Date: 14th January 2021
Module: GGX2203
Student Number: 10652934

Using QGIS to interpolate and map areas unfit for development to protect Stone Age and Iron Age buried land surfaces under a development site in Sneek, Netherlands.

Introduction

This report covers the GIS used to produce two risk assessment maps for a site of future development in Sneek, Netherlands. Two buried land surfaces, protected by the Malta Agreement, are present underneath the development site, an Iron Age surface, and a Stone Age surface. These are considered at risk when they are within 1.2m and 2m, respectively, of the surface. Risk areas are to be estimated and mapped to provide a guide that will prevent future developments from harming the protected surfaces.

This report will outline what can be drawn from the high-quality maps produced and conclude, through discussing the chosen methodology used to produce them, how applicable they are to their intended use.

Datasets and Methodology

The materials provided were, one Excel spreadsheet containing the location and depths of core samples for the Stone Age and Iron Age surfaces, an aerial photograph of the development site, the size of the development site (125 hectares), and a shapefile of a vector that outlined the site's extent. There was no data quality report or description of the sampling methodology provided.

First the data for each surface was separated into individual QGIS geopackages and evaluated. The Stone Age surface contained 1005 observations, whilst the Iron Age contained 1471. The data required alteration as some depths had been recorded as above the surface using greater than symbols (e.g. >1.4). No explanation of these observations was provided, and they were difficult to interpret. It was assumed that they represented depths deeper than the negative of the measured value. Using Python and the Pandas library, the greater than symbols were replaced by negative symbols and included in the following analyses. Then a Ryan-Joiner normality test and a visual inspection was carried out for each dataset, to provide further context.

Next an appropriate output scale was calculated for each protected surface. As the sampling method was unknown it was assumed, based on the results of the data evaluation, that random sampling was used. Due to this assumption, the following formula was chosen to calculate appropriate output resolutions for the interpolations of each dataset:

$$p = 0.25 \times \sqrt{\frac{A}{N}}$$

Where '*A*' represents the total area of the sample location and *'N'* represents the total sample population of the dataset. This random sampling scale formula was outlined in (Hengl, 2006) and is appropriate for randomly sampled vector data.

For choosing an appropriate interpolation method a criteria was created.

- The interpolation method had to be processable by the computer at hand. This criterion was chosen as some methods were tested and resulted in QGIS crashing.
- It also had to be effectively simple to ensure a fast turnaround and limit the possibility of user error in the parameter choices made.
- Finally, an exact interpolation method was required to ensure the sample measurements were honoured and not omitted from the final product as these were the only points where risk, or lack thereof, was empirically known.

IDW was chosen as the interpolation method, specifically, the GDAL Grid IDW with nearest neighbour searching tool was chosen on the basis that in practice it had a far faster processing time than the built in QGIS IDW tool.

According to (Liu *et al.*, 2020), the key influences on the resulting surface of the IDW method are the search radius and weighting/power. To identify the parameter choices that would yield the most accurate model for each surface, 4 distances and 3 weightings were compared, resulting in 12 interpolated surfaces for each dataset.

The first radius chosen was the minimum distance that would produce a surface that covered the entire extent of the development site, 200m for the Iron Age surface and 100m for the Stone Age surface. The maximum radius was the largest possible distance within the extent of the development site, 2300m, representing a global interpolation. And then a quarter (575m) and a half (1150m) of the maximum radius were chosen as the intermediate distance parameters. The three weightings chosen were 1, 2, and 3. These power intervals were chosen as they are intuitive and were used in a similar analysis by (Gia Pham *et al.*, 2016).

To measure the accuracy, RMSE was calculated for each model using the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n}}$$

This was chosen as it was a widely used statistical metric to measure the accuracy of an interpolation (Ris, Holthuijsen and Booij, 1999; Salekin *et al.*, 2018; Fuentes *et al.*, 2019; Srivastava *et al.*, 2019). Furthermore whilst it is “more sensitive to extreme outliers” than MAE (Salekin *et al.*, 2018), both datasets had a low range. 3.85m for the Stone Age surface and 2.7m for the Iron Age surface, and therefore RMSE was a suitable choice.

