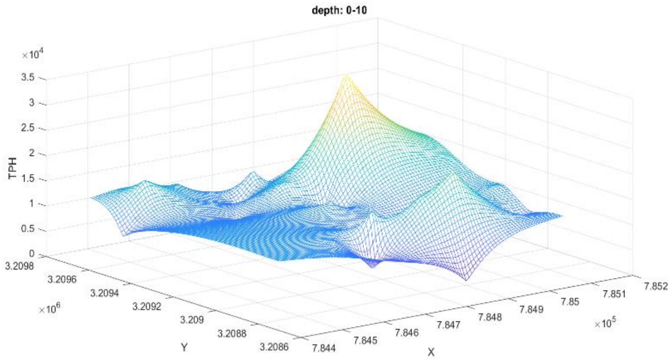
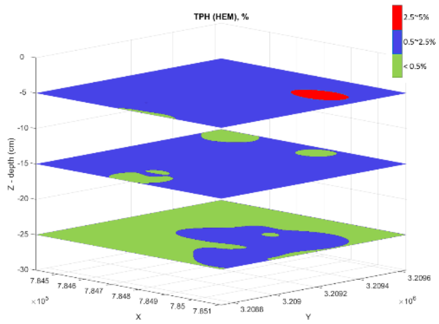
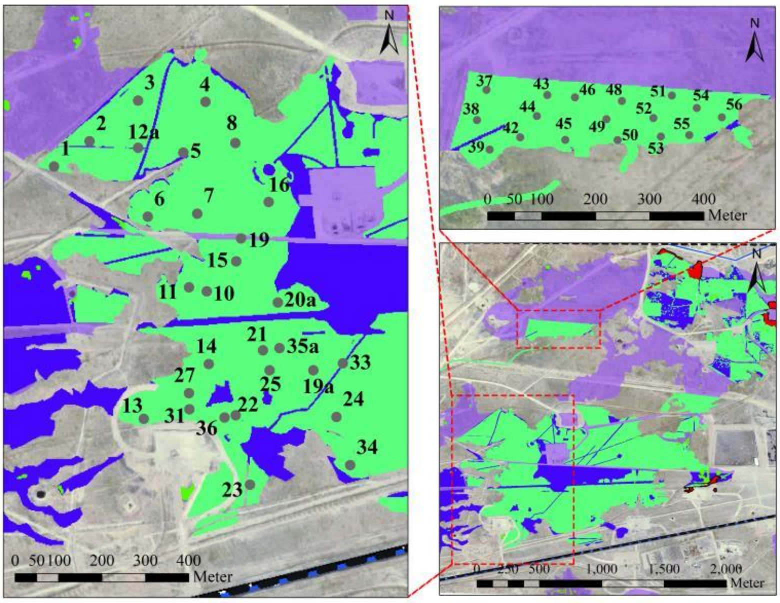
KOC Soil Contamination Information Tool

Version 1







School of Science, Engineering and Environment

University of Salford, Manchester UK

January 2024

Dr. Yu Wang

Contents

1. Introduction

KOC Soil Contamination Information Tool (KOC-SCIT) is a site-specific database software for soil contamination information of two oil lake areas in Kuwait. The information in the database tool is primarily supported by a 3D mapping analysis using a state-of-the-art geo-information analysis method, which uses a limited set of sampled data points to estimate the value of a variable over a continuous spatial field. KOC-SCIT couples the geo-information mapping and curve fitting methods together to predict truly three-dimensional soil contaminated situation from geo-locations to varied local depth, given by small number of sampling points.

KOC-SCIT runs using the Excel of the MS office package. The database contains all the analysed soil contamination properties, they are: TPH (HEM), TPH (C8-C16), TPH (C16-C35), pH, EC, Hazardous Daily Soil Ingestion, and PAHs, in the form of the data-tables for all the predicted properties in meshed geo-map at two to three sampling layers. A friendly user interface has been developed to facility to extract data by the request and definition of users for identified geo-locations and depths, and the looked contamination property levels. The data extraction within the data-tables uses both curve fitting and linear interpolation methods. A technician with basic knowledge and skills using MS Excel can operate the software.

2. The Geo-locations of All Sample Points

Table 1 lists out the location of all sample points adopted for the geo-statistical modelling. Figure 1 marks them in the two site maps.

Table 1. Geo-location of Sample Points

Sampling Plot 1										
No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**10**	**11**
X	784422	784504	784616	784771	784720	784638	784752	784840	784774	784733
Y	3209479	3209537	3209631	3209628	3209511	3209363	3209370	3209533	3209191	3209200
No	**13**	**14**	**15**	**16**	**19**	**21**	**22**	**23**	**24**	**25**
X	784629	784779	784842	784917	784853	784904	784841	784874	785073	784919
Y	3208897	3209023	3209260	3209397	3209313	3209055	3208905	3208745	3208901	3209009
No	**27**	**31**	**33**	**34**	**36**	**12a**	**19a**	**20a**	**35a**	
X	784733	784734	785088	785105	784815	784616	785020	784938	784942	
Y	3208956	3208919	3209025	3208790	3208900	3209522	3209009	3209165	3209060	
Sampling Plot 2										
No	**37**	**38**	**39**	**42**	**43**	**44**	**45**	**46**	**48**	**49**
X	784899	784885	784907	784956	784998	784982	785029	785043	785119	785095
Y	3210434	3210378	3210324	3210348	3210427	3210388	3210345	3210424	3210419	3210385
No	**50**	**51**	**52**	**53**	**54**	**55**	**56**			
X	785114	785200	785171	785184	785241	785230	785282			
Y	3210347	3210431	3210389	3210355	3210409	3210359	3210393			

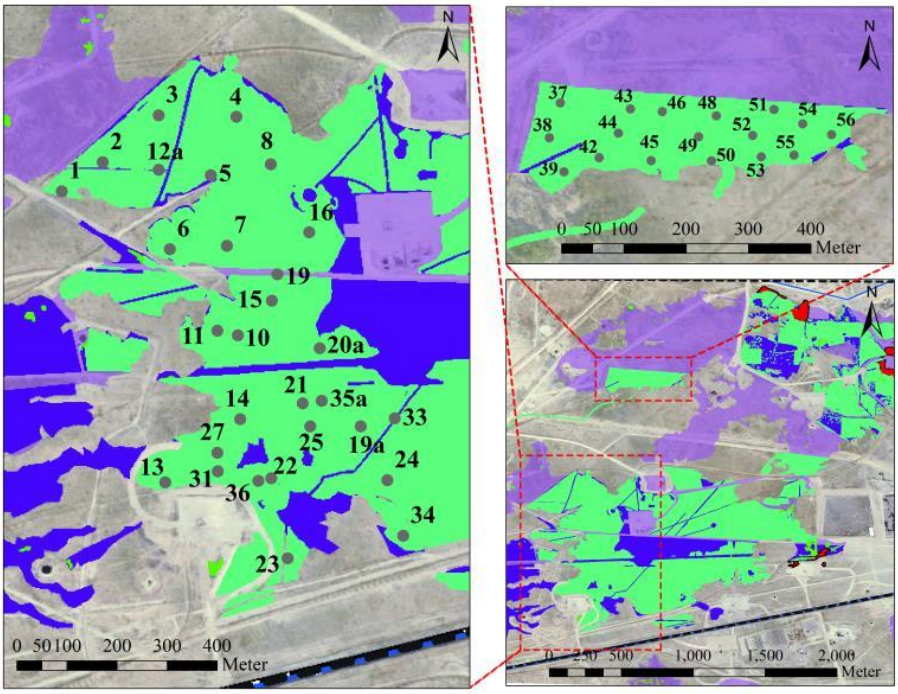


Figure 1. Location of sampling

3. KOC-SCIT Structure

KOC-SCIT is delivered in a form of file folder named **KOC-SCIT-2023**. In the folder there is one executable Excel file named **KOC_SCIT_2023_v1** and a sub-folder**, Backup**, as showed in Figures 2-3.

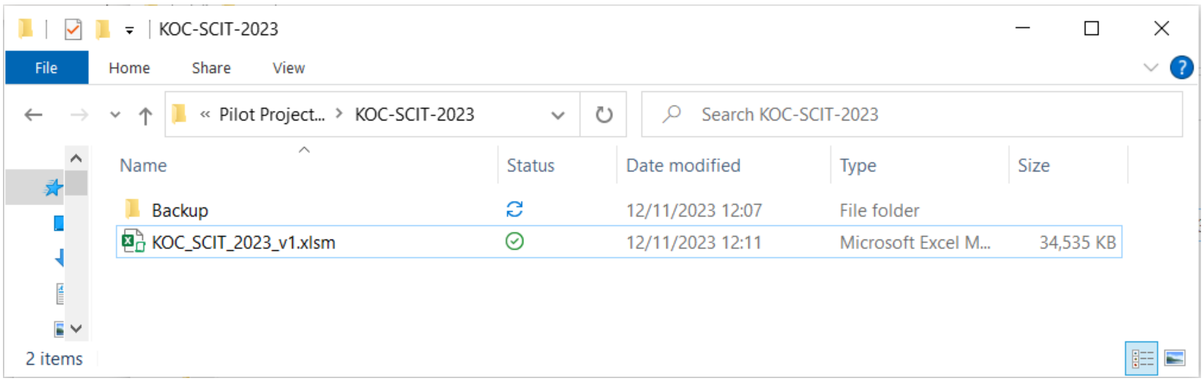


Figure 2. KOC-SCIT-2023 file structure

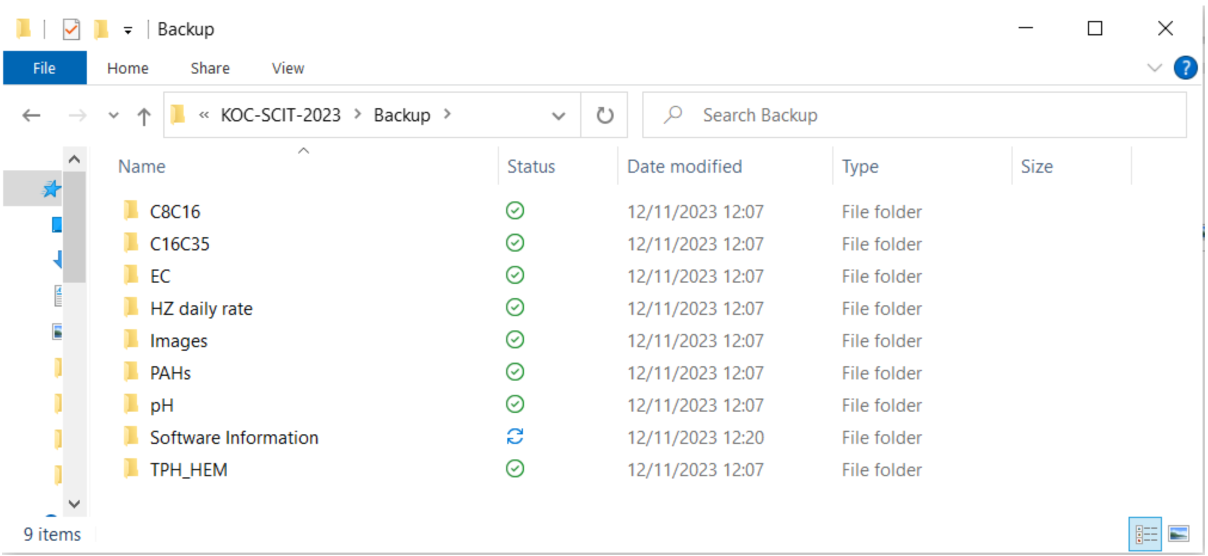
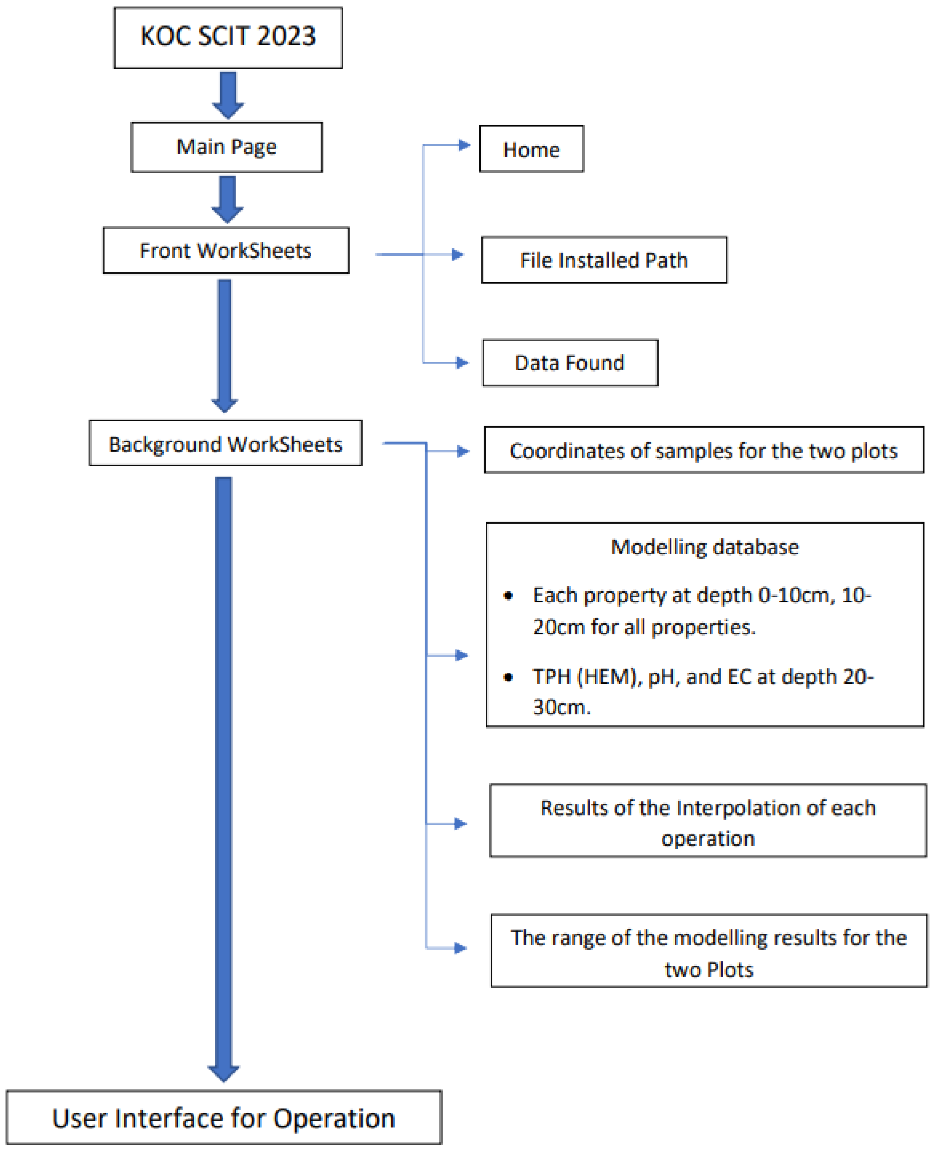


Figure 3. Files in sub-folder, **Backup**.

