



Title	Determining minimum data input levels for reliable three-dimensional soil profiling
Authors(s)	Laefer, Debra F., O'Neill, James, Williams, Derek
Publication date	2009-03
Publication information	Laefer, Debra F., James O'Neill, and Derek Williams. "Determining Minimum Data Input Levels for Reliable Three-Dimensional Soil Profiling." American Society of Civil Engineering (ASCE), March 2009. https://doi.org/10.1061/41022(336)42 .
Conference details	Presented at the 2009 International Foundation Congress and Equipment Expo, March 14-19, Orlando, Florida
Publisher	American Society of Civil Engineering (ASCE)
Item record/more information	http://hdl.handle.net/10197/2164
Publisher's version (DOI)	10.1061/41022(336)42

Downloaded 2026-05-02 00:26:16

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

Determining minimum data input levels for reliable three-dimensional soil profiling

Debra Laefer¹, James O'Neill², and Derek Williams³

1. Lead Investigator, Urban Modelling Group (UMG), School of Architecture, Landscape, and Civil Engineering (SALCE), University College Dublin (UCD), Newstead G25, Dublin 4 Ireland; phone 353-1-716-3226; debra.laefer@ucd.ie.
2. Research Assistant, UMG, SALCE, UCD.
3. Research Assistant, UMG, SALCE, UCD.

ABSTRACT

The ultimate goal of the project reported herein was to determine whether or not a readily quantifiable threshold data density could be established to accurately predict subsurface lithology. A study area in Dublin Ireland was selected for its high density of data (567 boreholes in less than 2km²) and for it being slated for the first phase of the upcoming Dublin metro. Analysis was conducted using the kriging function embedded in ArcGIS based on identifiable lithologies. Difficulties were encountered based on inconsistent labeling and the program's inability to cope with the presence of a single lithology appearing multiple times within a single soil profiles. In an area of 10,000m², a minimum data density of 8.0 boreholes predicted bedrock within 0.5m. Whereas in an area of 90,000m², a sampling density of 8.3 was per 10,000m² was needed to achieve the same accuracy, approximately double of British standards.

KEY WORDS

Soil boring, lithology, kriging, statistical analysis, bore hole spacing

INTRODUCTION

To predict soil strata and bedrock depth from discrete borehole data is challenging, especially if the study area is large, as soil strata are non-uniform and heterogeneous. The main question from a commercial perspective is how much data is necessary for reliable prediction, and thus, how frequently must a borehole or testpit be created. This study aimed to identify a data threshold for optimum borehole spacing/density, as well as an appropriate borehole distribution pattern for site investigation. This was done by progressively introducing portions of an existing but heretofore disparate data set into a geographic information system (GIS). The work was done in preparation for a larger study related to the installation of Ireland's first metro.

SCOPE OF STUDY

The study comprised an area of approximately 1.82 km² in the city center of Dublin, Ireland and encompassed the first section of the proposed Metro North line (currently under tender) starting from the western edge of St. Stephen's Green in the south and as far as O'Connell Bridge to the North. Within the study area, 567 subsurface sampling logs (Fig. 1) from the Geological Survey of Ireland (GSI) were available as part of a new, free, downloadable format (www.gsi.ie). Most were boreholes, some dating back several decades.

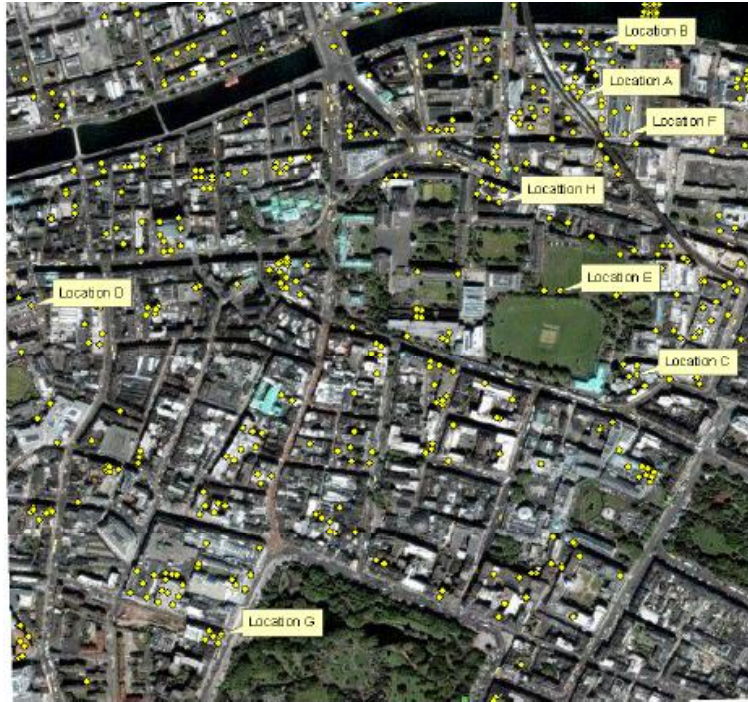


Figure 1. Study Area with 567 Sampling Locations and Boreholes A-H.