A random selection of 20% of the data was chosen using the QGIS random selection tool, and this was held to calculate the square error residuals. The remaining 80% was used to create an interpolate by the IDW tool. The held values were then joined with their corresponding estimates on the interpolated surface using a point sampling tool, generating an attribute table with a column for the samples and a column their corresponding estimates. The field calculator was then used to calculate the residuals for each point in a new column. The average of this column was taken and when square rooted, this produced the RMSE value for that surface.

Once RMSE was calculated for every interpolate, the most accurate one for each buried land surface was chosen to produce a map of areas safe and unsafe for development on the site.

The interpolated values extended beyond the boundaries of the site being evaluated, so a mask, generated using the development site vector and the QGIS clip raster by mask tool, was applied to the chosen interpolated surfaces. Resulting in 2 high quality maps of appropriate resolution, each representing one of the buried land surfaces, within the boundaries of the site.

All analysis was completed using QGIS, Minitab, and a desktop PC with an AMD Ryzen 1600x CPU, 16GB of RAM, and an Nvidia 1050TI GPU.

Results

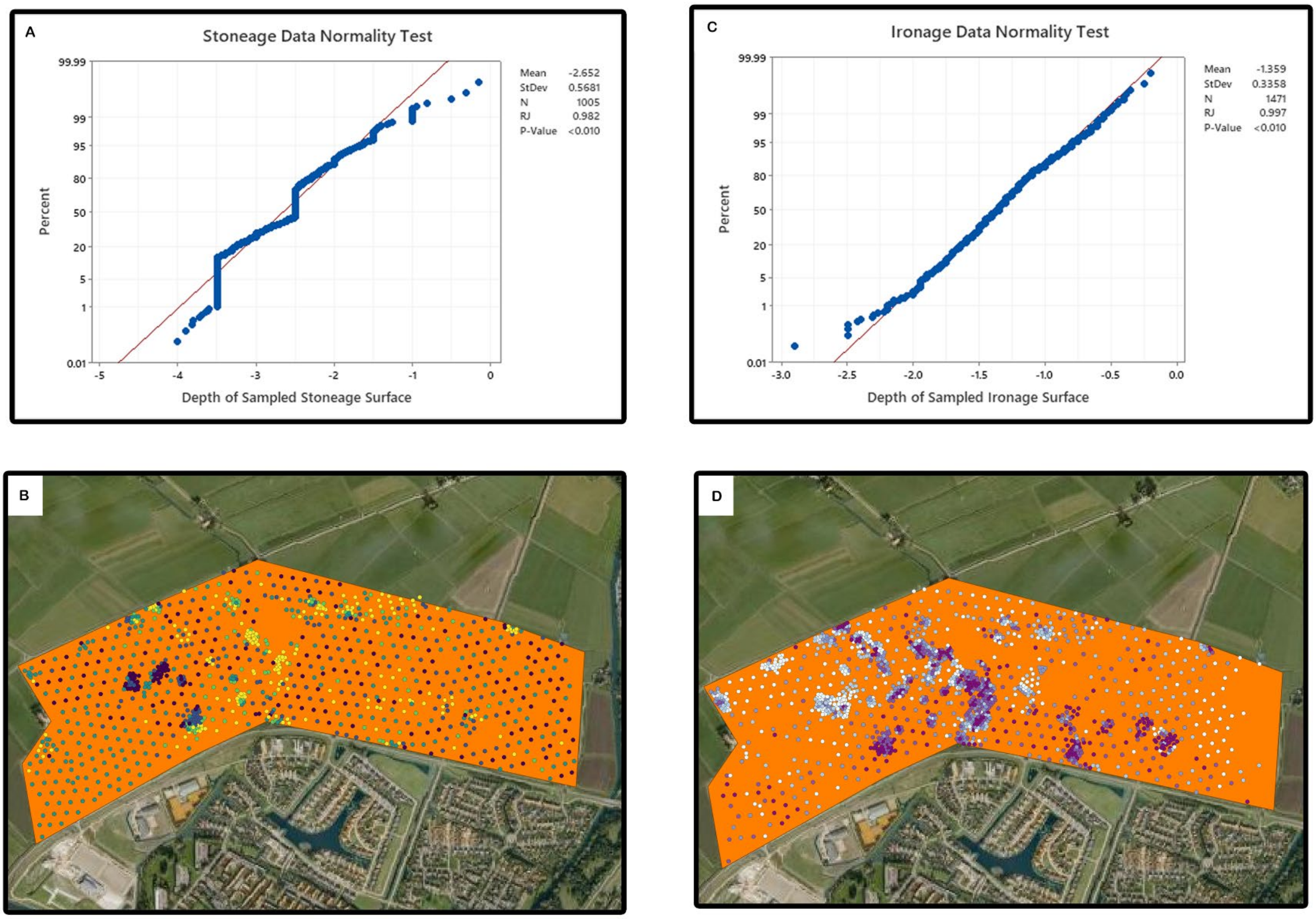

Figure 1: Evaluation of the Normality (A) and Structure (B) of the Stone Age Surface and the Normality (C) and Structure (D) of the Iron Age Surface

It can be seen in figure 1, the Stone Age and Iron Age surfaces achieved Ryan-Joiner scores of 0.982 (A) and 0.997 (C), respectively. Meaning under high confidence (p-value = <0.010), both datasets strongly followed a normal distribution. Furthermore, under visual inspection, there are gaps in the Iron Age point cloud (D) that are not present in the Stone Age point cloud (C) and some clustering is present. As both datasets were normally distributed and under visual inspection there is a clear lack of intuitive order in the sampling methods used, it was assumed the datasets were produced through random sampling.

