Scientific Review Engineering and Environmental Sciences (2023), 32 (4), 338–356

Sci. Rev. Eng. Env. Sci. (2023), 32 (4) https://srees.sggw.edu.pl

ISSN 1732-9353 (suspended) eISSN 2543-7496 DOI 10.22630/srees.4857

Received: 02.04.2023
Accepted: 16.08.2023

Alaa Nuri MERZA
Aram Mohammed RAHEEM✉ https://orcid.org/0000-0002-6889-3939
Ibrahim Jalal NASER https://orcid.org/0000-0001-8986-9711
Mohammed Omar IBRAHIM
Najat Qader OMAR https://orcid.org/0000-0003-2665-7978
University of Kirkuk, Civil Engineering Department, Kirkuk, Iraq

Implementing GIS and linear regression models to investigate partial building failures

Keywords: GIS, IDW technique, crack identification, linear single, multi-regression models

Introduction

The influence of soil on urban ecology is both direct and indirect (Ouabo, Sangodoyin & Ogundiran, 2020), and creating a digital soil map also has a significant impact on research in terms of time and preventing soil problems, consequently, there is a growing need for information about soil (Zijl, 2019). The data for 56 different soil samples collected from various depths in the city of Kirkuk were combined utilizing the inverse distance weighted (IDW) spatial analysis technique (Raheem & Omar, 2021). Soil behavior and the possible impact of soil issues on structures may be predicted using mathematical models and computer simulations (Addiscott, 1993). Based on prior project observations and data, empirical correlations may be applied to predict soil behavior. These correlations can be used to estimate the soil's

✉aram_raheem@uokirkuk.edu.iq; engaram@yahoo.com

swelling potential and the probable impact on structures (Cantillo, Market & Bird, 2017). Creating models and simulations of how soil would counter or predict some soil characteristics in a particular location requires the use of a geographic information system (GIS) by storing information and data, such as Atterberg limits (liquid limit – *LL*, plasticity index – *PI*), and various soil characteristics denoted by gravel content, sand content, silt content, clay content, gypsum content (*GYP*), total suspended solids (*TSS*), potential of hydrogen (pH), and organic content (*ORG*) components (Raheem, Omar, Naser & Ibrahim, 2022; Raheem, Naser, Ibrahim & Omar, 2023; Salahalden, Shareef & Al Nuaimy, 2023). This information may be utilized to develop predictions regarding soil stability, soil volume change, and other soil qualities. The approach typically involves data collection and mapping, the development of a mathematical model to define soil behavior, and simulation testing to validate the models and provide predictions concerning soil behavior in specific regions (Omer et al., 2018). The accuracy of the findings is determined by the dependability of the models used and the quality of the data used. Machine learning was used to predict continuous dependent soil characteristics from an assortment of independent factors using both single and multi-linear regression models (Matarira, Mutanga & Dube, 2021; Mohsin & Lone, 2021; Pentos, Mbah, Pieczarka, Niedbała & Wojciechowski, 2022).

This study stands out because it employs the IDW method to produce digital maps for several significant soil attributes in a Kirkuk City residential complex. More importantly, no digital maps for apartment buildings in the city of Kirkuk have ever been examined. Planners in Kirkuk, Iraq, will be able to more precisely classify the area's diverse soil sections with the use of these computerized topsoil maps. This will make it possible to put into practice various risk-reduction strategies for structural engineering and make better decisions regarding the potential cracking areas.

The main goal of this study is to utilize a spatial statistical approach determined by the inverse distance weighted (*IDW*) technique to incorporate all available information on the soils of the investigated residential complex in the city of Kirkuk, Iraq. The particular objectives can be represented as geographical data combining the proportions of *LL*, gravel, sand, silt, *GYP*, *TSS*, pH, and *ORG* for the current investigation residential complex in Kirkuk, Iraq. The aim was to investigate the relationship between *PI* and physical soil parameters using a linear single regression model with discrete correlations. The specific correlations between *PI* and the physical, chemical, and compounded physical and chemical soil properties were further examined using linear multi-regression models. The digital maps created using prediction models may be used to visualize the possibility of soil swelling and decrease the anticipated risks for future civil engineering projects.

Merza, A. N., Raheem, A. M., Naser, I. J., Ibrahim, M. O., Omar, N. Q. (2023). Implementing GIS and linear regression models to investigate partial building failures. *Sci. Rev. Eng. Env. Sci.*, *32* (4), 338–356. DOI 10.22630/srees.4857

The study area and problem identification

The site is in the city of Kirkuk, around 750 m from the Kasa Su River and 600 m from the Kirkuk ring motorway. Several apartment buildings are located near the site. The building is about 30,000 m^2 in size. The project entails the development of one and two-story residential apartments. Figure 1 illustrates the city of Kirkuk, including all governmental buildings, as well as the current investigated study area (Omar & Raheem, 2016). In many buildings in the inspected residential complex, numerous cracks of varied sizes and places have been noticed. Figure 2 illustrates various cracks in size and position for the investigated residential complex.

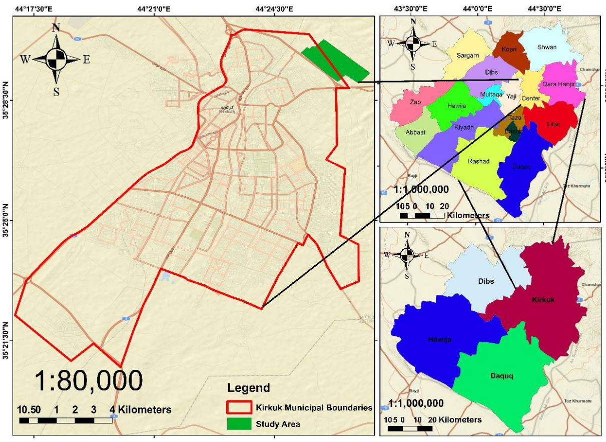

FIGURE 1. Detailed map of Kirkuk City
Source: own work.

FIGURE 2. Various cracks in shape and position in the studied residential complex in Kirkuk City/Iraq
Source: own work.

