Development of Predictive Model for 3-D Distribution of Key Soil Parameters

Dr. Yu Wayne Wang

Associate Professor in Civil & Environmental Engineering

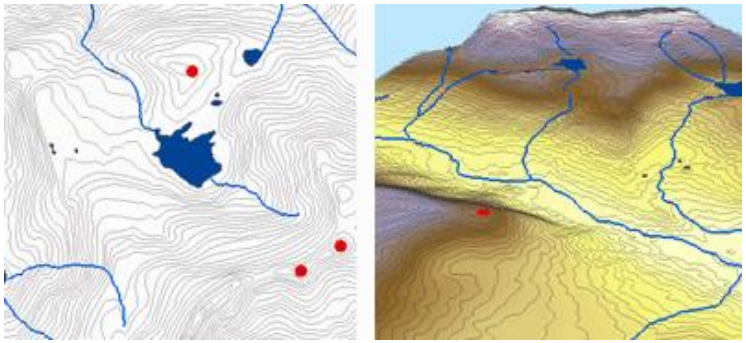
School of Science, Engineering & Environment
University of Salford, Manchester UK

Contents

1. Geostatistical interpolation techniques
2. Modelling and prediction for sampling plots
3. Software development
4. Software installation and operation

1. Geostatistical interpolation techniques

- Geostatistics is used to analyze and predict the values associated with spatial phenomena.
- Geostatistical methods not only provide interpolated values, but also the uncertainty for those values.
- Geostatistics is widely used in environmental science and engineering.

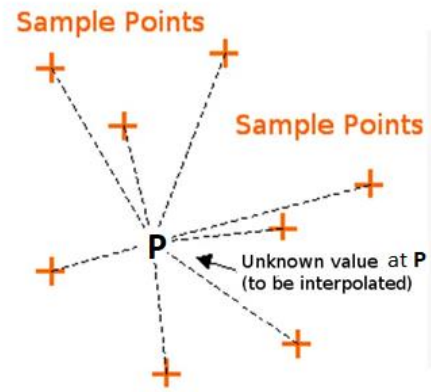


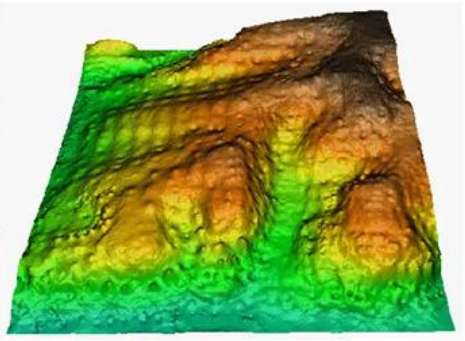
The spatial distributions of geo-physical properties, such as elevations, soil contamination, etc. can be approximated using spatial interpolation.

- Spatial Interpolation
  - Deterministic Interpolation
    - Global Interpolation: Global Polynomial Interpolation, GP
    - Local Interpolation
      - Inverse Distance Weighting, IDW
      - Local Polynomial Interpolation, LP
      - Spline Function, SF
  - Geostatistical Interpolation
    - Ordinary Kriging, OK
    - Simple Kriging, SK
    - Universal Kriging, UK
    - Disjunctive Kriging, DK
    - Cokriging, CK

An example of the deterministic interpolation - the Inverse Distance Weighting (IDW)

- IDW is simple and popular in practice.
- It assumes that the value of a geo-property at a spatial position relates to the data at other surrounding points.
- Their relation can be approximated in terms of their distance between.





$$Z(X_p) = \sum_{i=1}^{n} w_i \, Z(X_i)$$

$$w_i = \frac{\frac{1}{(X_p - X_i)^m}}{\sum_{j=1}^{n} \frac{1}{(X_p - X_j)^m}}$$

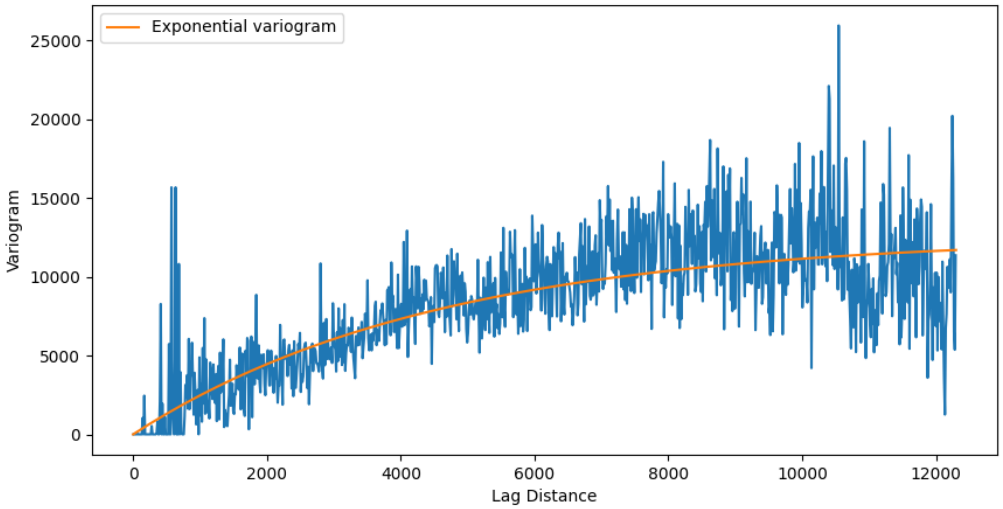
w_i : the weight factor at each sampling point i.

$X_p - X_i$: the distance between point p and i.

m : an order number for approximation.

The Geostatistical Analysis

Geostatistics introduces in a concept which considers random variation of geo-properties and represents them in terms of their spatial covariance.



For a spatially continuous geo-property, $Z(\boldsymbol{X})$, where $\boldsymbol{X}$ indicates the spatial coordinates (x, y), It has a deterministic mean value, $M(Z(\boldsymbol{X}))$. As a result, the geo-property can be expressed in a form of:

$$Z(\boldsymbol{X}) = M(Z(\boldsymbol{X})) + \Delta(\boldsymbol{X})$$

where Δ is called stochastic spatial residual – variation against the mean value, M.

Two Assumptions

I. the value, M, over a certain spatial region is the same as that within the whole area, $\boldsymbol{X} + h$,

h – all the distances of all the location $\boldsymbol{X}$ to all other points in the area. i.e.:

$$E\{ M(Z(\boldsymbol{X})) – M(Z(\boldsymbol{X}+h)) \} = 0$$

E – an error function stands for error or difference.

II. The variance the geo-property over the spatial area is auto-correlated to the distance, and the correlation can be evaluated using the semi-variance :

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

$\gamma(\boldsymbol{X}, h)$ – the semi-variance at distance, h, referring to the value at position, $\boldsymbol{X}$.

For certain number of collection of data (samples), an average semi-variance (also simply called the **semi-variance**) for all is also defined as:

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

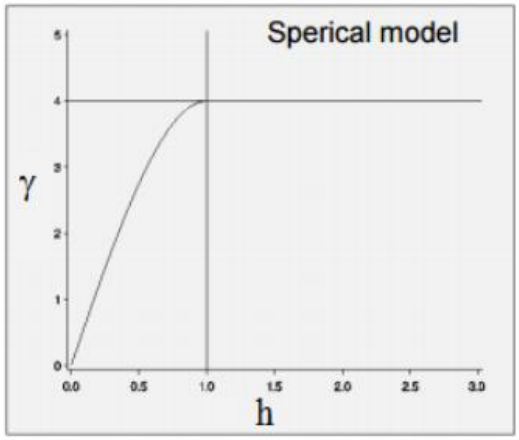
i – the individual sampling points

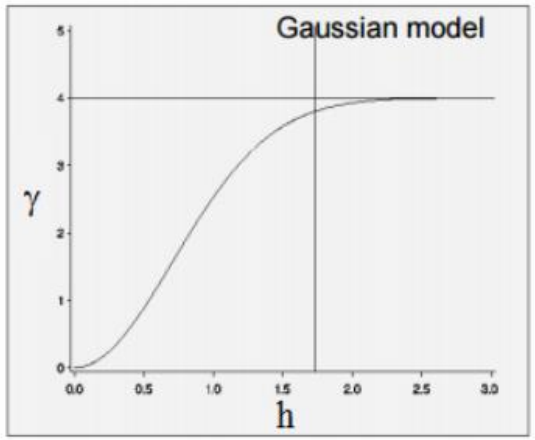
n – the total number of sampling points

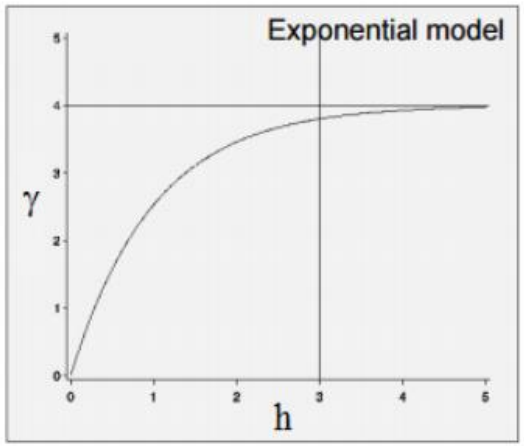
Now $\gamma(h)$ is not relevant to a specific position but for all points in the area.