Stoneage Surface Output Resolution	Ironage Surface Output Resolution
$p = 0.25 \times \sqrt{\frac{1250000\ metres}{1005\ samples}}$	$p = 0.25 \times \sqrt{\frac{1250000\ metres}{1471\ samples}}$
$p = 7.29\ metres$	$p = 8.82\ metres$

Figure 2: Calculation of appropriate output resolutions (to 3 sig. figs.) for each buried land surface under an assumption of random sampling.

Figure 2 shows the estimated appropriate resolutions for each surface were 8.82 metres and 7.29 metres, for the Iron Age and Stone Age surfaces, respectively. By default, the chosen GDAL Grid IDW tool produced surfaces with a resolution of 6.9 metres, this was deemed an appropriate output scale as it was slightly finer than the appropriate resolutions estimated.

Table 1: Ironage Surface RMSE Results of Different Weighting and Search Radius Parameters Tested Within the GDAL Grid IDW Tool

Search Radius (metres)	Weighting	RMSE
100	1	0.51103047387
100	2	0.48772843261
100	3	0.56527320877
575	1	0.60148020115
575	2	0.548956653586
575	3	0.52804776491
1150	1	0.53202611743
1150	2	0.49853109755
1150	3	0.49429917314
2300	1	0.57281511032
2300	2	0.4588598047
2300	3	0.56117644607

Table 2: Stoneage Surface RMSE Results of Different Weighting and Search Radius Parameters Tested Within the GDAL Grid IDW Tool

Search Radius (metres)	Weighting	RMSE
200	1	0.2652197621
200	2	0.27565808976
200	3	0.30569785605
575	1	0.29336634508
575	2	0.27737111481
575	3	0.27657409913
1150	1	0.29593777357
1150	2	0.28831212669
1150	3	0.25544441311
2300	1	0.30598807899
2300	2	0.29312678821
2300	3	0.24888836005

Figure 3: RMSE Results of Different Weighting and Search Radius Parameters Tested Within the GDAL Grid IDW Tool for Iron Age Buried Land Surface (Table 1) and Stone Age Buried Land Surface (Table 2)

Figure 3 shows the RMSE values for each of the interpolated surfaces. The most accurate parameters for the Iron Age surface were a search distance of 2300m and a weighting of 3. For the Stone Age surface, a search radius of 2300m and a weighting of 2 produced the lowest RMSE.

