

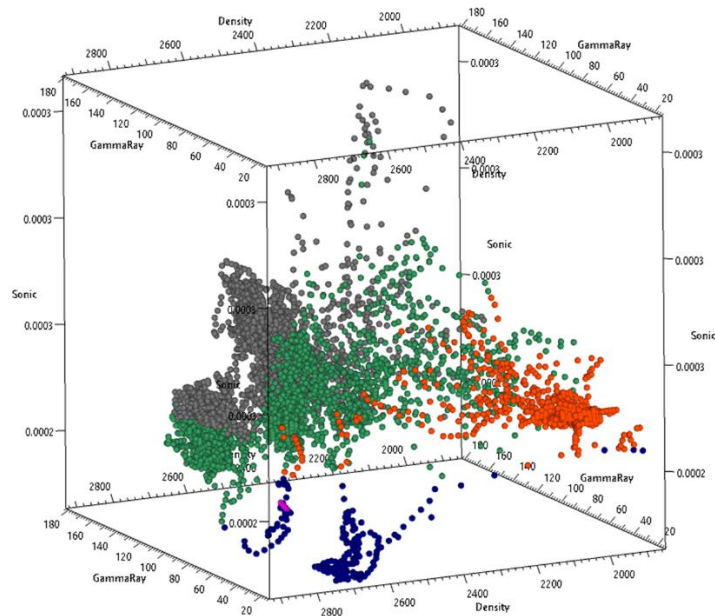


British
Geological
Survey

Derivation of lithofacies from geophysical logs: a review of methods from manual picking to machine learning

National Geoscience Programme

Open Report OR/21/006



BRITISH GEOLOGICAL SURVEY

NATIONAL GEOSCIENCE PROGRAMME

OPEN REPORT OR/21/006

The National Grid and other
Ordnance Survey data
© Crown Copyright and
database rights 2020.
Ordnance Survey Licence
No. 100021290 EUL.

Keywords

Geophysical logs, lithofacies,
stratigraphy, classification.

Front cover

Geophysical log parameters
plotted in 3 dimensions to
reveal lithofacies clusters.

Bibliographical reference

NEWELL ANDREW J, WOODS
MARK A, GRAHAM L ROMAINE &
CHRISTODOULOU VYRON 2021.
Derivation of lithofacies from
geophysical logs: a review of
methods from manual picking
to machine learning. *British
Geological Survey Open
Report, OR/21/006*. 43pp.

Copyright in materials derived
from the British Geological
Survey's work is owned by
UK Research and Innovation
(UKRI) and/or the authority
that commissioned the work.
You may not copy or adapt
this publication without first
obtaining permission. Contact
the BGS Intellectual Property
Rights Section, British
Geological Survey, Keyworth,
e-mail ipr@bgs.ac.uk. You
may quote extracts of a
reasonable length without
prior permission, provided a
full acknowledgement is given
of the source of the extract.

Maps and diagrams in this
book use topography based
on Ordnance Survey
mapping.

Derivation of lithofacies from geophysical logs: a review of methods from manual picking to machine learning

Andrew J Newell, Mark A Woods, Romaine L Graham & Vyron
Christodoulou

Contributor/editor

Andrew Kingdon

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of UK Research and Innovation.

British Geological Survey offices

**Nicker Hill, Keyworth,
Nottingham NG12 5GG**

Tel 0115 936 3100

BGS Central Enquiries Desk

Tel 0115 936 3143

email enquiries@bgs.ac.uk

BGS Sales

Tel 0115 936 3241

email sales@bgs.ac.uk

**The Lyell Centre, Research Avenue South,
Edinburgh EH14 4AP**

Tel 0131 667 1000

email scotsales@bgs.ac.uk

**Natural History Museum, Cromwell Road,
London SW7 5BD**

Tel 020 7589 4090

Tel 020 7942 5344/45

email bgs_london@bgs.ac.uk

**Cardiff University, Main Building, Park Place,
Cardiff CF10 3AT**

Tel 029 2167 4280

**Maclean Building, Crowmarsh Gifford,
Wallingford OX10 8BB**

Tel 01491 838800

**Geological Survey of Northern Ireland, Department of
Enterprise, Trade & Investment, Dundonald House,
Upper Newtownards Road, Ballymiscaw,
Belfast, BT4 3SB**

Tel 01232 666595

www.bgs.ac.uk/gsni/

**Natural Environment Research Council, Polaris House,
North Star Avenue, Swindon SN2 1EU**

Tel 01793 411500

Fax 01793 411501

www.nerc.ac.uk

**UK Research and Innovation, Polaris House,
Swindon SN2 1FL**

Tel 01793 444000

www.ukri.org

Website www.bgs.ac.uk

Shop online at www.geologyshop.com

Contents

Contents.....	v
1 Introduction.....	1
1.1 The importance of geophysical logs	1
1.2 Aim of the report.....	1
1.3 The problem of deriving lithofacies from geophysical log data.....	1
2 Manual picking	2
2.1 Basic principal and rationale.....	2
2.2 Manual classification of lithofacies example: The Chalk Group.....	3
2.3 Pros and cons of manual facies interpretation on geophysical logs in carbonate-dominated systems.....	13
3 Lithology determination using log cut-off values	13
3.1 Introduction	13
3.2 Required input data	14
3.3 Normalisation of well log data.....	14
3.4 Method.....	15
3.5 Selecting cut-off values	16
3.6 The importance of working within well regions.....	22
4 Mineral composition analysis by linear inversion of log data	23
4.1 Background.....	23
4.2 Practical application	24
5 Unsupervised cluster analysis	25
5.1 Background.....	25
5.2 Practical implementation of k-means clustering.....	26
5.3 Example of results.....	27
6 Facies classification using supervised machine learning	28
6.1 Background.....	28
6.2 Input data	29
6.3 Running the classification.....	29
6.4 Example of results.....	31
7 Conclusions.....	32
References.....	34

FIGURES

Figure 1. Some issues in the derivation of discrete lithofacies from geophysical log data	2
Figure 2. Typical track set-up for manual lithofacies digitising within a geophysical log package (WellCAD)	3
Figure 3. Gamma-ray and sonic log responses to the increase in marl content in the Flamborough Chalk Formation.....	4