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

$$\gamma(h) \approx f(h)$$





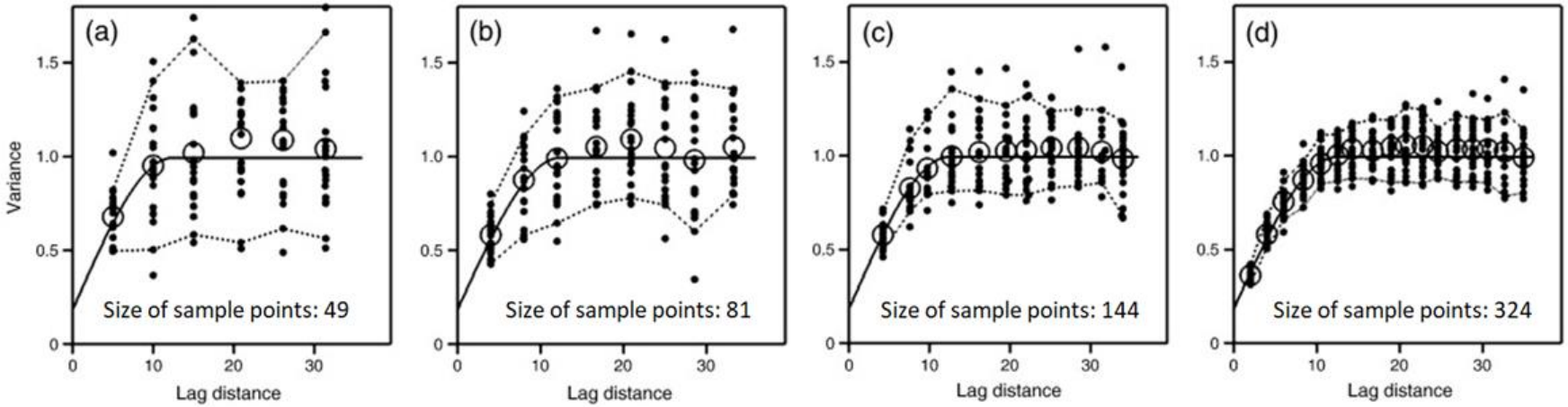


$$\gamma(h) = \begin{cases} c_0 \left[\frac{3}{2}\frac{h}{a_0} - \frac{1}{2}\left(\frac{h}{a_0}\right)^3\right], & \text{for } h \le a_0 \\ c_0, & \text{for } h > a_0 \end{cases}$$

$$\gamma(h) = c_0 \left[1 - \exp\left(-\frac{h^2}{a_0^2}\right)\right]$$

$$\gamma(h) = c_0 \left[1 - \exp\left(-\frac{h}{a_0}\right)\right]$$

The more the number of sample points the higher the accuracy of the variogram model representation.



Using the **semi-variance** function to represent the stochastic spatial residual value, $\Delta(\boldsymbol{X})$, we have:

$$Z(\boldsymbol{X}) = M\big(Z(\boldsymbol{X})\big) + \gamma(h)$$

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

The Geostatistical interpolation – Kriging

In term of the interpolation principle:

$$Z(\boldsymbol{X}_p) = \sum_{i=1}^{n} w_i \, Z(\boldsymbol{X}_i) \qquad \text{Eq. (1)}$$

and the Geostatistical theory:

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

We have :

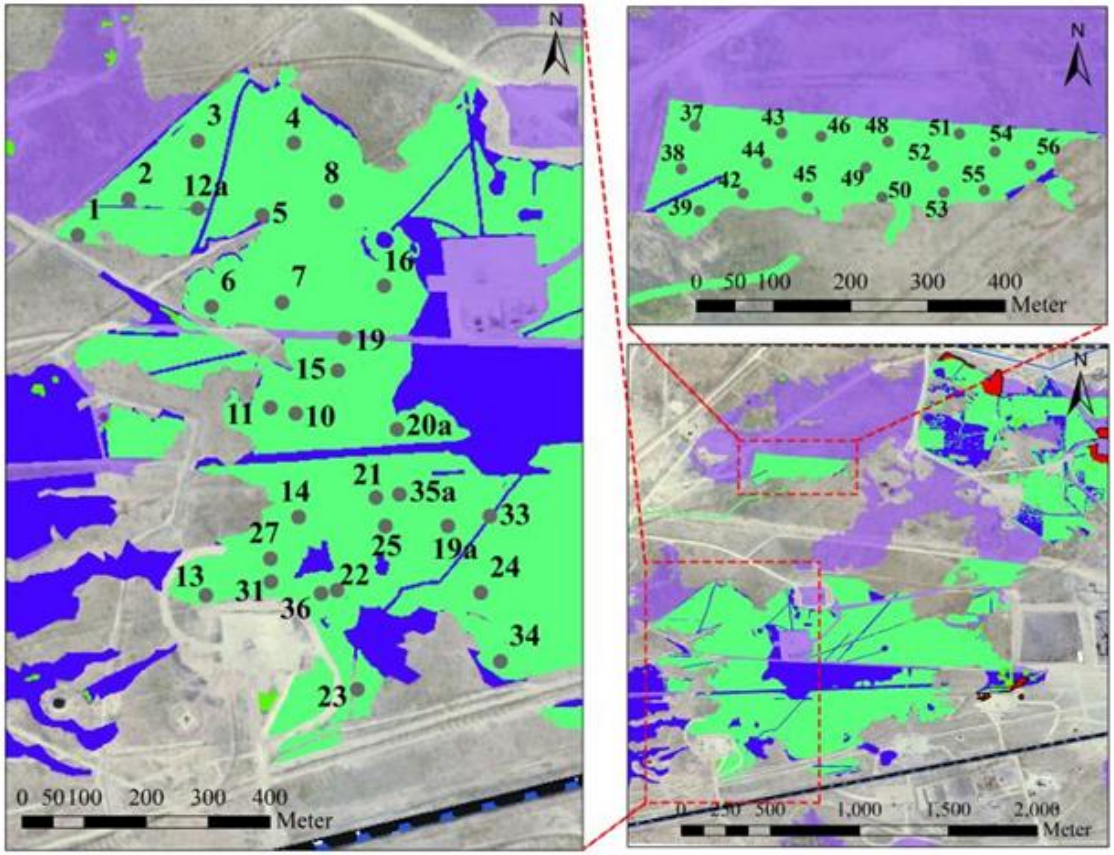
$$M\left(Z(\boldsymbol{X})\right) + \gamma\left(\boldsymbol{X}_p - \boldsymbol{X}_j\right) = \sum_{i=1}^{n} w_i \, Z(\boldsymbol{X}_i)$$

where, $j = 1,2, \cdots, n.$

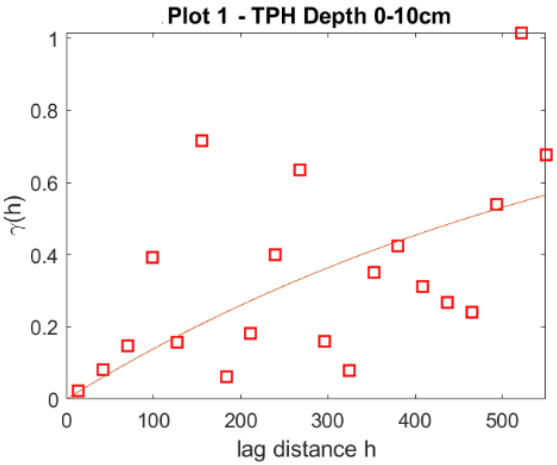
At last, the weighting coefficient, w_i, for each sampling points can be determined solving the produced equation system:

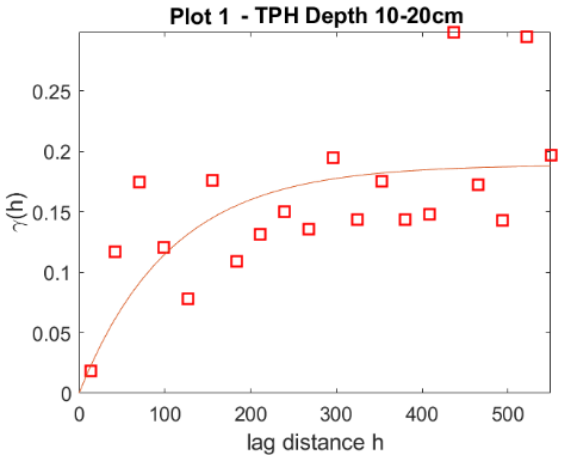
$$[\boldsymbol{A}] \begin{Bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{Bmatrix} = \begin{Bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{Bmatrix}$$

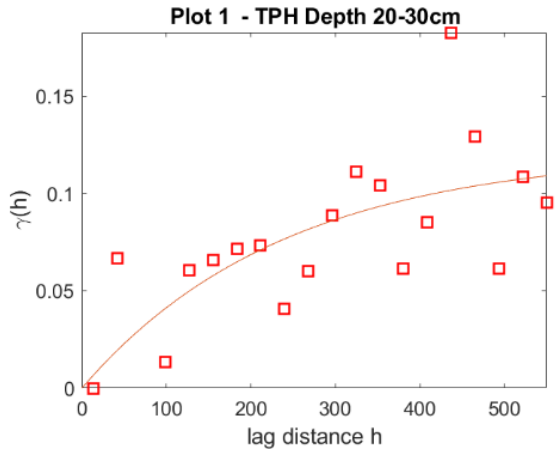
2. Modelling and prediction two for plots