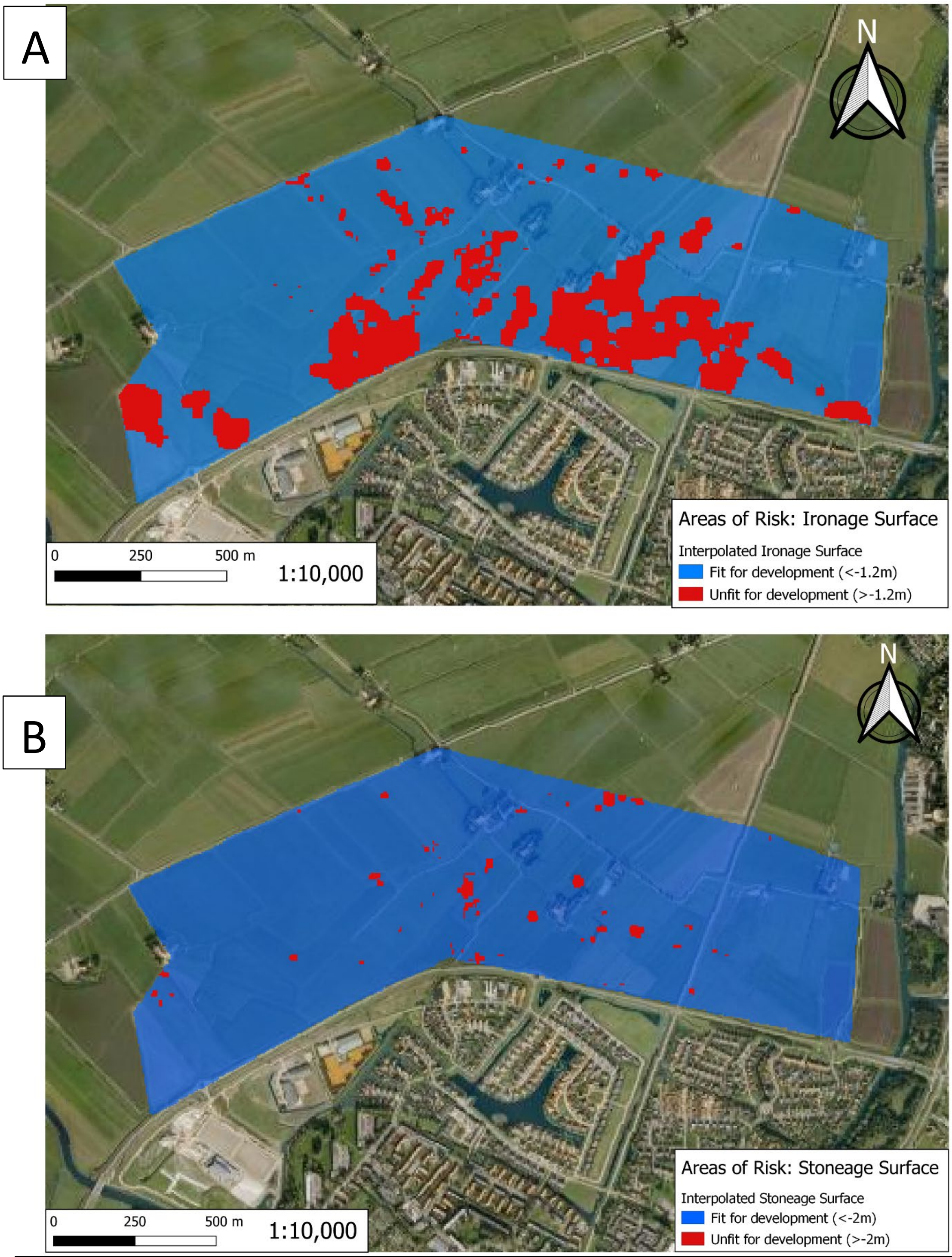

Figure 4: Two high quality maps of appropriate resolution, (A) Areas where development could put the Iron Age surface at risk and (B) Areas where development could put the Stone Age surface at risk.

Figure 4 shows the final products, 2 high quality maps that represent models of the Stone Age and Iron Age buried land surface and areas where those surfaces are too close to the surface for development to safely occur.

Discussion

Uncertainty was present in this model from the outset as key information about the data, such as the sampling method used, who produced the data, and any data quality standards implemented was not provided. This limited the effectiveness of the methods used to manage uncertainty in the produced models.

Specifically, the precision, in this case output resolution, of the final products were limited in the fact random sampling was assumed but not explicitly stated. If the true sampling method had been provided, processing time could have been saved, as a more appropriate and possibly coarser resolution could have been chosen that would have still provided a precise product (Hengl, 2006). More importantly, the large scale of the sample populations could have been wasted if a sampling method requiring a finer scale was used (Hengl, 2006).

Moreover, it is unclear whether the values with greater than symbols were treated properly. Every interpolation within this project assumes that replacing the greater than symbols of positive depths with a negative was an appropriate treatment. However, the intention the creator of each dataset had for these values is not known and it is possible that these are areas where the protected buried land surfaces protrude the surface of the development site. If this is the case, these areas might not be protected within each map.