Methodology

Geographic information system and data collection

Geographic information system applications are used to compile geographic information into databases to save time and money. The data-gathering procedure involves inputting updated data into the GIS application; however, data integration includes adjusting and protecting data integrity. For data acquired arbitrarily, the GIS application enables the evaluation of global auto-correlation (Olmo, Guervos & Rocha, 2019). Seven boreholes in the analyzed area provided a variety of soil data, which is adequate for such an explored area. Physical and chemical soil characteristics such as *PI*, *LL*, gravel, sand, silt, clay, *GYP*, *TSS*, pH, and *ORG* content were obtained. Table 1 provides illustrations of geographical, physical, and chemical soil characteristics for the residential complex in Kirkuk City.

TABLE 1. Geographical and physical soil characteristics of sample examples of the studied residential complex in Kirkuk City/Iraq

FID	BH	North	East	Depth [m]	Gravel content [%]	Sand content [%]	Silt content [%]	Clay content [%]	*LL*	*PI*	*GYP* [%]	*TSS* [%]	pH	*ORG* [%]
0	1	3928584.755	448672.088	1.0	10	26	38	26	52	26	1.02	2.81	7.98	0.31
1	2	3928578.675	448641.835	1.0	6	48	24	22	17	2.67	1.95	2.64	7.94	1.42
2	3	3928638.499	448575.699	1.0	54	24	14	8	35	14	3.23	3.53	7.91	2.1
3	4	3928545.745	448585.963	1.0	4	8	49	39	48	26	2.3	2.12	7.92	1.77
4	5	3928580.663	448548.569	1.0	19	25	36	20	41	18	3.02	3.57	7.88	1.05
5	6	3928728.284	448476.17	1.0	40	15	27	18	38	14	2.89	3.9	7.95	1.76
6	7	3928742.614	448462.409	1.0	38	26	17	19	39	15	2.34	4.02	7.88	1.67

BH – borehole, *LL* – liquid limit, *PI* – plasticity index, *GYP* – gypsum content, *TSS* – total suspended solids, pH – potential of hydrogen, *ORG* – organic content.

Source: own work.

Individual data acquired for a specific soil is altered from an Excel file with latitude and longitude to a themed map utilizing Arc Map capabilities. Using the specified coordinates as a starting point, resampling maps were produced using the inverse distance weighted (IDW) interpolation technique. On the studied maps, each borehole position is represented by a small blued point.

Inverse distance weighted (IDW) technique

The IDW technique is simple, comprehensive, and one of the most important approaches to interpolation. A linear sequence of surrounding points is used, which is normalized by an inverse connection of distance between the examined and tested locations (Billah, 2018). Eq. (1) demonstrates how to recognize the IDW approach:

$$Y_o = \sum_{i=1}^{N} \frac{y_i d_1^{-n}}{d_1^{-n}}, \tag{1}$$

where Y_o is the required quantity of the z factor at the I point; y_i is the known quantity in the I point, d_1 is the Euclidian range between the desired and provided values; N is the weighting factor based on distance; n is the degree of inverse distance weighting.

The retrieved digital maps were utilized to define a surface that is a function of a specified attribute and neighborhood approach. It should be emphasized that the IDW approach conveniently interprets maximum and minimum field values.

Linear single regression model

A linear single regression model was used to estimate the soil *PI* attribute using field information on the physical and chemical properties of the soil. In Eq. (2), the represented *PI* soil properties in the linear single regression model are as follows:

$$PI\ [\%] = L \cdot \text{physical or chemical soil property}\ [\%] + M, \tag{2}$$

where L and M are model parameters.

Parameters such as *LL*, gravel content, sand content, silt content, clay content, gypsum content, *TSS*, pH, and organic content percentages are physical and chemical model factors that may be directly associated with *PI* content. However, *PI* requires a rigorous laboratory analysis and may be the primary cause of soil swelling. As a result, the suggested linear single regression model can estimate *PI* using physical and chemical soil parameters.

Linear multi-regression model

A linear multi-regression model was used to calculate the *PI* soil attribute using field measurements for physical and chemical soil factors. The linear multi-regression model represents the following soil characteristics: *LL*, gravel, sand, silt, clay, gypsum, *TSS*, pH, and organic percentages. Three distinct linear multi-regression models were employed. Equations (3), (4), and (5) reveal the model forms as follows:

$$PI\,[\%] = A \cdot LL + B \cdot \text{gravel}\,[\%] + C \cdot \text{sand}\,[\%] + D \cdot \text{silt}\,[\%] + E \cdot \text{clay}\,[\%] + F, \quad (3)$$

$$PI\,[\%] = G \cdot GYP\,[\%] + H \cdot TSS\,[\%] + I \cdot \text{pH}\,[\%] + J \cdot ORG\,[\%] + K, \quad (4)$$

$$PI\,[\%] = N \cdot LL + O \cdot \text{gravel}\,[\%] + P \cdot \text{sand}\,[\%] + Q \cdot \text{silt}\,[\%] + \\ + R \cdot \text{clay}\,[\%] + S \times GYP\,[\%] + T \cdot TSS\,[\%] + U \cdot \text{pH}\,[\%] + V \cdot ORG\,[\%] + W, \quad (5)$$

where the model parameters are *A*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, *N*, *O*, *P*, *Q*, *R*, *S*, *T*, *U*, *V*, and *W*.

The *PI* component can be predicted using varying percentages of physical and chemical soil variables based on the specified multi-regression model characteristics.

Results

Physical soil distribution

The IDW method was used to estimate the gravel distribution for the examined residential complex in the city of Kirkuk, as shown in Figure 3a. The percentage of gravel is accurately divided into five regions: very low, low, medium, high, and very high. The maximum and minimum gravel concentrations obtained by the IDW interpolation approach were used to characterize these zones. The ranges were as follows: from 4.016 to 13.997, 13.998 to 23.979, 23.980 to 33.961, 33.962 to 43.943, and 43.944 to 53.925. The first region has a maroon color and appears as a single spot near one of the field boreholes. The second zone is identified by its red color and may be found in the analyzed residential complex's northern and western zones. The third region is designated by the pink color and situated in the center of the investigated residential complex. The fourth region is designated by a light pink color and located in the southeastern zone of the investigated residential complex. The fifth zone is designated by a white color and divided into tiny areas around two of the boreholes.

