# An Application of Machine Learning in Modelling the Rockhead, a 3D Geological Surface


Marieta Garcia-Bajo
*British Geological Survey*
Keyworth, United Kingdom
mgarc@bgs.ac.uk

Lothar Piepmeyer
*Furtwangen University*
*Faculty Computer Science*
Furtwangen, Germany
lothar.piepmeyer@hs-furtwangen.de



***Abstract* —The British Geological Survey (BGS) holds much information on the depth of the rockhead surface (the transition between Quaternary and Bedrock geological units) in the UK. This information has been extracted from more than one million borehole paper logs and has been used to create the BGS rockhead surface model. A difficulty arises when different interpretations of the rockhead depth from the paper logs are introduced into the database, and a selection as to which interpretation to use needs to be made. Here, we outline the application of machine learning (ML) methodologies in automatising the selection of one rockhead interpretation per borehole based on previous decisions and therefore saving a huge amount of manual checking work. This reduces the selection process from weeks to minutes. The outcomes were quality controlled with subset examples where results were known. This showed that using just 5% of the complete data, the resulting error was less than 10%. The final results showed that in 5 out of 100 conflicting cases, the ML algorithm favours a different interpretation than that selected by a geologist. This is an acceptable rate because only 5% of the entire set of boreholes have more than one interpretation.**



***Keywords—rockhead, bedrock depth, Superficial Deposits, Quaternary, machine learning***


## I. Introduction

Boreholes provide one of the most important sources of information on the geology and structure of the subsurface. Together with information for surface outcrops, they provide the information upon which we base our interpretation of the geology, from where we create geological maps and 3D geological models ([5], [6], [7], [9]). Boreholes are essential to the majority of geological disciplines including environmental monitoring, resource evaluation and waste management. However, certain aspects of the subsurface geology proved by a borehole cannot simply be derived from the log. A prominent example is the contact between the bedrock (pre-Quaternary) and superficial (Quaternary) deposits, which is called geological rockhead surface (Fig. 1; [9]). The bedrock is normally the pre-Quaternary succession and generally comprises fully-consolidated lithotypes. By contrast, superficial deposits are usually of Quaternary age, and are typically unconsolidated due to their lack of burial (Fig. 1). Quaternary deposits are of particular societal importance as they influence our environment and landscape.

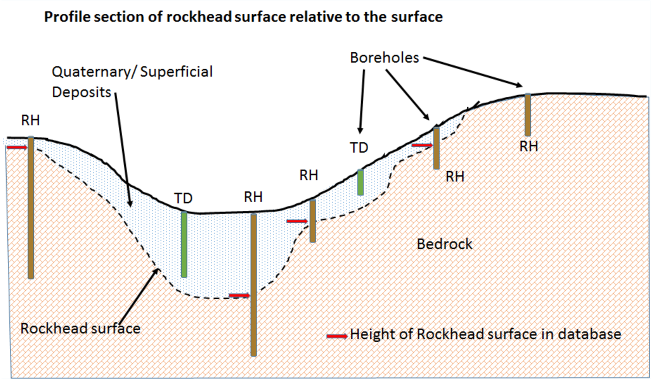


Fig. 1. TD boreholes are boreholes proving the existence of superficial deposits but not reaching Bedrock (green). RH boreholes that reach Bedrock and prove the existence of the Rockhead surface (brown). Point at which the RH borehole crosses the rockhead surface as per the database.

Assuming that the solid geology has been penetrated, in analogue or digital borehole data, the depth to rockhead is frequently difficult to identify. In practice, geologists use a number of criteria to identify the depth to rockhead from borehole logs. Distinguishing the position of the rockhead surface when modelling is problematical if several geologists have made different interpretations, and a decision between the different depths is required.

The British Geological Survey (BGS) has developed a raster model, the Superficial Deposit Thickness Model (SDTM), to show the thickness variations of the Quaternary/Superficial Deposits throughout Great Britain (Fig. 2; [8]). For vertical (non-deviated) boreholes, the depth of bedrock is precisely equivalent to the thickness of the Quaternary deposits. In the SDTM, this parameter was used to derive the thickness of Superficial Deposits throughout the whole country. In [10] borehole data was used besides other information to train a machine learning (ML) model to obtain a global model for bedrock depth.