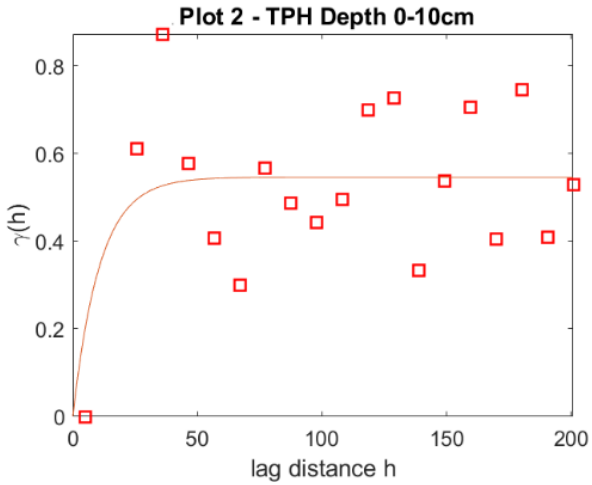


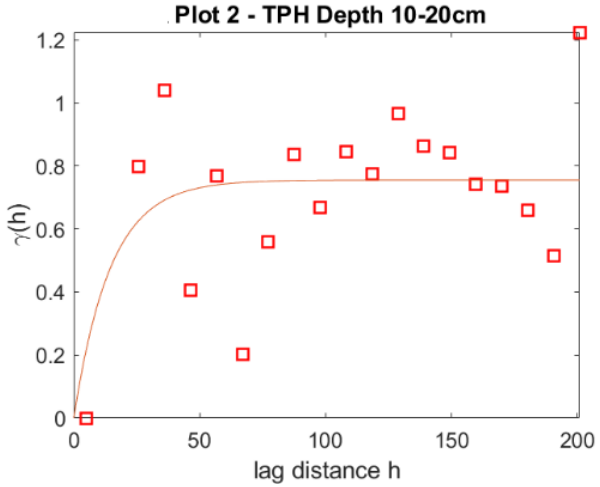
The semi-variance of TPH



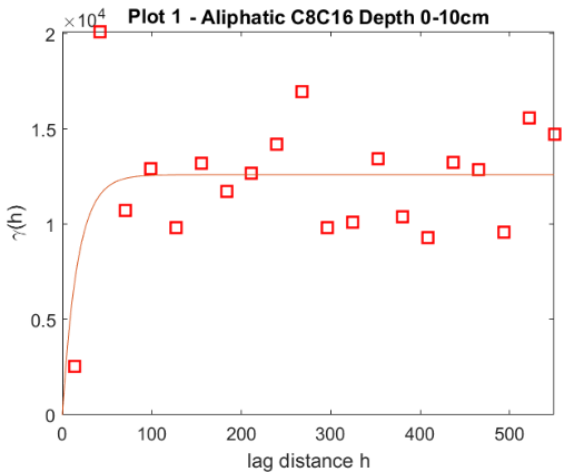


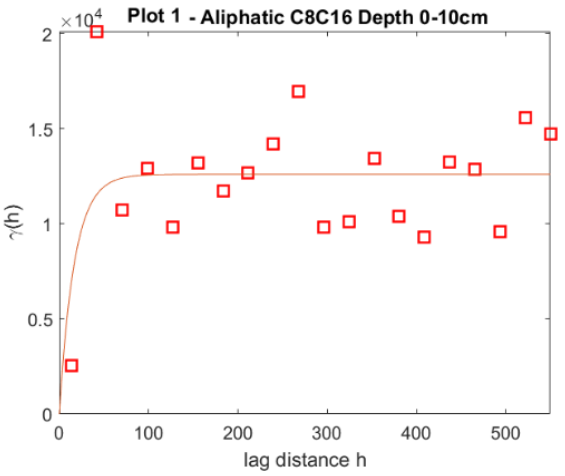


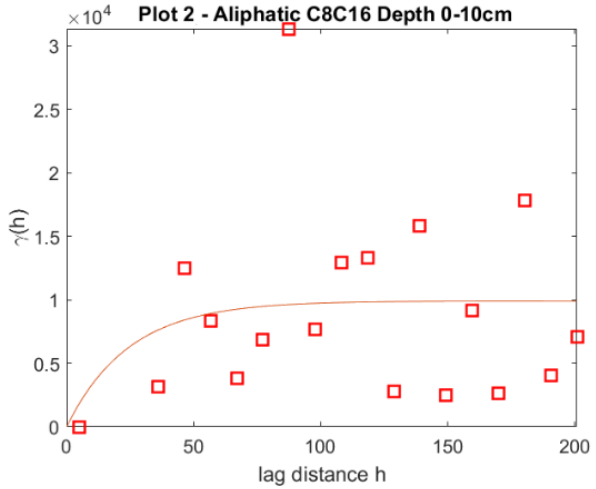


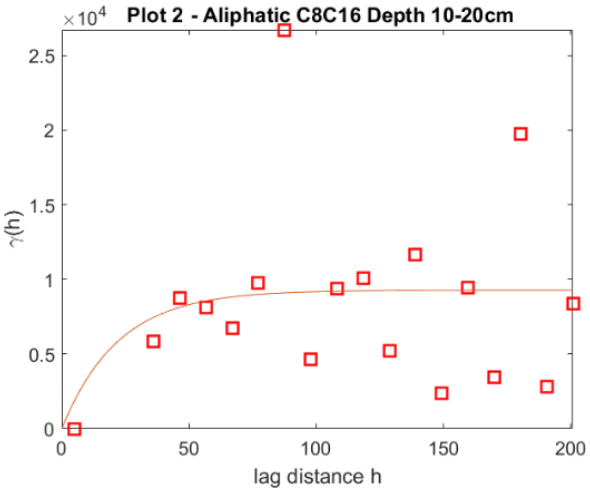


The semi-variance of Aliphatic C8-C16

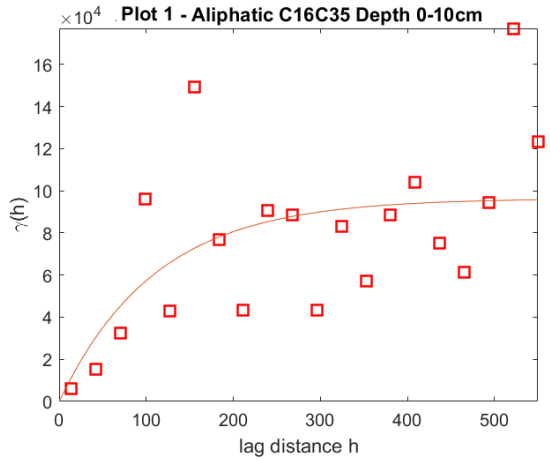


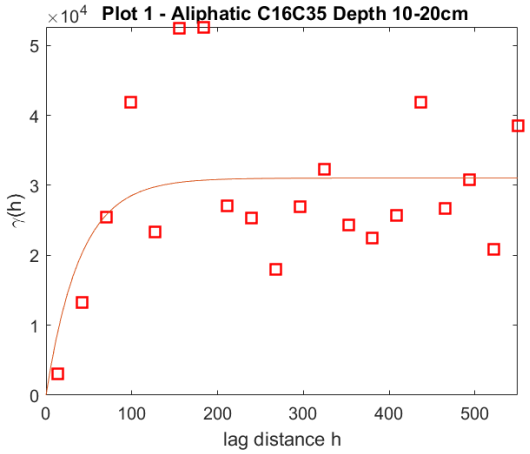


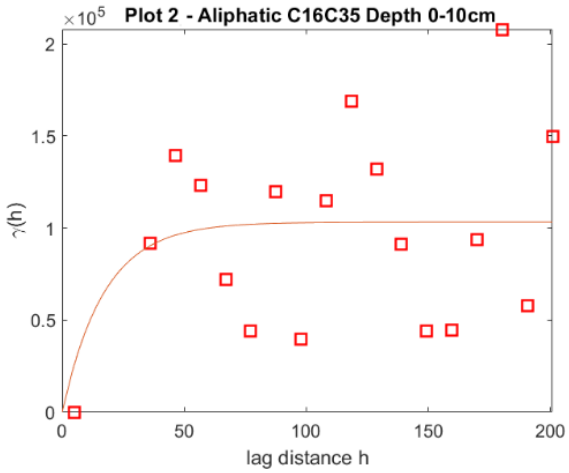


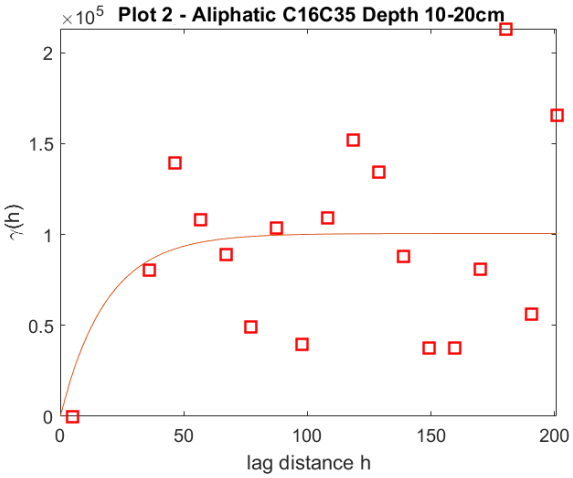


The semi-variance of Aliphatic C16-C35

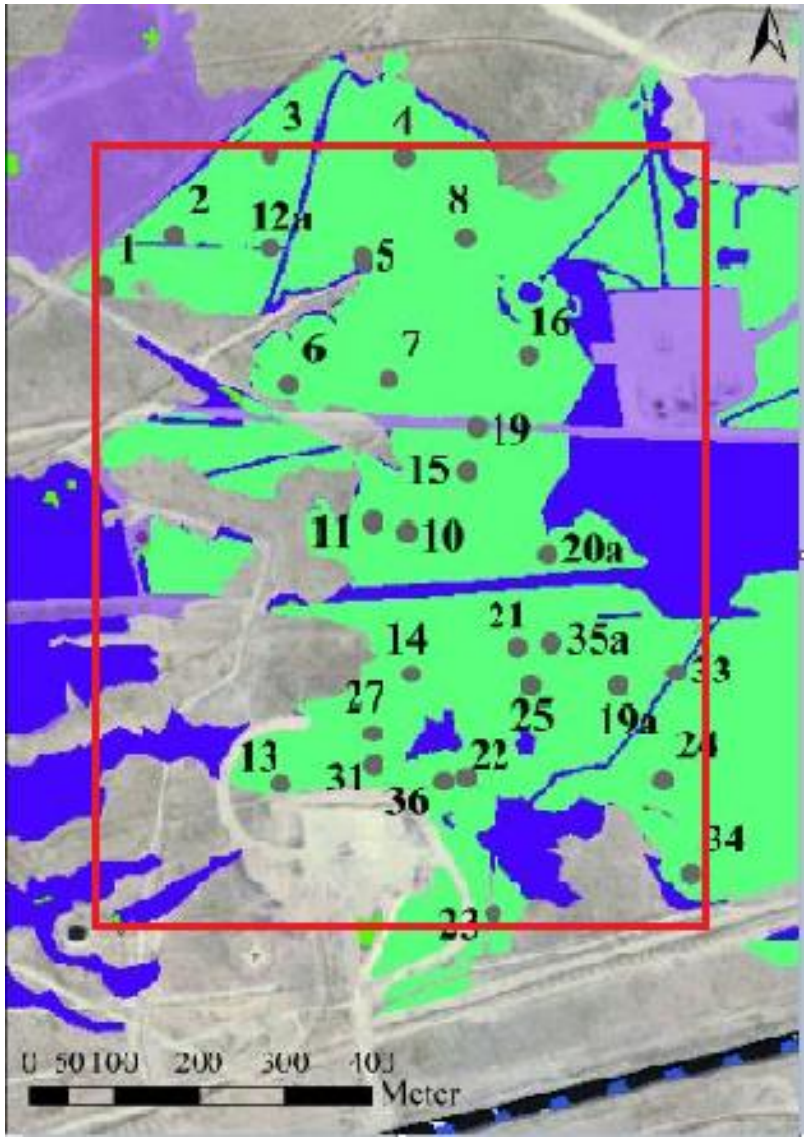