Figure 4. Gamma-ray and sonic log responses to hardgrounds (1) and marl seams (2) in the Lewes Nodular Chalk Formation	5
Figure 5. Gamma-ray and resistivity log responses to major flint seams in the basal part of the Burnham Chalk Formation. *Note no scale available for Killingholme DG1 gamma-ray log. 6	6
Figure 6. Marly chalk facies in the Grey Chalk Subgroup indicated by significant increase in average gamma-ray log values.	7
Figure 7. Out-of-phase oscillations of gamma-ray and sonic logs that is characteristic of interbedded limestone/mudstone facies in the lower (Early - Mid Cenomanian) part of the Grey Chalk Subgroup.	8
Figure 8. Gamma-ray and sonic log responses to sandy chalk facies (with glauconite) at the base of the Chalk Group	8
Figure 9. Sonic and gamma-ray log responses to inferred phosphatic conglomerate (grey highlight) at the base of the Chalk Group	9
Figure 10. Strong inflection patterns across the boundary of the Grey Chalk and White Chalk subgroups form a characteristic 'eventbundle' for anchoring interpretations and correlations. 10	10
Figure 11. (a): Sonic and gamma-ray log 'event bundle' characterising Late Turonian Chalk. The bundle consists of a marl-rich, low resistivity/slow sonic interval, overlain by a succession that progressively builds in resistivity/sonic velocity, interrupted by inflections representing marl seams. The bundle is capped by a high resistivity/high sonic velocity hardground-rich interval; (b): Transitional log responses across the boundary of the Holywell Nodular Chalk and New Pit Chalk formations, marked by declining resistivity and increasing gamma-ray log values. Care is needed to consistently identify boundary levels within these transitional responses.	10
Figure 12. Differences in inflection detail between single point resistivity logs (red line) and continuous resistivity logs.	12
Figure 13. Chalk Group units recognised by geophysical log inflection patterns in the North Sea. 12	12
Figure 14. Illustration of how a log cut off value on a gamma-ray curve can be used to automatically subdivide a sequence of halite-rich strata (H) from halite-poor mudstones (M) in the Mercia Mudstone Group of the Winterborne Kingston borehole.....	14
Figure 15. Facies calculator in SKUA-GOCAD 2019.....	16
Figure 16. A lithology log (with a simple two-fold classification) created for a short section of Carboniferous Millstone Grit using a gamma-ray log. A histogram highlights the bimodal distribution created by interbedded sandstone and mudstone. The histogram shows distinct modes at 40API and 90API. A cut-off value of 70 API produces the lithology log shown in A. Note that the cut-off is a rather arbitrarily placed 'wall' in the lithofacies-component histogram.....	17
Figure 17. (A) Gamma-ray log of the Triassic St Bees Sandstone in NW England (B) histogram of the gamma-ray values (C) Outcrop of the St Bees Sandstones showing thinly interbedded sandstones and mudstones (D) gamma-derived lithology log.....	18
Figure 18. Common logs ranked from lowest to highest vertical resolution as follows. The absolute resolution of a given log varies depending on the tool, the sampling rate, logging speed, and processing methods.	18
Figure 19. Cross-plot of gamma-ray and sonic logs for the Mercia Mudstone Group of the Winterborne Kingston Borehole (see Fig. 7) reveals four high density (darker red) clusters more effectively than logs plotted as individual distributions.....	19
Figure 20. (A) Histogram of density log data for the Mercia Mudstone and Penarth Groups of the Winterborne Kingston borehole. The typical densities for a range of rock materials are	

shown on the plot. Cross-plot of sonic and gamma-ray logs with points coloured up by the density log. This reveals a cluster of high density anhydrite which was not apparent before including the additional log information on the 2D cross-plot.	20
Figure 21. 3D scatter plot of density, sonic and gamma-ray for the Mercia Mudstone Group of the Winterborne Kingston borehole with the points coloured by derived lithology information.	21
Figure 22. Resultant lithology log (central track, see A for key to colours).....	22
The main difficulty of using simple cut-off analysis is discriminating lithologies when there is large overlap in their log properties. Figure 23 illustrates an example from the Winterborne Kingston borehole in South Dorset where the dolomitic siltstones and mudstones of the Blue Anchor Formation show a remarkably similar log response to the Sherwood Sandstone at greater burial depth. While effort could be expended to discriminate these rock types by introducing additional logs or by applying more sophisticated clustering techniques a simple solution is to restrict the analysis to a shorter well region that excludes the overlapping rock types.	22
Figure 24. Gamma-ray and sonic log from Triassic strata in the Winterborne Kingston Borehole in South Dorset. The interbedded mudstones and dolomitic siltstones of the Blue Anchor Formation show a large overlap in their gamma-ray and sonic log response with the interbedded sandstones and mudstones of the Sherwood at greater depth. Rather than attempting to discriminate lithofacies by introducing more log data it is often simpler and more efficient to work in smaller well regions e.g. B rather than A.	23
Figure 25. K-means cluster analysis of normalised gamma-ray and sonic log data for the Mercia Mudstone Group of the Winterborne Kingston borehole. Seven centroids have been pre-selected corresponding to the inflection point on the elbow plot.	26
Figure 26. K-mean clustering as implemented in Mira Geoscience geoapps Jupyter-Notebook. Inbuilt Plotly visualisation tools provide a highly interactive environment. 2D and 3D cross-plots show normalised gamma-ray, sonic and density curve data for the Mercia Mudstone Group of the Winterborne Kingston Borehole.	27
Figure 27. Results of k-means cluster analysis (right-hand track) performed in geo-apps and imported to SKUA-GOCAD. Note both similarities and differences between the cut-off analysis and cluster analysis.	28
Figure 28. Part of the Jupyter notebook for facies classification (https://github.com/brendonhall/facies_classification) showing the typical mixture of runnable Python code, text providing instructions and commentary and interactive tabular output (or graphical plots).	29
Figure 29. Splitting the training dataset into training data and test data	30
Figure 30. 2D example of a linear boundary that maximises the margin between the closest pair of data points belonging to two classes. The support vectors are the points on the dashed lines. Modified from Wilimitis (2018).	31
Figure 31. Example of SVM classifier	32

TABLES

Table 1. Typical logging-tool response values for a selection of common rock types (after Rider, 2000).	13
Table 2. Quality checklist for geophysical log data in order to avoid working with erroneous data 14	14
Table 3. Table showing a section of the library of petrophysical properties used in the inverse modelling process. (DENS density, DT Compressional Delta Time (Slowness), DTS Shear Delta Time, VP P-Wave velocity), VS S-Wave velocity, PE Photoelectric Index (barns/electron).....	24

Table 4.	Graphical user interface of MinInversion with input data and results from the Mercia Mudstone Group of the Winterborne Kingston Borehole. The availability of only two input logs (Density and Sonic) restricts the analysis to two specified mineralogical components (halite and clay) and inverted porosity.....	25
Table 5.	Example format of training data. In this case a subset of the geophysical measurements (INPUT_FACIES) is used to classify the entire stratigraphic interval (SVM_CLASS). The training data would often be based on expert descriptions of core.....	29
Table 6.	Table summarising some of the advantages and disadvantages of the different methods trialled.	33

1 Introduction

1.1 THE IMPORTANCE OF GEOPHYSICAL LOGS

Geophysical logs are continuous recordings of a physical parameter along a borehole. Many different types of parameter can be recorded such as natural radioactivity (the gamma-ray log), acoustic slowness, electrical resistivity and bulk density (see Rider and Kennedy (2011) for a complete discussion). They also include imaging tools which record images of the internal circumferential surface of the boreholes measured using light or false colour images derived from resistivity, acoustic pulses or measurement of active and passive radioactivity.

Geophysical logs are routinely used for the characterisation of rock formations in the subsurface where they may be important as aquifers, hydrocarbon reservoirs or potential sites for the storage of hydrogen or carbon dioxide. Geophysical logs are necessary because only a small fraction of boreholes are sampled by mechanical rock coring. For reasons of time and cost, coring may be limited to short intervals of a formation of interest, discrete side-wall cores or often omitted entirely. In many boreholes the only physical record of the subsurface geology comes from drill-bit cuttings which are retrieved from the circulating drilling fluid. Cuttings samples, however, leave a very imprecise record of the formations. Geophysical logs bridge the gap between cuttings and core by bringing additional 'remotely-sensed' data which can be used to reconstruct rock type and other rock properties such as fracturing at a high resolution. The relatively low cost of geophysical logging relative to continuous rock coring means that these data are routinely collected.

Geophysical logs are probably the single largest source of continuous, quantitative stratigraphical data in the British Geological Survey archives and their potential as stratigraphic tools has long been known (Whittaker et al. 1985). They are, however, still relatively underexploited for this purpose particularly given the huge advances in log handling and analysis that has been made since these pre-digital studies. Geophysical logs offer the potential to not only pick gross lithostratigraphic contacts but to generate refined information on the internal lithological (or lithofacies) composition of a formation. The latter is the primary focus of this report.