A flow-chart of the of the components of KOC-SCIT is illustrated below.



The components of KOC-SCIT

4. User Interface and Functions

Run the executable file, **KOC_SCIT_2023_v1**, to start the software to enter the **Home** page as shown by Figure 4.

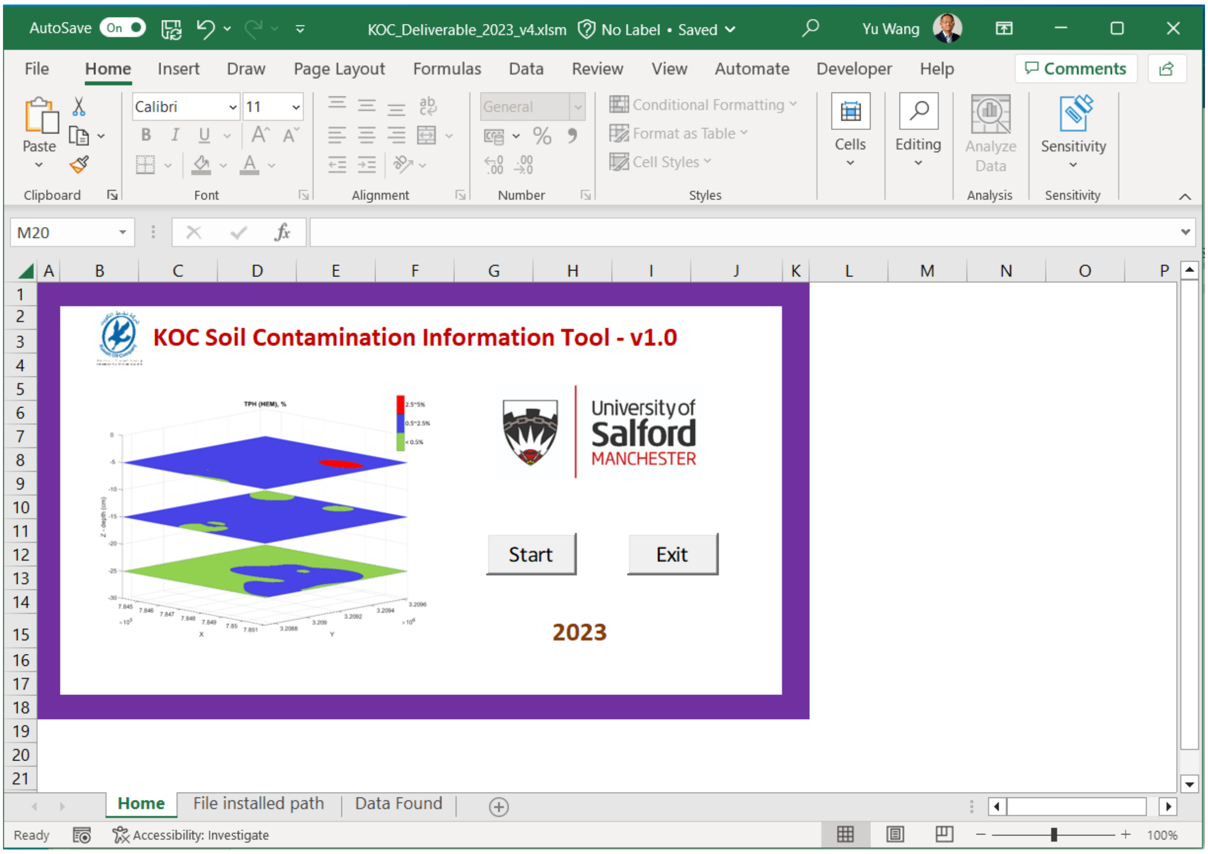


Figure 4. The **Home** page of the KOC-SCIT-2023-v1

There are other two Excel worksheets in line with the **Home** page worksheet, they are **File installed path** and **Data Found**, as displayed in Figures 5-6.

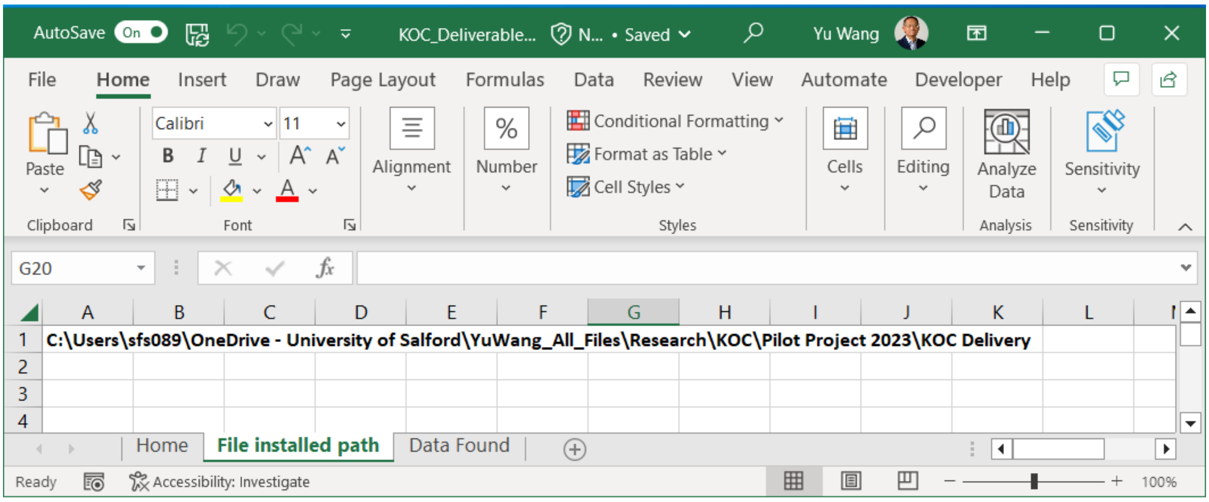


Figure 5. Information in the worksheet, **File installed path**

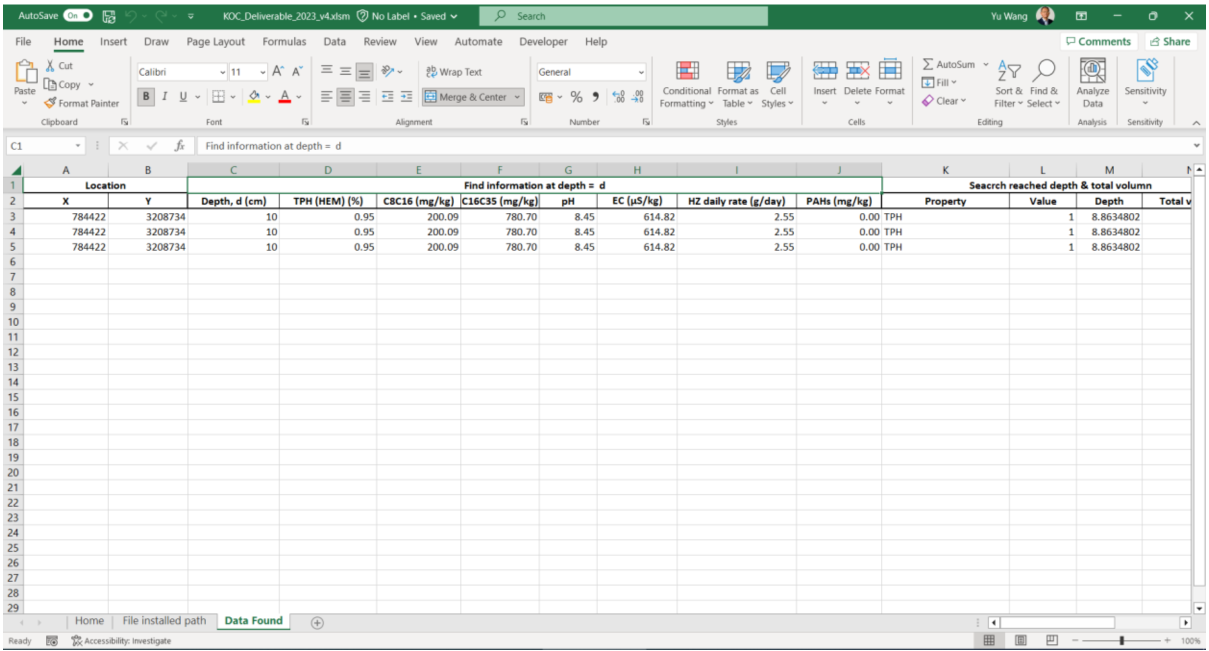


Figure 6. Information in the worksheet, **Data Found**

Click the **Start** button in the **Home** page to enter the User Interface (UI), as presented in Figure 7, for data search and extraction.

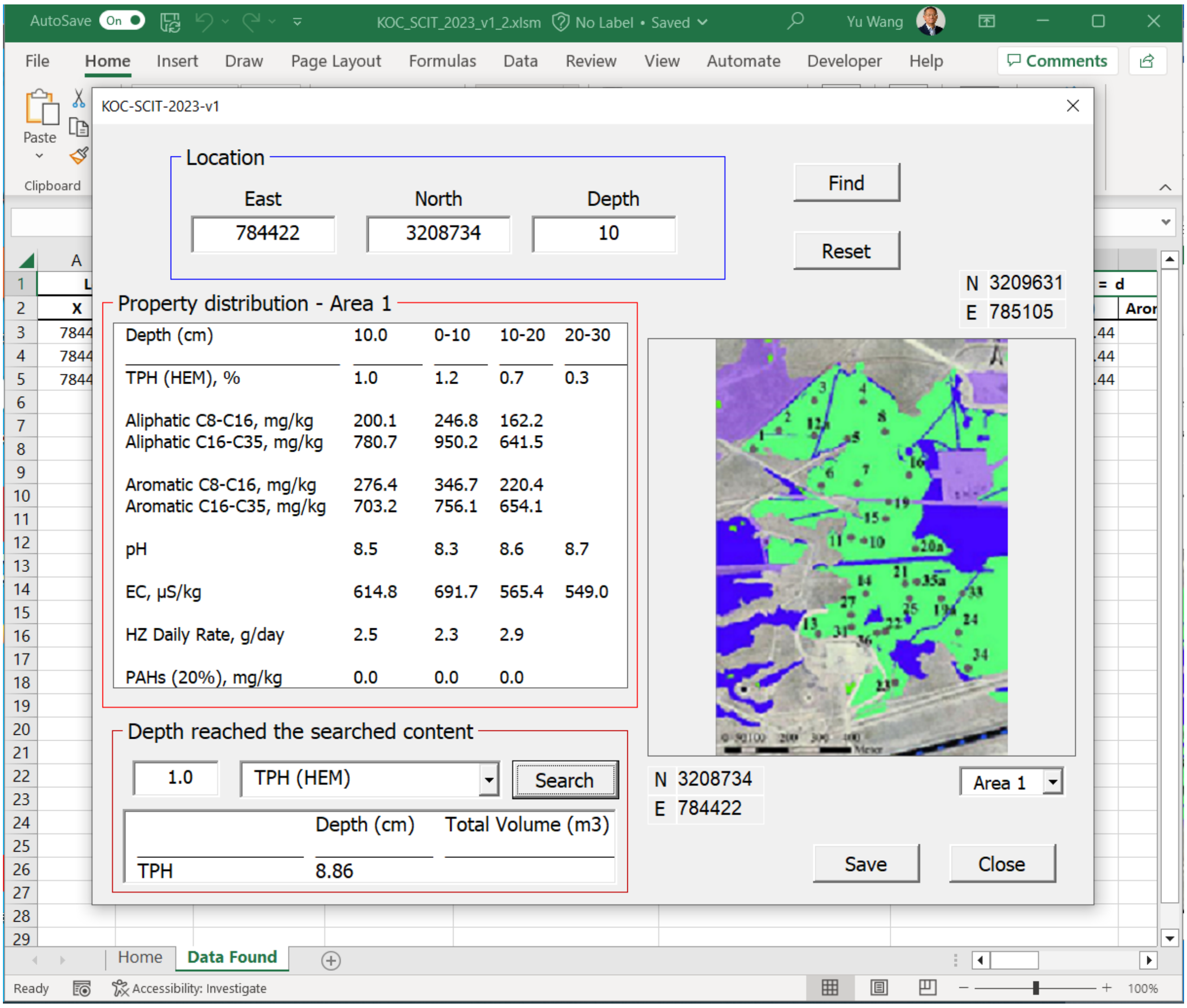


Figure 7. the User Interface for data search and extraction.