The borehole logs were in a spreadsheet format and contained the following data:

Borehole ID: a unique identification code assigned to every borehole documented by the GSI.

Borehole Coordinates: the exact easting and northing of the boreholes location referenced directly to the Irish National Grid.

Depth of Hole: exact excavation depth.

Depth of Bedrock: whether bedrock was encountered and at what depth.

Investigation Type: investigation information (e.g. cable percussion based bore hole, trial/observation pit).

Major Lithology: geological description of primary constituents of lithology encountered and depth and thickness of layer.

Minor Lithology: geological description of secondary constituents of lithology encountered and depth and thickness of layer.

Borehole ID	Report ID	Company Hole Name	Easting	Northing	Hole Depth	Depth to Bedrock	OD Malin	Investigation Type
51347	127	1	315909	233824	4.2	4	11.01	Cable Percussion (Shell and Auger)
51348	127	2	315925	233814	4	3.8	11.01	Cable Percussion (Shell and Auger)
54022	251	1	315953	233745	7.01	>7.01	8	Drilling Type Unspecified
55392	347	1	315873	233695	3.12	3.05	5.82	Cable Percussion (Shell and Auger)
55393	347	2	315893	233692	4.88	3.35	5.82	Cable Percussion (Shell and Auger)
55394	347	4	315877	233679	3.96	2.44	5.82	Cable Percussion (Shell and Auger)
59935	601	1	316132	233730	3.175	>3.175	7.5	Drilling Type Unspecified
59936	601	2	316141	233736	5.57	5.029	7.5	Drilling Type Unspecified
59937	601	3	316142	233745	5.029	4.267	7.5	Drilling Type Unspecified
60376	646	1	316223	233843	13.106	10.058	6.74	Rotary Percussion (Wagon Drilling)
60378	646	3	316174	233775	7.925	>7.925	7.64	Cable Percussion (Shell and Auger)
60379	646	4	316167	233715	9.754	>9.754	7.48	Rotary Percussion (Wagon Drilling)

Figure 2. Sample of input data

ArcMap, a component of the Environmental Systems Research Institute's (ESRI) ArcGIS software package was used to store and analyze the data. ArcMap allows for any data geographically referenced to be manipulated and edited and to be displayed by way of maps and charts. The borehole data could be edited and separated as desired, when imported into ArcGIS. Particular information could be selected by attributes such as major lithology or location, or whether or not bedrock was found.

As the main aim of this project was to determine the level of data density required to make accurate predictions of the subsurface lithology, eight boreholes (shown as A-H in Fig. 1) were randomly removed from the data set and set aside without any review of their contents. Then analysis began using only 10% of the records, which were randomly selected, using a random number generator. Upon completion of the initial analysis an additional 10% of the data records was randomly selected and incorporated. This procedure was repeated until all data were incorporated. Prior to the data's introduction, the various layers were labeled according to major lithologies.

Interpolation relied on the raster form of the GIS image file, which stored its information in cells called pixels that are more commonly associated with images. The information examined was the lithology layers within each sampling. When the data were interpolated to a raster, every pixel was assigned a negative position. The results each had colors assigned to them based on depth (fig. 3). By using this function, a raster image or map of the area displaying the subsurface elevations based on various soil layers could be created. Rasters could be viewed and explored in both two-dimensional (2D) and three-dimensional (3D) formats. The system used kriging (one of three main geostatistical analysis techniques) to interpolate between discrete data points.

