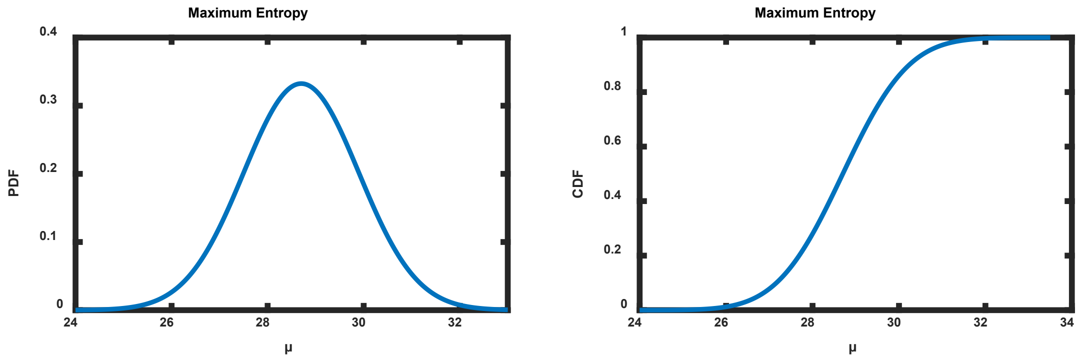



**Figure 8. Informative prior distribution.**

### *3.4. Maximum entropy*

The prior probability distribution was obtained based on the maximum entropy method. With this method, the uncertainty of the prior probability distribution can be measured using information entropy $H_I$. The prior distribution can then be found by maximizing the information entropy subject to constraints or prior knowledge. This method can be used when the data is small but can offer a probability constraints.

The maximum entropy approach was also applied on the mean value of $C_u$. Fig. 9 shows the PDF and CDF distribution of the maximum entropy. In addition to the maximum entropy approach and regression analysis, the prior probability distribution was determined using the mean and standard deviation of the undrained cohesion data, respectively 28.7 and 1.2 kPa, as shown in Fig. 9.




**(a)** **(b)**

**Figure 9. Informative prior distribution.**

### *3.5. Subjective probability*

Table 4 summarizes the evidence related to the parameters $C_u$, which include mean value $(\mu)$, standard deviation $(\sigma)$, and correlation length $(\lambda)$. The evidences were obtained from eight sets of items, each of which was categorized by strength and weight.

For $C_u$, the evidence is based on three item groups: (I), (II), and (III). Item Group (I) includes four guides with strong strength and moderate weight (evidence 1, 6, 8, 13). Item Group (II) includes a single guide with moderate strength and weight (evidence 7). Item Group (III) includes evidence 2, 9, and 14, which are characterized by weak strength and moderate weight. The equations (16) and (17) are used.

$$X = \frac{X_{\max} + X_{\min}}{2}; \quad (16)$$

$$w_x = \frac{X_{\max} - X_{\min}}{6}. \tag{17}$$

$X$ is the mean of $x$; $w_x$ is the standard deviation of $x$; $x$ is the cohesion.

The $C_u$ rates were calculated for Item Group (I), resulting in the following ranges:

$\mu$ = 32, $\sigma$ = 4, range: 20–44;

$\mu$ = 37, $\sigma$ = 1.6, range: 32.4–42.2;

$\mu$ = 21.3, $\sigma$ = 0.25, range: 20.5–22;

$\mu$ = 30, $\sigma$ = 1.6, range: 25.5–35.

Similar calculations were performed for the rest of the Table contents.