The UI designed for two specific areas, i.e., Plot 1 and Plot 2, where the sampling was conducted, and survey data were obtained. The UI provides two types of data search in the database, they are the **Find** for all the contaminating properties at a specific point, when given the geo-**Location** coordinates, X (**East**) and Y (**North**), and the **Depth** down the surface. The obtained data are listed out in the **Property Distribution** window, as illustrated in the Figure 8. The **N** and **E** at the plot map indicate the effective region of the area recoded in the database.

There are 4 series of data listed out in the **Property Distribution** table. That in the 2nd column are the predicted property values at the specified **Depth**. That in the other three columns are the estimated average values at three depth ranges, i.e., **0-10cm**, **10-20cm**, and **20-30cm**, the 3D mapping results on the given sampling.

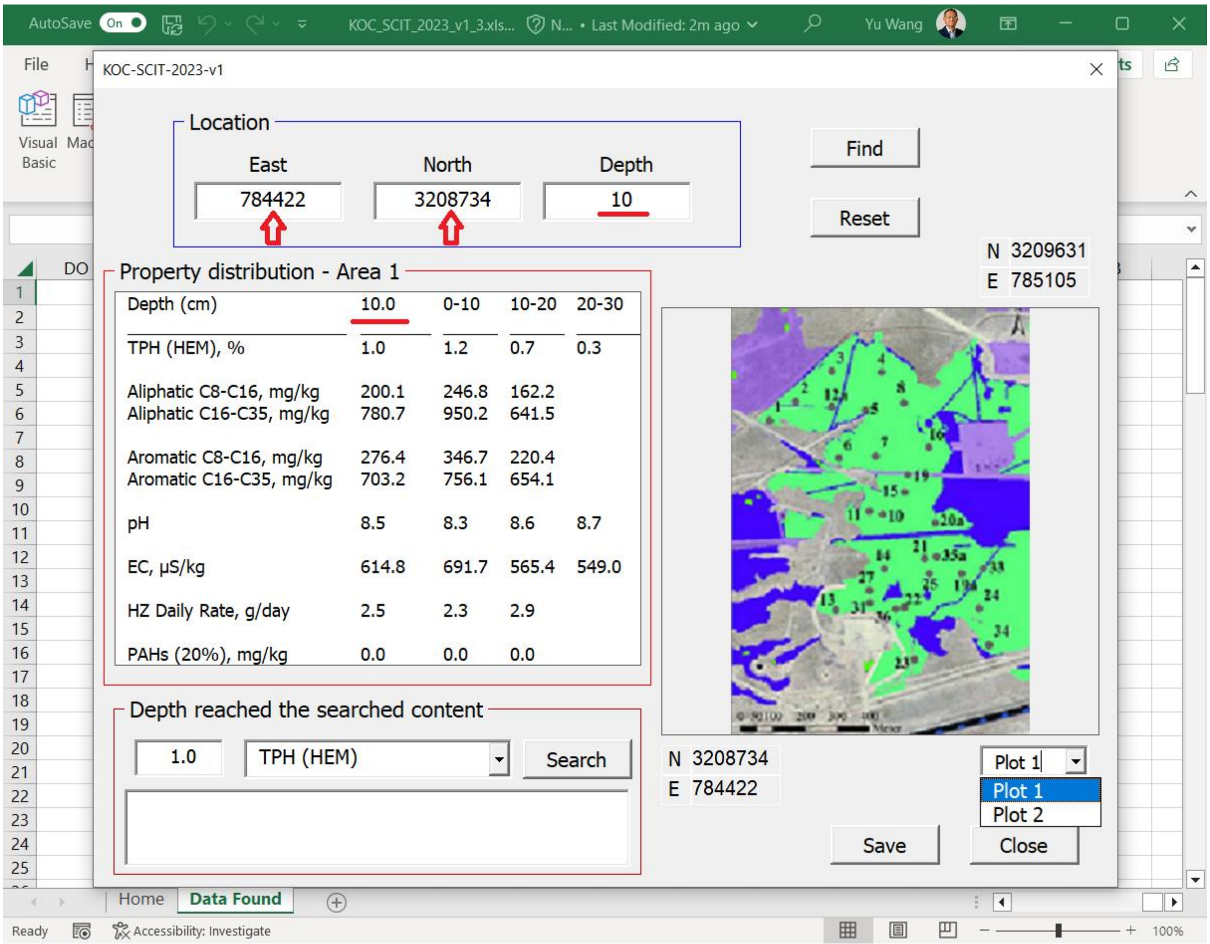


Figure 8. The total contaminating information at a specific point with X, Y and Depth

The second search task is to find the contaminated **Depth** for each property when given the specified property value by the user (Figure. 9). The estimated **Total Volume** of the contaminated soil in the area by the property criterion can also be given out (A function to be delivered by further confirmation whether it is in the contract).

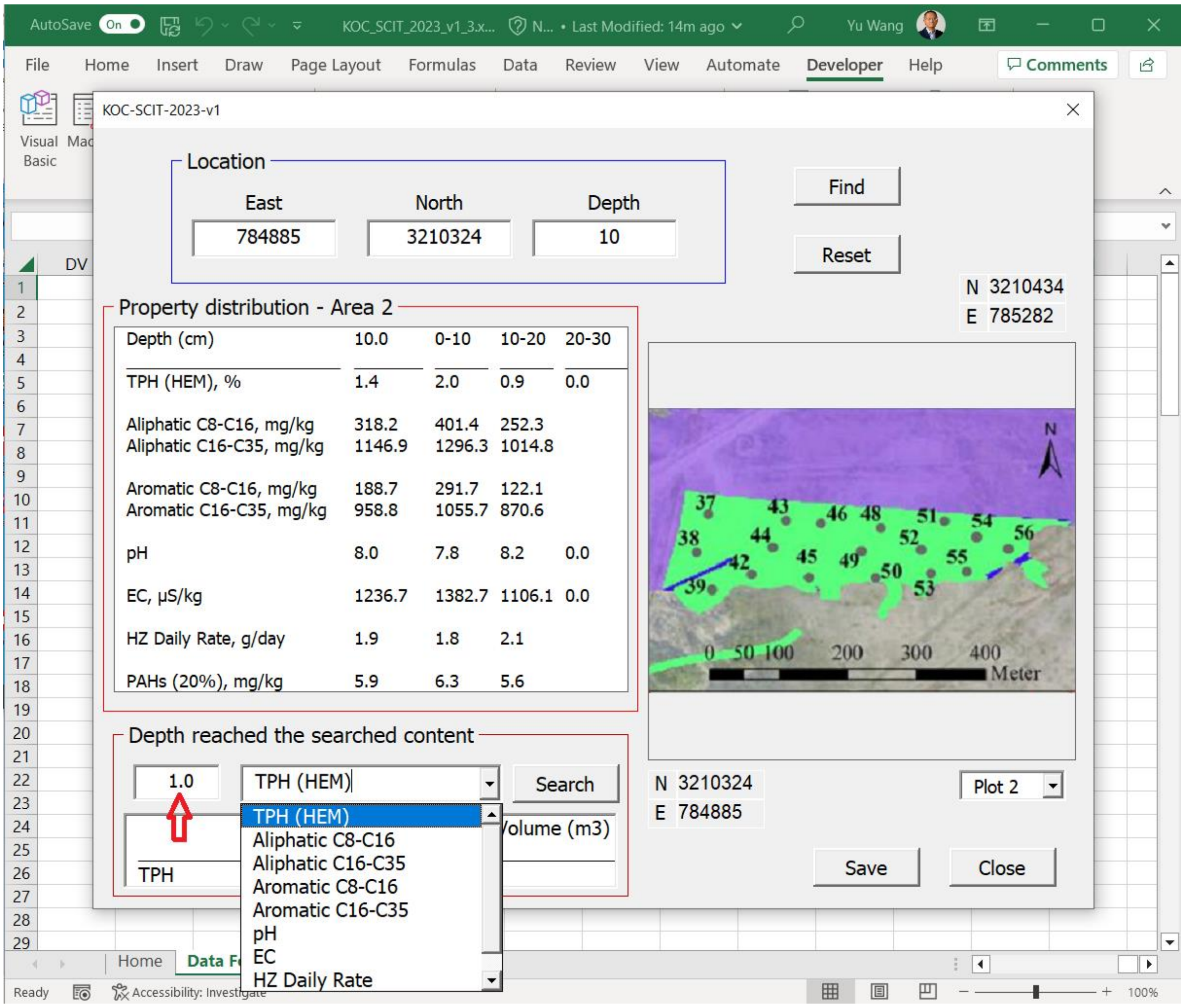


Figure 9. The contaminated depth for a specified property given the criterial value

Each of the data search activities and their corresponding extracted out data information are recorded by the search history table in worksheet, **Data Found**. Click the **Save** button in the UI to export and save all the search history and obtained data to a local file in the same form as the table in the worksheet, **Data Found**, as illustrated in Figure 10.

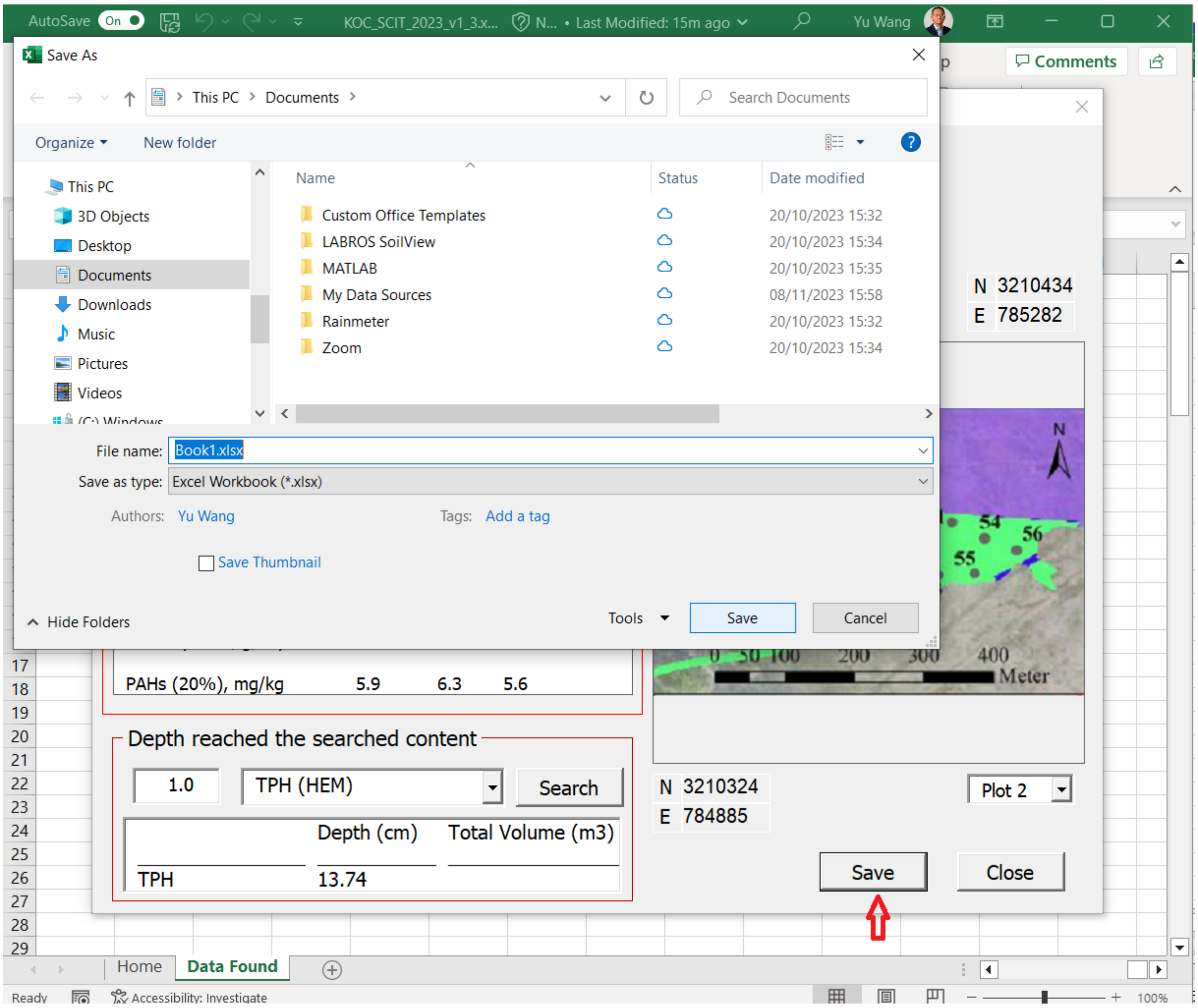
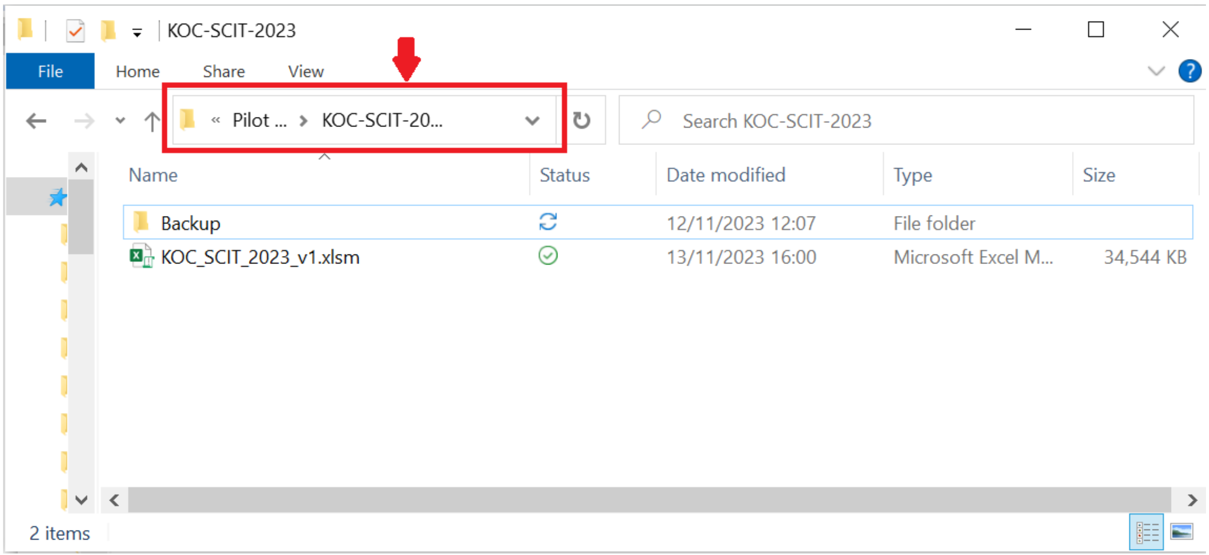