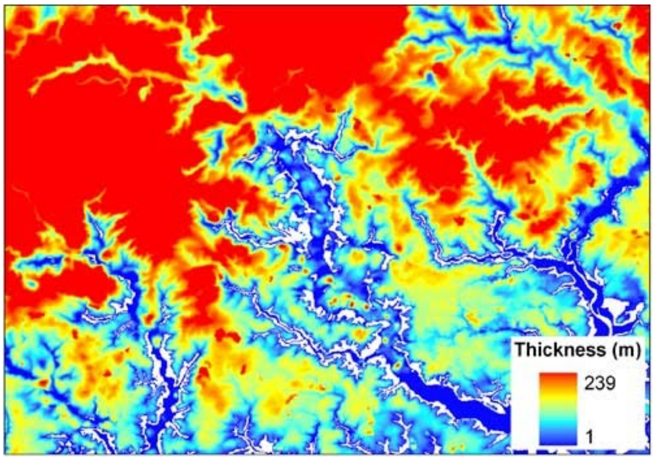


Fig. 2. Cut out of the Superficial Thickness Model.

## II. USING BOREHOLES TO MODEL THE ROCKHEAD SURFACE

The BGS maintains a collection of over one million borehole paper logs from all forms of drilling and site investigations, and significant numbers of new records are added every year (Fig. 3; [3]). Boreholes range from one metre to several kilometres deep; one of the earliest is from 370 AD, but the majority of them postdate 1950. The borehole paper logs are produced from the observations of geologists or surveyors of the rock samples extracted. These paper logs are then interpreted by BGS geologists, who classify the different rock units and enter them into a relational database management system (RDBMS) called Borehole Geology.

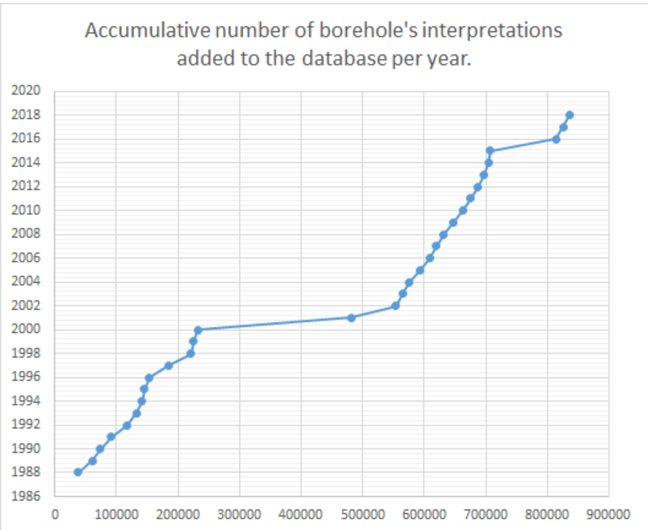


Fig. 3. Cumulative curve of the number of boreholes interpretations being added to the borehole geology database since its creation in 1985.

From the Borehole Geology database, information is extracted about the depth of bedrock to create the BGS national rockhead surface model. However, individual geologists have different interpretations of the depth of bedrock, resulting in multiple interpretations for the same borehole. This may be due to several reasons, including the continuously evolving knowledge of the geology of an area, variations in expertise of log interpretation, the differing complexities of geology area-by-area and different interpretations of what has been written on the logs. Of the approximately 800k borehole records in the Borehole Geology database, around 40% have duplicate interpretations. Some interpretations are the same, and some differ by several metres and some interpretations, lack a value for the depth of bedrock altogether, due to human error.

The proportion of multiple interpretations being substantially high, and the increasing rate at which new borehole interpretations are being entered into the database (an average of 15k interpretations a year) makes it impossible to continue the screening manually.

In summary, the manual selection of interpretations of rockhead in boreholes requires a vast amount of work. This is increasing every year due to the growing number of interpretations by different geologists, failure to code the rockhead and the ever-increasing addition of new boreholes into the database. In this paper, we use a machine learning (ML) approach to automate the manual identification and selection process of the rockhead depth in borehole records.

.

## III. TERMINOLOGY RELATING TO THE MODEL

Several different interpretations of a borehole log may exist in the database. The borehole log contains details such as lithology, depth or thickness of the units that were drilled. This information is not always entered in the database, however the minimum interpretation consists of at least one layer of information. Each interpretation will have a terminal depth (TD) value, which is the depth at which the borehole ended. Some of the interpretations will also have a layer labelled ‘RH’ to indicate the depth of the rockhead. It is assumed that a borehole not labeled with ‘RH’ is entirely composed of Superficial Deposits, therefore the TD is selected as a lower bound for RH depth. An interpretation is of type ‘RH’ if an RH label exists, and of type ‘TD’ in all other cases. The rockhead surface is therefore penetrated in ‘RH’ boreholes and not in ‘TD’ boreholes (Fig. 1). A set of interpretations of a borehole is termed homogeneous if the interpretations are either all type ‘RH’ or all type ‘TD’. A set of interpretations which is not homogeneous is termed heterogenous. A set of interpretations is called conflicting if it is:

- heterogenous or
- homogeneous, and the difference of their depths is greater than 50 cm.