1.2 AIM OF THE REPORT

The aims of this report are to document:

1. A range of methods that are currently used by the BGS stratigraphers to extract lithological information from geophysical logs (includes manual classification, cut-off analysis, mineral composition by linear inversion).
2. Alternative methods which, at present, are not routinely applied but are sufficiently practical and accessible that they could become important, including unsupervised (k-mean clustering) and supervised machine learning approaches.

The report does not aim or claim to be a complete inventory of all possible methods to derive lithological information from geophysical logs. The authors welcome correspondence and information on any additional methods that are available or emerging.

1.3 THE PROBLEM OF DERIVING LITHOFACIES FROM GEOPHYSICAL LOG DATA

The derivation of rock lithology or lithofacies using geophysical log data is complicated by a number of factors including (Figure 1) –

1. Lithological (or lithofacies) classes are assumed to be discrete categories but in reality often represent a continuum e.g. sandstone merging to a mudstone by a progressive decrease in sand content/increase in mud content.

2. Downhole sampling of lithologies by geophysical logging tools is imperfect and signals are often mixed (convolved) from multiple rock types where boundaries are crossed or the bed thickness is less than the relatively coarse resolution of the logging tools.
3. Logging tools operate in difficult environments where variations in borehole diameter, drilling muds, pore-filling fluids, logging speed, tool calibration, high temperatures and pressures, borehole wall caving which affect contact between the lithology and some logging tools and other factors introduce many uncertainties in the process of determining lithology from logs signals.
4. Geophysical log archives assembled over many decades are extremely heterogeneous in the quantity and quality of information they contain.
5. Like all analytical equipment geophysical logs have ranges of optimum measurement resolution. Lithologies that record outside that range (such as very low gamma-ray values in very pure limestones, like the White Chalk Subgroup or high gamma-ray values in highly radioactive mudstone like the Kimmeridge Clay Formation) become much less reliable and measurements suitable for only qualitative rather than quantitative techniques.

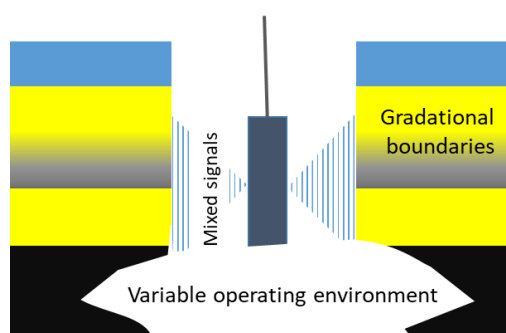


Figure 1. Some issues in the derivation of discrete lithofacies from geophysical log data

2 Manual picking

2.1 BASIC PRINCIPAL AND RATIONALE

The manual approach to lithofacies classification of geophysical logs is a digitising process. This is generally (and most efficiently) undertaken within log interpretation software such as WellCAD where lithofacies boundaries can be inserted, dragged and edited using a mouse within a lithofacies track (*Figure 2*). The lithofacies track is placed adjacent (or overlain) on other log tracks such as gamma-ray and sonic transit time. These input logs form the basis for picking the lithofacies boundaries. The input logs are not (necessarily) required for any quantitative analysis so it is possible to use scaled raster scans of the logs in place of the digital versions if required. This can be advantageous in negating the need for log digitisation where digital versions are unavailable

The primary advantage of this manual approach is that the geologist can place the lithofacies boundaries where they wish. This brings into play all of the experience and tacit knowledge that the geologist possesses on the formation under investigation. This might include the typical range of lithofacies that are found, the usual thickness of beds and any known recurring patterns of lithofacies that define cycles and rhythms within the formation. Moreover because the logs are not being used quantitatively the quality (and quantity) of the log data is less critical than other methods. The geologist can use their experience to interactively recognise, correct and account for poor log quality throughout the classification process. The disadvantages of this technique is that the geologist's preconceived ideas of what they expect to see can influence their decision making and can lead to over interpretation that identifies subtle variations that are not effectively demonstrated at the resolution of logging tools or conversely missing subtle facies variations. Such issues can be effectively remediated by robust peer-review.

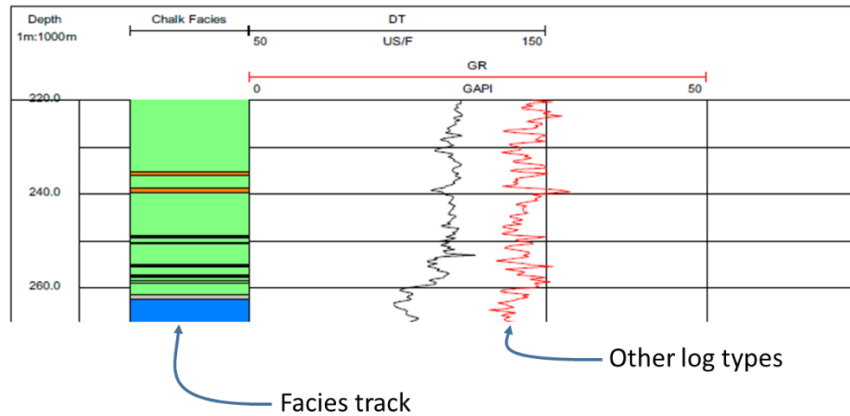


Figure 2. Typical track set-up for manual lithofacies digitising within a geophysical log package (WellCAD)

The technique can be particularly applicable to some superficially 'homogeneous' carbonate-dominated systems such as the Chalk Group where log responses are often subtle compared to other types of formation (e.g. siliciclastic). Moreover units such as the Chalk Group of the UK are penetrated by boreholes drilled for many different purposes (hydrocarbon, water, geotechnical, research) resulting in extremely heterogeneous geophysical log suites of vastly varying quality. This can hamper blanket automated recognition of lithofacies within the Chalk. To illustrate some of these issues and demonstrate the value of manual approaches in these types of scenario the Chalk is considered in more detail below.

2.2 MANUAL CLASSIFICATION OF LITHOFACIES EXAMPLE: THE CHALK GROUP

2.2.1 Optimal geophysical log combinations

The most widely available geophysical logs that are useful for interpreting stratigraphy and facies in the Chalk are gamma, resistivity and sonic. Induction logs are also valuable for deriving resistivity profiles for air-filled borehole intervals that cannot be recorded by resistivity tools, and digital image logs (where they exist) allow direct feature observation and refine understanding of corresponding geophysical responses. Ideally, a combination of gamma-ray *and* resistivity/sonic logs is desirable for maximum confidence of interpretation, but this rarely exists for the total depth range of Chalk in boreholes. In deep boreholes drilled for hydrocarbons exploration, the only logs typically run in the Chalk Group are caliper and low-resolution (often through casing) gamma-ray. For these boreholes, the availability of digital log data and ability to adjust the scale range is key to optimising interpretation value. For shallower boreholes (typical those less than 150 m deep), which in the Chalk Group have usually been drilled in connection with groundwater or major civil engineering investigations, there is usually complete coverage of gamma-ray logs and partial coverage of resistivity/sonic logs, the latter typically being restricted by the location of the water table and/or borehole construction (particularly borehole lining). However these often drilled by civil engineering focussed contractor recording in non-standardised units with quality standards below those adopted for oil industry work making quantitative assessment very difficult. In some instances, caliper logs may be a useful guide to stratigraphical boundaries (particularly where these correspond to changes in cementation), and for identifying particular types of feature (e.g. fracturing).