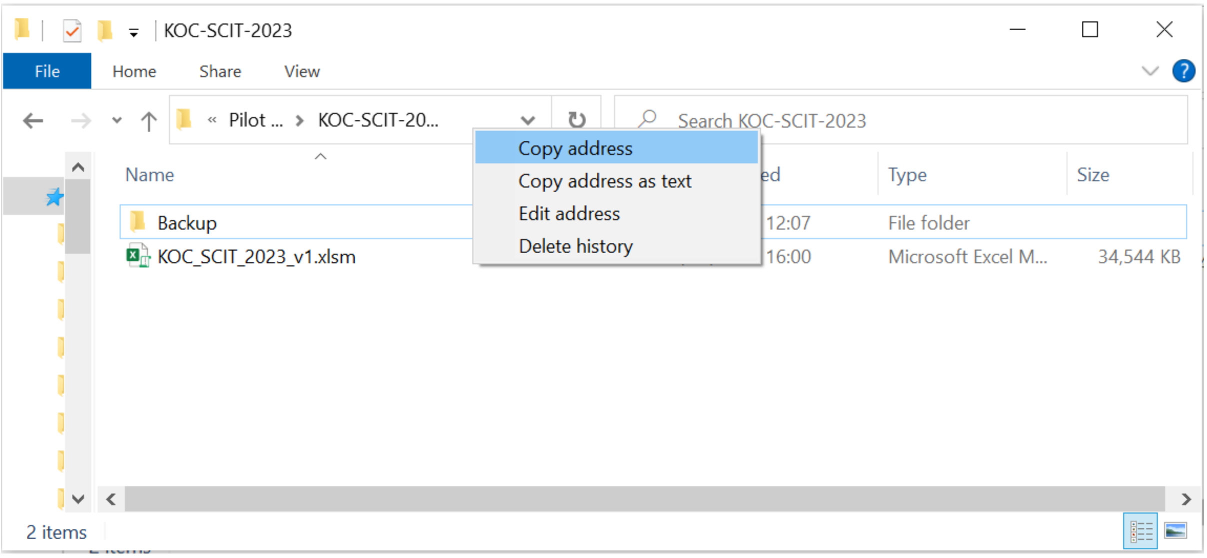
Figure 10. Export and save the search history and data found

5. Software Shipping and Installation

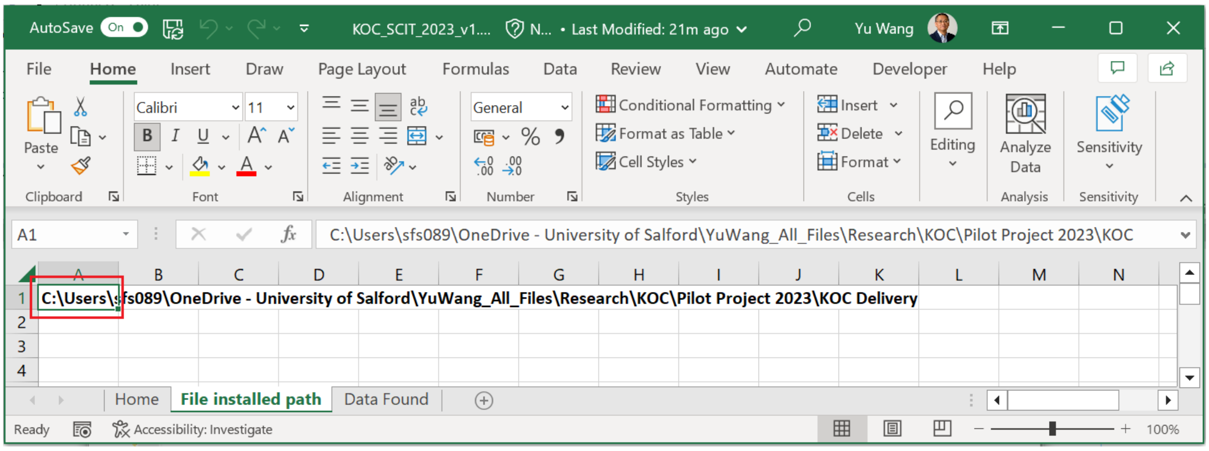
The software is shipped in a compressed file. Save the file in a created new folder with user given name. Extract all files in the folder which now the software folder, **KOC-SCIT-2023**. Open the **KOC-SCIT-2023** folder, there is the Excel file, **KOC_SCIT_2023_v1.xlsm**, and a sub-folder, **Backup**, as illustrated by Figure 2.

Copy the file path of **KOC_SCIT_2023_v1.xlsm**, as illustrated below.

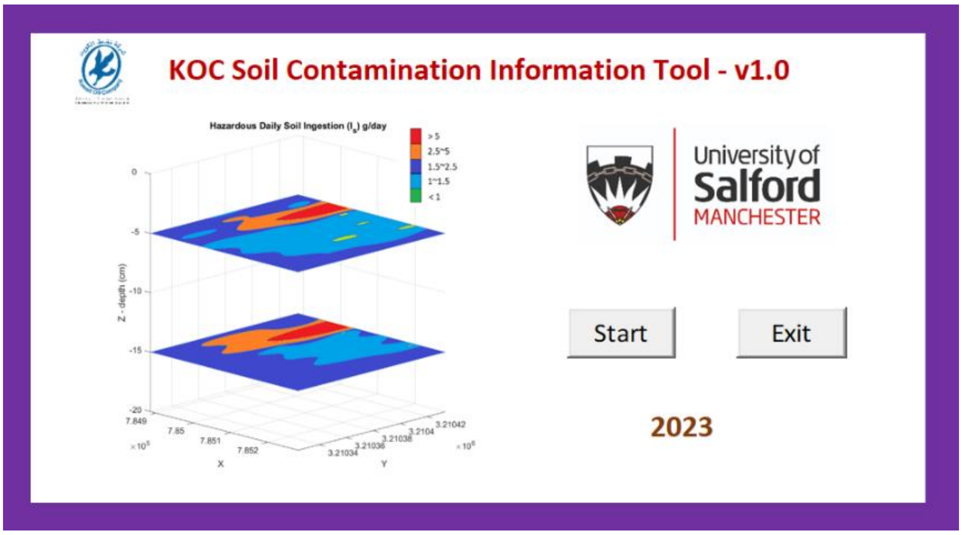




Then paste the copied file path to the cell **"A1"** in the worksheet, File installed path.



Now you can run the software by click the button, **Start**.



Appendix A – The Theory of Geo-statistic Mapping

3-D Mapping of the spatial distributions of geo-physical properties, such as elevations, land uses, soil properties, soil moisture contents etc. can be approximated using the functions of locations, and spatial interpolation. Spatial interpolation refers to the process estimating the unknown property values at specific locations in terms of the measured/observed data at other locations, the sampling points. Many interpolation and approximation methods have been developed, which have been successfully used to predict values of spatial phenomena and the data transformations between different discrete and continuous representations of spatial and spatiotemporal fields.

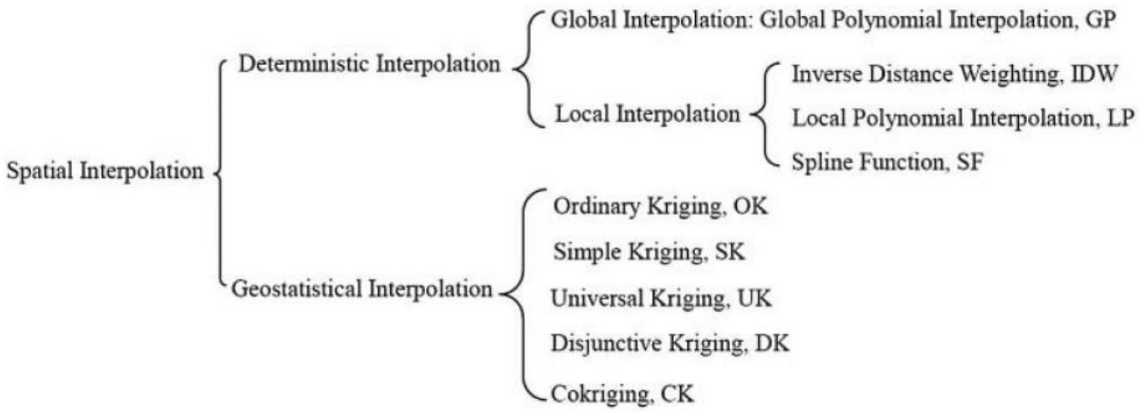


Figure A-1. An overview of spatial interpolation methods (Adapted from Chen et al., 2016)

Geostatistical Analysis

The principles of geostatistical interpolation have been described in a large body of literature. Kriging refers in general to all such geostatistical interpolation methods, which predicts the value at an unobserved location taking a linear combination of the values at surrounding locations, using weights according to a model that describes the spatial correlation (Knotters et al. 2010). In other words, seeing that the special variation of any continuous attributes is generally too complicated to be represented using a simple, smooth deterministic mathematical function, Kriging instead is based on a concept using a random function to represent the geo-parameters/geo-properties, which have a certain spatial covariance. For example, for a spatially continuous geo-property, Z(***X***), where ***X*** indicates the spatial coordinates (x, y), we may find a deterministic value, *M(Z(**X**))*, which is the mean of all Z(***X***). As a result, the geo-property can be expressed in a form of:

$$Z(\boldsymbol{X}) = M(Z(\boldsymbol{X})) + \Delta(\boldsymbol{X}) \qquad \text{Eq. (1),}$$

where Δ is called stochastic spatial residual (variation against the mean value, *M*).

There are two assumptions, they are:

- the mean value, *M*, over a certain spatial region is the same as that within the whole area, ***X***+*h*, where h is the assemble of all the distances of all the location ***X*** to all the points in area. This can be expressed as:

$$E\{ M(Z(\boldsymbol{X})) - M(Z(\boldsymbol{X}+h))\} = 0 \quad \text{Eq. (2)}$$

where, *E* is an error function stands for error or difference between the $M(Z(\boldsymbol{X}))$ and $M(Z(\boldsymbol{X}+h))$.

- The variance the geo-property over the spatial area is auto-correlated to the distance, i.e., points closer to each other have more similar values, and the correlation can be evaluated using the semi-variance as defined below:

$$\gamma(\boldsymbol{X},h) = \frac{1}{2}\,[Z(\boldsymbol{X}) - Z(\boldsymbol{X}+h)]^2 \quad \text{Eq. (3)}$$

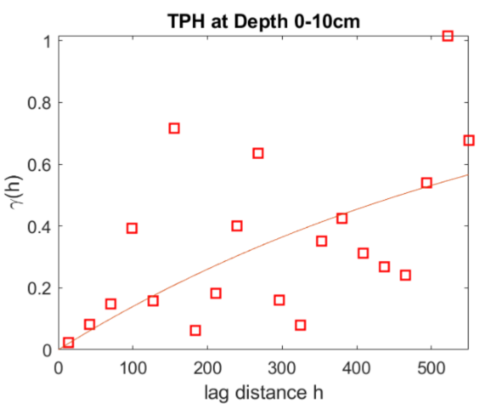
where $\gamma(\boldsymbol{X}, h)$ is called the semi-variance of the geo-property at the distance, *h*, referring to the value at position, $\boldsymbol{X}$.