From the exploratory results of the data, kriging could have been a better suited and more accurate method for this project. The data was irregularly spaced, which according to (Yang *et al.*, no date) , kriging is well suited for. However, IDW was chosen as it met the criteria outlined in methodology. It is simple to implement, whilst still being documented as satisfactorily accurate and having a reasonable calculation time (Maleika, 2020)**.** Furthermore, this method has been widely used (Liu *et al.*, 2020; Maleika, 2020), and the more complicated procedure for creating a kriging interpolation represented a higher chance of user error when choosing parameters and would have taken significantly longer to implement due to the theoretical background required to use it effectively and with confidence**.**

Furthermore, over the course of this project it became apparent that QGIS did not make full use of the available processing power of the computer used. It relied almost entirely on the CPU instead of the GPU for analyses and the QGIS IDW tool took over 3 hours to execute but failed to use more than a quarter of the CPU processing power. This was the main reason the GDAL Grid IDW was chosen. Whilst there is evidence that using the GPU would lead to faster rendering times in a GIS analysis it could possibly sacrifice the quality of the final product (Senthilnath, Sindhu and Omkar, 2014). However, the limited use of the CPU hindered to scope of the project and anecdotal evidence showed others had experienced a similar issue (*Getting QGIS to use dedicated graphics card (GPU) over integrated?*, 2016; *How to use more CPU while processing in QGIS*, 2016).

Date: 14th January 2021
Module: GGX2203
Student Number: 10652934

Conclusion

This report has covered choosing a suitable method for modelling protected buried land surfaces from coring data and mapping areas where those surfaces are at risk of disturbance. The choice and implementation of the most suitable method, based on the chosen criteria, to produce two maps using QGIS was summarised.

The dataset provided and methods used to model the buried land surfaces were evaluated and the resulting uncertainty within the model stemming from the lack of context surrounding the data was discussed.

The applicability of other interpolation methods was explored, and the final choice was justified despite the possibility of sacrificing the accuracy of the final maps.