All other sets of interpretations are called non-conflicting. Conflicting interpretations are where human intervention is necessary: A geologist has to judge, and choose the interpretations he/she considers correct. A conflicting set of interpretations that has been considered by a geologist is called resolved, all other conflicting sets are called unresolved. The result of the resolution consists of at least one accepted interpretation, and at least one rejected interpretation. To guarantee a high quality of the thickness model mentioned above, it has to be ensured that only non-conflicting or accepted interpretations are used. For some boreholes, the interpreted rockhead depths may differ only by a small amount: a tolerance of 50 cm was chosen to decrease the number of interpretations to review. Therefore, several accepted interpretations may exist for a single borehole. Not only boreholes of type RH, but also those of type TD are of some value for the model, since the total depth of a borehole is a lower limit of the thickness of Quaternary deposits at this point, even if rockhead was not reached.

## IV. Data structure and examples

The borehole data is currently spread across several tables in the BGS database, and these contain many attributes necessary for the interpretation of a borehole. Some of these attributes are explained below:

- interpretationID: unique identifier of the borehole interpretation. Duplicate values do not exist.
- boreholeID: identifier of the borehole. It is not unique, since different interpretations of the same borehole may exist in the database.
- type: type of interpretation with possible values for RH and TD.
- thickness: thickness of the Quaternary deposits in case of type RH, or the terminal depth of the borehole in case of TD.
- x, y: coordinates of the borehole according to the British National Grid.
- interpreter: identifier of the geologist who created the interpretation.
- project: identifier of the BGS project under which the interpretation was created.

Table I concerns three boreholes with two interpretations per well and some attributes. The interpretations of borehole 4711 are non-conflicting since they are both of type RH and their thicknesses differ by only 34 cm. The interpretations for boreholes 4712 and 4713 are by contrast both conflicting: the interpretations of borehole 4712 are heterogenous; and although the interpretations of borehole 4713 are homogeneous their thicknesses differ by more than 50 cm. Therefore the interpretations of both these boreholes need resolving.

TABLE I.

| ***boreholeID*** | ***interpretationID*** | ***type*** | ***thickness*** |
|---|---|---|---|
| 4711 | 23 | RH | 1066 |
| 4711 | 24 | RH | 1100 |
| 4712 | 42 | RH | 310 |
| 4712 | 43 | TD | 310 |
| 4713 | 78 | TD | 1256 |
| 4713 | 79 | TD | 1200 |

Fig. 4. Example data from the borehole geology database

## V. Feature Engineering

Borehole data with non-conflicting interpretations can be used by the SDTM. Boreholes with conflicting interpretations require additional processing. The goal is to train a machine learning model using interpretations which have already been manually reviewed. The model can then be applied to interpretations that have not yet been reviewed to predict the acceptance decision of the geologist.

In almost all cases, it is inadvisable to use the set of all attributes for a machine learning model. Feature engineering describes the process of selecting attributes and deriving new attributes from the set of all known existing attributes. We show here examples of selecting and deriving features from the attributes discussed in the previous section.

It is unsurprising that some geologists produce interpretations which are more reliable than others. Indeed, existing data proves that the quality of interpretations are heavily dependent on the attributes interpreter and project. We nevertheless decided to include none of these attributes in the set of features. This is because there are over one hundred interpreters and projects in the borehole geology database. The standard methodology in machine learning requires one-hot-encoding, i.e. one additional boolean column per interpreter and project, resulting in hundreds of additional columns. This increase in dimensionality however results in a substantial decrease in information density, which is known as the ‘curse of dimensionality’ [1]. We observed the existence of the ‘curse of dimensionality’ in this case by experiments with borehole data: the prediction quality is much better without the attributes interpreter and project.

An additional improvement in quality is obtained if conflicting interpretations are grouped by boreholeID and groupwise aggregates are computed. In addition to standard aggregates like minimum, maximum and average thickness, we also included attributes like isHomogeneous, which is 1 and 0 if the interpretations of the borehole are homogeneous and heterogeneous respectively. Aggregates introduced more information about the relation between the different interpretation of each borehole, which improved prediction quality. Table II shows data, which is derived from Table I. Borehole 4711 was not included, since it does not contain any conflicts. The attribute isAccepted is 0 or 1 depending on the result of the manual review of the interpretations.

