Evaluating the Use of Statistical and GIS software as Modelling Tools for Analysis of Soil Samples

Morga Mateusz (1705592)

GIS & Data Analysis, University of Wales Trinity Saint David

1.0 Introduction

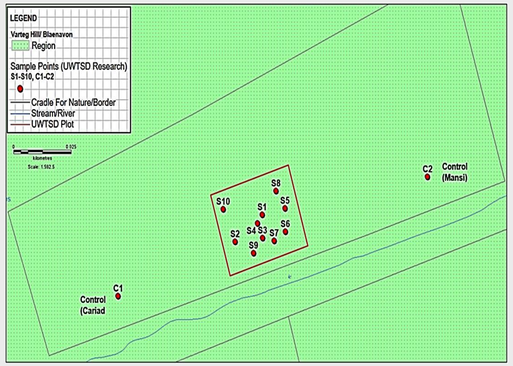
The subject of the research is an analysis of the soil samples collected from a post-mining site located in Varteg Hill, Blaenavon (South Wales). Due to the increased tree mortality occurring in the reclaimed site, ten soil samples were collected from the UWTSD research plot in 2019 to measure the concentration of 15 heavy metal in soil. The aim was to locate toxic hotspots by comparing high tree mortality and high vitality areas and test their correlation with the plant's mortality.

Once the samples were professionally analysed by NRW Lab (Swansea), results were tabulated in Microsoft Excel and statistically tested, including graphical presentation. Also, GIS software packages like Surfer 17.0 and MapInfo 16.0 were used to present the sampling point and the spread of inorganics visually as well as for further investigation and comparison purposes.

2.0 Methodology

Firstly the sapling points (geo-referenced locations) were highlighted using MapInfo software with inclusion of OS Topography map (DigiMap) (Fig.1) used as background map layer for all MapInfo data analysis (UWTSD research site 2019).

Figure1 UWTSD Site Sampling Points Locations (S1-S10) Including Control Sample Points (C1, C2) (Cradle for Nature Project) (OS Vector Map, Map Info Pro 16.0 Maps).



Next, a GIS software Surfer 17.0 (Golden Software) was used to visually present the obtained results in the form of contour maps (Fig. 2, 3). Additionally, a visual assessment of spread patterns was also undertaken in order to pinpoint areas of multiple heavy metal concentrations, causing a toxic effect on plants.

Then, MapInfo Pro 16.0 (also GIS) software package was utilised to perform a more in-depth analysis of the spread patterns (including a visual presentation) in contaminated soil within the UWTSD perimeter. Data querying and several thematic mapping (including heatmaps) were implemented (Fig. 4,5, 6,7).

Lastly, data were analysed with the use of descriptive and mathematical statistics provided by Microsoft Excel (ToolPak) package. The correlation coefficient was tested, and results were then graphed using cluster column charts and scatter graphs (including trend lines and R^2 values) (Fig. 8, 9).

Additionally, the obtained results were compared with initial research findings as well as with other academic literature. Both GIS software packages were commented on regarding their usability and application.

3.0 Results Analysis

3.0.1 Surfer 17.0 (Visual Assessment)

The two example contour maps show the spread of Manganese (Mn) (Fig. 2) and Cadmium (Cd) (Fig. 3) contamination in the UWTSD plot. The visual analysis shows that contamination hotspots occur in the middle and upper sections (the highest tree mortality was also observed in mid-upper zones). For example, Cd is found to be very toxic to plants) (Saraswat and Rai, 2011).
Contour maps (Fig.2, 3) obtained through Surfer software allowed for indicating spread patterns of potentially toxic elements in soil (e.g., Cd, As) and help to locate areas of several condensed heavy metal loadings where synergistic reactions between elements are more frequent (Desai, 2013).

Figure2 Manganese (Mn) Contour Map (Surfer) (UWTSD Plot).

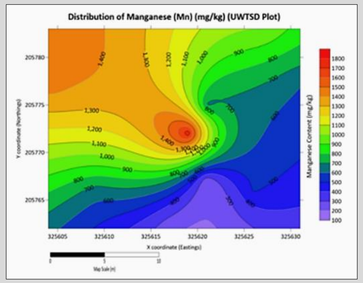
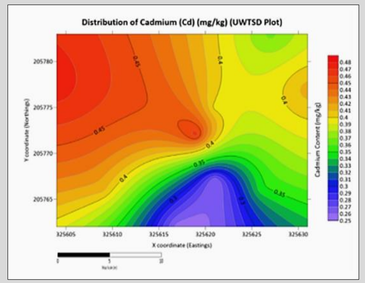


Figure 3 Cadmium (Cd) Contour Map (Surfer) (UWTSD Plot).