Figure 3b illustrates the findings of an IDW technique study utilizing a GIS to evaluate the distribution of sand in a residential complex in the city of Kirkuk. The distribution of sand follows the same five zones as the distribution of gravel. The interpolation by IDW method was used to calculate the maximum and minimum sand contents for each zone. The very low, low, medium, high and very high regions started from 12.005 to 17.202, 17.203 to 22.400, 22.401 to 27.598, 27.599 to 32.795, and 32.796 to 37.993, respectively. The first zone is maroon in color and dispersed as a spot near one of the

Merza, A. N., Raheem, A. M., Naser, I. J., Ibrahim, M. O., Omar, N. Q. (2023). Implementing GIS and linear regression models to investigate partial building failures. *Sci. Rev. Eng. Env. Sci.*, *32* (4), 338–356. DOI 10.22630/srees.4857

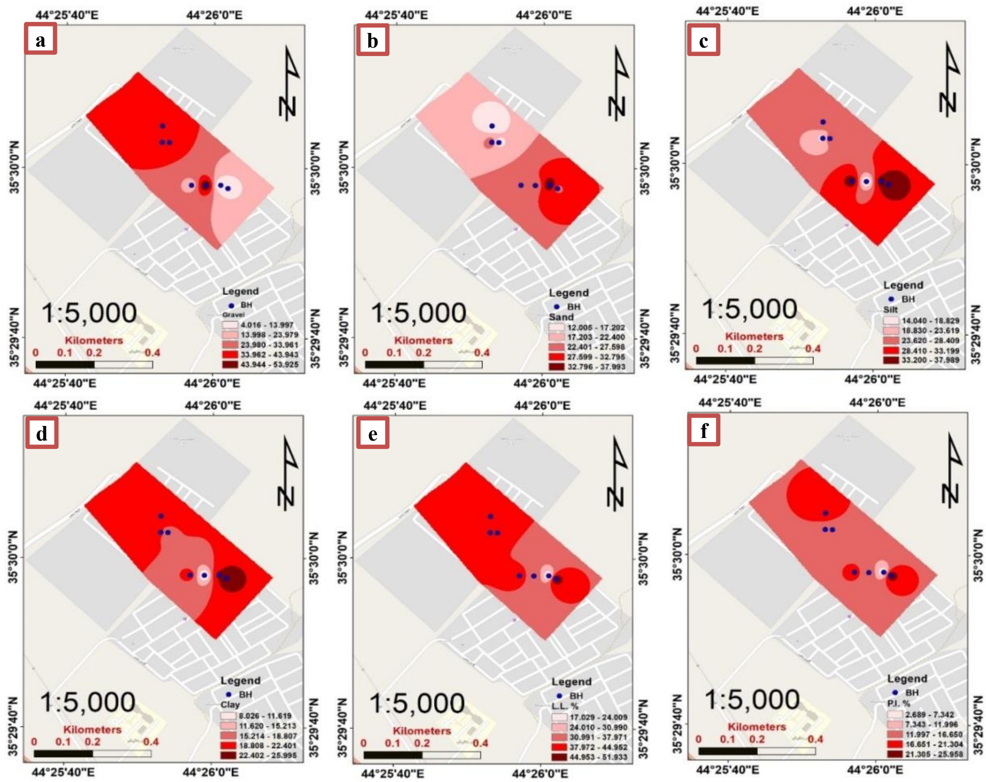

FIGURE 3. Inverse distance weighted technique for physical soil distributions in the studied residential complex in Kirkuk City/Iraq: a – gravel, b – sand, c – silt, d – clay, e – liquid limit (*LL*), f – plasticity index (*PI*)

Source: own work.

boreholes. The second zone is located in the complex's southeast corner. The third region extends from the complex's geographic center to the south. The fourth region is located in the examined residential complex's center to the northern zone. The fifth zone is separated into little portions in the complex's central northern zone.

As demonstrated in Figure 3c, the silt distribution has been consistently evaluated using the IDW approach combined with the GIS technology for the residential complex in the city of Kirkuk. The silt has been divided into five zones, ranging from extremely low to very high. The IDW interpolation method was used to compute the highest and lowest silt contents in each zone. The first zone is maroon in color and scattered throughout the complex's south and east zones. The second zone is positioned in the complex's center, to the south. The third section stretches from the complex's geographic center to the north. The fourth area is represented by locations in the complex's center and north. The fifth zone is divided into discrete locations across the complex.

Using IDW with the GIS technique, we mapped out the residential complex of Kirkuk clay distribution as identified in Figure 3d. The clay has been divided into five zones ranging from extremely low to very high. To calculate the maximum and minimum clay concentrations for each zone, the IDW interpolation method was utilized. Starting from 8.026 to 11.619, 11.620 to 15.213, 15.214 to 18.807, 18.808 to 22.401, and 22.402 to 25.995, consecutively, the extremely low, low, medium, high, and very high regions were constructed. The first zone is maroon in color and appears as a spot in the complex's southern zone. The second zone covers up the majority of the complex's space. The third segment extends westward from the complex's geographic center. A site close to one of the boreholes in the examined complex represents the fourth area. The fifth zone is represented by a little piece in the complex's southern zone.

We mapped out the residential complex of Kirkuk *LL* distribution using IDW and the GIS approach, as shown in Figure 3e. The *LL* is separated into five zones ranging from extremely low to extremely high. The IDW interpolation method was used to compute the maximum and minimum *LL* concentrations for each zone. From 17.029 to 20.009, 24.010 to 30.990, 30.991 to 37.971, 37.972 to 44.952, and 44.953 to 51.933, the extremely low, low, medium, high, and very high zones were built in sequence. The first zone is a maroon-colored region in the complex's southern zone. The second zone covers much of the complex's north space. The third segment encompasses the complex's southern portion. The fourth area is located near one of the boreholes in the investigated complex. A little section in the complex's southern zone represents the fifth zone.