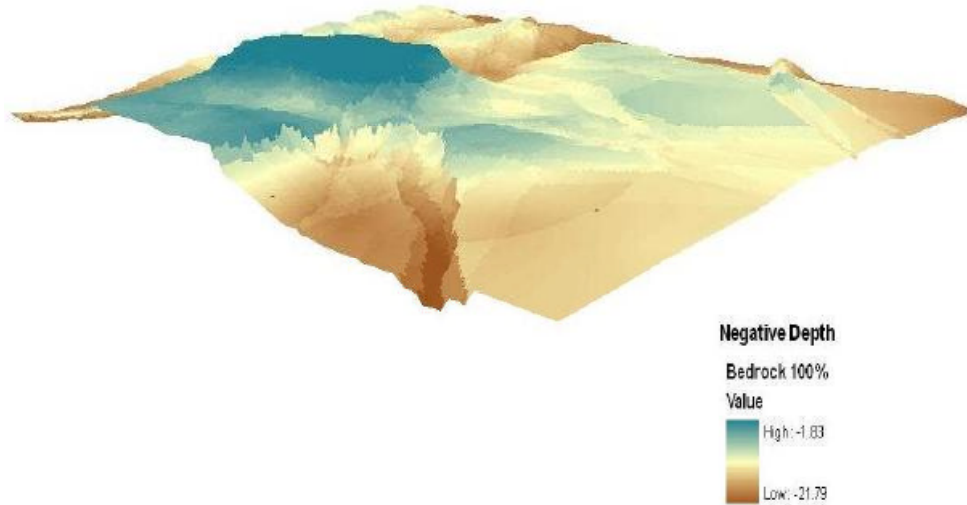


Figure 3. Sample Raster output from ArcGIS for 100% data for Bedrock layer.

Kriging employs geostatistical models that include autocorrelation, which is the spatial relationship among the measured points. Kriging fits a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location (Robinson and Metternicht 2006). The method is essentially a two-step process. Firstly the data is fitted to the model using spatial variance or semi variance techniques (the estimation of the spatial autocorrelation of the data). Spatial variance is calculated using eqn (1)

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (\text{eqn 1})$$

The spatial variance is half of the average squared difference between the value at $z(x_i)$ and the value at $z(x_i + h)$, h is the separation distance and $N(h)$ is the number of data pairs within a given class of distance and direction (Robinson and Metternicht 2006). Secondly the predictions of the unknowns are made in a similar way to Inverse Distance Weighting (IDW) by forming weights surrounding the known values to measure the unknowns. The measured values that are closest to the unknown values have the most influence, just like IDW, however the weights formed by Kriging are much more sophisticated. Kriging weights take into account the semi variance that was developed by looking at the spatial nature of the data, and not just by distance as with IDW.

Based on examination of the raster images, predicted locations of each of the lithologies were plotted as a function of percentage of data included versus the actual known lithology, as taken from the eight previously unexamined boreholes.

Sample Findings

The majority of cases failed to consistently show a trend of increasing accuracy

with improved data density (Fig. 4), and in some cases lithologies were predicted that did not actually occur at that location. A major error source was from the fact that when a lithology occurred more than once in a borehole, the software took a mean value. Another problem was that over the half-century in which the data was collected there was not rigorous consistency in lithology labelling. As the bedrock appeared only once in each of the subsurface records and since there was a lower chance for disagreement with respect to its labelling, a strong correlation between increased data density and prediction was in evidence as shown in the bottom two lines in Fig. 4, where the solid line is the actual bedrock depth.

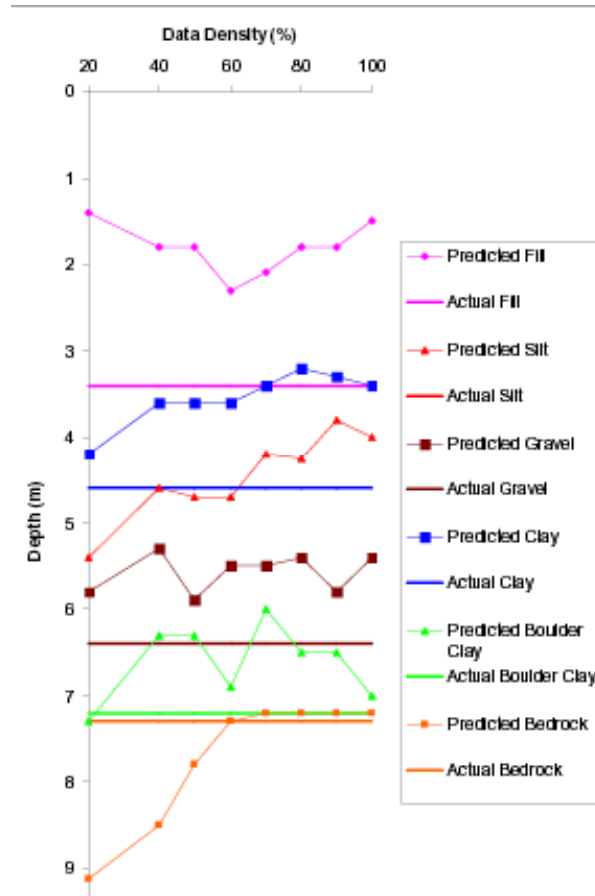
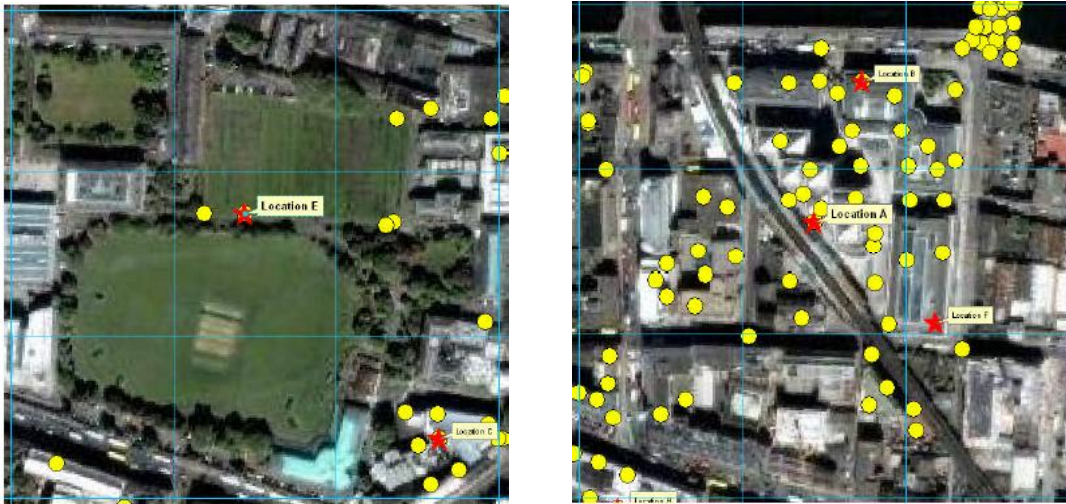