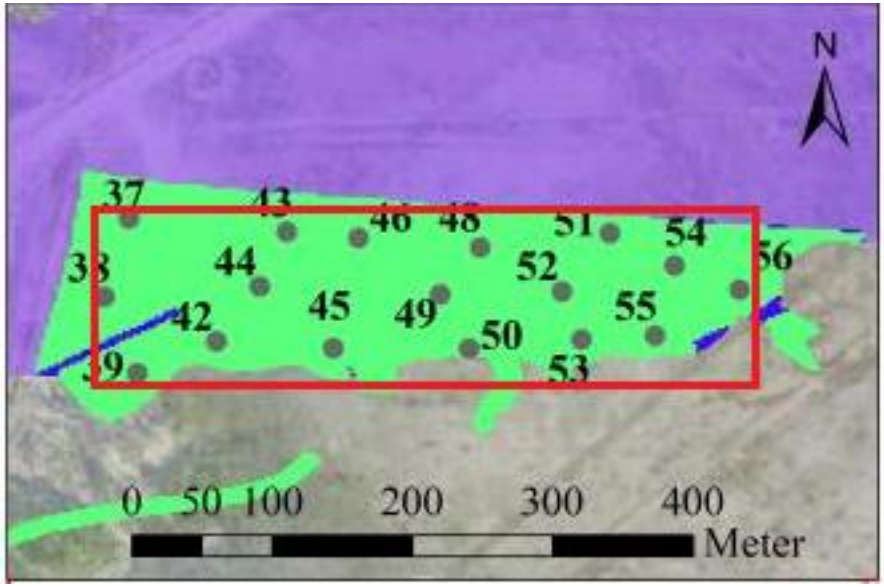






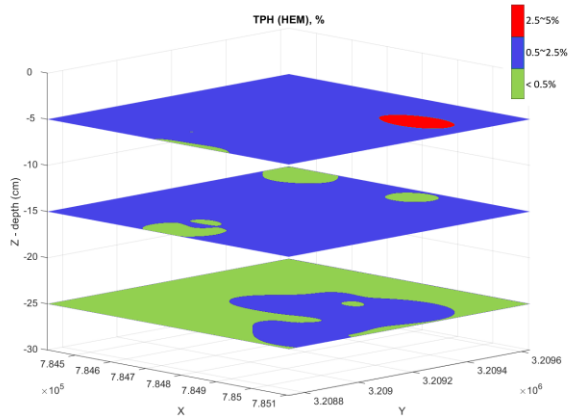
The predicted areas in the two plots

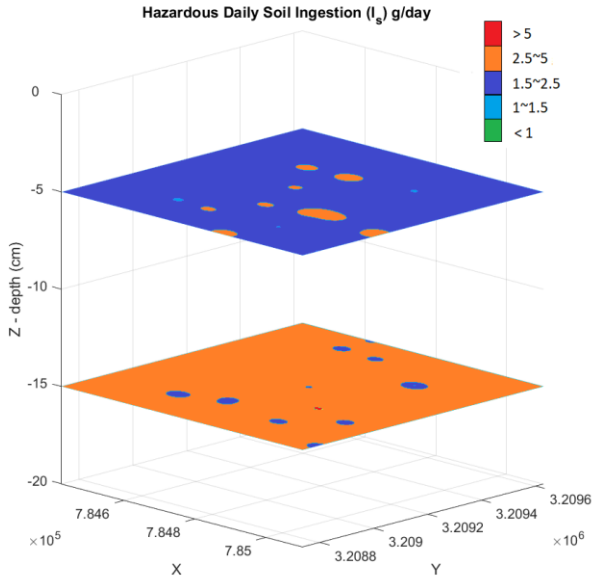




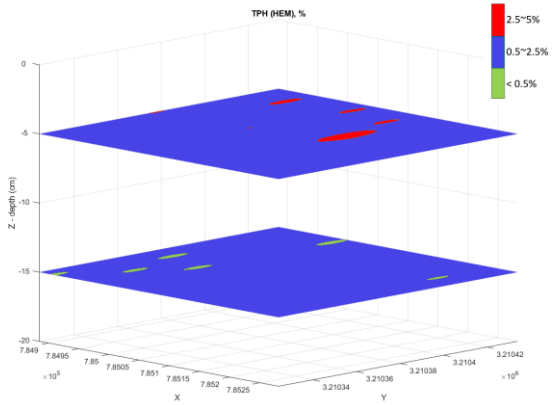
The Modelling results

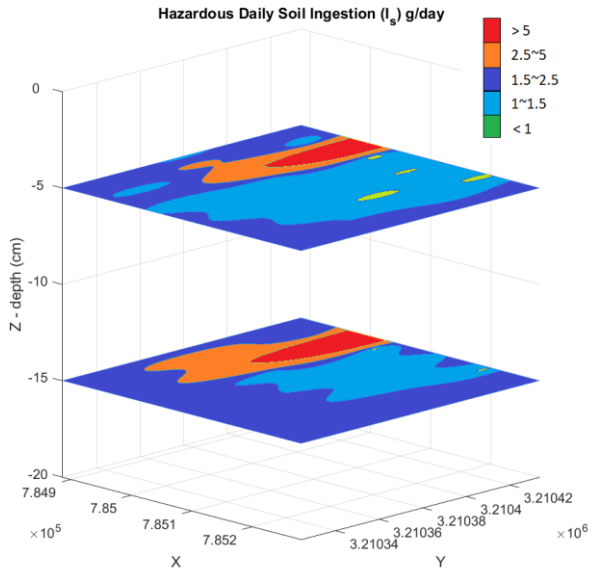
Plot 1



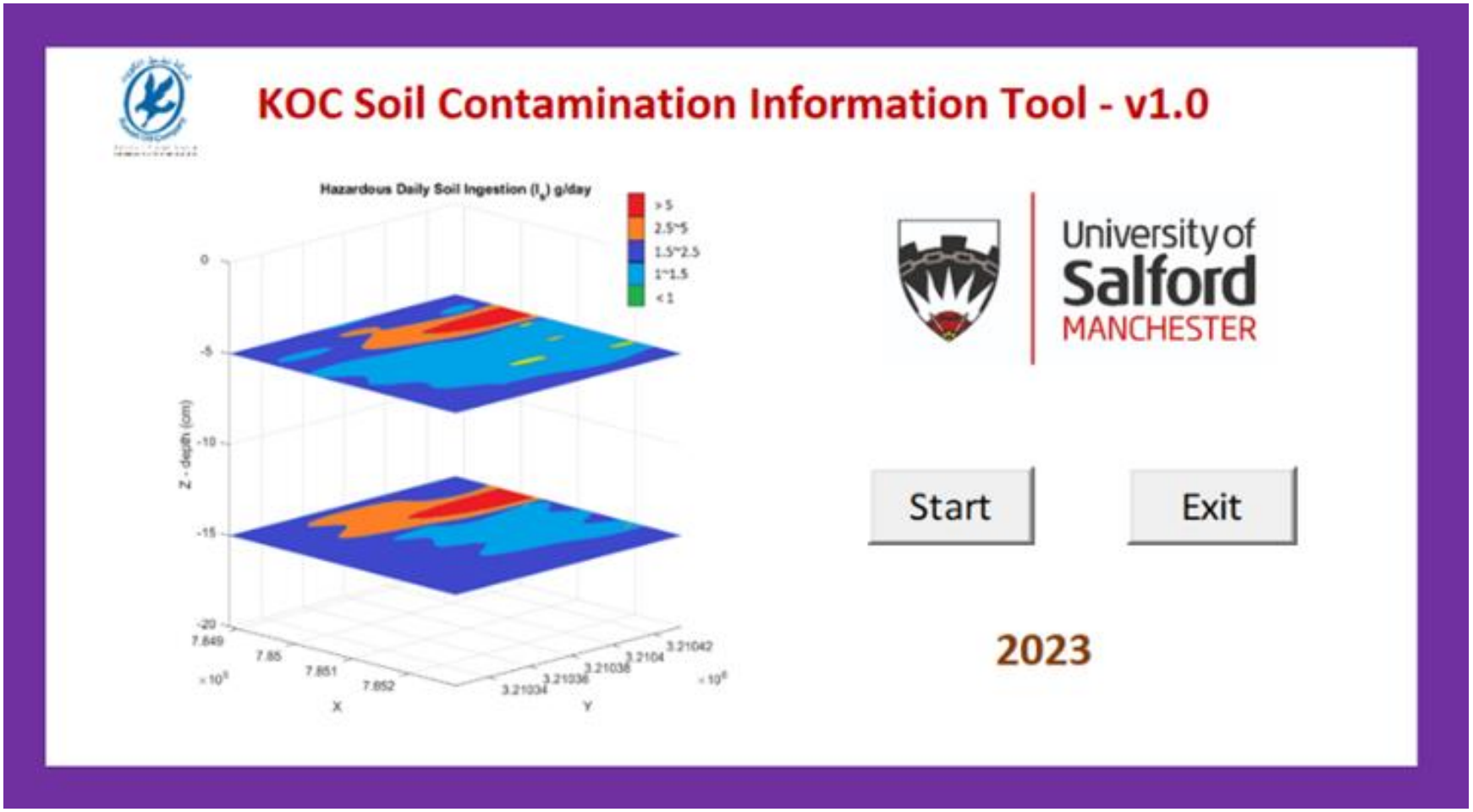


Plot 2



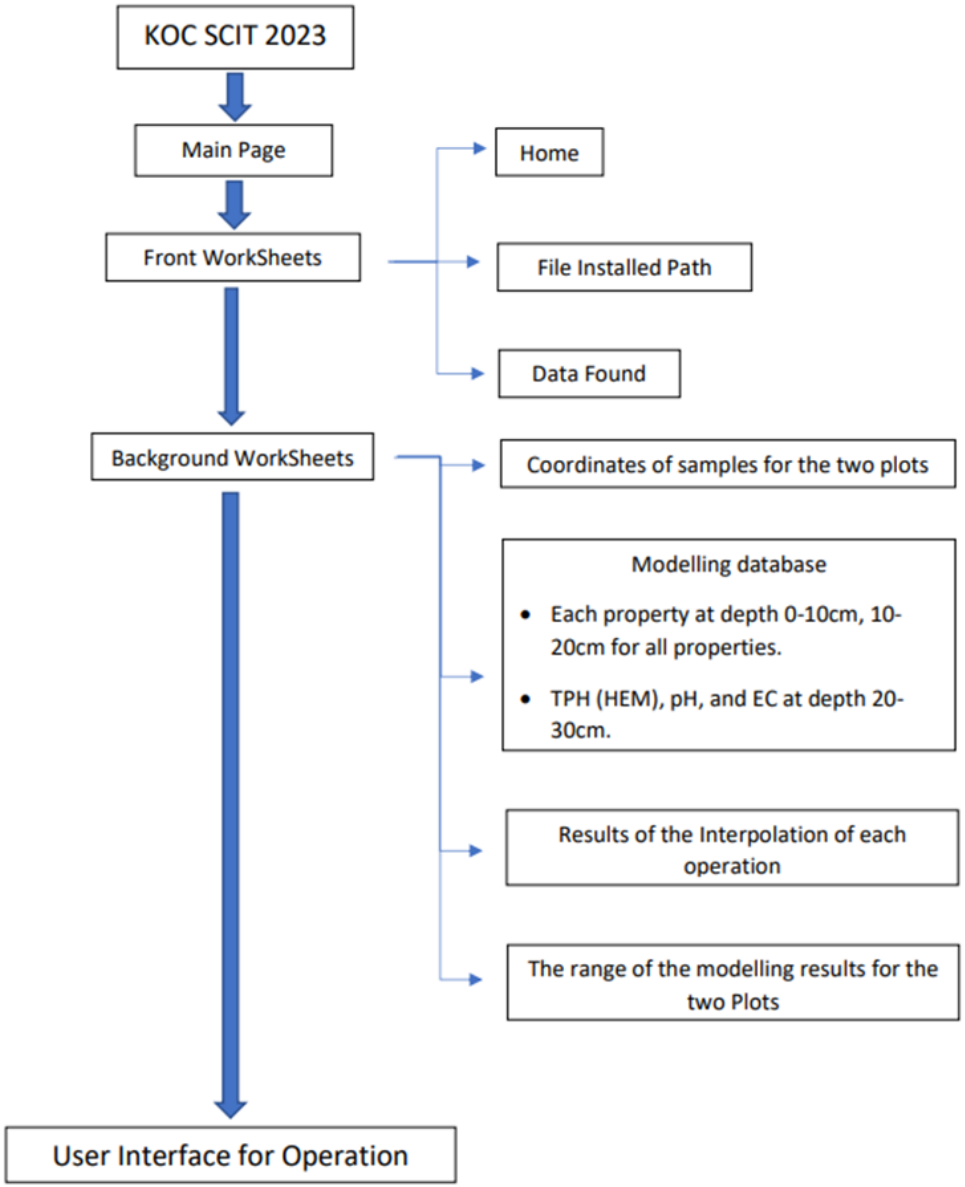


3. Software development

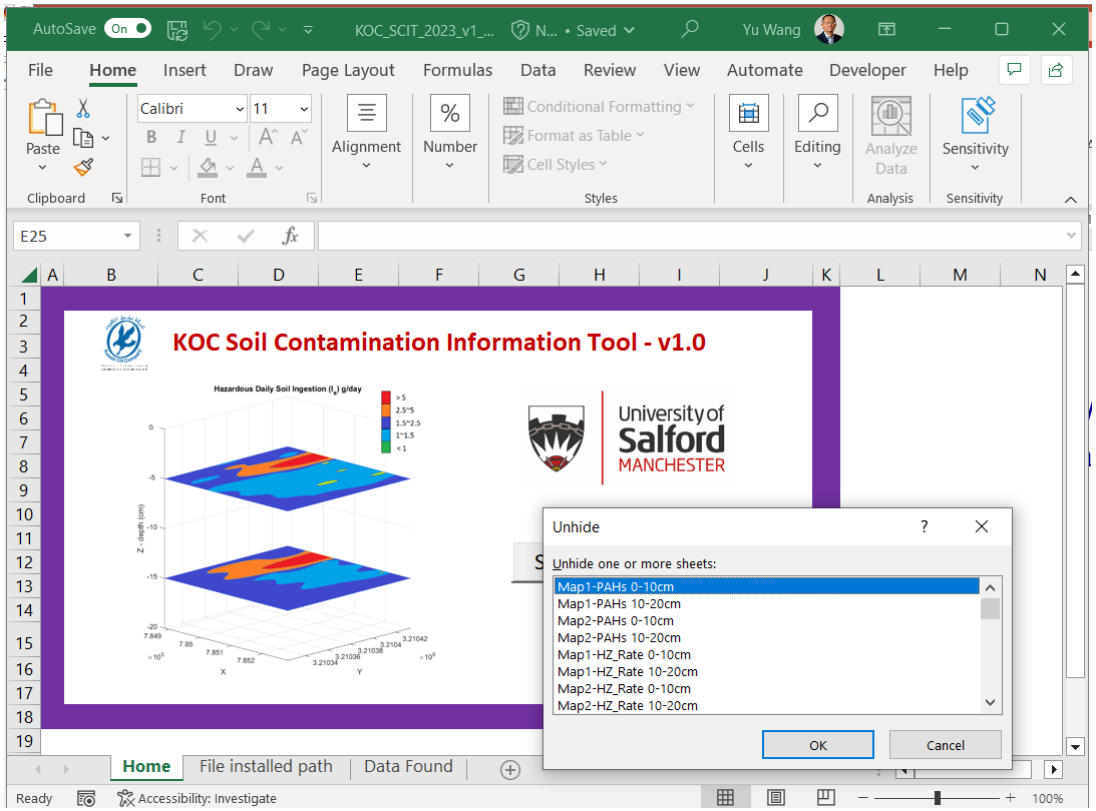


- KOC Soil Contamination Information Tool (KOC-SCIT) is a database software for two specific oil lake areas in Kuwait.