The residential complex of Kirkuk *PI* distribution was mapped out using IDW and the GIS technique, as illustrated in Figure 3f. The *PI* is divided into five zones, which range from extremely low to very high. The highest and lowest *PI* concentrations for each zone were calculated using the IDW interpolation method. Extremely low, low, medium, high, and very high zones were developed in order from 2.689 to 7.342, 7.343 to 11.996, 11.997 to 16.650, 16.651 to 21.304, and 21.305 to 25.958. The first zone is a maroon patch in the complex's southern zone. The second zone is represented by three circular shapes in red that are scattered over the study area. The third portion covers the majority of the site. The fourth region is near one of the studied complex's boreholes. The fifth zone is represented by a small part of the complex's southern zone.

It should be noted that the analyzed maps have shown that the maximum values for both silt and clay concentrations are higher than the corresponding values for both gravel and sand values. Thus, attributed geotechnical complications such as soil swelling is expected that may lead to partial building cracks in the absence of required appropriate engineering protections.

Chemical soil distribution

As shown in Figure 4a, the IDW approach was utilized to estimate the *GYP* distribution for the investigated residential complex in the city of Kirkuk. The proportion of *GYP* is classified into five categories: very low, low, medium, high, and very high. These zones were defined using the maximum and minimum *GYP* concentrations determined using the IDW interpolation method. The ranges were as follows: from 1.023 to 1.464, 1.465 to 1.906, 1.907 to 2.348, 2.349 to 2.790, and 2.791 to 3.232. The first section is maroon in color and appears as a single circular pattern in the complex's southern zone. The second zone is distinguished by its red color and placed in the complex's center. The pink color represents the third region, which is located in the northern and southern parts of the examined residential complex. The fourth region is represented by little spots in the northern and southern zones of the analyzed residential complex and marked by a light pink color. The fifth zone is identifiable as a little spot in the complex's southern zone with a white color.

As shown in Figure 4b, we used IDW and the GIS technique to map out the Kirkuk TSS distribution residential complex. The *TSS* is divided into five zones ranging from extremely low to extraordinarily high. The greatest and lowest *TSS* concentrations for each zone were calculated using the IDW interpolation method. Starting from 2.670 to 2.940, 2.941 to 3.211, 3.212 to 3.482, 3.483 to 3.753, and 3.754 to 4.024, consecutively, the extremely low, low, medium, high, and very high regions were constructed. The first zone, which is maroon in color, may be observed in the complex's central zone. The second zone encompasses the majority of the center zone. The third portion extends from the complex's northern zone to its southern region. The fourth location in the investigated complex is towards the complex's southern corner. The fifth zone is represented by white spots in the complex's northern and southern regions.

We mapped out the Kirkuk pH distribution residential complex using IDW and the GIS approach, as shown in Figure 4c. The pH is composed of five zones, from very low to very high. The highest and lowest pH concentrations for each zone were calculated using the IDW interpolation technique. The extremely low, low, medium, high, and very high zones were developed systematically from 7.850 to 7.875, 7.876 to 7.901, 7.902 to 7.927, 7.928 to 7.953, and 7.954 to 7.979. A spot in the complex's southern zone is designated as the first zone, which is symbolized by the maroon color. The complex's southern zone includes one circular shape that

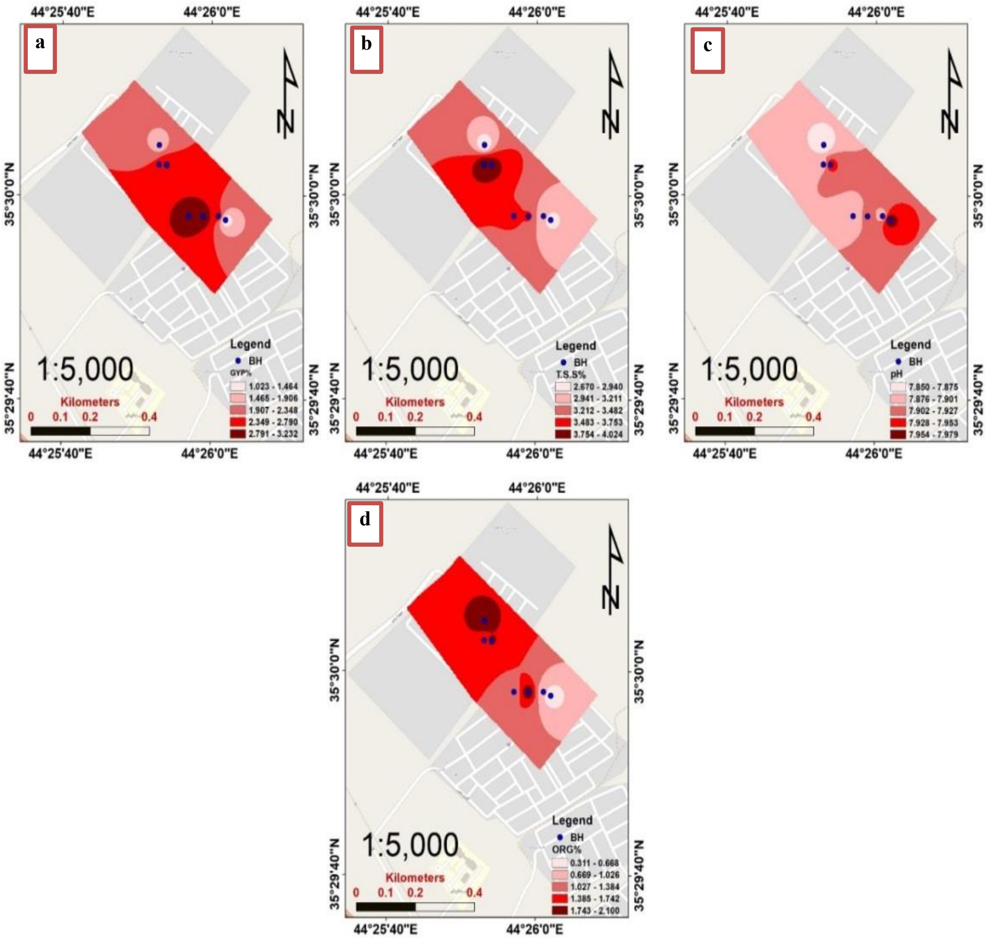

FIGURE 4. IDW technique for chemical soil distributions in the studied residential complex in Kirkuk City/Iraq: a – gypsum content, b – total suspended solids, c – pH, d – organic content

Source: own work.

represents the second zone, which is designated by a red color. The third and fourth zones cover the majority of the southern and northern zones of the study area respectively. The fifth zone is represented by two spots in the complex's northern and southern zones.