Overarching approach to classifying a thick carbonate formation

For the purpose of geophysical log interpretation, the stratigraphy of the Chalk (Mortimore, 1986) can be thought of as comprising broad-shifts in mud content and/or cementation, within which is embedded a detailed framework of marker-beds. The marker-beds include the following:

- Very thin calcareous mudstones (marls) including bentonites (typically 50 – 100 mm thick)

- Hardgrounds, locally enriched with pyrite, glauconite and phosphate
- Sponge beds (cemented, typically with sponges preserved as iron-rich mesh-works)
- Flint bands (named flints typically ca. +30 cm & closely spaced/semi-continuous)

The marker-beds and facies units each have particular geophysical log features that allow their recognition, and these combine into packages of markers and facies defining lithostratigraphical units with distinctive associations of inflection patterns.

2.2.2 Recognition of key marker-beds

The methodology for interpreting marker-beds in the Chalk is described below:

Marls: Typically seen as localised sharply developed peaks in gamma-ray log values, corresponding with localised sharp drops in the resistivity log profile, and localised sharp increases in interval transit time on sonic logs. Geophysical log response varies with marl-type, from relatively stronger, higher amplitude signatures in solid closed marl seams and bentonitic marls, to weaker responses in anastomosing plexus marls that intercalate significant chalk sediment, and thin marl wisps and coatings on stylolitic surfaces. In Northern England, the development of high concentrations of thin marl seams in the Flamborough Chalk is signalled by an increase in “spikiness” of gamma-ray logs, and corresponding response of sonic logs (Woods, 2018; Figure 3).

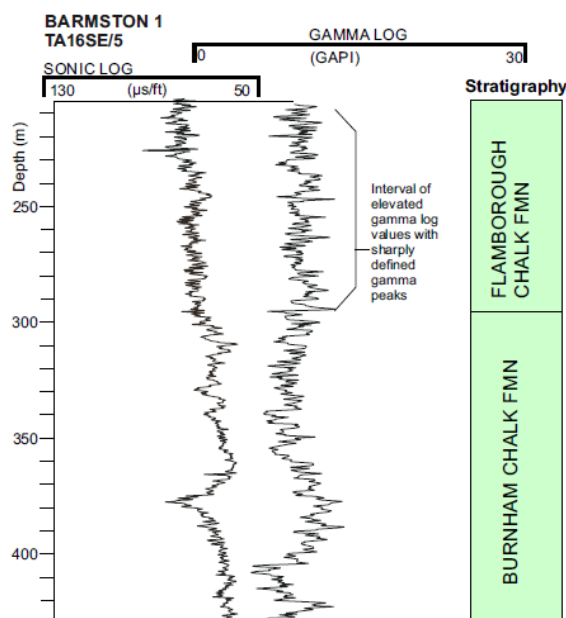


Figure 3. Gamma-ray and sonic log responses to the increase in marl content in the Flamborough Chalk Formation

Hardgrounds: Hardgrounds correspond to localised, very sharp increases in resistivity log values and localised sharp reductions in interval transit time. Corresponding gamma-ray logs typically show sharp increases because of the tendency of these features to concentrate iron and phosphate minerals (e.g. glauconite & apatite). Where both gamma-ray and resistivity/sonic logs are available, then the coupled response described above is strong evidence for hardground identification (Figure 4). When only a single log type is available, significantly more caution is required, and interpretations should take account of the stratigraphical context provided by interpretations of adjacent intervals and any laterally related boreholes. The shape of the peak in gamma-ray log values is typically somewhat different from marl seams, usually broader and blunter (reflecting the fact that hardgrounds are typically thicker (dm) than marls (cm), with much higher amplitude. Sharp peaks on resistivity and sonic logs can also be produced by very thick flint seams (typically with low gamma-ray values), or where flint seams are highly concentrated, and might be difficult to distinguish from hardgrounds if no gamma-ray log is available.

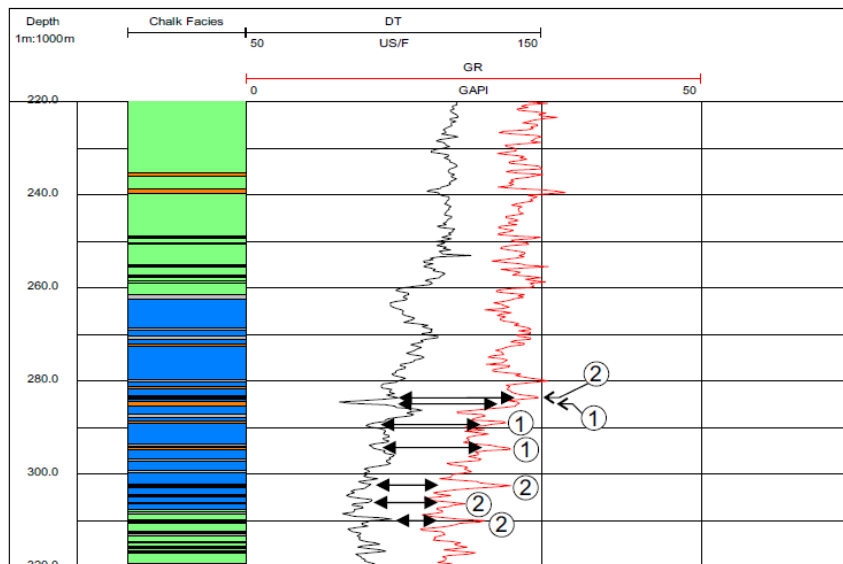


Figure 4. Gamma-ray and sonic log responses to hardgrounds (1) and marl seams (2) in the Lewes Nodular Chalk Formation

Hardgrounds imply significant stratigraphical *omission*, but other thin, hard cemented units in the Chalk that are not hardgrounds, and likely represent reductions in the *rate* of sedimentation (e.g. sponge beds), may have a similar geophysical log response. Typically both the gamma-ray and resistivity/sonic responses are lower amplitude compared to hardgrounds (because cementation is less extreme & mineral enrichment is limited to thin films of iron-oxide/hydroxide associated with sponge preservation), but understanding this relative difference can be difficult if undoubted hardground signatures are not available for comparison. Again, the stratigraphical context provided by interpretation of adjacent intervals and laterally related boreholes will be an important guide to feature discrimination.

Flints: Although flint is widely distributed in the Chalk, and there are a number of important named flint marker-beds that can be traced laterally for 100s km. However, it is often difficult to recognise their individual signatures on geophysical logs. This may partly reflect the brittle character of flint, and its tendency to fragment during coring rather than providing a clean, solid surface for instrument detection. There are a few examples of where recognition has proved possible, particularly in the Chalk of northern England, where thick (+30cm) bands of laterally continuous flint form a series of named markers at the base of the Late Turonian Burnham Chalk Formation (Fig. 3). These flints can be identified by sharply defined high resistivity / 'fast' sonic (low interval transit time) peaks and low gamma-ray log values.