For certain number of geo-position points, i.e., a collection of sampling data, an average semi-variance for all can be defined as:

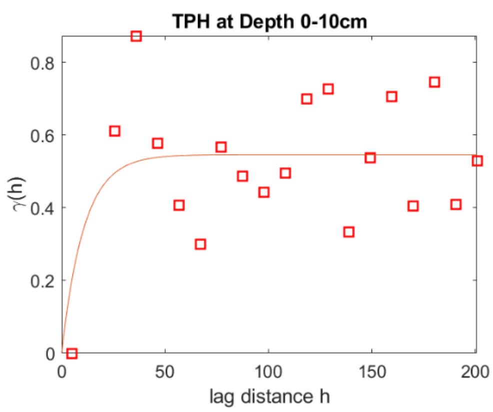
$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}[Z(\boldsymbol{X}_i) - Z(\boldsymbol{X}_i - h)]^2 \quad \text{Eq. (4)}$$

where *i* indicates the individual sampling points and n is the total number of sampling points. The average semi-variance for all is commonly simply called the semi-variance.

The Eq. (4) stands for the fundamental principle of the Kriging method. The Fig. A-2 below shows an example of the semi-variance values, γ, versus distance values, *h*, for the two sets of the sampling data (TPH) used for the KOC-SCIT site specific modelling.



(a) TPH in Plot 1



(b) TPH in Plot 2

Figure A-2. The variation of the semi-variance with distance of the sample points in the Plot 1 and 2 in Figure 1. (*h* is normalized relative distance)

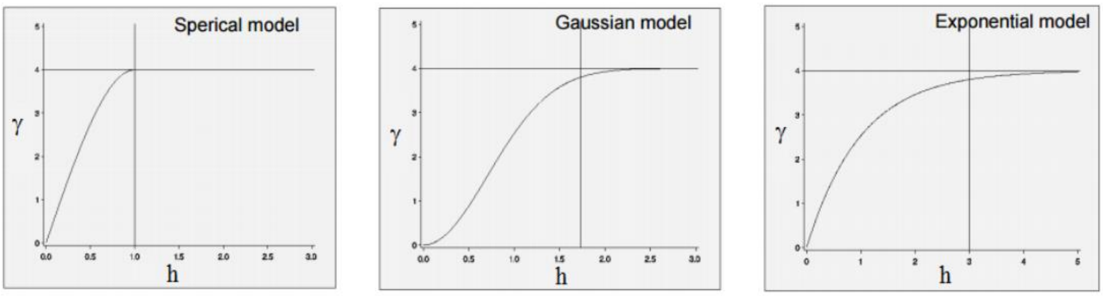
The relationship of the semi-variance, γ, against the distance, *h*, is called the variogram. Mathematical functions (called mathematical models) are employed to represent the relationship, i.e.,

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}[Z(\boldsymbol{X}_i) - Z(\boldsymbol{X}_i - h)]^2$$

and

$$\gamma(h) = f(h)$$

where *f* stands for the models (mathematical functions) as illustrated by the blue curve in Fig. A-2. The typical models popularly used geo-statistic mapping include: Spherical, Guassian, Exponential, and etc, as demonstrated in the Figure A-3 below.



$$\gamma(h) = \begin{cases} c_0\left[\frac{3}{2}\frac{h}{a_0} - \frac{1}{2}(\frac{h}{a_0})^3\right], & \text{for } h \le a_0 \\ c_0, & \text{for } h > a_0 \end{cases} \qquad \gamma(h) = c_0\left[1 - \exp\left(-\frac{h^2}{a_0^2}\right)\right] \qquad \gamma(h) = c_0\left[1 - \exp\left(-\frac{h}{a_0}\right)\right]$$

Figure A-3. The proposed variogram models

For purely modelling aspect, the accuracy depends upon the number of the sample points. Figure A-4 is the experimental variograms obtained from a field by repeated sampling with different sizes of sample points. It demonstrates that the more the number of sample points the higher the accuracy of the variogram model representation.

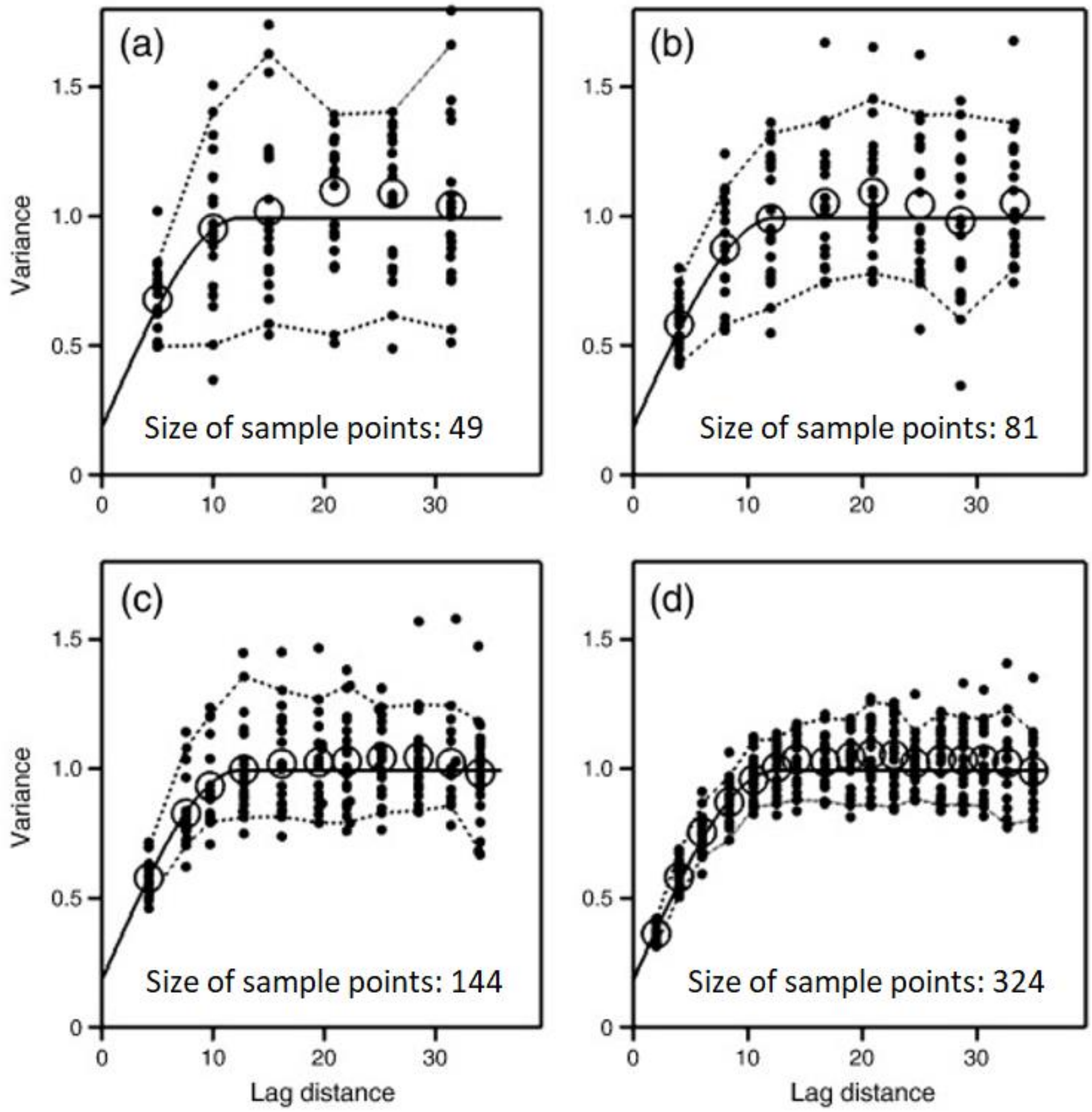


Figure A-4. An example of the size of sample points on the effect of variogram representative model

(The dashed lines join the 5% and 95% quantiles; The circles are the mean values; The solid lines are the spherical model fitted to the variogram) (Oliver & Webster 2014)

With these two assumptions and using the semi variance function to represent the stochastic spatial residual value Δ(***X***), we can rewrite Eq. (1) in the form below:

$$Z(\boldsymbol{X}) = M\big(Z(\boldsymbol{X})\big) + \gamma(h) \qquad \text{Eq. (6)}$$

The Eq. (6) expresses that a geo-property at a specific site, ***X***, equals to the mean value of the geo-property, $M\big(Z(\boldsymbol{X})\big)$, in the region plus the semi-variance, $\gamma(h)$, of the geo-property in the whole region against site, ***X***.

Geostatistical interpolation – Kriging

Kriging is a geostatistical interpolation technique that considers both the distance and the degree of variation between known data points when estimating values in unknown areas.

In term of the interpolation principle, the data value of a geo-property at a random point, $\boldsymbol{X}_p$, can be estimated by giving weights to the sampling points and taking a general formula as:

$$Z(\boldsymbol{X}_p) = \sum_{i=1}^{n} w_i Z(\boldsymbol{X}_i)$$

As:

$$Z(\boldsymbol{X}_p) = M(Z(\boldsymbol{X})) + \gamma(\boldsymbol{X}_p - \boldsymbol{X}_j), \ (j = 1,2,\cdots,\mathrm{n}).$$

$$Z(\boldsymbol{X}_i) = M(Z(\boldsymbol{X})) + \gamma(\boldsymbol{X}_i - \boldsymbol{X}_j), \ (j = 1,2,\cdots,\mathrm{n}).$$

So we have:

$$M(Z(\boldsymbol{X})) + \gamma(\boldsymbol{X}_p - \boldsymbol{X}_j) = \sum_{i=1}^{n} w_i[M(Z(\boldsymbol{X})) + \gamma(\boldsymbol{X}_i - \boldsymbol{X}_j)], \ (j = 1,2,\cdots,\mathrm{n})$$

As:

$$\sum_{i=1}^{n} w_i = 1$$

So, we have:

$$\gamma(\boldsymbol{X}_p - \boldsymbol{X}_j) = \sum_{i=1}^{n} w_i[\gamma(\boldsymbol{X}_i - \boldsymbol{X}_j)], \ (j = 1,2,\cdots,\mathrm{n})$$

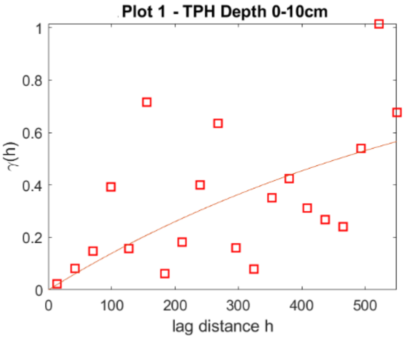
At last, we can work out the weighting coefficient, w_i, for each sampling points, by solving the produced equation system:

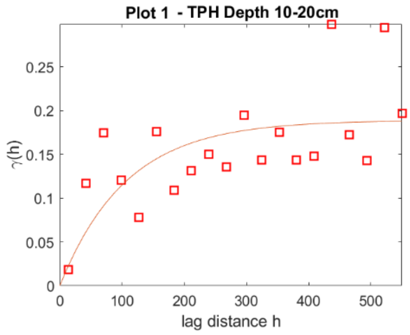
$$[\boldsymbol{A}]\begin{Bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{Bmatrix} = \begin{Bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{Bmatrix},$$

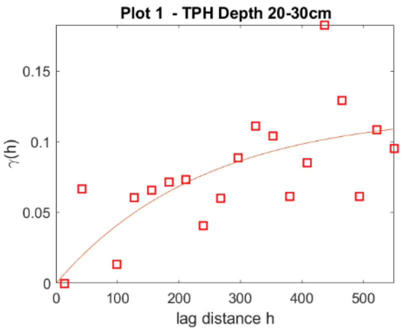
where, **[A]** is an n-by-n matrix.