As illustrated in Figure 4d, we mapped the Kirkuk *ORG* distribution residential complex using IDW with the GIS method. Five zones, ranging from incredibly low to very high, constitute the *ORG*. Using the IDW interpolation technique, the highest and lowest *ORG* concentrations for each zone were determined. The extremely low, low, medium, high, and very high regions were built sequentially starting from 0.311 to 0.668, 0.669 to 1.026, 1.027 to 1.384, 1.385 to 1.742, and 1.743 to 2.100. The first zone, which is maroon in color, may be observed in the complex's north-central zone. The second zone encompasses the majority of the northern zone. The third portion occupies the southern zone of the complex. The fourth location in the investigated complex is located towards the complex's southeastern corner. The fifth zone is represented by a white spot in the complex's southern region.

It should be observed that the highest values for both gypsum and organic concentrations are greater than the equivalent values for pH and *TSS* values, according to the evaluated maps. Hence, zones with high gypsum and organic contents are more susceptible to geotechnical difficulties such as voids development due to gypsum melting and excessive settlement because of weak soil with high organic content. Consequently, partial building cracks can be caused due to the presence of such problematic soils.

Correlation between plasticity index and soil characteristics

The relationships between *PI* and soil characteristics such as *LL*, gravel, sand, silt, clay, gypsum, *TSS*, pH, and organic components and *PI* are summarized in Table 2.

TABLE 2. Correlations between plasticity index and soil characteristics for studied residential complex in Kirkuk City/Iraq

×	*PI* [%]	*LL* [%]	Gravel [%]	Sand [%]	Silt [%]	Clay [%]	*GYP* [%]	*TSS* [%]	pH	*ORG* [%]
PI [%]	1.000	–	–	–	–	–	–	–	–	–
LL [%]	0.345	1.00	–	–	–	–	–	–	–	–
Gravel [%]	–0.298	0.063	1.00	–	–	–	–	–	–	–
Sand [%]	–0.427	–0.369	0.159	1.00	–	–	–	–	–	–
Silt [%]	0.323	0.316	–0.750	–0.611	1.00	–	–	–	–	–
Clay [%]	0.431	0.116	–0.716	–0.700	0.692	1.00	–	–	–	–
GYP [%]	–0.379	–0.026	0.451	–0.021	–0.180	–0.337	1.00	–	–	–
TSS [%]	–0.234	0.125	0.299	–0.231	0.011	–0.086	0.577	1.00	–	–
pH	0.022	0.084	0.091	0.145	–0.147	–0.137	0.106	0.037	1.00	–
ORG [%]	–0.410	–0.138	0.612	–0.061	–0.415	–0.281	0.412	0.304	0.211	1.00

PI – plasticity index, *LL* – liquid limit, *GYP* – gypsum content, *TSS* – total suspended solids, pH – potential of hydrogen, *ORG* – organic content.

Source: own work.

Positive and negative correlations have been identified between the *PI* characteristic and various soil parameters, with degrees of correlation ranging from –0.427 to 0.431. Furthermore, correlations between the *LL* and gravel, sand, silt, clay, *GYP*, *TSS*, pH, and *ORG* varied from –0.369 to 0.316. Moreover, satisfactory relationships between gravel content and the amounts of sand, silt, clay, *GYP*, *TSS*, pH, and *ORG* substances are observed, with degrees of correlation ranging from –0.750 to 0.612. Positive and negative relationships between sand content and proportions of silt, clay, *GYP*, *TSS*, pH, and *ORG* substances have also been identified, with degrees of correlation ranging from –0.700 to 0.145. Likewise, relationships between silt content and clay, *GYP*, *TSS*, pH, and *ORG* material percentages have been established, with degrees of correlation ranging from –0.415 to 0.692. Clay content has been found to have negative associations with chemical soil percentages, with degrees of association ranging from –0.337 to –0.086. Positive relationships exist between soil chemical concentrations, with degrees of correlation ranging from 0.037 to 0.577. It is essential to establish accurate correlations between *PI* and physical and chemical soil contents that may be utilized to determine the critical swelling feature indirectly with no cost or effort.

Linear single regression model

Table 3 summarizes the properties of the proposed linear single regression model [Eq. (2)]. The least squares approach was used to solve the proposed linear single regression model. Table 3 shows the model parameters (*L* and *M*) as well as R^2 values for the suggested model for all of the analysed situations. The model's coefficients *L* and *M* have comparable ranges from –4.851 to 3.587 and –8.650 to 28.573. Furthermore, the R^2 ranges from 0.001 to 0.186.

The variation of *PI* with the physical soil properties has been illustrated in Figure 5a–e. Different negative and positive associations have been observed between *PI* and physical soil contents. The *PI* has demonstrated positive correlations with *LL*, silt content, and clay content, whereas the *PI* has revealed negative associations with both gravel and sand contents. It is obviously indicated that the *PI* has positive associations with fine particle contents and negative associations with coarse particle substances. A decent relationship between *PI* and *LL* is observed for *LL* values higher than 30, with slight changes in gravel concentration over 10%. In addition, reasonable correlations between *PI* with sand, silt and clay contents have been observed except for sand and silt values in the range of 20–30% and for clay content in the range of 30–40%.

TABLE 3. Linear regression model analysis for swelling soil characteristics of studied residential complex in Kirkuk City/Iraq

Swelling property [%]	Physical and chemical properties [%]	*L*	*M*	Equation	R^2
PI	*LL*	0.277	8.255	$PI = 0.277 \cdot LL + 8.255$	0.120
PI	gravel	–0.136	22.679	$PI = -0.136 \cdot \text{gravel} + 22.679$	0.089
PI	sand	–0.236	25.486	$PI = -0.236 \cdot \text{sand} + 25.486$	0.182
PI	silt	0.197	13.879	$PI = 0.197 \cdot \text{silt} + 13.879$	0.104
PI	clay	0.263	13.550	$PI = 0.263 \cdot \text{clay} + 13.550$	0.186
PI	*GYP*	–4.851	28.573	$PI = -4.851 \cdot GYP + 28.573$	0.144
PI	*TSS*	–2.415	26.635	$PI = -2.415 \cdot TSS + 26.635$	0.055
PI	pH	3.587	–8.650	$PI = 3.587 \cdot \text{pH} - 8.650$	0.001
PI	*ORG*	–4.003	23.888	$PI = -4.003 \cdot ORG + 23.888$	0.169

Source: own work.