***Table 4. Summary of the evidence undrained cohesion for soil in Nasiriyah.***

| Variable | No. of evidence | Item | Previous uncertain evaluation | Strength | Weight |
|---|---|---|---|---|---|
| Cohesion (kPa) | (1) | (I) | 20.5–44 | Strong | Intermediate |
| | (6) | | 32.36–42.16 | Strong | Intermediate |
| | (8) | | 20.5–22 | Strong | Intermediate |
| | (13) | | 25.5–35 | Strong | Intermediate |
| | (7) | (II) | 14.55–62 | Intermediate | Intermediate |
| | (2) | (III) | 13.75–28.15 | Weak | Intermediate |
| | (9) | | 19–24 | Weak | Intermediate |
| | (14) | | 12–62 | Weak | Intermediate |
| μ | (10) | (IV) | 31.55 | Strong | Intermediate |
| | (5) | (V) | 49 | Intermediate | Intermediate |
| | (11) | (VI) | 22.05 | Weak | Intermediate |
| | (12) | | 20 | Weak | Intermediate |
| | (15) | | 37.2 | Weak | Intermediate |
| | (16) | | 26.67 | Weak | Intermediate |
| | (17) | | 18.29 | Weak | Intermediate |
| σ | (19) | (VII) | 5.02–20.3 | Weak | Intermediate |
| λ | (18) | (VIII) | 6–8 | Intermediate | Intermediate |

Depending on the self-assessment solution stages and the data shown in Table 4, $\mu$ and $\sigma$ were obtained as shown in Table 5.

The average cohesion and standard deviation were 37 and 8.8 kPa, respectively (Table 5). Figure 10a shows the probability distribution based on the SPAF. The value of the probability distribution reaches 0.043. It follows from the obtained results that there is a significant difference in the mean and standard deviation value in this method compared with other methods. Other probability distribution of $C_u$ is presented in Fig. 10: a, b for $\mu$; c, d for $\sigma$; e, f for $\lambda$.

***Table 5. The prior percentiles of the mean,*** $\mu$, $\sigma$, $\lambda$**.**

| Cumulative of probability | 0.01 | 0.25 | 0.5 | 0.75 | 0.99 |
|---|---|---|---|---|---|
| Mean μ | 12 | 24.5 | 37 | 49.5 | 62 |
| Standard deviation $\sigma$ | 0.25 | 5.3 | 10.3 | NA | 20.3 |
| Length of the correlation λ | 6 | NA | NA | NA | 8 |