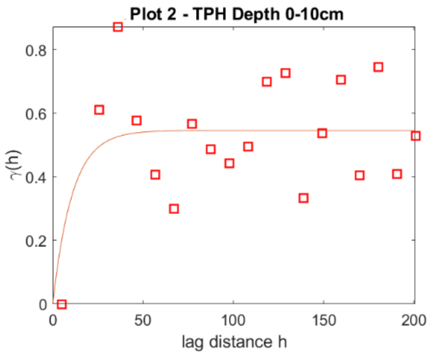
Appendix B – The Semi-Variance of the Sample Properties

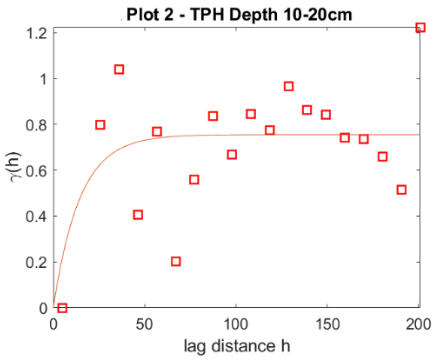
1. TPH (HEM)



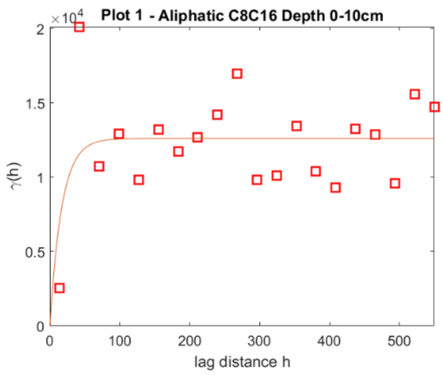


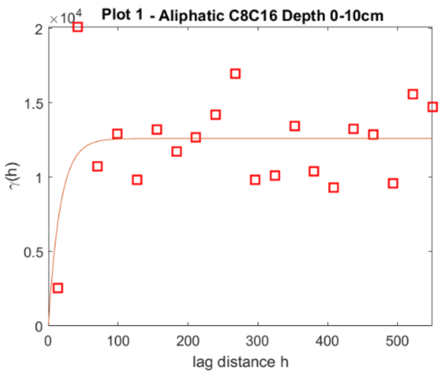


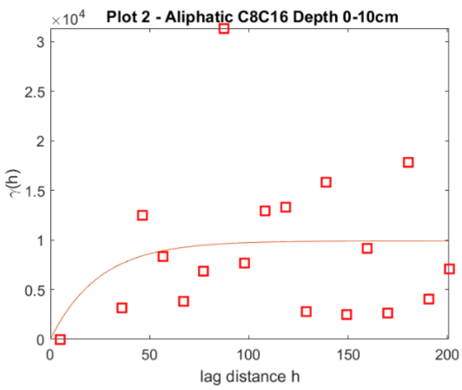


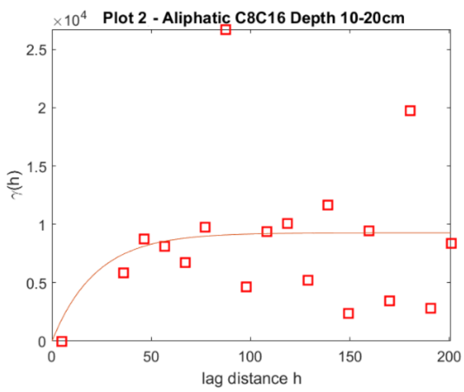


2. Aliphatic C8-C16

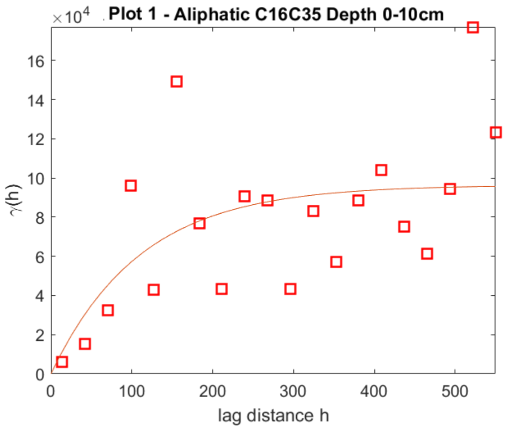


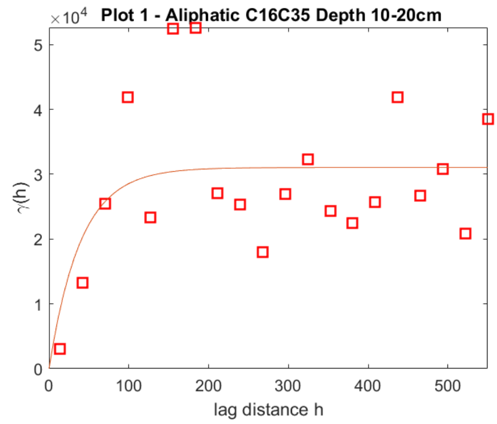


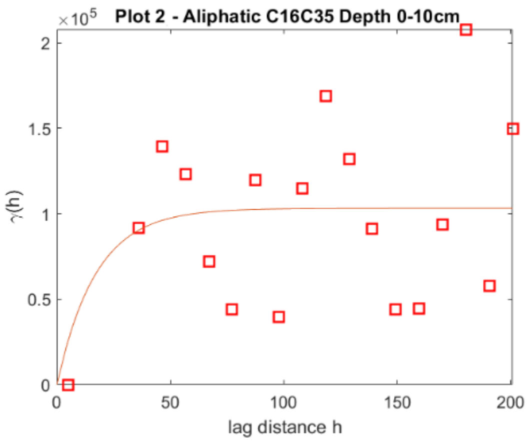


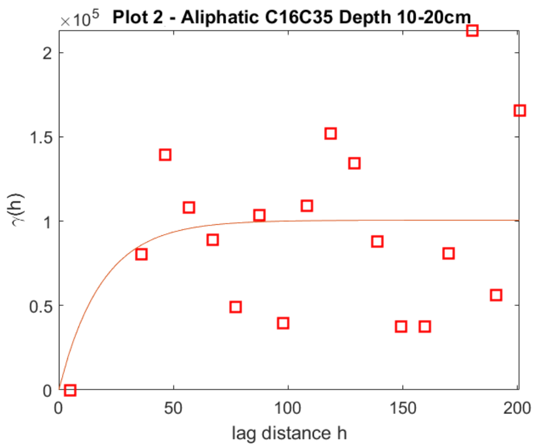


3. Aliphatic C16-C35

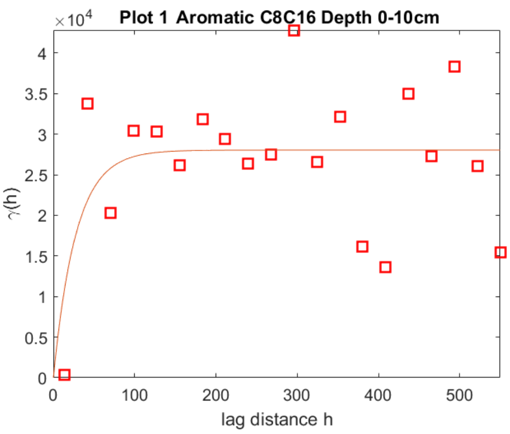


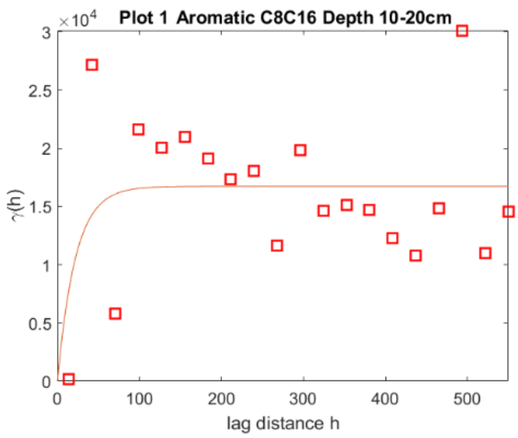


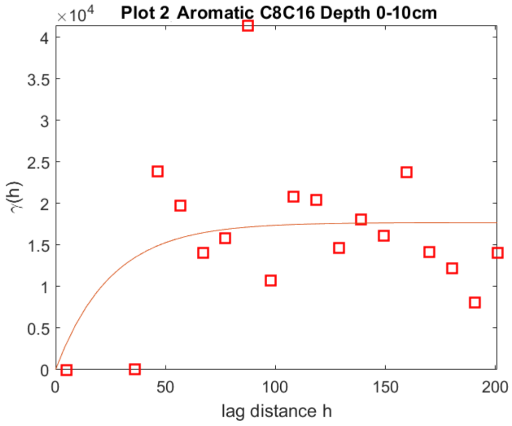


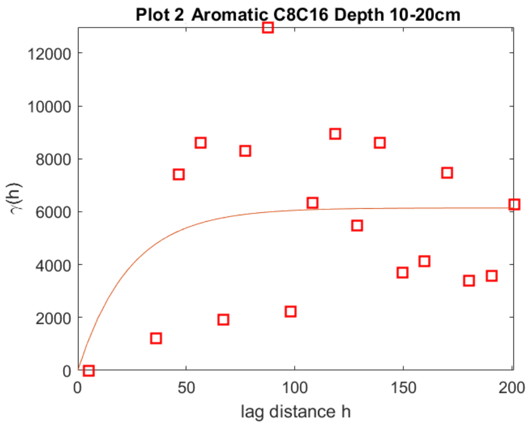


4. Aromatic C8-C16

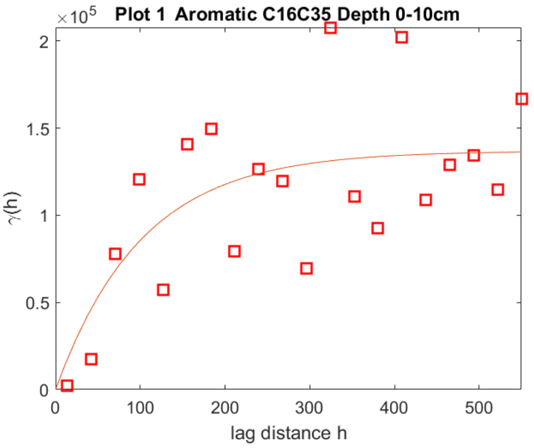


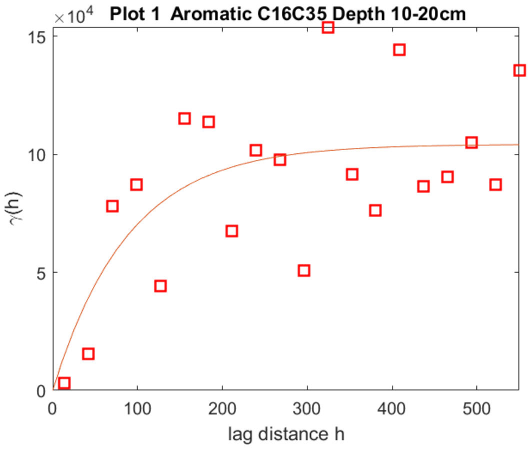


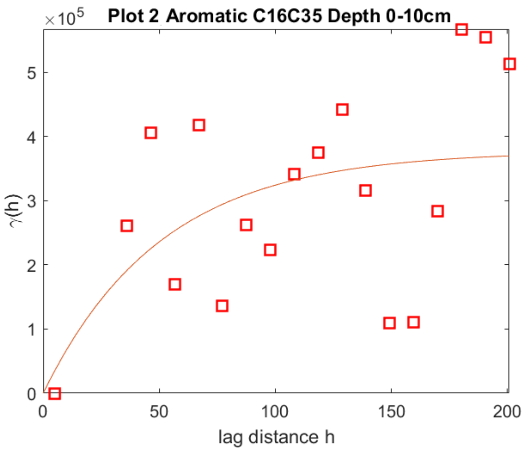


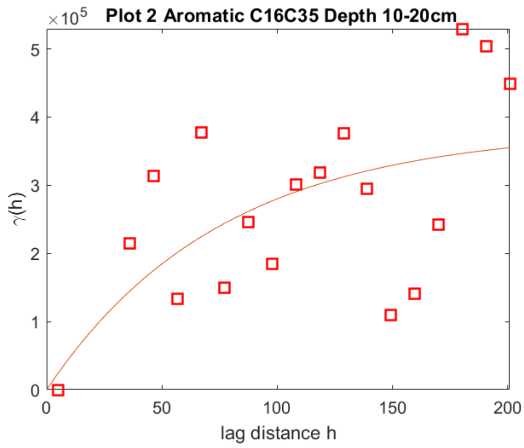


5. Aromatic C16-C35

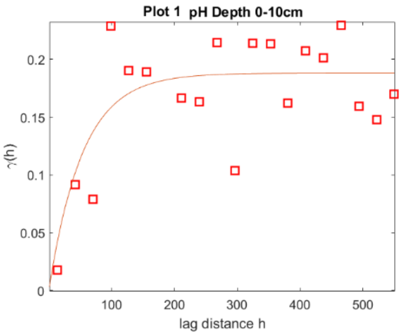


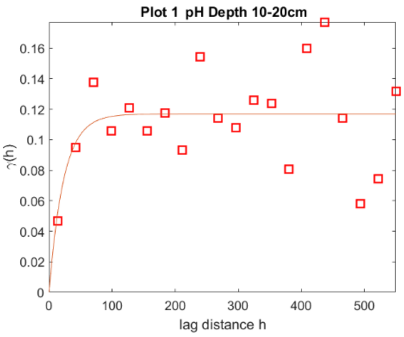


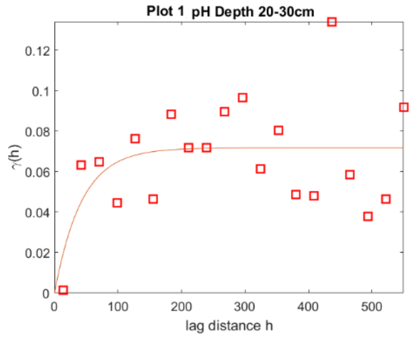


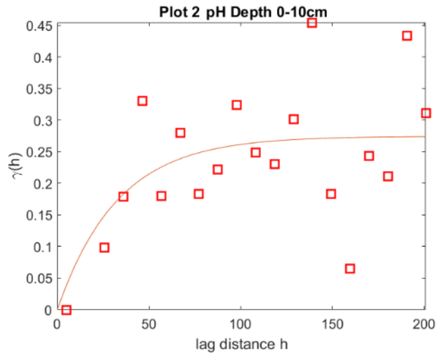


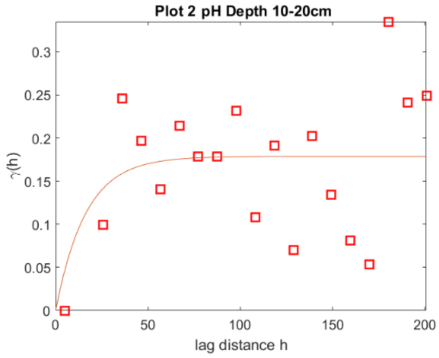
6. pH


Plot 1 pH Depth 0-10cm
γ(h)
lag distance h
0 0.05 0.1 0.15 0.2
100 200 300 400 500