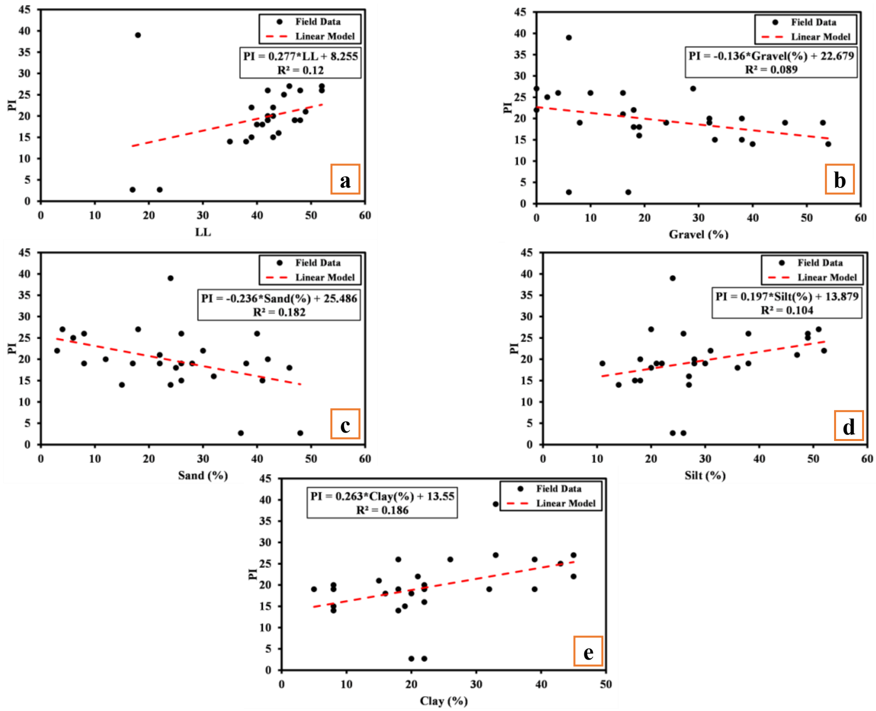

FIGURE 5. The variation of plasticity index content with the physical soil contents of studied residential complex in Kirkuk City using linear single regression model: a – correlation between plasticity index (*PI*) and liquid limit (*LL*), b – correlation between plasticity index (*PI*) and gravel content [%], c – correlation between plasticity index (*PI*) and sand content [%], d – correlation between plasticity index (*PI*) and silt content [%], e – correlation between plasticity index (*PI*) and clay content [%]

Source: own work.

The variation of *PI* with chemical soil properties has been demonstrated in Figure 6a–d. The *PI* has revealed negative correlations with *GYP* [%], *TSS* [%], and *ORG* [%] contents while a positive correlation between *PI* and pH has been noticed. The *PI* values have shown good correlations with both *GYP* [%] and *TSS* [%] for most values greater than 1% and 2% respectively. A reasonable variation between *PI* and pH values has been noticed except for a pH value close to 7.9, whereas *PI* and *ORG* values were correlated rationally for *ORG* values greater than 0.5.

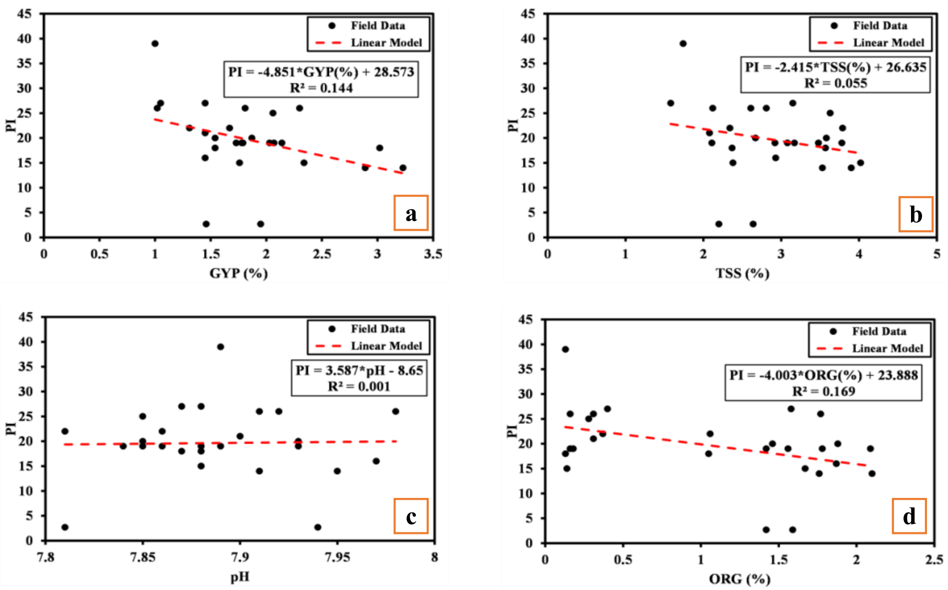

FIGURE 6. The variation of plasticity index content with the chemical soil contents of studied residential complex in Kirkuk City/Iraq using linear single regression model: a – correlation between plasticity index (*PI*) and gypsum content (*GYP*) [%], b – correlation between plasticity index (*PI*) and total suspended solids (*TSS*) [%], c – correlation between plasticity index (*PI*) and pH, d – correlation between plasticity index (*PI*) and organic content (*ORG*) [%]

Source: own work.

Linear multi-regression model

Information about the linear multi-regression framework [Eqs (3), (4) and (5)] is provided in Table 4. The least squares approach was used to solve the linear multi-regression equations. Table 4 shows the proposed model parameters (*A*–*W*) as well as the predicted values of the multiple *R* model. The multiple *R* value ranges from 0.485 to 0.921.