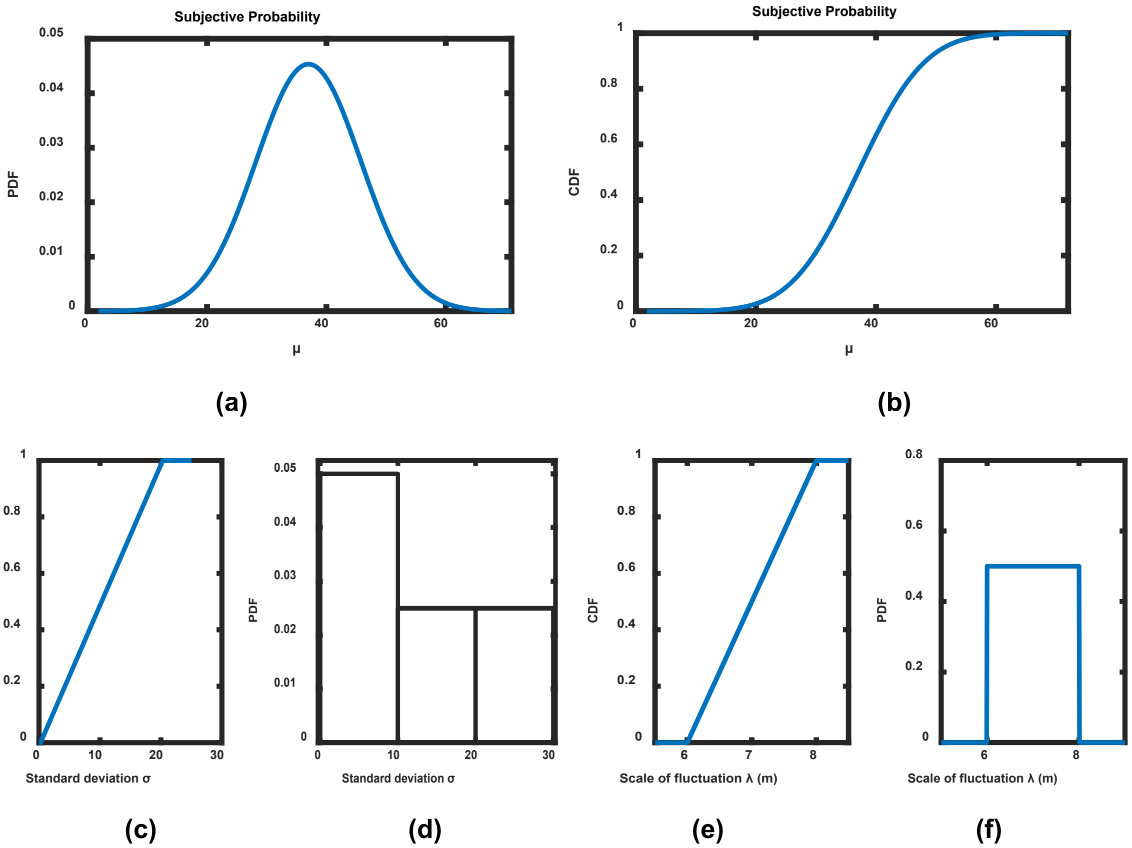



**Figure 10. Informative prior distribution of μ, σ, λ.**

# *4. Conclusion*

This research examines the indicating of the prior probability distribution of undrained cohesion parameter based on a collected database of prior knowledge about this parameter in Nasiriyah, southern Iraq. About 152 $C_u$ values from different boreholes presented in various geotechnical investigation reports were used in this research. A wide range of conditions for selecting the suitable probability distribution was considered. The following points can be outlined based on the results of this study.

1. The range of cohesion values based on database of 152 value was between 28 kPa and 46 kPa. The uniform distribution is simple and easy to use when there is little prior knowledge or when all outcomes are considered equally likely. It can be used as a non-informative prior in Bayesian analysis when no other information is available. Since it implies all cohesion values within the given range are equally likely, it may not be the case in reality.

2. The Jeffreys prior provided a prior probability distribution based on the given mean and standard deviation. The effectiveness of Jeffreys prior might be limited to 152 data points, and the prior could dominate the posterior distribution, leading to biased results. The Jeffreys prior is particularly useful when dealing with transformation-based parameters.

3. The maximum entropy can be used for the least informative while satisfying given constraints. In this case, the mean of cohesion was given as 28.7 kPa, and the standard deviation was 1.2 kPa. By applying the principle of maximum entropy, we can find a probability distribution that is consistent with these constraints. The maximum entropy distribution fits the available information (mean and standard deviation) while being as unbiased as possible. The resulting distribution is expected to be relatively smooth and not make strong assumptions about the underlying data.

4. Regression analysis can be effective when dealing with larger data sets because it allows relationships between variables to be modeled and predictions to be made based on those relationships. With 152 data points, it may be necessary to increase the data, especially when there are significant analysis challenges, to provide useful and reliable results.

5. When working with a limited number of data points, the subjective probability assessment approach, which considers expert knowledge and judgment, might be the most suitable method. However, it is essential to be transparent about the assumptions and uncertainties introduced by the subjective assessment and to carefully interpret and validate the results against any available data.

Additionally, if further data becomes available, it could be useful to reassess the analysis using methods that are better suited for larger datasets, such as regression analysis. The subjective probability assessment heavily relies on the knowledge and judgment of the expert providing the assessment. It can be subjective and may lead to varying results depending on different expert opinions.