- The database contains all the analysed soil contamination properties for two rectangular areas in the two plots.
- The data extraction uses both curve fitting and linear interpolation methods within the database.
- KOC-SCIT runs on the MS Excel. A technician with basic knowledge and skills using MS Excel can operate the software.

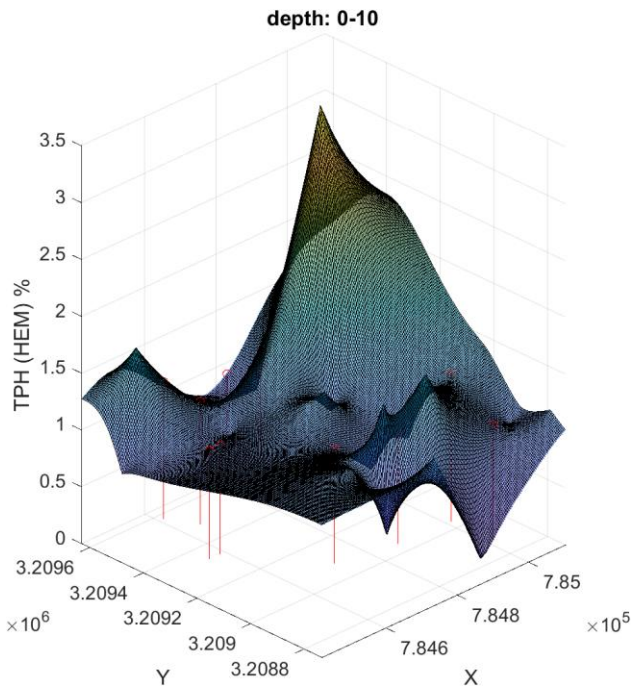
The software structure

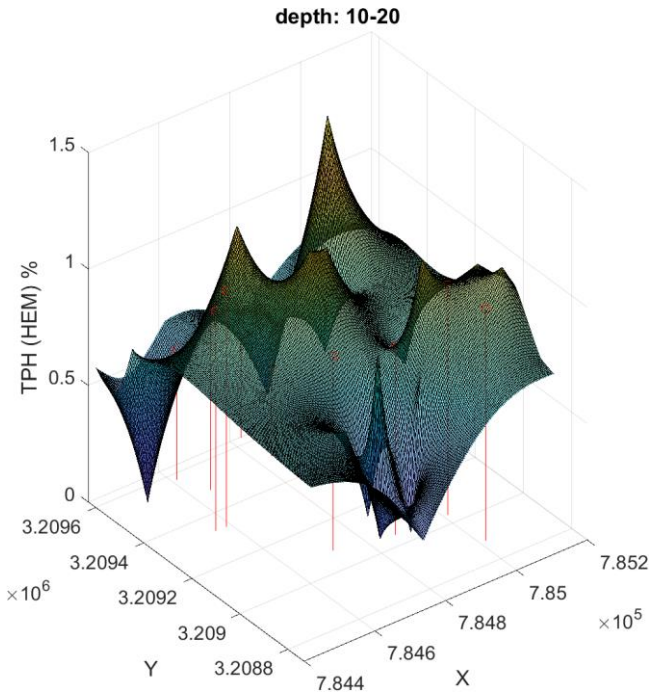


KOC-SCIT has three operating worksheet in front, they are **Home**, **File Installed Path**, **Data Found**, and numbers of data worksheets in background.