TABLE 4. Linear multi-regression model analysis for swelling soil characteristics of studied residential complex in Kirkuk City/Iraq

Swelling property [%]	Equation number	Linear multi-regression coefficients						Multiple *R*
		A	*B*	*C*	*D*	*E*	*F*	
PI	3	0.637	–2.661	–2.645	–2.625	–2.560	254.478	0.921
		G	H	I	J	K		
PI	4	–3.369	0.141	19.801	–3.232	–127.336		0.485
		N	O	P	Q	R		
PI	5	0.647	–2.688	–2.665	–2.645	–2.590		
		S	T	U	V	W		
		–0.261	0.010	–4.952	0.259	295.696		0.921

Source: own work.

The variations in the expected and actual soil *PI* characteristics using the multi-linear regression model [Eqs (3)–(5)] are displayed in Figure 7a–c. The proposed linear multi-regression analysis of Eqs (3) and (5) estimate *PI* values successfully in Figures 7a and 7c, with multiple *R* values of 0.921.

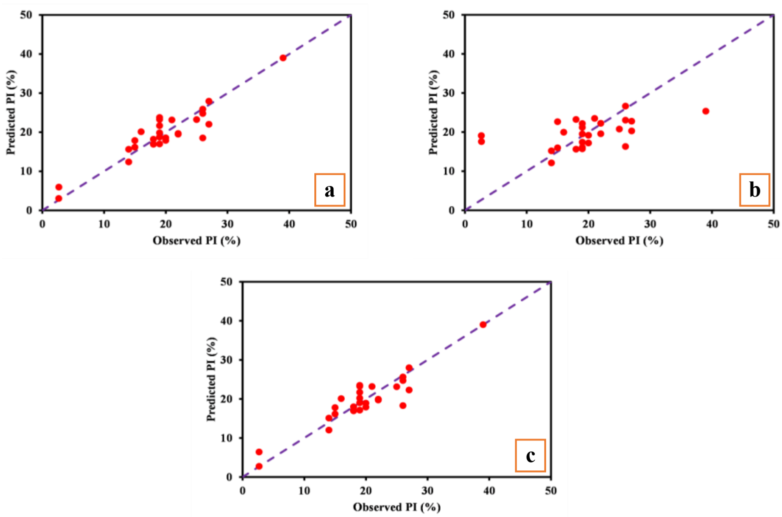

FIGURE 7. The variation of *PI* content with the chemical soil contents of studied residential complex in Kirkuk City/Iraq using linear single regression model: a – correlation between plasticity index (*PI*) and gypsum content (*GYP*) [%], b – correlation between plasticity index (*PI*) and total suspended solids (*TSS*) [%], c – correlation between plasticity index (*PI*) and pH, d – correlation between plasticity index (*PI*) and organic content (*ORG*) [%]

Source: own work.

However, a lower multiple *R* value of 0.485 has been noticed for the predicted *PI* values of Eq. (4) as presented in Figure 7b. It is evident that *PI* values can be approximated more conveniently based on physical or integrated physical and chemical soil features rather than chemical attributes alone. Moreover, it has been demonstrated that *PI* values, which are the primary cause of cracks, may be predicted more accurately using multi-regression models than single-regression models.

Discussion

The arrangement of soil granules is a key factor in assessing the possibility of soil swelling. When contrasted to soil with a more consistent particle size distribution, known as poorly graded soil, soil with a broader range of sizes has greater drainage and a slighter propensity for swelling. It's due to the fact that properly graded soil facilitates the ability for water to drain, which lowers the risk of soil swelling carried on by excessive water absorption. Poorly graded soil, on the contrary, has a larger water-holding capacity and may not drain as efficiently, which might increase the probability of swelling. The particulate distribution might provide useful data for analyzing the possibility of soil swelling and contributing to the design of suitable foundation alternatives. Soils with a substantial clay concentration have a significant possibility for swelling compared to soils with low clay content. Although clay particles are tiny and have a high surface area to volume ratio, they may absorb and retain a lot of water. Since water is absorbed by the soil, the clay particles acquire additional water and expand in volume, leading the soil to swell. The swelling would be determined by the type of clay content, the proportion of clay in the soil, and the amount of water absorbed. As a result, while assessing the possibility of soil swelling, the clay concentration of the soil is an important component to be evaluated. The soil type content in neighboring locations where samples are collected can be predicted using the GIS procedure and IDW approach. It is clear that the *TSS*, *GYP*, and *ORG* concentrations are insufficient to be active throughout the entire examined region. The *LL* and *PI* can be utilized to predict the possibility of soil swelling. These two factors provide useful information regarding soil behavior when subjected to moisture fluctuations. Swelling is more likely in soils with high *LL* and *PI* levels. This is due to the fact that soils with high *LL* and *PI* may hold more water and have a greater tendency to alter volume concerning moisture variations. Such soils can swell as they retain moisture, causing volume changes and possible collapse. According to *LL* and *PI*, most regions have swell potentials ranging from moderate to high.

Conclusions

The geographical analysis in this work was carried out using the distance weighted (IDW) approach, based on information received from seven field boreholes obtained from a residential complex in the city of Kirkuk. The physical and chemical properties of field soil samples were used to perform investigations on the existing cracks in the structures of the examined complex. In particular, the plasticity index and soil characteristics relationships, as well as linear single and linear multi-regression models were used in statistical analysis. Based on the available data and findings of this investigation, the following research conclusions have been proposed:

1. Gravel soil is mostly spread in three zones across the examined residential complex, according to the IDW approach combined with the GIS technique.
2. Using the GIS and IDW analysis, the examined residential complex has the greatest sand and silt percentages in the northern part.
3. According to the IDW approach combined with the GIS procedure, clay soil covers the majority of the analyzed residential complex.
4. The investigated residential complex has the highest gypsum concentration in the central part, according to GIS and IDW examinations.
5. The majority of the soil in the examined residential complex area has liquid limit and plasticity index values between 37.972 and 44.952 and 11.997 and 16.650, respectively, according to the GIS and IDW analyses.
6. The analyzed residential complex has the highest total suspended solids in the central and western regions, according to GIS and IDW assessment.
7. All of the soil in the investigated residential complex is alkaline, with small variations in pH values between 7.850 and 7.979, according to the GIS approach and IDW technique.
8. Most of the soil between the north and center of the analyzed residential complex has organic content values between 1.385% and 1.742%, according to the GIS methodology and IDW approach.
9. Significant positive and negative associations between the physical and chemical parameters of the soil and the plasticity index as a swelling indicator have been observed.
10. The presented linear multi-regression models of the physical and integrated physical and chemical soil properties successfully predicted the plasticity index values, with multiple *R* values of 0.92 for both models.
11. The proposed statistical models can thus offer comprehensive geographic and mechanical interpretations for the crack origins in the analyzed residential complex.