## References


1. Gelman, A. Prior distribution. Encyclopedia of Environmetrics. 2002. 3. Pp. 1634–1637.
2. Contreras, L.F., Serati, M., Williams, D.J. Bayesian approach for the assessment of sufficiency of geotechnical data. Slope Stability 2020: Proceedings of the 2020 International Symposium on Slope Stability in Open Pit Mining and Civil Engineering. 2020. Pp. 609–624. DOI: 10.36487/ACG_repo/2025_37
3. Cao, Z.-J., Wang, Y., Li, D.-Q. Probabilistic approaches for geotechnical site characterization and slope stability analysis. Heidelberg: Springer Berlin, 2017. 190 p. DOI: 10.1007/978-3-662-52914-0
4. Phoon, K.-K., Kulhawy, F.-H. Evaluation of geotechnical property variability. Canadian Geotechnical Journal. 1999. 36 (4). Pp. 625–639. DOI: 10.1139/t99-039
5. Cao, Z.-J., Wang, Y., Li, D.-Q. Site-specific characterization of soil properties using multiple measurements from different test procedures at different locations – A Bayesian sequential updating approach. Engineering Geology. 2016. 211. Pp. 150–161. DOI: 10.1016/j.enggeo.2016.06.021
6. Jeffreys, H. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society A. Mathematical, Physical and Engineering Sciences. 1946. 186 (1007). Pp. 453–461. DOI: 10.1098/rspa.1946.0056
7. Baecher, G.B., Christian, J.T. Reliability and statistics in geotechnical engineering, 1st ed. NJ: John Wiley & Sons, 2003. 616 p.
8. Vick, S.G. Degrees of belief: Subjective probability and engineering judgment. USA: Amer Society of Civil Engineers (ASCE), 2002. 472 p.
9. Cao, Z.-J., Wang, Y., Li, D.-Q. Quantification of prior knowledge in geotechnical site characterization. Engineering Geology. 2016. 203. Pp. 107–116. DOI: 10.1016/j.enggeo.2015.08.018
10. Sivia, D., Skilling, J. Data analysis: a Bayesian tutorial, 2nd ed. UK: Oxford University Press, 2006. 246 p.
11. Ching, J., Phoon. K.-K., Li, K.-H., Weng, M.-C. Multivariate probability distribution for some intact rock properties. Canadian Geotechnical Journal. 2019. 56 (8). Pp. 1080–1097. DOI: 10.1139/cgj-2018-0175
12. Zhao, T., Wang, Y., Xu, L. Efficient CPT locations for characterizing spatial variability of soil properties within a multilayer vertical cross-section using information entropy and Bayesian compressive sensing. Computers and Geotechnics. 2021. 137. Article no. 104260. DOI: 10.1016/j.compgeo.2021.104260
13. Li, C., Wang, W., Wang, S. Maximum-entropy method for evaluating the slope stability of earth dams. Entropy. 2012. 14 (10). Pp. 1864–1876. DOI: 10.3390/e14101864
14. Siu, N.O., Kelly, D.L. Bayesian parameter estimation in probabilistic risk assessment. Reliability Engineering & System Safety. 1998. 62 (1–2). Pp. 89–116. DOI: 10.1016/S0951-8320(97)00159-2
15. Carlin, B.P., Louis, T.A. Bayes and empirical Bayes methods for data analysis, 2nd ed. UK: Chapman and Hall/CRC, 2000. 440 p.
16. Feng, Y. Development of Bayesian approaches for uncertainty quantification in in situ stress estimation. Ph.D. thesis. University of Toronto, Canada, 2021.
17. Cao, Z.-J., Wang, Y., Li, D.-Q. Bayesian perspective on geotechnical variability and site characterization. Engineering Geology. 2016. 203. Pp. 117–125. DOI: 10.1016/j.enggeo.2015.08.017
18. Xu, J., Zhang, L., Wang, Y., Wang, C., Zheng, J., Yu, Y. Probabilistic estimation of cross-variogram based on Bayesian inference. Engineering Geology. 2020. 277. Article no. 105813. DOI: 10.1016/j.enggeo.2020.105813
19. Wang, Y., Zhao, T., Cao, Z. Bayesian perspective on ground property variability for geotechnical practice. Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR). 2019. Pp. 63–74. DOI: 10.3850/978-981-11-2725-0-key5-cd
20. Otake, Y., Honjo, Y. Challenges in geotechnical design revealed by reliability assessment: Review and future perspectives. Soils and Foundations. 2022. 62 (3). Article no. 101129. DOI: 10.1016/j.sandf.2022.101129
21. Phoon, K.-K., Cao, Z.-J., Ji, J., Leung, Y.F., Najjar, S., et al. Geotechnical uncertainty, modeling, and decision making. Soils and Foundations. 2022. 62 (5). Article no. 101189. DOI: 10.1016/j.sandf.2022.101189
22. Zhang, P., Yin, Z.-Y., Jin, Y.-F. Bayesian neural network-based uncertainty modelling: application to soil compressibility and undrained shear strength prediction. Canadian Geotechnical Journal. 2022. 59 (4). Pp. 546–557. DOI: 10.1139/cgj-2020-0751
23. Zhao, Z., Congress, S.S.C., Cai, G., Duan, W. Bayesian probabilistic characterization of consolidation behavior of clays using CPTU data. Acta Geotechnica. 2022. 17 (3). Pp. 931–948. DOI: 10.1007/s11440-021-01277-8
24. Gong, W., Zhao, C., Juang, C.H., Zhang, Y., Tang, H., Lu, Y. Coupled characterization of stratigraphic and geo-properties uncertainties – A conditional random field approach. Engineering Geology. 2021. 294. Article no. 106348. DOI: 10.1016/j.enggeo.2021.106348
25. Muhsin, D., Shakir, R.R. Evaluation of soil classification methods based on CPT data for soil in Nasiriyah, Iraq. AIP Conference Proceedings. 2023. 2830. Article no. 030018. DOI: 10.1063/5.0157813
26. Adel, R., Shakir, R.R. Evaluation of static pile load test results of ultimate bearing capacity by interpreting methods. IOP Conference Series: Earth and Environmental Science. 2022. 961. Article no. 012013. DOI: 10.1088/1755-1315/961/1/012013
27. Ewaid, S.H., Abed, S.A., Al-Ansari, N. Crop water requirements and irrigation schedules for some major crops in Southern Iraq. Water. 2019. 11 (4). Article no. 756. DOI: 10.3390/w11040756
28. Ali, H.M. Characteristic Values Of Geotechnical Data And Bearing Capacity Mapping For Soils In Nasiriyah. [Online]. System requirements: AdobeAcrobatReader. URL: https://dspace.utq.edu.iq/items/cdac6587-e97c-49ea-9500-c78420e02037 (Accessed: 02.12.2024).
29. Shakir, R.R. Probabilistic-based analysis of a shallow square footing using Monte Carlo simulation. Engineering Science and Technology, an International Journal. 2019. 22 (1). Pp. 313–333. DOI: 10.1016/j.jestch.2018.08.011