Fig. 4. Lithology depth prediction for location C as a function of data density; note that no silt was present in actual borehole.

These results confirmed the conclusions of McCarthy and Graniero (2006) that standardisation would promote better management and maintenance so that the “user can spend more time performing analysis tasks and less time restructuring data, thus increasing productivity and reducing the risks associated with a series of data modification cycles”. Furthermore, any attempts at electronic mapping must overcome the legacy of many geologists and civil engineers storing and manipulating their borehole and test pit data in spreadsheets. Although spreadsheets are tabular in structure, often they do not easily conform to the strict relational database structure requirements of software packages geared towards large enterprise scale organizations such as that developed by Chang and Park (2004).



(a) Borehole E

(b) Borehole A

Fig. 5. Non-uniform density and distribution of boreholes in study area

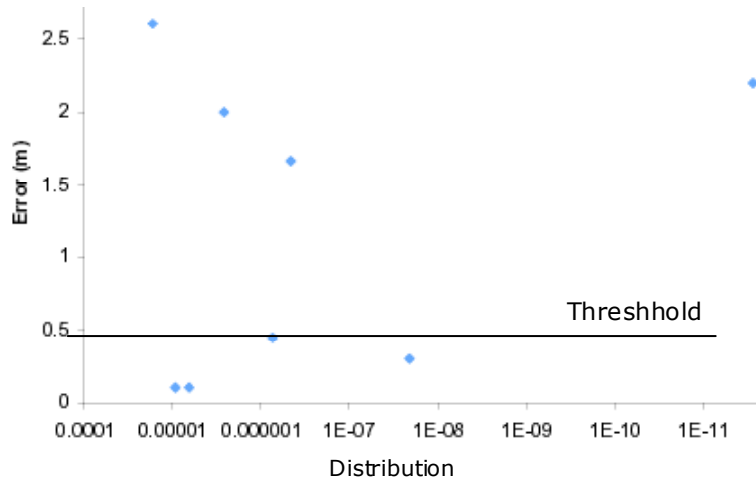


Fig. 6. Error in depth prediction vs borehole distribution

Using only the rock layer, distribution issues were investigated by superimposing a 300m x 300m grid divided into 9 separate 100m x 100m squares atop each of the 8 verification boreholes described above (Fig. 5).