References

Addiscott, T. M. (1993). Simulation modelling and soil behaviour. *Geoderma*, *60* (1–4), 15–40. https://doi.org/10.1016/0016-7061(93)90016-E

Billah, M. M. (2018). Mapping and monitoring erosion-accretion in an alluvial river using satellite imagery – the river bank changes of the Padma river in Bangladesh. *Quaestiones Geographicae*, *37* (3), 87–95. https://doi.org/10.2478/quageo-2018-0027

Cantillo, V., Market, V. & Bird, C. (2017). Empirical correlations for the swelling pressure of expansive clays in the city of Barranquilla, Colombia. *Earth Sciences Research Journal*, *21* (1), 45–49. https://doi.org/10.15446/esrj.v21n1.60226

Matarira, D., Mutanga, O. & Dube, T. (2021). Landscape scale land degradation mapping in the semi-arid areas of the Save catchment, Zimbabwe. *South African Geographical Journal*, *103* (2), 183–203. https://doi.org/10.1080/03736245.2020.1717588

Mohsin, S. & Lone, M. A. (2021). Modeling of reference evapotranspiration for temperate Kashmir Valley using linear regression. *Modeling Earth Systems and Environment*, *7*, 495–502. https://doi.org/10.1007/s40808-020-00921-8

Olmo, J. C., Guervos, R. C. & Rocha, I. M. (2019). The spatial effects of violent political events on mortality in countries of Africa. *South African Geographical Journal*, *101* (3), 285–306. https://doi.org/10.1080/03736245.2019.1612770

Omar, N. Q. & Raheem, A. M. (2016). Determining the suitability trends for settlement based on multi criteria in Kirkuk, Iraq. *Open Geospatial Data, Software and Standards*, *1* (10), 1–9. https://doi.org/10.1186/s40965-016-0011-2

Omer, M., Shanableh, A., Mughieda, O., Arab, M., Zeiada, W. & Al-Ruzouq, R. (2018). Advanced mathematical models and their comparison to predict compaction properties of fine-grained soils from various physical properties. *Soils and Foundations*, *58* (6), 1383–1399. https://doi.org/10.1016/j.sandf.2018.08.004

Ouabo, R. E., Sangodoyin, A. Y. & Ogundiran, M. B. (2020). Assessment of ordinary Kriging and Inverse Distance Weighting methods for modeling Chromium and Cadmium soil pollution in E-Waste sites in Douala, Cameroon. *Journal of Health and Pollution*, *10* (26), 1–20. https://doi.org/10.5696/2156-9614-10.26.200605

Pentos, K., Mbah, J. T., Pieczarka, K., Niedbała, G. & Wojciechowski, T. (2022). Evaluation of multiple linear regression and machine learning approaches to predict soil compaction and shear stress based on electrical parameters. *Applied Sciences*, *12* (17), 8791. https://doi.org/10.3390/app12178791

Raheem, A. M. & Omar, N. Q. (2021). Investigation of distinctive physico-chemical soil correlations for Kirkuk city using spatial analysis technique incorporated with statistical modeling. *International Journal of Geo-Engineering*, *12* (18), 1–21. https://doi.org/10.1186/s40703-021-00147-2

Raheem, A. M., Naser, I. J., Ibrahim, M. O. & Omar, N. Q. (2023). Inverse distance weighted (IDW) and kriging approaches integrated with linear single and multi-regression models to assess particular physico-consolidation soil properties for Kirkuk city. *Modeling Earth Systems and Environment*, *2023*, 1–23. https://doi.org/10.1007/s40808-023-01730-5

Raheem, A. M., Omar, N. Q., Naser, I. J. & Ibrahim, M. O. (2022). GIS implementation and statistical analysis for significant characteristics of Kirkuk soil. *Journal of the Mechanical Behavior of Materials*, *31* (1), 691–700. https://doi.org/10.1515/jmbm-2022-0073

Salahalden, V. F., Shareef, M. A. & Al Nuaimy, Q. A. M. (2023). Characterization of the Chemical Properties of Deposited Red Clay Soil Using GIS Based Inverse Distance Weighted Method in Kirkuk City, Iraq. *Ecological Engineering & Environmental Technology*, *24* (7), 46–60. https://doi.org/10.12912/27197050/169571

Zijl, G. (2019). Digital soil mapping approaches to address real world problems in southern Africa. *Geoderma*, *337* (1), 1301–1308. https://doi.org/10.1016/j.geoderma.2018.07.052

Summary

Implementing GIS and linear regression models to investigate partial building failures. One of the most dangerous field problems in the civil engineering discipline is the suddenly developed cracks in the building, which could be caused by the swelling of the subsurface soil. Thus, this work has focused on employing a procedure in the geographic information system known as the inverse distance weighted (*IDW*) technique, to analyze the extent of cracks in a residential complex in the city of Kirkuk in Iraq using the physical and chemical soil data for seven boreholes from the field of the study. Physical soil parameters such as liquid limit (*LL*), gravel, sand, silt and clay percentages were characterized first, followed by chemical properties such as gypsum content (*GYP*), total suspended solids (*TSS*), potential of hydrogen (pH), and organic content (*ORG*). Furthermore, statistical studies such as plasticity index (*PI*) and soil characteristics association, linear single, and various linear multi-regression models were used. The data analysis shows that there are significantly positive and negative relationships between *PI* as a swelling indicator and the physical and chemical soil properties, although weak to moderate correlations were observed between *PI* and these variables. The *PI* values were accurately predicted by the proposed linear multi-regression models of the physical and integrated physical and chemical soil characteristics, with multiple *R* values of 0.92 for both models. As a result, the suggested statistical models can provide complete geographic and mechanical explanations for the crack sources in the investigated residential complex.