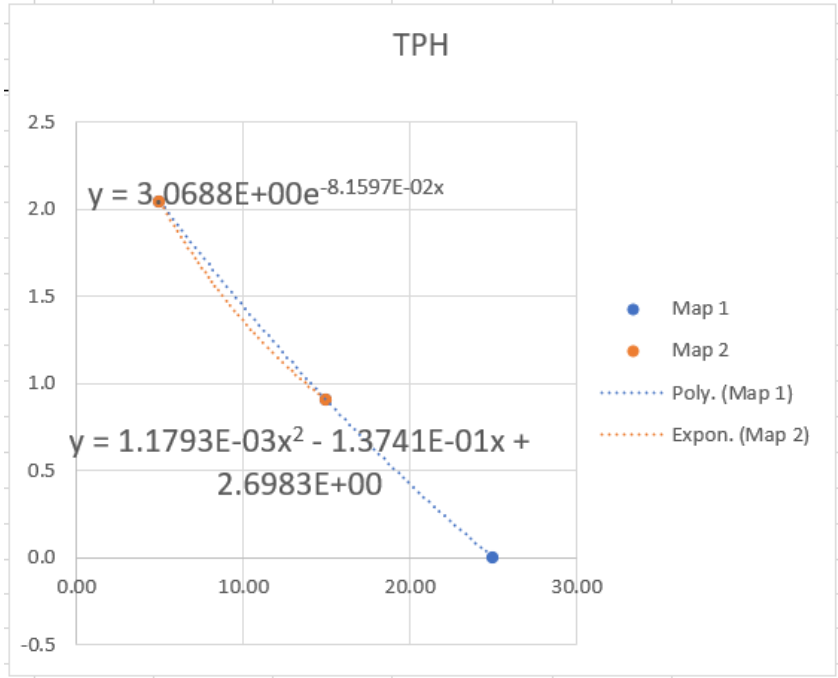


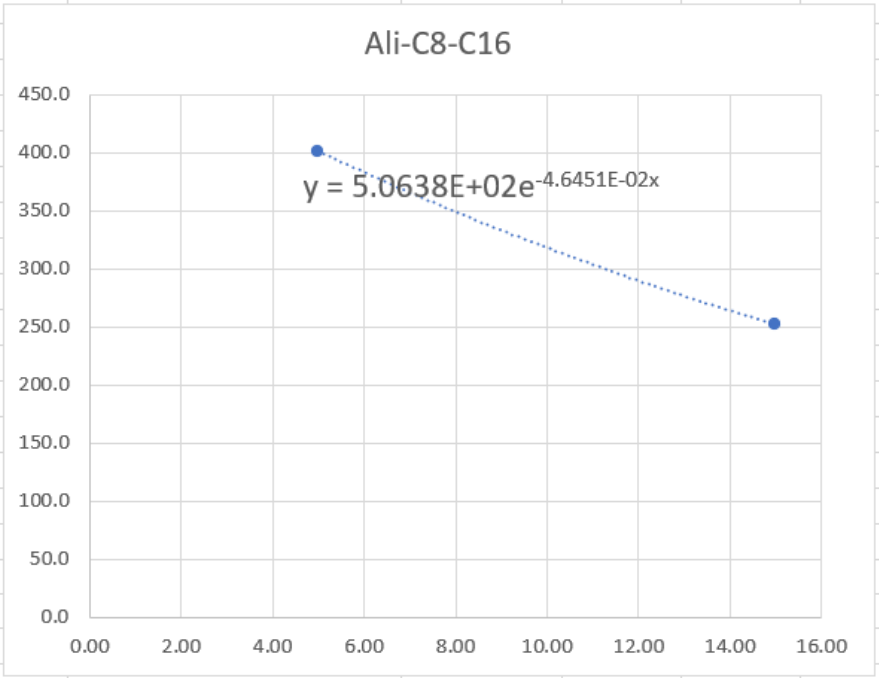
Each of the data worksheet contains the geo-statistical modelling results computed using an in-house Kriging solver at 40,401 locations of a mesh for a sampling layer.



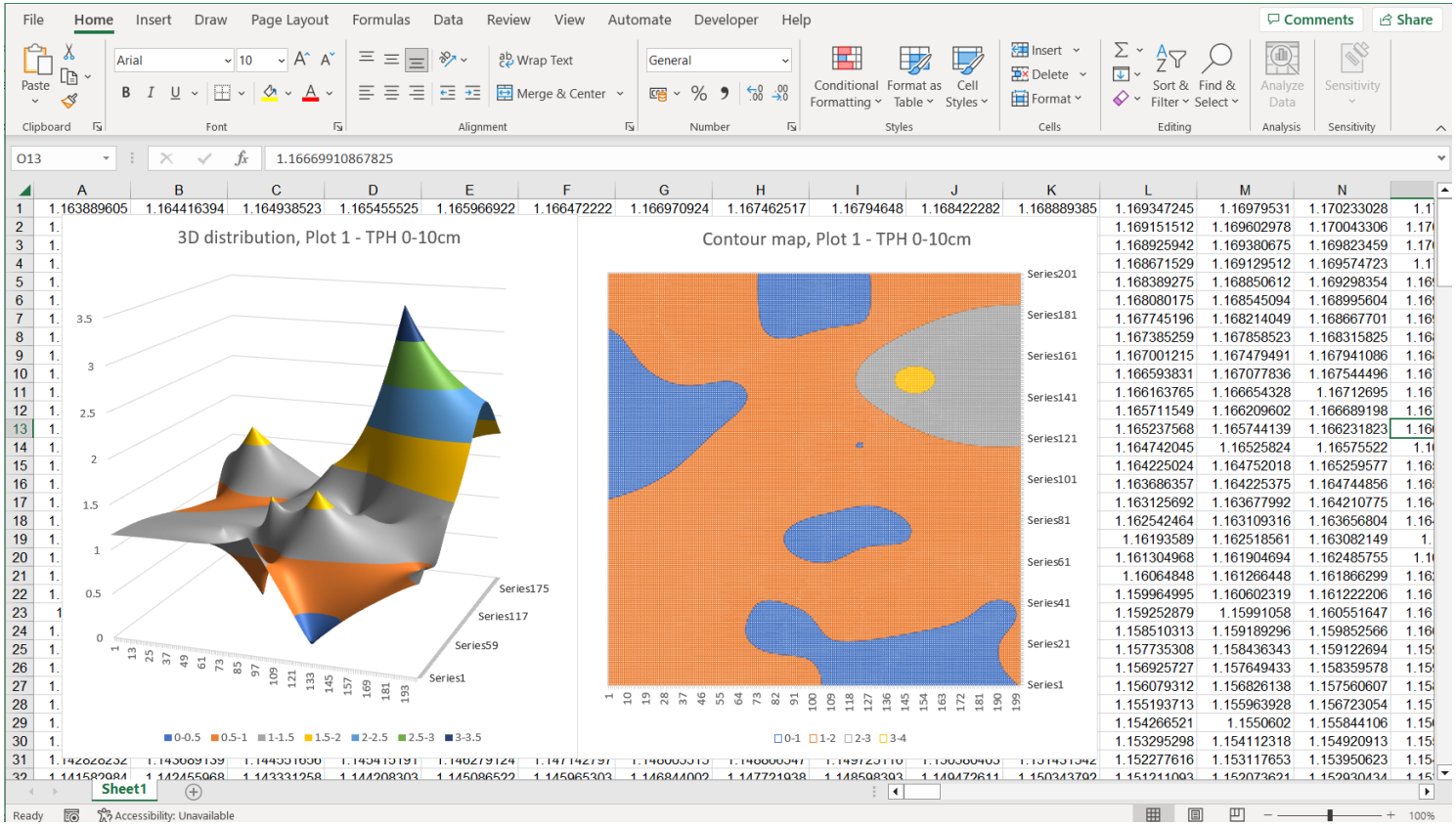


In depth direction, KOC-SCIT adopts curve fitting to predict local vertical contamination profile.