30. Ali, H.M., Shakir, R.R. Geotechnical map of Thi Qar governorate using geographical information systems (GIS). Materials Today: Proceedings. 2022. 60 (3). Pp. 1286–1296. DOI: 10.1016/j.matpr.2021.09.138
31. Thajeel, J.K., Adel, R., Ali, H.M., Shakir, R.R. Effect of the covariance matrix type on the CPT based soil stratification utilizing the Gaussian mixture model. Journal of the Mechanical Behavior of Materials. 2022. 31 (1). Pp. 492–500. DOI: 10.1515/jmbm-2022-0049
32. Shakir, R.R. Selecting the probability distribution of cone tip resistance using moment ratio diagram for soil in Nasiriyah. Geotechnical and Geological Engineering. 2019. 37 (3). Pp. 1703–1728. DOI: 10.1007/s10706-018-0716-3
33. Fienberg, S.E. Does it make sense to be an “objective Bayesian”? (Comment on articles by Berger and by Goldstein). Bayesian Analysis. 2006. 1 (3). Pp. 429–432. DOI: 10.1214/06-BA116C
34. Fisher, R.A. Theory of statistical estimation. Mathematical Proceedings of the Cambridge Philosophical Society. 1925. 22 (5). Pp. 700–725. DOI: 10.1017/S0305004100009580
35. Shannon, C.E., Weaver, W. The mathematical theory of communication. USA: The University of Illinois Press, 1964. 131 p.
36. Niven, R.K., Andresen, B. Jaynes’ maximum entropy principle, Riemannian metrics and generalised least action bound. Complex Physical, Biophysical and Econophysical Systems. 2010. Pp. 283–317. DOI: 10.48550/arXiv.0907.2732
37. Kanwar, N.S., Deng, J. A comparison of probabilistic distributions of undrained shear strength of soils in Nipigon River, Canada. IOP Conference Series: Earth and Environmental Science. 2019. 351. Article no. 012024. DOI: 10.1088/1755-1315/351/1/012024

### *Information about the authors:*

***Ressol R. Shakir***,
*E-mail: rrshakir@utq.edu.iq*

***Zainab Abd AL-Haleem***,
*E-mail: zainababd96@utq.edu.iq*

*Received: 15.02.2023. Approved after reviewing: 19.02.2024. Accepted: 05.08.2024.*