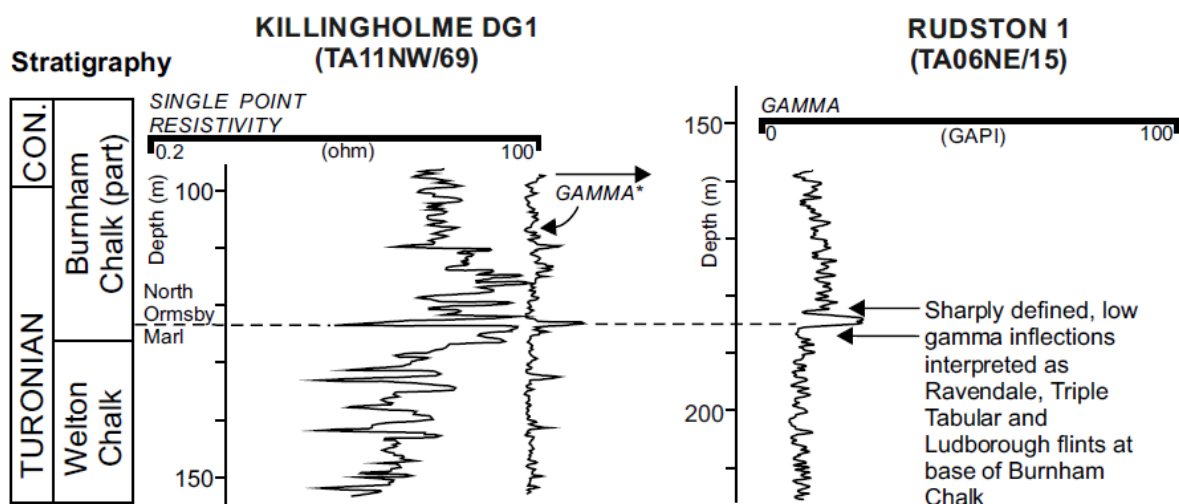


Figure 5. Gamma-ray and resistivity log responses to major flint seams in the basal part of the Burnham Chalk Formation. *Note no scale available for Killingholme DG1 gamma-ray log.

2.2.3 Manual log interpretation of facies units

The creation of a regional geological model of the Chalk across southern England (Woods et al., 2016; Newell et al., 2018) included information about chalk facies units derived from borehole geophysical logs. In addition to facies types represented by marker-beds (see above), these included:

- Hard chalk
- Marly chalk
- Mudstone
- Limestone
- Phosphatic chalk

In the above list, 'Limestone' in this context refers to cemented carbonate units that are not dominated by nanofossils and lack typical chalk fabrics. There is a close association of the limestone and mudstone facies, and in this report (below) they are discussed as 'Interbedded limestone/mudstone facies'. Although not forming part of the previous facies modelling work, it is also possible to identify very flint-rich chalk, also described below.

Hard Chalk: Strong lateral shifts in resistivity and sonic logs to higher values (resistivity) and fast interval transit times (sonic), that are sustained over 10s of metres of strata, can generally be used to infer the presence of hard chalk. In combination with marker-bed information, and knowledge of coeval successions at outcrop and in boreholes, inference of 'nodular' fabrics is possible. Interpretational ambiguity can be caused by high concentrations of flint, which collectively may cause similar but usually less marked shifts in resistivity and sonic profiles. If the shift itself occurs gradually, over 10s of m, then in the context of the known mechanisms that typically influence broad patterns of hardness in the Chalk (relative sea level change and basin architecture), a response to flint content should be suspected. Additionally, regional knowledge and stratigraphical context can provide valuable supporting evidence about the potential influence of flint on geophysical logs.

Flint-rich chalk: Some intervals in the Chalk contain high concentrations of flint over narrow stratigraphical intervals that are laterally persistent, particularly in the Late Turonian (e.g. Brandon Flint Series; Mortimore & Wood, 1986). Individual flints and flint-rich intervals typically correspond with very low background gamma-ray log values, serving to distinguish the corresponding fast sonic and elevated resistivity log responses from highly cemented chalk intervals.

Marly Chalk: This facies represents chalk with a greater background mud content, but is distinct from a marl seam where there is greater separation of mud and chalk phases. On geophysical logs, marly chalk is represented by an increase in average gamma-ray log values with correspondingly lower resistivity/slower sonic responses. In the Chalk, the admixture of mud and chalk usually results in a relatively poorly cemented lithology. In contrast, carbonate systems that are not dominated by nannofossils, the equivalent facies (wackestone) is typically highly cemented. The Late Cenomanian Grey Chalk Subgroup provides a good example of marly chalk facies (*Figure 6*), in which massive-bedded units of creamy-grey chalk have significantly higher gamma-ray log values than the purer chalk deposited in the Early Turonian.

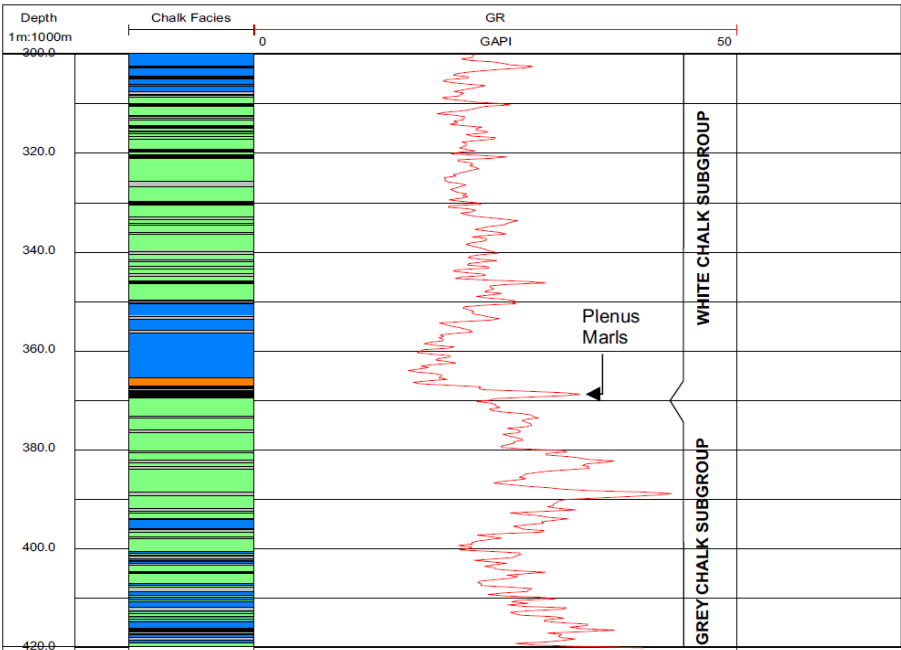


Figure 6. Marly chalk facies in the Grey Chalk Subgroup indicated by significant increase in average gamma-ray log values.

Intercalated mudstone/limestone facies: This facies is typical of mixed carbonate systems, its development in the lower part of the Chalk Group (Grey Chalk Subgroup) reflecting the influence of Milankovitch cycles immediately prior to deep flooding of continental shelves (and corresponding restriction of clastic input) in the earliest Turonian (Gale et al., 1999). Gamma, resistivity and sonic logs show a pronounced, high frequency, out-of-phase oscillations over wavelengths of 1 – 5 m, with high gamma-ray log intervals corresponding with low resistivity/fast sonic response (Figure 7).