We used SQL-scripts to create two new tables with selected and derived features: one with already reviewed interpretations and another with non-reviewed data, which of course does not contain the column isAccepted.

TABLE II.

| ***bore hole*** | ***interpreta tion*** | ***type*** | ***thickness*** | ***isHomo genous*** | ***avg*** | ***isAccep ted*** |
|---|---|---|---|---|---|---|
| 4712 | 42 | RH | 310 | 0 | 310 | 1 |
| 4712 | 43 | TD | 310 | 0 | 310 | 0 |
| 4713 | 78 | TD | 1256 | 1 | 1228 | 0 |
| 4713 | 79 | TD | 1200 | 1 | 1228 | 1 |

Fig. 5. Examples of derived features

## VI. Machine learning (ML)

Although the BGS borehole collection includes more than one million records, here we only include those which are in regions known to be underlain by Quaternary deposits. We began with 687 073 interpretations of 527 615 different boreholes. For only 5% of the boreholes, conflicting interpretations occured. The conflicting interpretations of these 22 427 boreholes were already resolved manually, but for 3 561 boreholes the interpretations are yet unresolved.

Before ML was applied to the resolution of unresolved interpretations, we ran exhaustive tests on previously resolved interpretations in order to understand the quality of the predictions of the, so-called, response variable isAccepted. Cross-validation is the part of the standard methodology of ML. We used 5-fold cross-validation: the set of boreholes is split into five almost equal sized parts, resulting in five different 80%-20% splits of the data. The largest part of each split was used to train the model, the smaller part was used to test the model. The predictions of isAccepted are compared to the actual values. The average error rate of all five tests is a reasonable estimate that can be expected for unresolved interpretations where the actual value of isAccepted is unknown.

Gradient boosting is a ML technique, which has been successfully used recently. We applied Extreme Gradient Boosting [4] with the logistic regression learning objective and isAccepted as response variable. Each prediction of isAccepted is a number which indicates the probability that an interpretation would be accepted by a geologist. From each group of interpretations of the same borehole, the one with the highest response value is chosen as the interpretation that would be accepted by a geologist. Table III contains some example data with only four attributes. The isAccepted column contains the prediction value created by ML. The higher number represents the most likely to be the correct interpretation.

Therefore borehole 4712 would be considered to be of type RH with thickness 310 cm, while borehole 4713 would be considered to be of type TD with thickness 1200 cm.

TABLE III.

| *bore hole* | *interpreta tion* | *type* | *thickness* | *isHomo genous* | *avg* | *isAccep ted* |
|---|---|---|---|---|---|---|
| 4712 | 42 | RH | 310 | 0 | 310 | 0.8 |
| 4712 | 43 | TD | 310 | 0 | 310 | 0.4 |
| 4713 | 78 | TD | 1256 | 1 | 1228 | 0.1 |
| 4713 | 79 | TD | 1200 | 1 | 1228 | 0.2 |

Fig. 6. Sample predictions of the ML algorithm.

## VII. RESULTS

- We obtained an error rate of 5.3% using 5-fold cross-validation. This means that in 5 out of 100 conflicting cases, the ML algorithm favours a different interpretation from that chosen by a geologist. This does not necessarily mean that the algorithm's decision is wrong. At least for the SDTM, this is an acceptable rate considering that only 5% of the boreholes have conflicts (section VI).
- For 5-fold cross-validation, 80% of the manually resolved borehole interpretations are considered. We observed that the error rate quickly drops when the quantitative base of the model decreases. Fig. 7 shows that the error rate already drops to under 10% if only 5% of the resolved interpretations are used. This means that only a small proportion of conflicting boreholes has to be manually reviewed to obtain acceptable results.
- It is surprising, that the interpretationID has high significance. If only these interpretationIDs together with the boreholeIDs are used to train the model, an error rate of 10% is obtained. The importance of the interpretationID can be explained by the way information is entered into the database: The interpretationID is increased by one with every new interpretation. Geologists do not normally enter one interpretation at a time, but batches of borehole interpretations for a specific project. This means that the interpretationID contains hidden information like the geologist or the project properties which are significant (section V).
- As outlined in the previous paragraph, the interpretationID is a surrogate primary key, which is a significant feature for the ML model. However, the database may be physically reorganized resulting in the primary keys being updated, and therefore the interpretationID would lose its significance. In a robust database design the primary key should therefore not be tainted with any kind of semantics. Although the interpretationID is very important, the collection of all other features is sufficient for the ML model. Even if the interpretationID is omitted from the ML model, 5-fold cross-validation yields an error rate of 7.7% instead of 5.3% (see above) which is still acceptable.