At Location E [Fig. 5(a)] there was very low local data density relative to that at Location A [Fig. 5(b)]. Furthermore, at Location E, the data was heavily concentrated to the left grid, while at Location A the distribution was fairly even. Data distribution was assessed using a Chi-Squared Goodness of Fit Test. If in a selected area there were 18 boreholes in the 9, a totally random distribution would expect 2 boreholes within each square, thus achieving a value of 1 by the Chi-Squared test. Figure 6 displays the absolute error of the final prediction of bedrock level in the eight study locations. Some correlation between the evenness of the distribution of the data around the unknown location and the error in the prediction was seen. The more evenly distributed the data, the less error occurred in the predicted results, but by removing the three locations with the lowest local data densities, a vastly improved correlation was achieved (Fig. 7 – removed data shown as struck through in Table 1.), leading to the

conclusion that a minimum data threshold exists around 8 boreholes per 10,000m². Repeating the exercise for an area of 90,000m² increases the threshold to 8.3 boreholes per 10,000m², which is more than twice current British Standards (1999)

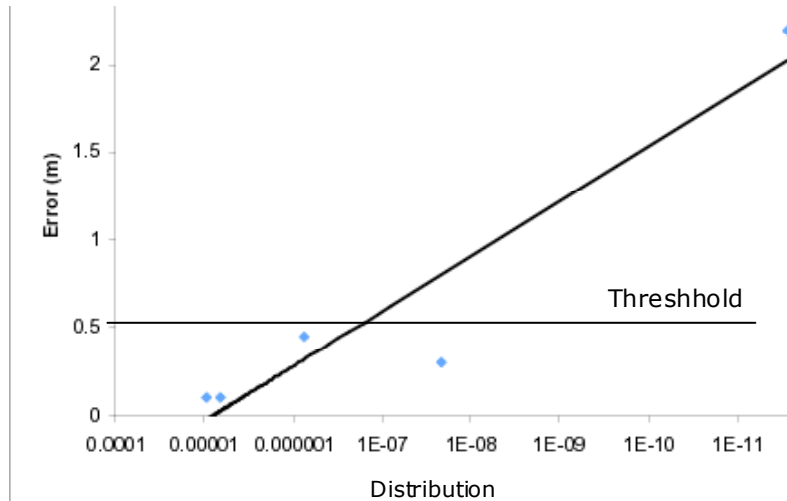


Fig. 7. Absolute Error versus data distribution

Table 1. Error measures, distribution and density of each verification location

Location Designation	Absolute Error in Bedrock Depth (m)	Distribution	No. of Boreholes per 10,000m ³
A	0.1	6.6143E-06	13
B	0.3	2.12E-08	21
C	2.2	2.6395E-12	7
D	2.6	1.6851E-05	4
E	1.7	4.6453E-07	2
F	2.0	2.6176E-06	3
G	0.1	9.3575E-06	9
H	0.45	7.4519E-07	8

RECOMMENDATIONS

To obtain the maximum benefit from previous and future subsurface explorations, several actions are needed. These involve the creation of publicly maintained databases for which the voluntary and cooperative submission of subsurface data would be achieved through the introduction of a value added feature: namely that a contractor or consulting engineer would enter their new information and that the system would generate a series of probable soil layers through kriging, splining, and/or one of the other methods regularly employed in such interpolation. The larger geotechnical community would then cumulatively benefit through inclusion of each new piece of information, and the contributor would benefit by having access to large quantities of historical data and a publicly maintained system that provides analysis based on such content. Critical to this, however, is the standardization of new and to some extent existing data collections.

CONCLUSIONS

In summary, the heterogeneous nature of soil in the study area could not be fully interpolated or visualized from discrete borehole data. The investigative technologies employed and data used failed to represent the occurrence of random pockets of material. Additionally, the variability in the actual geological descriptions of soil lithology at the investigation stage lead to high levels of inconsistency in labeling choices, which compromised the data's reliability. These issues generally did not emerge when predicting the bedrock level. For the bedrock, a definite interdependency between the depth prediction and both the level of data density and the evenness of the data's distribution emerged. Furthermore, there was a definite correlation between the evenness of the distribution and the accuracy of the prediction. What was concluded was that when assessing an area of 10,000m² around an unknown borehole, a data density greater than 0.0008 boreholes/m² (8 boreholes per 10,000m²) ensured an error level less than 0.5m in the prediction of the level of bedrock, and when assessing an area of 90,000m² around an unknown borehole a data density greater than 0.00083 boreholes/m² (8.3 boreholes per 10,000m²) was needed to achieve the same level of accuracy for bedrock prediction.

ACKNOWLEDGMENTS

Data provided by the Irish Geological Survey.

REFERENCES

BS 5930:1999 Code of practice for site investigations.

Chang, Y.-S. and Park, H.-D. (2004). Development of a web-based Geographic Information System for the management of borehole and geological data. *Computers and Geosciences* 30(8): 887-97.

Geological Survey of Ireland (GSI) as part of a new, free downloadable format (www.gsi.ie).

McCarthy, J.D. and Graniero, P.A. (2006). A GIS-based borehole data management and 3D visualization system. *Computers & Geosciences* 32(10): 1699-1708.

Robinson, T.P. and Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture* 50(2): 97-108.