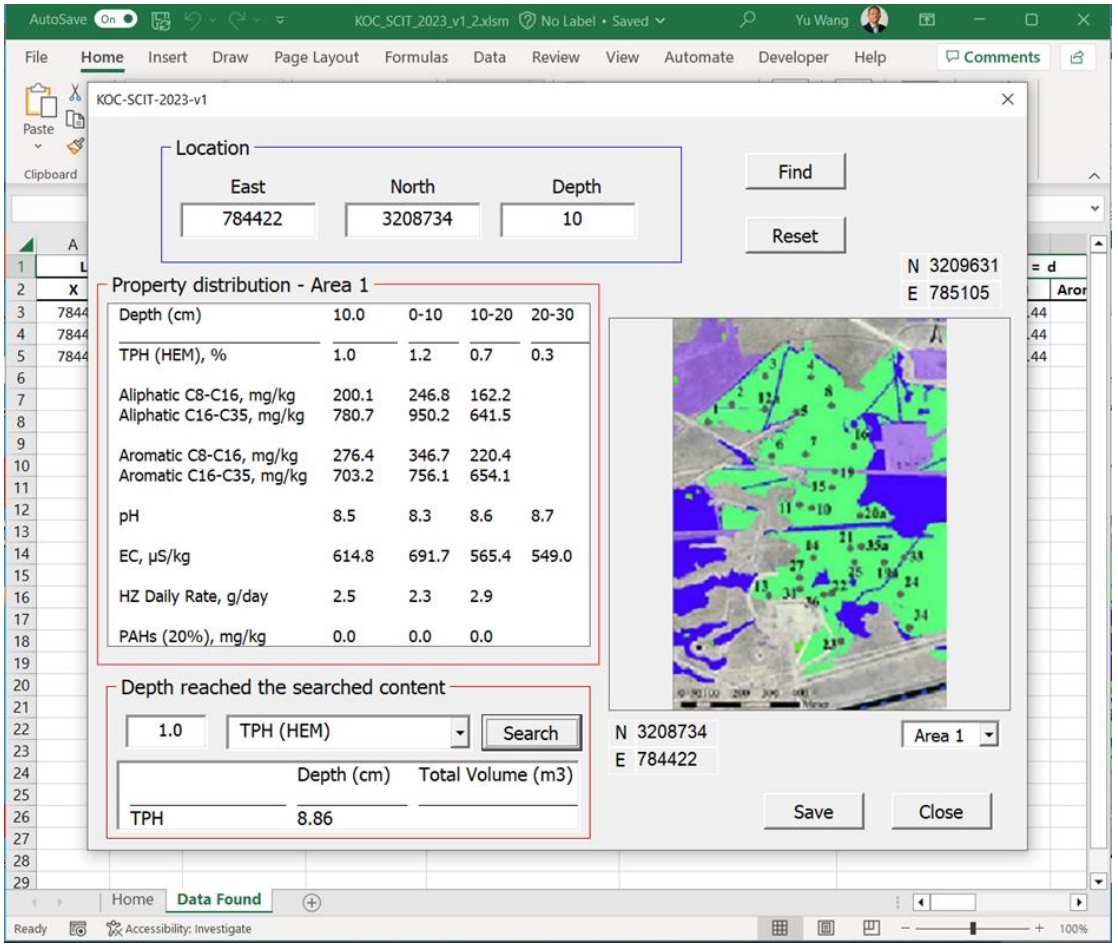




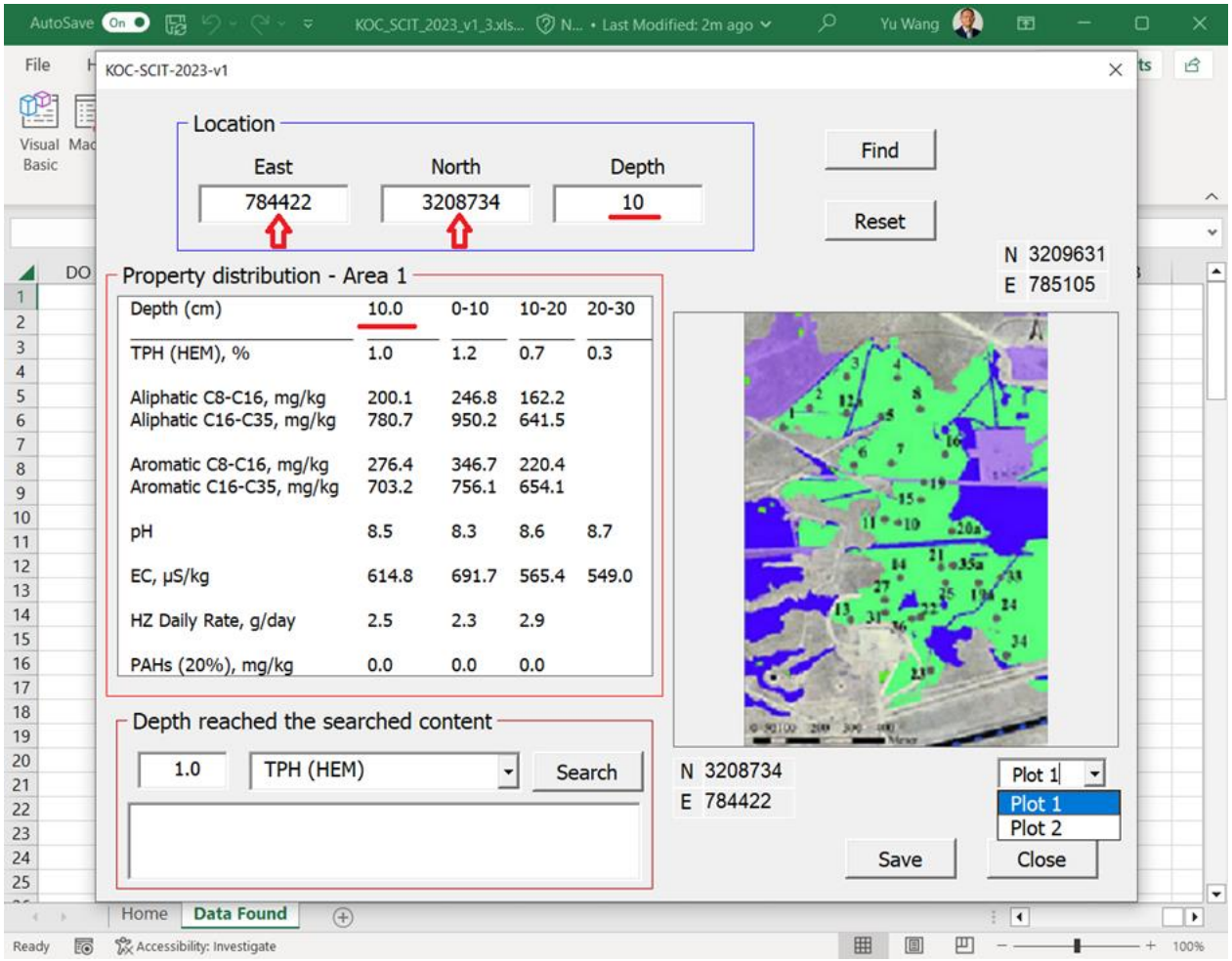
Each modelling task result has been recorded in mesh map format



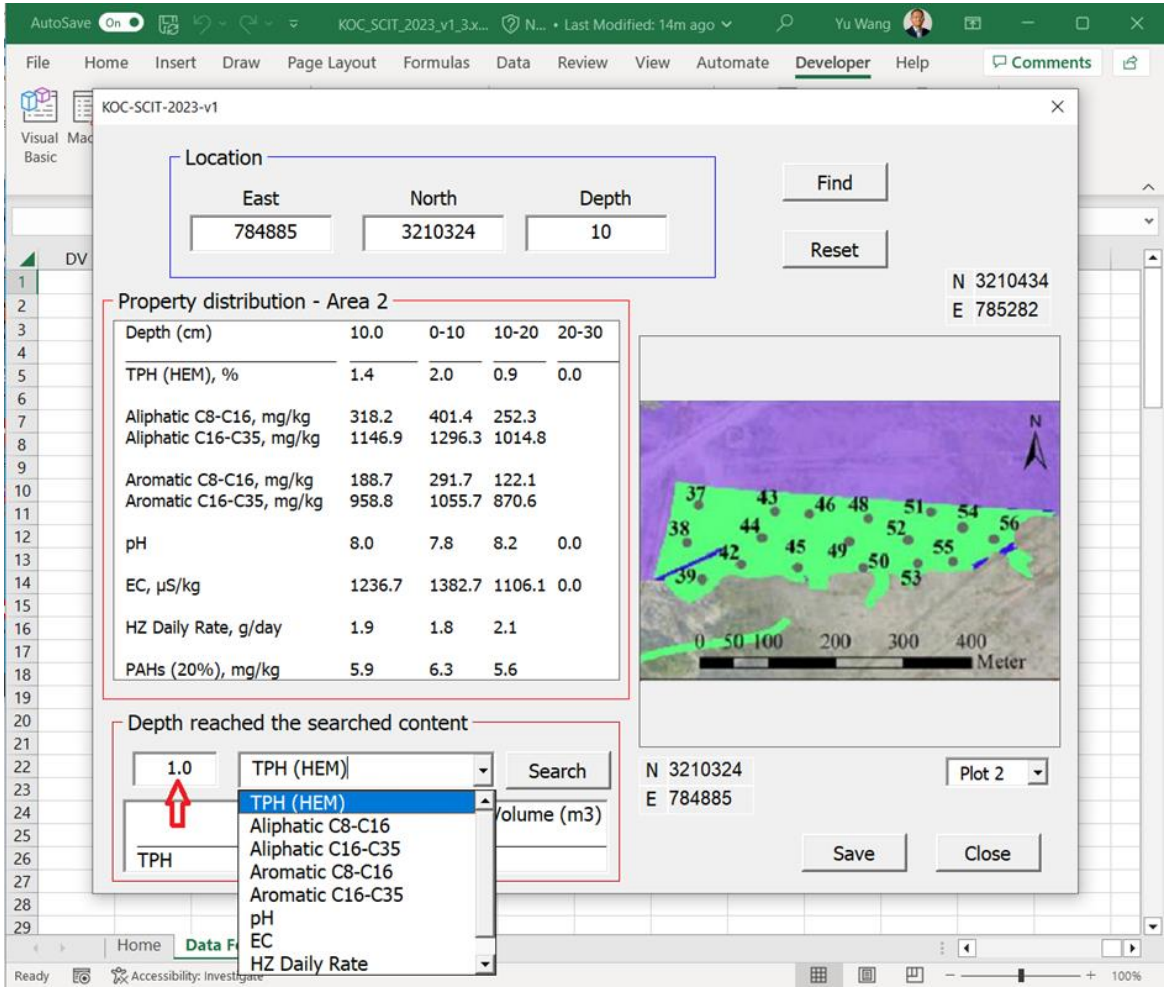
A user interface has been developed to facility to extract data by the request and definition of users.