3.0.2 MapInfo Pro 16.0 (In-Depth Analysis)

MapInfo, apart of being used for assigning geo-referenced points to map locations, it can be used for analysing of more-complex data set. Figure 4 shows results obtained through data querying (SQL Select) function which shows the As hotspots (As>10 mg/kg) located in the mid-sections (S3, S4) (In line with other research findings); As can be detrimental to trees in concentrations above 10mg/kg(Desai, 2013).
Figure 5 shows data being analysed with Thematic mapping (Bar chart) options. The result suggests that there are high concentrations of toxic elements accumulating within S5 (i.e. Lithium, Copper and Chromium) which according to literature favours the occurrence of synergies between metals further damaging trees (Medda and Mondal, 2017). The graphical visualisation of data offered by thematic mapping was beneficial when analysing spatial distributions of multiple heavy metal loadings in contaminated soil.

Figure 4 Data Querying Showing As>10mg/kg in S3 and S4.

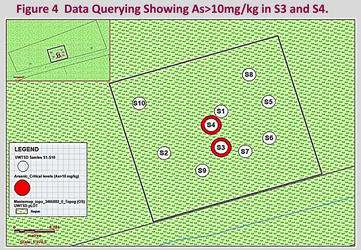
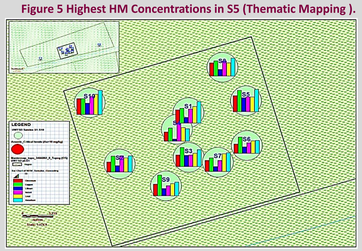


Figure 5 Highest HM Concentrations in S5 (Thematic Mapping).



3.0.3 MapInfo Pro 16.0 (Heat Maps)

Figure 6 Heatmap Showing Highest As Loadings in S3, S4 (Middle).

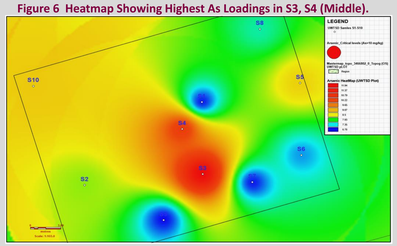
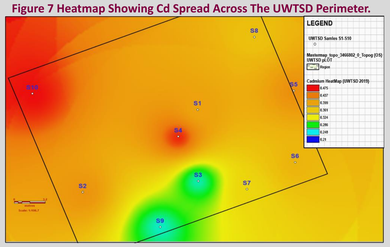


Figure 7 Heatmap Showing Cd Spread Across The UWTSD Perimeter.



3.0.4 Microsoft Excel Analysis (Statistical ToolPak)

The Excel software allows for both visual and statistical testing of the quantitative data collected during the scientific research (e.g. soil sampling and its inorganics content levels). Figure 8 Cluster column chart is comparing two extreme sample data-sets (S3-highest tree mortality and S10-highest survival rates). The visual assessment of the graph reveals some irregularities in individual samples regarding the toxic contents. The suspected differences in contaminant loadings in S3 (compared to S10) were statistically analysed and resulted 'not significant' (Varetg Hill Dissertation 2020).

Figure 8 Comparison of S3 (Highest Mortality) and S10 (Highest Vitality) Samples.

HM (mg/kg)	S3	S10
Al	15280	16290
As	11.94	8.85
Ba	270	192
Cd	0.259	0.475
Cr	24.2	32.6
Cu	38.6	32.7
Fe	25520	21050
Li	23.56	24.00
Hg	0.12	0.057
Mn	218	1442
Ni	24.2	37.0
Pb	27.9	41.5
Sr	118	105
V	34.4	49.8
Zn	80.6	84.3

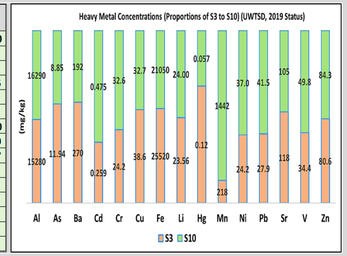
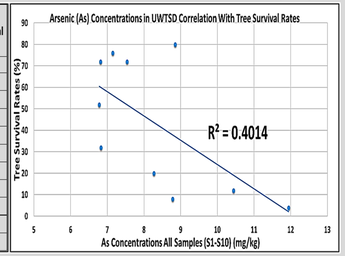