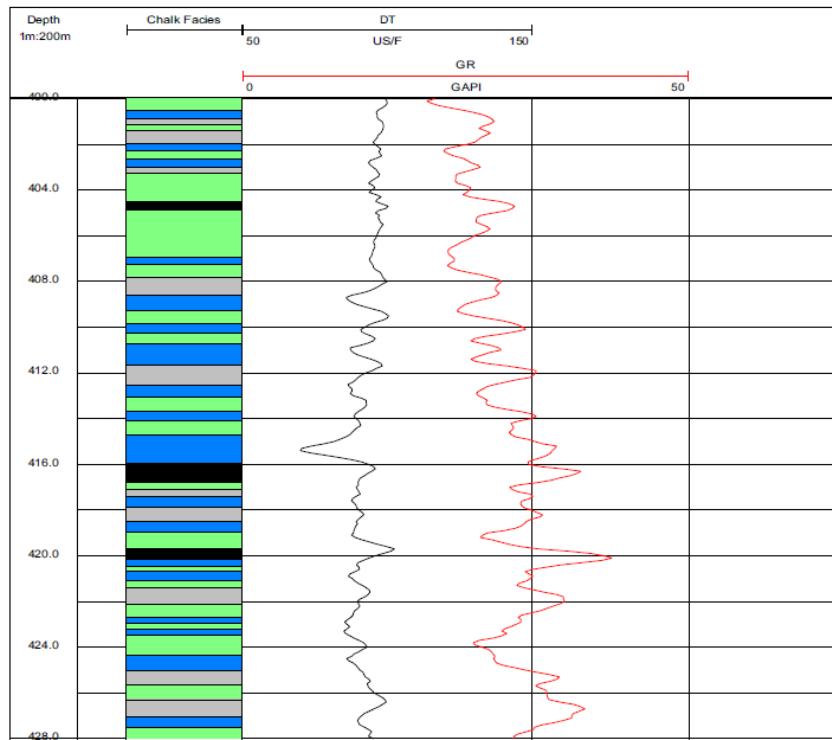


Figure 7. Out-of-phase oscillations of gamma-ray and sonic logs that is characteristic of interbedded limestone/mudstone facies in the lower (Early - Mid Cenomanian) part of the Grey Chalk Subgroup.

Sandy (including glauconitic) chalk: This facies is exemplified by intervals in the basal and lower parts of the Chalk Group (Grey Chalk Subgroup), and corresponds with significantly elevated gamma-ray log values that may extend for 5 – 10 m. The strong gamma-ray log response is driven by glauconite and dispersed mud content, usually boosted by significant quantities of phosphatic clasts. Patterns of log serration are a response to the piping-down of more chalky sediment infilling burrow systems that are usually pervasive. Moderately elevated resistivity and intermediate interval transit times on sonic logs reflect the significant sand content (*Figure 8*), but the responses of these logs are generally less amplified than for cemented chalk and hardground intervals.

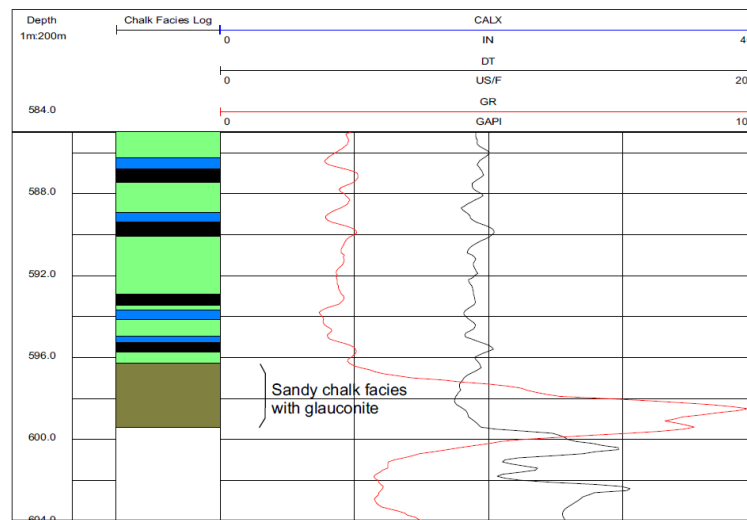


Figure 8. Gamma-ray and sonic log responses to sandy chalk facies (with glauconite) at the base of the Chalk Group

Phosphatic Chalk: In the Chalk, this facies may be developed as a conglomerate of phosphatic nodules and phosphate-encrusted limestone cobbles, usually less than a metre thick, or as intervals of fine granular phosphate, sometimes more than 10 m thick. Where developed as a phosphatic conglomerate, the geophysical log response can be hard to separate from that of a hardground, since in both gamma-ray log values tend to be sharply elevated and coincide with high resistivity/fast sonic log responses (*Figure 9*). In many ways these intervals resemble hardgrounds, being associated with condensed sedimentation with nodules/cobbles bored and encrusted with marine serpulids and molluscs. A distinction is that erosion and winnowing appear to dominate over in-situ sediment lithification that characterises hardgrounds. Understanding the stratigraphical context of these inflection patterns, and knowledge about the likely distribution of erosion surfaces, provides a valuable guide to their correct interpretation.

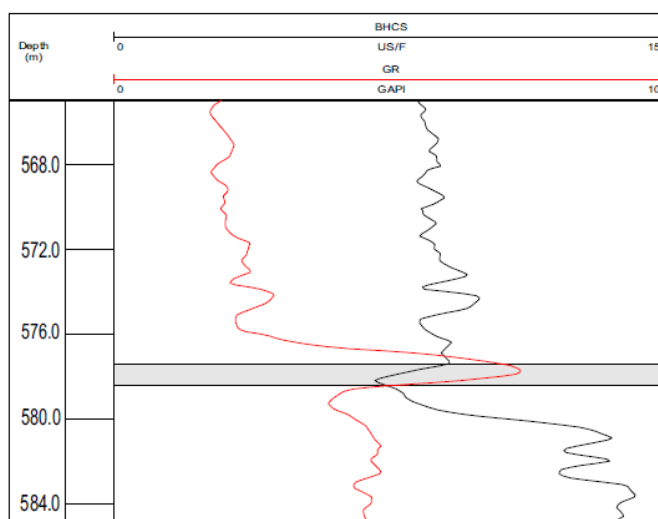


Figure 9. Sonic and gamma-ray log responses to inferred phosphatic conglomerate (grey highlight) at the base of the Chalk Group

Relatively thick intervals (+15 m) of granular phosphatic chalk have recently been reported and geophysically logged along the alignment of a proposed road tunnel adjacent to Stonehenge (Mortimore, 2014, fig. 3.26c). The gamma-ray log signature rapidly builds out over a few metres, from background values of 40 API in the host chalk, to 140 API in the phosphatic interval. The signature shows high frequency oscillations and longer wavelength variability, the latter probably reflecting detail of the internal sedimentary geometry. In Chalk, there is usually a close association between the development of these features and patterns of channelling developed in response to local structure (Mortimore et al., 2017). These may be evident on seismic profiles (Evans and Hopson, 2000).

2.2.4 Manual log interpretation of facies packages that define formational units

Onshore UK Chalk Group stratigraphy was rationalised by Rawson et al. (2001), and comprises a highly distinctive arrangement of marker-beds and facies units (Mortimore, 1986; Wood & Smith, 1978; Whitham, 1991, 1993) into broader packages with particular geophysical log patterns. These facies packages correspond with Chalk formational units, and the methodology for their recognition is described below.

1. Look for any broad shifts in log profile that act to anchor the interpretation at one or more levels. A good example is the Cenomanian/Turonian boundary, at the base of the Holywell Nodular Chalk, where the mud-rich Plenus Marls are overlain by the highly cemented Melbourn Rock. This stark contrast in physical properties is marked by a major inflection in both resistivity/sonic and gamma-ray logs, and gamma-ray values below the Plenus Marls are consistently higher than above because of higher overall mud content (*Figure 10*). Another distinctive 'event bundle' occurs in Late Turonian Chalk successions, where a closely spaced successions of marls and hardgrounds in the lower part of the Lewes Nodular Chalk produce a series of sharp oscillations in resistivity/sonic and gamma-ray

log signatures (Figure 11A). Confirm that this is consistent with any local or regionally available control data (e.g. published log interpretations, SOBI borehole data, and regionally relevant thickness data).