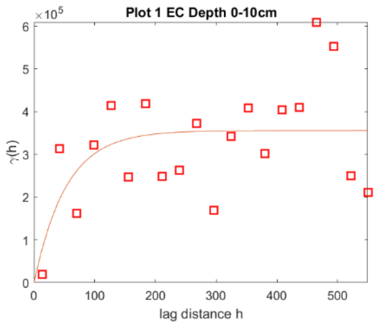

Plot 1 pH Depth 10-20cm
γ(h)
lag distance h
0 0.02 0.04 0.06 0.08 0.1 0.12 0.14 0.16
0 100 200 300 400 500

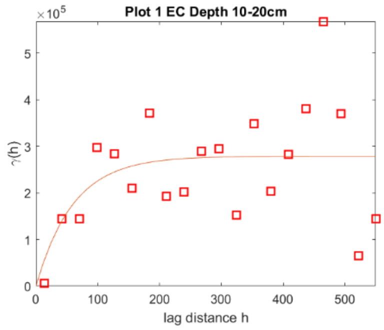

Plot 1 pH Depth 20-30cm
γ(h)
lag distance h
0 0.02 0.04 0.06 0.08 0.1 0.12
0 100 200 300 400 500

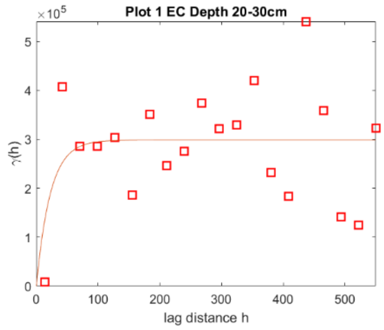

Plot 2 pH Depth 0-10cm
γ(h)
lag distance h
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45
0 50 100 150 200

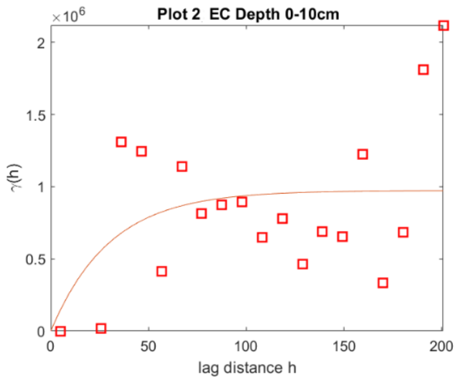

Plot 2 pH Depth 10-20cm
γ(h)
lag distance h
0 0.05 0.1 0.15 0.2 0.25 0.3
0 50 100 150 200

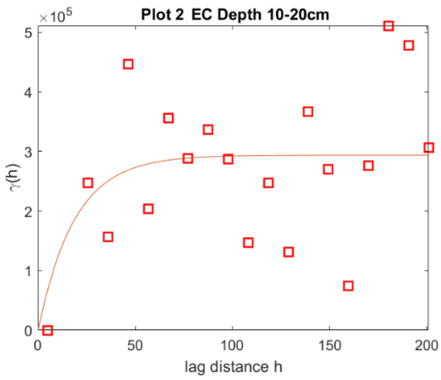
7. EC


Plot 1 EC Depth 0-10cm
×10^5
γ(h)
lag distance h
0 1 2 3 4 5 6
0 100 200 300 400 500

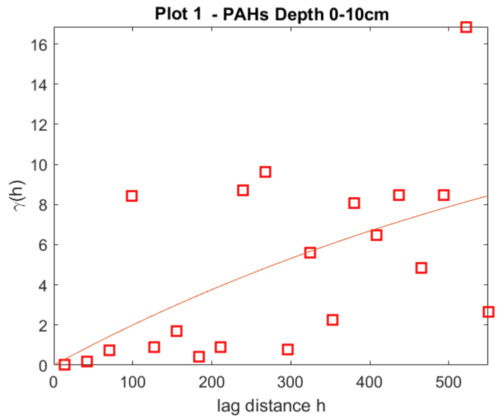

Plot 1 EC Depth 10-20cm
×10^5
γ(h)
lag distance h
0 1 2 3 4 5
0 100 200 300 400 500

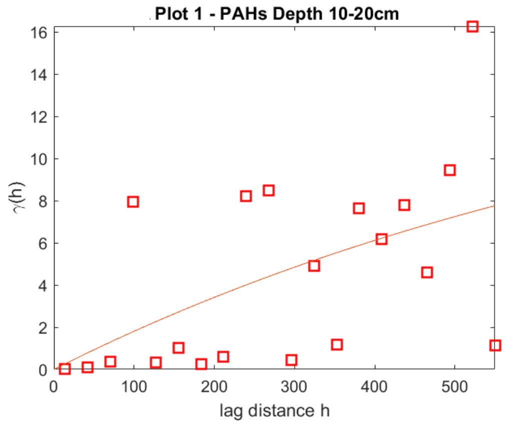

Plot 1 EC Depth 20-30cm
×10^5
γ(h)
lag distance h
0 1 2 3 4 5
0 100 200 300 400 500

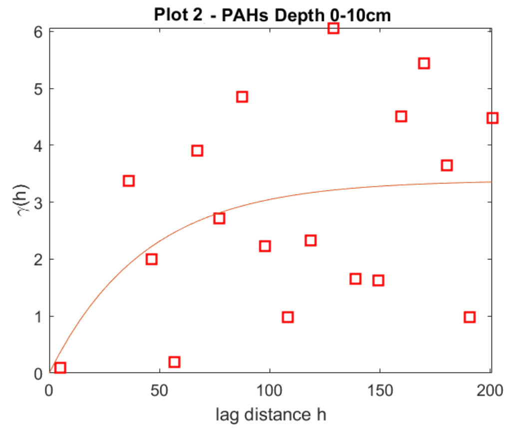

Plot 2 EC Depth 0-10cm
×10^6
γ(h)
lag distance h
0 0.5 1 1.5 2
0 50 100 150 200

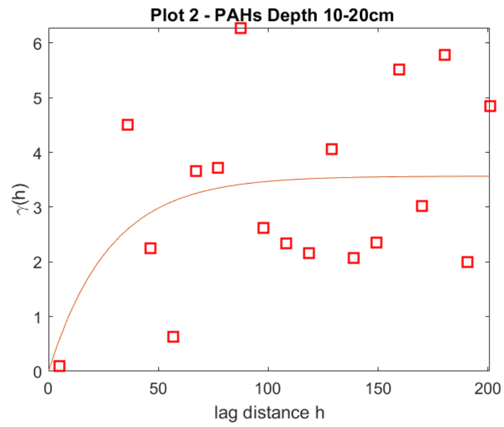

Plot 2 EC Depth 10-20cm
×10^5
γ(h)
lag distance h
0 1 2 3 4 5
0 50 100 150 200

8. PAHs


Plot 1 - PAHs Depth 0-10cm
γ(h)
lag distance h
0
2
4
6
8
10
12
14
16
0
100
200
300
400
500


Plot 1 - PAHs Depth 10-20cm
γ(h)
lag distance h
0
2
4
6
8
10
12
14
16
0
100
200
300
400
500


Plot 2 - PAHs Depth 0-10cm
γ(h)
lag distance h
0
1
2
3
4
5
6
0
50
100
150
200


Plot 2 - PAHs Depth 10-20cm
γ(h)
lag distance h
0
1
2
3
4
5
6
0
50
100
150
200

References

Chen, S., & Guo, J. (2016). Geomatics, Natural hazards and risk, 1947-5713.

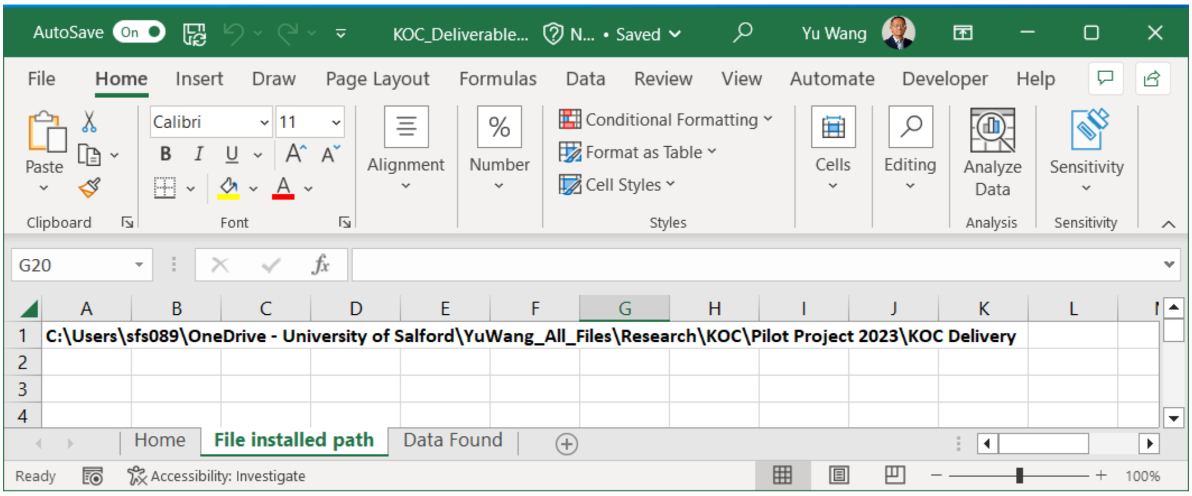
Knotters M., Heuvelink G.B.M., Hoogland T., Walvoort D.J.J., A disposition of interpolation techniques, Werkdocument 190 − juni 2010.

Oliver M.A., Webster R., A tutorial guide to geostatistics: Computing and modelling variograms and kriging, CATENA, 113, 2014, 56-69.

KOC-SCIT-2024

KOC-SCIT-2024 is a revised version of the KOC-SCIT-2023 following the request from KOC for the integration of the GeoModelling software and the KOC_SCIT database. It keeps the basic structure of the KOC-SCIT-2023, but can load in the result data file generated by GeoModelling. Once loaded in the GeoModelling data file, users can search for specific local information for the main soil contamination properties as same as the previous KOC_SCIT-2023 version.

Comparing with KOC-SCIT-2023, KOC-SCIT-2024 has no more need to edit the **File installed path.** So The **File installed path worksheet** has been deleted from KOC-SCIT-2024.



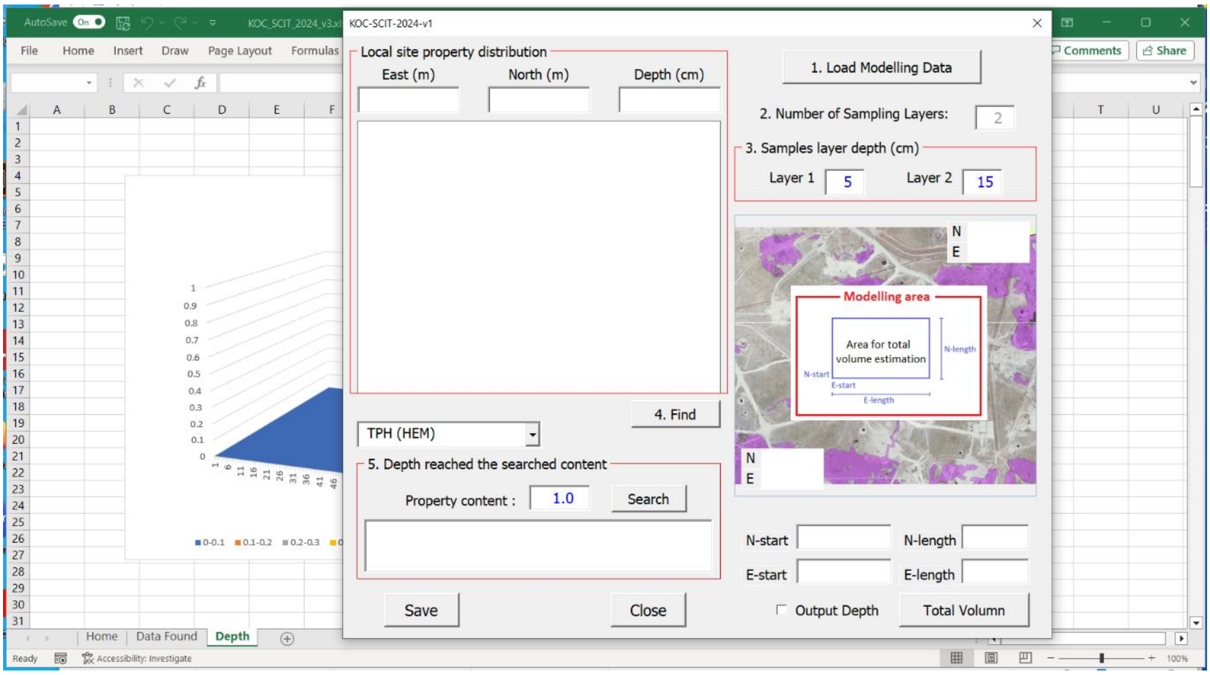


Fig. 1. The main window of the KOC-SCIT-2024

1. Load data information

Click the Load Modelling Data button to load in the modelling data from the GeoModelling software.