The final two maps are successful as they have modelled both buried land surfaces and clearly identified the areas where these surfaces are at risk of disturbance. Furthermore, they are reasonably precise and accurate models of the Stone Age and Iron Age land surfaces, within the constraints imposed by the chosen criteria, uncertainty imposed by the datasets, and the finite processing capabilities of QGIS. However, if this project were repeated, multiple exact interpolation methods would be compared, instead of trying to achieve the most accurate possible representation from a single method. As seen in figure 3, the range of RMSE is quite limited so evaluating multiple interpolation methods may provide larger benefits in accuracy. Finally, the impact of maintaining the greater than values would be measured by comparing the interpolated surfaces of data before and after treatment.

References

Fuentes, Padarian, Van Ogtrop, and Vervoort (2019) 'Comparison of Surface Water Volume Estimation Methodologies that Couple Surface Reflectance Data and Digital Terrain Models', *Water*. MDPI AG, 11(4), p. 780. doi: 10.3390/w11040780.

Getting QGIS to use dedicated graphics card (GPU) over integrated? (2016) *GIS Stack Exchange*. Available at: https://gis.stackexchange.com/questions/191905/getting-qgis-to-use-dedicated-graphics-card-gpu-over-integrated (Accessed: 14 January 2021).

Gia Pham, T., Huynh Van, C., Tran, P. and Thi Minh Chau, T. (2016) 'Impact of power value in IDW interpolation method on accuracy of the soil organic matter (SOM) mapping', in *Proceedings of International Symposium on Earth Sciences and Suitainable Geo. Resources Development (ESASGD2016)*, Vietnam.

Hengl, T. (2006) 'Finding the right pixel size', *Computers & Geosciences*, 32(9), pp. 1283–1298. doi: 10.1016/j.cageo.2005.11.008.

How to use more CPU while processing in QGIS (2016) *GIS Stack Exchange*. Available at: https://gis.stackexchange.com/questions/180003/how-to-use-more-cpu-while-processing-in-qgis (Accessed: 14 January 2021).

Liu, H., Chen, S., Hou, M. and He, L. (2020) 'Improved inverse distance weighting method application considering spatial autocorrelation in 3D geological modeling', *Earth Science Informatics*, 13(3), pp. 619–632. doi: 10.1007/s12145-019-00436-6.

Maleika, W. (2020) 'Inverse distance weighting method optimization in the process of digital terrain model creation based on data collected from a multibeam echosounder', *Applied Geomatics*. Springer Science and Business Media LLC, 12(4), pp. 397–407. doi: 10.1007/s12518-020-00307-6.

Ris, R. C., Holthuijsen, L. H. and Booij, N. (1999) 'A third-generation wave model for coastal regions: 2. Verification', *Journal of Geophysical Research: Oceans*. American Geophysical Union (AGU), 104(C4), pp. 7667–7681. doi: 10.1029/1998jc900123.

Salekin, S., Burgess, J., Morgenroth, J., Mason, E. and Meason, D. (2018) 'A Comparative Study of Three Non-Geostatistical Methods for Optimising Digital Elevation Model Interpolation', *ISPRS International Journal of Geo-Information*. MDPI AG, 7(8), p. 300. doi: 10.3390/ijgi7080300.

Senthilnath, J., Sindhu, S. and Omkar, S. N. (2014) 'GPU-based normalized cuts for road extraction using satellite imagery', *Journal of Earth System Science*. Springer Science and Business Media LLC, 123(8), pp. 1759–1769. doi: 10.1007/s12040-014-0513-1.

Srivastava, P. K., Pandey, P. C., Petropoulos, G. P., Kourgialas, N. N., Pandey, V. and Singh, U. (2019) 'GIS and Remote Sensing Aided Information for Soil Moisture Estimation: A Comparative Study of Interpolation Techniques', *Resources*. MDPI AG, 8(2), p. 70. doi: 10.3390/resources8020070.

Yang, C.-S., Kao, S.-P., Lee, F.-B. and Hung, P.-S. (no date) *Twelve Different Interpolation Methods: A Case Study of Surfer 8.0*.