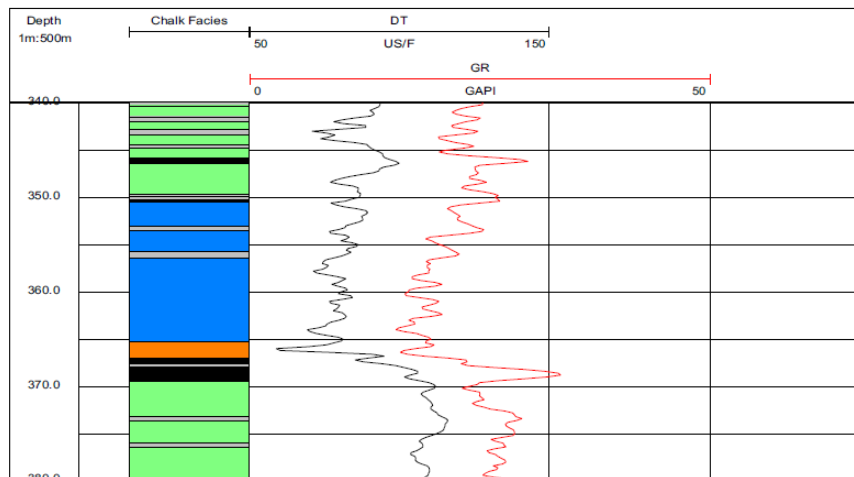


Figure 10. Strong inflection patterns across the boundary of the Grey Chalk and White Chalk subgroups form a characteristic 'event-bundle' for anchoring interpretations and correlations.

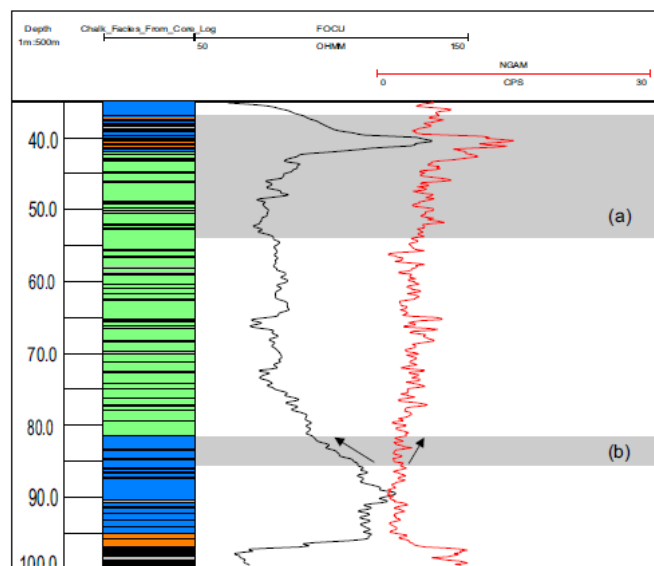


Figure 11. (a): Sonic and gamma-ray log 'event bundle' characterising Late Turonian Chalk. The bundle consists of a marl-rich, low resistivity/slow sonic interval, overlain by a succession that progressively builds in resistivity/sonic velocity, interrupted by inflections representing marl seams. The bundle is capped by a high resistivity/high sonic velocity hardground-rich interval; (b): Transitional log responses across the boundary of the Holywell Nodular Chalk and New Pit Chalk formations, marked by declining resistivity and increasing gamma-ray log values. Care is needed to consistently identify boundary levels within these transitional responses.

2. Look for finer-scale marker-bed event bundles. These second-order features are usually slightly less obvious because they are either changes to a single log type, or trends in the pattern or amplitude of log values through a broader depth range. Typically these events only become more meaningful for interpretation in the context of the parameters set by the first order log features. Examples of second-order events are: 1) increase in gamma-ray log serrations 15 – 30 m above the base of the Plenius Marls (indicative of the marl-rich New Pit Chalk); 2) significant low-resistivity punctuations coincident with high gamma-ray peaks in the middle and higher part of the New Pit Chalk, coinciding with increased

frequency and thickness of marl seams; 3) discrete series of typically 6 gamma-ray log peaks just above the base of the Seaford Chalk.

3. Where interpreting a series of spatially associated boreholes, explore the detailed pattern of inflections between the anchor points formed by marker-beds in adjacent boreholes. This has three important functions: 1) patterns that closely match enhance confidence of interpretation; 2) trends in the spatial organisation of these inflections provides refined understanding of where sedimentary packages are expanded or condensed locally within successions; 3) where there is variability in the quality or availability of log data for different boreholes, but high confidence in likely correlation, then flattening adjacent boreholes on a common stratigraphical datum can provide increased information about the likely position of stratigraphical boundaries.
4. Confirm that stratigraphical picks are consistent with regional structural trends and up-to-date geological map data from the borehole vicinity. Consider the availability of outcrop biostratigraphical data to constrain near-surface log interpretations. If inconsistencies are apparent, determine if there are other data to support a stratigraphical/structural explanation for this. If not, re-evaluate correlation and decide if the error is likely to be with first or second-order features. If first-order features are suspected to be incorrectly assigned, consider the wider implications of this for other borehole correlations and if necessary re-evaluate the correlation of all first-order features in all boreholes.
5. Refine interpretation of stratigraphical boundaries to ensure consistency. In the Chalk Group, consistency of interpretation is affected by two issues: 1) the transitional nature of facies changes at stratigraphical boundaries, and 2) log resolution. In stratotype sections the boundaries of Chalk units are named marker-beds, but these are not always easily recognisable as discrete entities on borehole logs, and the associated shifts in facies patterns that they signal are rarely precisely coincident. In such cases, trends in log patterns, for example increasing/declining resistivity/gamma-ray profiles become important for boundary interpretation, and a consistent decision needs to be made about where along the trend line a boundary is likely to be located.

Inability to identify stratigraphy-defining marker-beds can be associated with geophysical data of particular vintage. A significant number of single point resistivity logs in BGS archives record data at widely spaced depth increments. Although the trends are still meaningful for stratigraphical interpretation, multiple marker-beds that are usually separately resolved on continuous logs may be covered by a single inflection on vintage logs. This problem can be mitigated if gamma-ray logs are also available. A good example of facies transition at a stratigraphical boundary is the junction of the Holywell Nodular Chalk and New Pit Chalk (Figure 11B), and Figure 12 compares the difference in marker-bed resolution revealed by vintage single point and modern continuous resistivity logging. Therefore log age and quality is a proxy of data uncertainty; borehole with less uncertain data should be prioritised and the most uncertain data considered for exclusion.

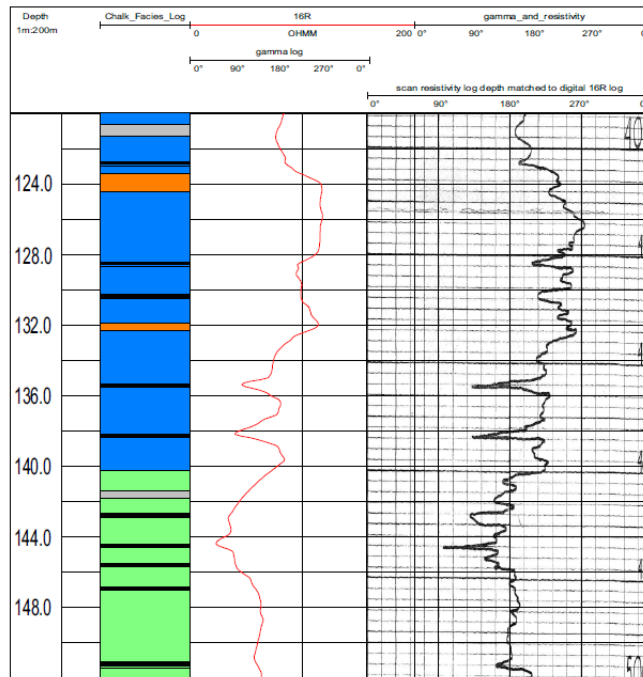


Figure 12. Differences in inflection detail between single point resistivity logs (red line) and continuous resistivity logs.