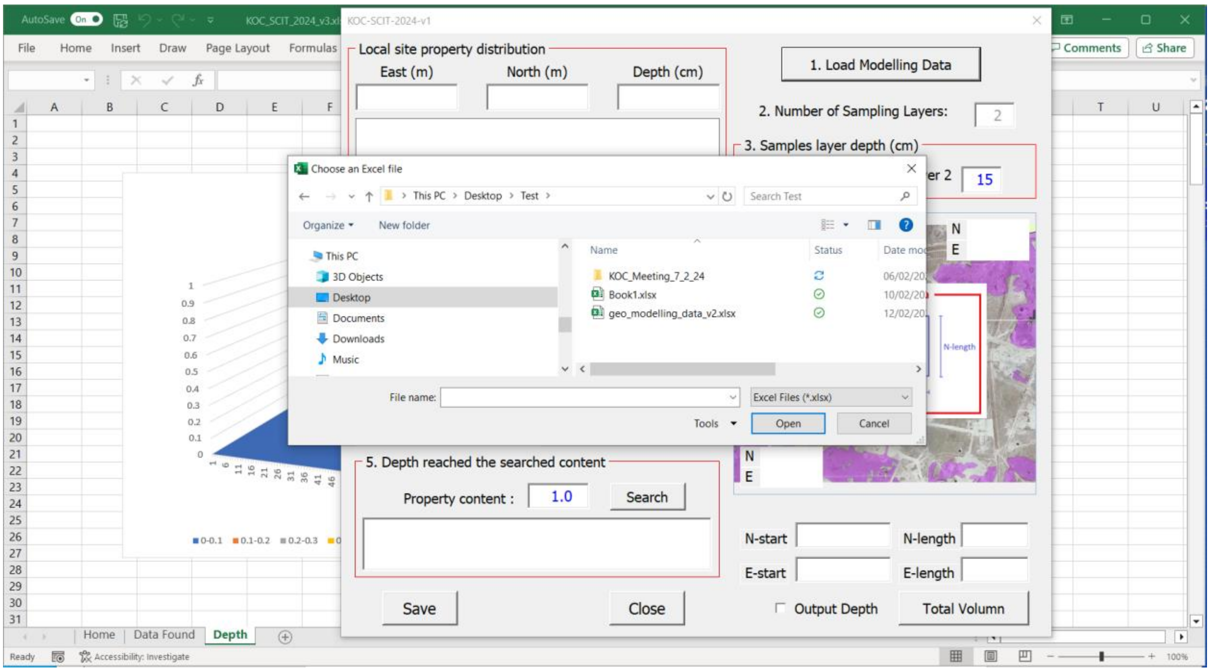


Fig. 2. Open the folder where the GeoModelling result file is, such as the example, modelling_test.xls

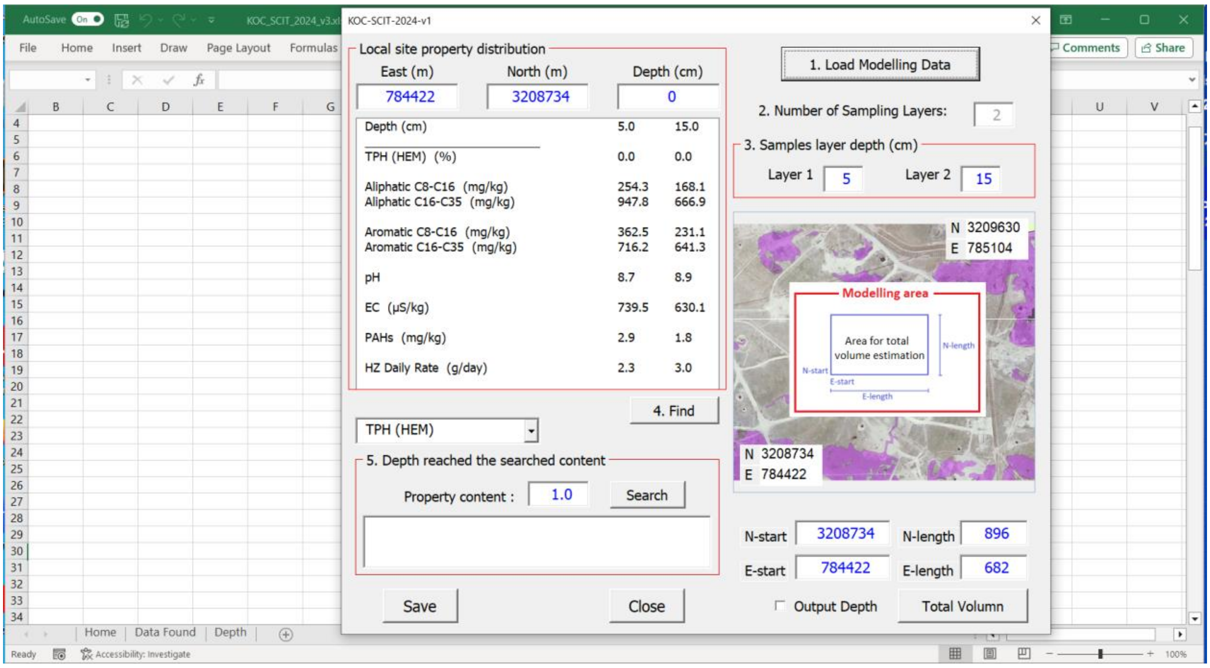
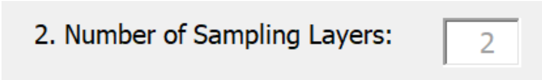


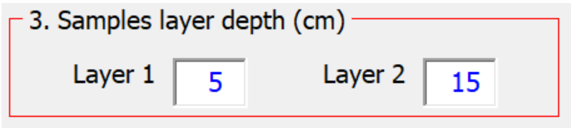
Fig. 3. GeoModelling result file has been successfully loaded in

2. Number of layers



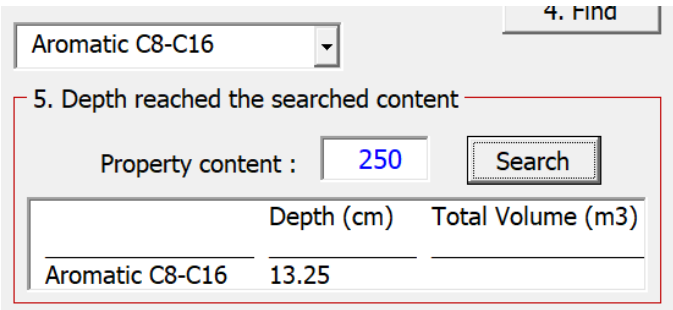
KOC-SCIT-2024 has predefined number of the sampling layers, which is 2 and is not editable.

3. Depth of the two layers



Users need to give certain definition for the depth of the two layers Inaccurate values will affect the local interpolation and the total volume calculation.

4. Set searched property criterion

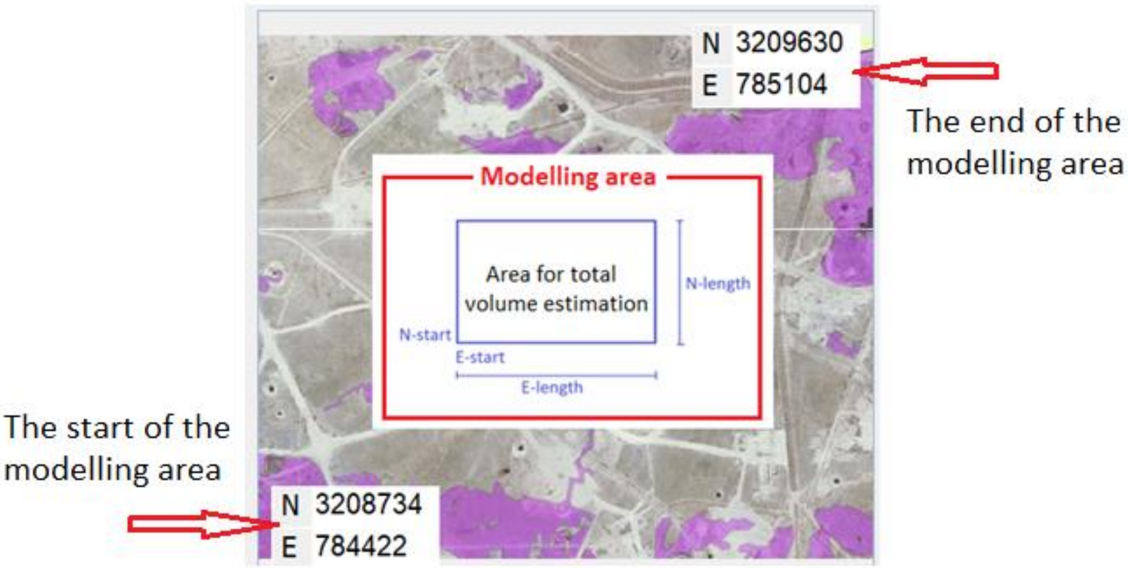


Select the property and give the property value (content) to search for the contaminated depth.

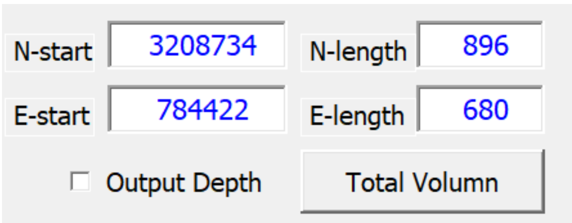
5. Calculate the total contaminated soil volume

User can define a sub-rectangular area within the effective sampling region (Modelling area) to calculate the total contaminated soil volume in the sub-area for a given criterion, which includes a certain soil property and the value of content.

- The modelling area and sub-rectangular area

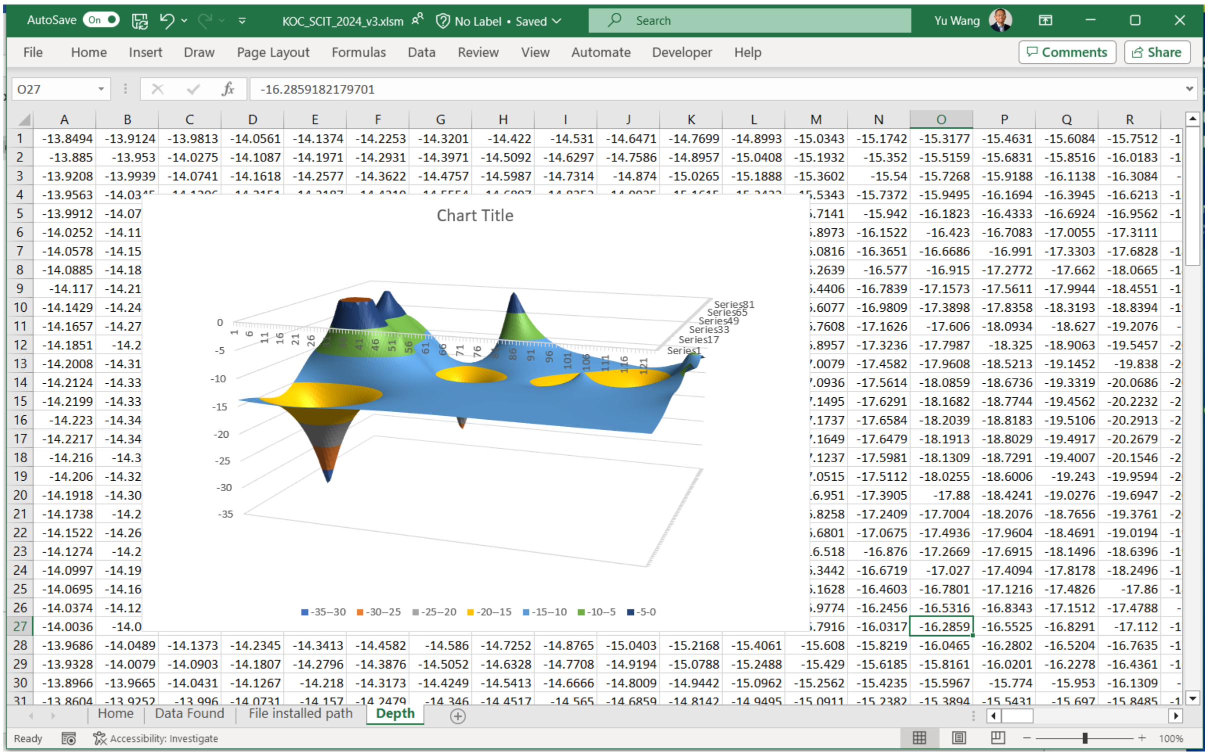


- Define the sub-area for volume estimation



- Output the volume distribution (depth) map

Click the Output Depth will export the bottom surface of the contaminated volume in the area for the given property criterion. The exported data are in Excel format.



6. Save/Output the volume distribution

Click the Save button to out the volume distribution in the form of Excel worksheet.