Figure 9 shows a correlation coefficient tested between As concentrations in the UWTSD plot related to tree survival rate. Data were plotted on the Scatter Diagram, including Trend line and R^2 value. The graph revealed a negative correlation (with decreasing As concentrations, the tree survival rate increases). The ToolPak from Microsoft has helped to determine one of the key findings showing that As is affecting the planted trees and it can have moderate (R^2>0.4) (Martin and Bateson 1993) impact on vitality.

Figure 9 Correlation of As (S1-S10) and Tree Survival Rates at UWTSD Site (2019).

Sample	Heavy Metal	Tree Survival Rate
Nr.	As (mg/kg)	(%)
S1	6.83	32
S2	8.27	20
S3	11.94	4
S4	10.44	12
S5	8.78	8
S6	7.15	76
S7	6.82	72
S8	7.54	72
S9	6.78	52
S10	8.85	80



Based on the above examples, it must be stated that both Microsoft Excel and its Statistical ToolPak are handy when analysing various datasets (even the smallest ones) which are very common for BSc research. In presented example, scatter graphs proved to be very successful in finding the most toxic elements present in the UWTSD testing plot.

Figures 6, 7 show As and Cd spread in UWTSD plot presented in the form of Heatmaps. MapInfo offers Grid Map function which allows for importing larger and more complex data sets. The highlighted As hotspots are located in the middle of the plot (in line with the other research finding), which confirms the correlation between As and increased mortality of trees. Moreover, the Cd concentrations are shown to be spreading across the site perimeter, with highest concentrations in mid and upper sections (also in line with the research findings).
Both colour-codded GIS systems (Surfer and MapInfo) are widely applied by various researchers to graphically present and analyse data (Dundas Data Visualisation, 2018).
When comparing Cd contour map (Surfer, Fig. 3) with Cd heat map (MapInfo, Fig. 7), it can be inferred that Heatmapping offers much smoother and more precise visualisation of spatially changing metal loadings compared to Surfer contours. Therefore MapInfo can be an excellent alternative to Surfer software, especially when working with larger, more complex data sets.

4.0 Conclusions

Scientific research usually produces raw numbers which represent no value on their own and can be challenging to analyse.

Software packages like Microsoft Excel, including various GIS packages, are excellent support which not only helps to visualize various datasets but also enable performing detail, in-depth analysis.

The GIS software, as well as the Excel package has largely contributed to the presented soil sampling analysis. Some key findings and their significance were determined, and several hypotheses were tested and proved/disproved thanks to the statistics.

Also, it was concluded that MapInfo heatmaps are very useful in assessing of larger, more complex dataset (due to their precision). Therefore, they are highly recommended for more serious, advanced studies (e.g. MSc or PhD).

Additionally, the graphical modelling software has helped to compare/contrast the findings with other academic research and supported formulating sound research conclusions.

Lastly, creating and presenting the obtained results on the scientific poster would not be possible without utilizing the data processing software.

5.0 references

Desai, M. A. (2013) *Phytoremediation: A tool for restoring land degraded due to opencast coal mining*.

Dundas Data Visualisation (2018) *When (and Why) to use Heat Maps*, *Dundas Inc.* Available at: https://www.dundas.com/resources/blogs/when-and-why-to-use-heat-maps (Accessed: 15 May 2020).

Martin, P. and Bateson, P. (1993) Measuring behavior: An introductory guide. Second. Cambridge: Cambridge University Press.

Medda, S. and Mondal, N. K. (2017) 'Chromium toxicity and ultrastructural deformation of Cicer arietinum with special reference of root elongation and coleoptile growth', *Annals of Agrarian Science*. Elsevier Ltd, 15(3), pp. 396–401. doi: 10.1016/j.aasci.2017.05.022.

Saraswat, S. and Rai, J. P. N. (2011) 'Prospective application of leucaena leucocephala for phytoextraction of Cd and Zn and nitrogen fixation in metal polluted soils', *International Journal of Phytoremediation*, 13(3), pp. 271–288. doi: 10.1080/15226514.2010.483259.

Further information

All data analysis including graphs, charts, maps and full statistical testing presented in the research report are available upon request.