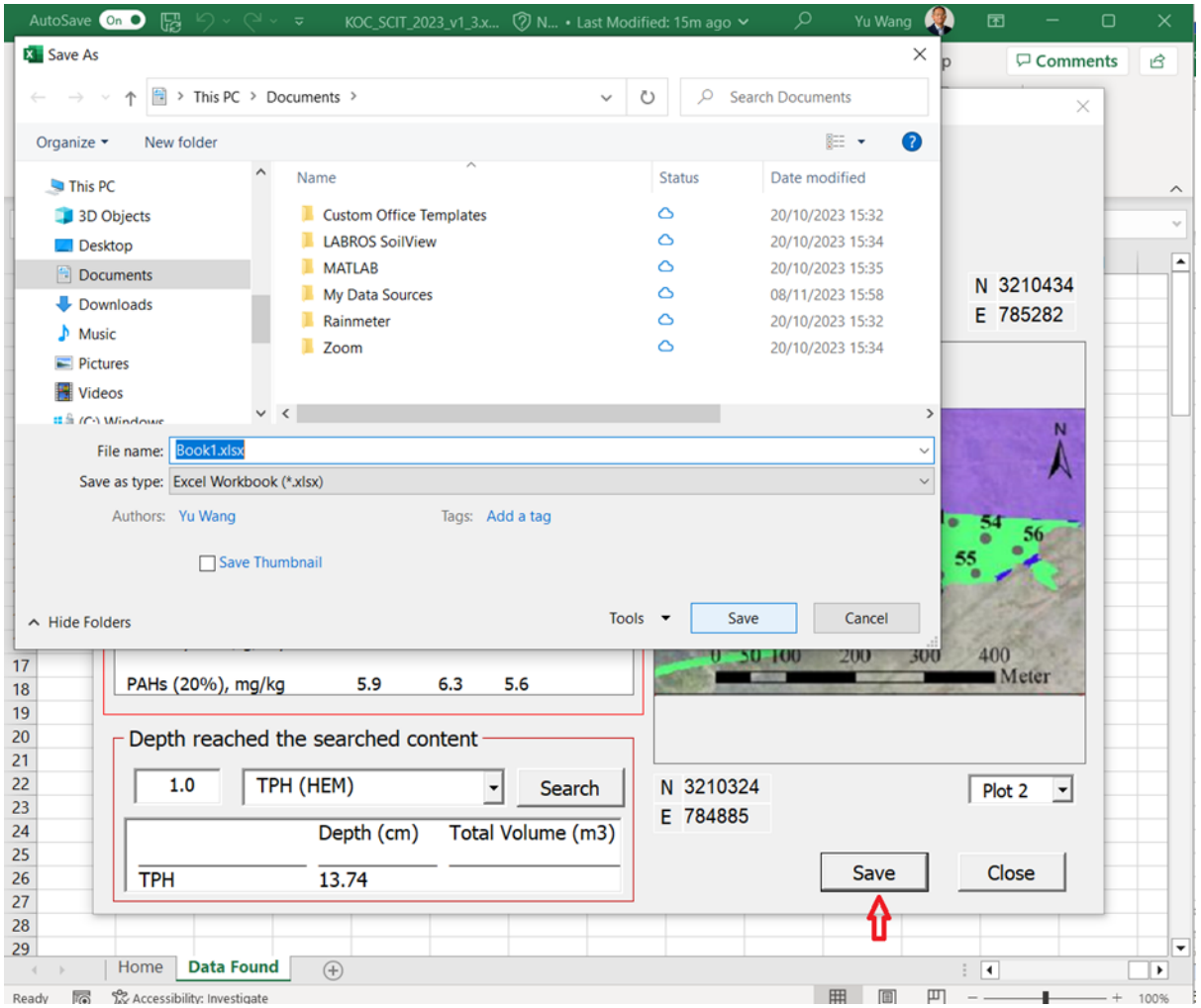
All local contaminating information is listed out in a single table



Find the local depth given the contaminating degree of properties

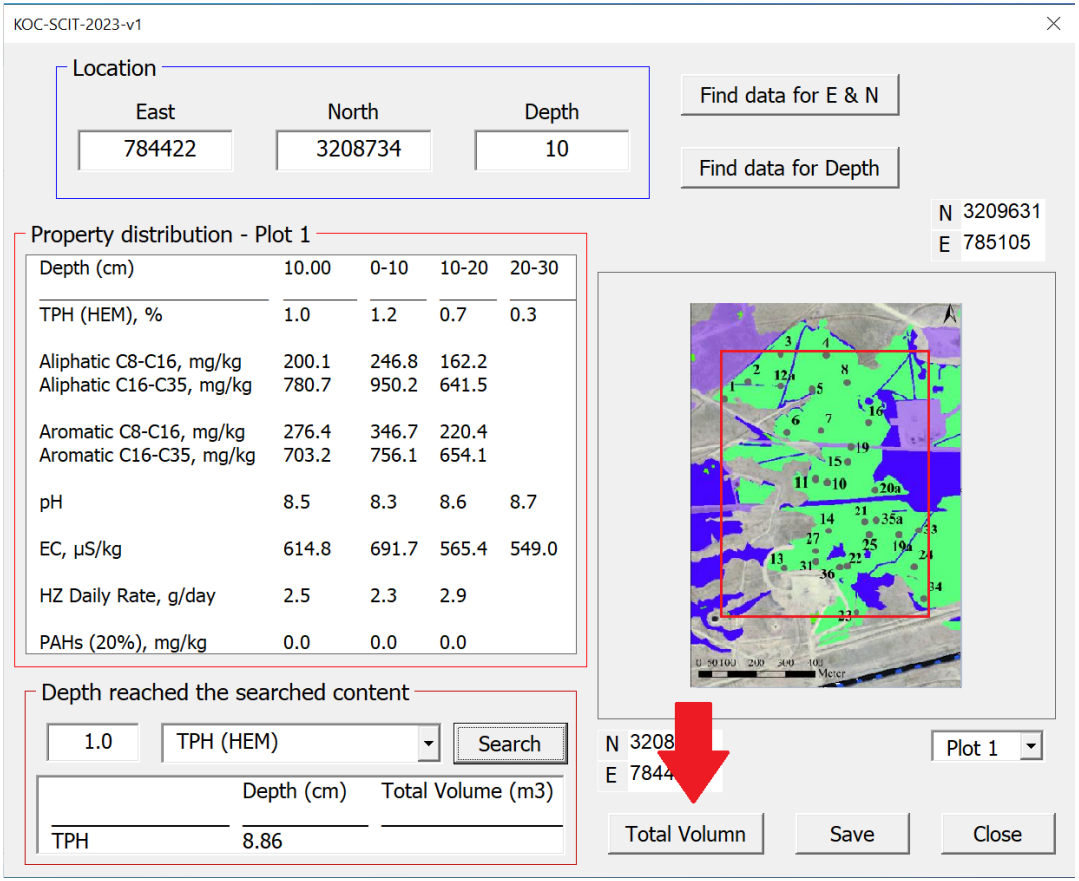


Record and output all the search history



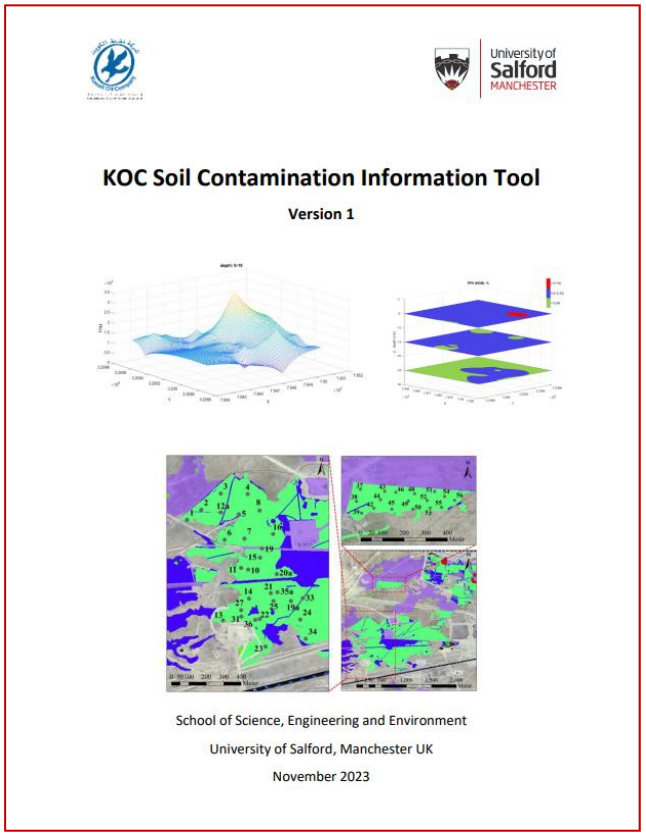
Find the total contaminated volume given a criterion

– function has not finished yet.



4. Software installation and operation

Information is provided by the software document



KOC Soil Contamination Information Tool

Version 1

School of Science, Engineering and Environment

University of Salford, Manchester UK

November 2023