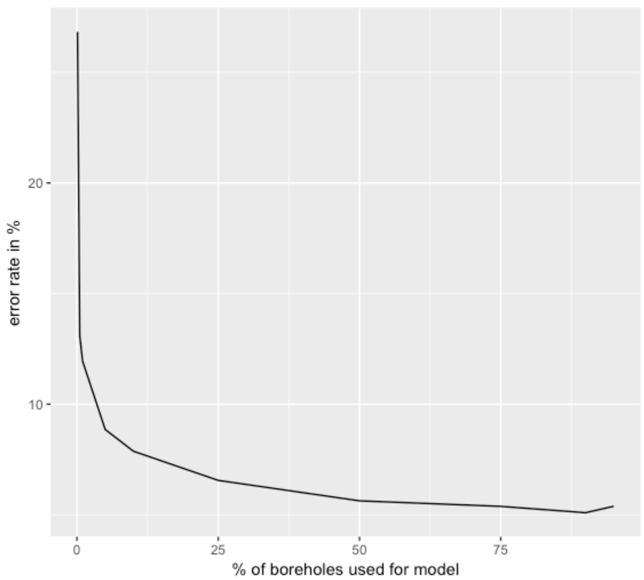


Fig. 7. TD Error rate in dependency of the fraction of used reviewed interpretations

## VIII. CONCLUSIONS AND FUTURE WORK

The ML model developed here has allowed the rapid creation of the rockhead surface, and allows the faster addition of new borehole data with a calculated risk. The ML quality is comparable with manual data entry, but the former is substantially faster and has a very low error rate (Fig. 7).

As explained in section I, geologists have to label borehole records 'RH' if rockhead was reached. This label may not have been entered or simply be wrong. This for example, creates a TD borehole instead of a RH borehole. Boreholes like these become an error source when used by the SDTM, giving a deeper value for the rockhead. Borehole logs could become simpler and more reliably coded, if the labelling is automated with a similar ML approach as the one discussed herein.

## COMPUTER CODE AVAILABILITY

The source code for the borehole selection algorithm described here is available from the authors and can be downloaded from https://github.com/lpiepmeyer/Geology-Drilling-Log-Classification.

## ACKNOWLEDGMENT

This paper is published with the permission of the Executive Director of the British Geological Survey (Natural Environment Research Council).

.

## REFERENCES


[1] Bellman R.E., 1957. Dynamic programming. Princeton University Press, Princeton.

[2] Borehole Records Collection, British Geological Survey, http://www.bgs.ac.uk/discoverymetadata/13480265.html

[3] Bowie, R. C., 1995. Data management in the National Geological Records Centre. In: Giles, J.R.A. (editor). Geological Data Management. Geological Society Special Publication No. 97, p. 117-125.

[4] Chen, T., Guestrin, C., XGBoost , 2016. A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, p. 785-794, San Francisco, CA, 2016.

[5] Gakis A., Cabrero P., Entwisle D.C., Kessler H., 2016. 3D geological model of the completed Farringdon underground railway station, in Crossrail Project: Infrastructure design and construction, v. 3: London, Institution of Civil Engineers Publications, p. 431–446.

[6] Kessler, H., Mathers, S., Sobisch, H.-G., 2009. The capture and dissemination of integrated 3D geospatial knowledge at the British Geological Survey using GSI3D software and methodology. Computers and Geosciences 35, p. 1311–1321.

[7] Kessler H., Mathers, S. J., 2004. From geological maps to models - finally capturing the geologists' vision. Geoscientist, 14/10 , p. 4-6.

[8] Lawley, R., Garcia-Bajo, M., 2009. The National Superficial Deposit Thickness Model. (Version 5). Nottingham, UK, British Geological Survey, Open Report, OR/09/049.

[9] Strange, P. J., Booth S. J.,. Ellison, R. A., 1998. Development of ‘rockhead’ computer-generated geological models to assist geohazard prediction in London. Geological Society, London, Engineering Geology Special Publications, 15, p. 409-414, 1998, https://doi.org/10.1144/GSL.ENG.1998.015.01.41

[10] Wei, S., Hengl, T., Mendes de Jesus, J., Yuan, H., Dai, Y., 2017. Mapping the global depth to bedrock for land surface modeling, Journal of Advances in Modeling Earth Systems 9. p. 65-88