2.2.5 Manual facies interpretation in offshore Chalk successions (North Sea)

Manually picked geophysical log responses underpin the stratigraphical nomenclature for the Chalk in the North Sea (here including strata of earliest Palaeocene age) (Johnson & Lott, 1993; Lott & Knox, 1994; Gradstein & Waters, 2018; *Figure 13*). Formations are recognised by broad increases or decreases in gamma-ray and sonic log values, and by recognition of particular patterns of inflections in a series of reference boreholes, with microfossils (foraminifera, dinoflagellate cysts & nannofossils) where available providing a correlatable framework of age-related marker-horizons. Patterns of intraformational log variability (that are not stratigraphically formalised) have been described in the context of facies variation, for example, significantly high gamma-ray log responses in the basal part of the Ekofisk Formation, corresponding with a mud-rich chalk interval (Johnson & Lott, 1993). The logs illustrated by Johnson and Lott (1993) and Lott and Knox (1994) suggest significant potential for identifying facies subdivisions, and Mortimore (2014, figs 5.5 – 5.9) highlighted some of these on logs that he illustrated. In the context of the history of North Sea Chalk sedimentation (Mortimore, 2014), these facies are likely to be significantly more variable and diachronous than for onshore Chalk successions.

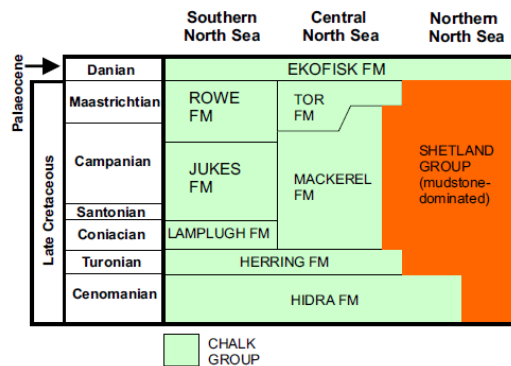


Figure 13. Chalk Group units recognised by geophysical log inflection patterns in the North Sea.

2.3 PROS AND CONS OF MANUAL FACIES INTERPRETATION ON GEOPHYSICAL LOGS IN CARBONATE-DOMINATED SYSTEMS

For carbonate systems like the Chalk, the main advantage of manually interpreting carbonate facies on geophysical logs is that it overcomes the problem of subtle log-responses to facies contrasts, and the difficulty this creates for establishing appropriate cut-off values for automated systems of facies interpretation. For the Chalk Group, recognising changes in the frequency and overall pattern of low amplitude geophysical log responses is the most important aspect of facies and stratigraphy interpretation. This conclusion likely reflects the fact that the Chalk was deposited on a deeply flooded shelf, with more subtle facies responses to sea level oscillation than might be predicted for a shallow carbonate platform or ramp setting. The main disadvantages of the manual picking methodology are that it is usually time-consuming; requires detailed knowledge of stratigraphical trends and regional facies variation, and may be guided by a conceptual model that is flawed and over-looks the significance of log responses that are inconsistent with this model. Where large-scale log responses are developed in carbonate facies systems, there is a potential role for using threshold cut-off values to highlight first-order anchor points for log interpretation and correlation.

3 Lithology determination using log cut-off values

3.1 INTRODUCTION

The geophysical log response of different rock types generally occurs within a defined range of values that (partly) reflects the bulk chemical composition of the rock framework and cementing materials. Table 1 shows a selection of typical log values for some common rock types taken from the more complete listing of Rider (2000). Although the range of log response values for a particular rock type are often large and there are frequently substantial overlaps in response between different rock types, it is often possible to compute a continuous lithology log in a semi-automated way by using defined ranges on one or more geophysical logs. The ranges are defined by one or more 'cut-off' values.

Table 1. Typical logging-tool response values for a selection of common rock types (after Rider, 2000).

Rock type (matrix only)	Gamma ray (API)	Resistivity (ohm m ² /m)	Sonic transit time (µs/ft)
Halite	0	<10 ⁴ - 10 ¹⁴	66.7-67
Anhydrite	0-12	10 ⁴ - 10 ¹⁰	50
Limestone	18-100	80-6 x 10 ³	47.6-53
Sandstone	18-160	Up to 1000	53-100
Mudstone	24-1000	0.5-1000	60-170

Figure 14 provides a simple example where a cut-off value of 30 API on a gamma-ray curve could be used to subdivide halite-rich sedimentary rocks (with a low gamma-ray response) from those that are predominantly mudstone. The cut-off value creates a continuous lithological log using a method that does not require any manual digitising.

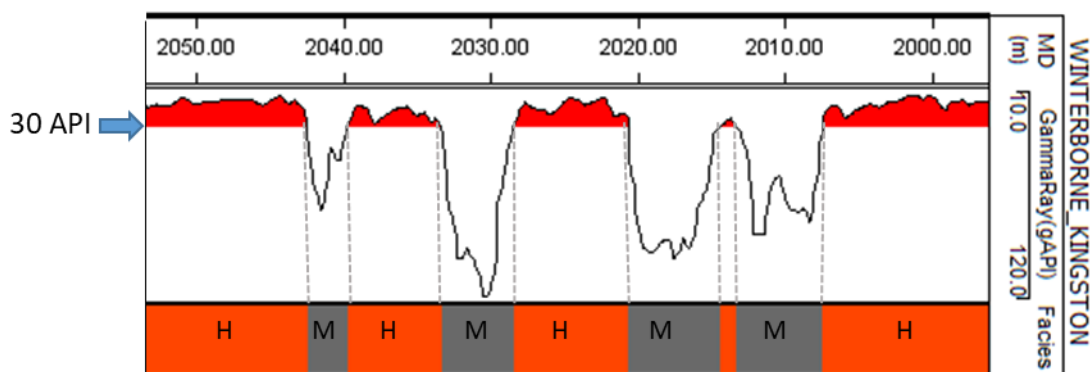


Figure 14. Illustration of how a log cut off value on a gamma-ray curve can be used to automatically subdivide a sequence of halite-rich strata (H) from halite-poor mudstones (M) in the Mercia Mudstone Group of the Winterborne Kingston borehole.

3.2 REQUIRED INPUT DATA

Cut-off methods can be undertaken on almost any geophysical log data, from a single curve which may have been digitised from a vintage paper record to a modern suite of digital logs. While the input logs should ideally pass the quality control checks listed in Table 2 this is often not possible and, with due care, does not necessarily preclude the application of cut-off lithological classification. Most projects undertaken in BGS unavoidably combine log data that has been acquired across many decades, was run to meet the needs of many different geoscience sectors (e.g. water, oil, coal), is of vastly varying quality and quantity, and may have been received and archived with little supporting metadata.

Table 2. Quality checklist for geophysical log data in order to avoid working with erroneous data

Mud Resistivity Properties (Rm, Rmf) and corresponding temperature reported	✓
Proper logging scales chosen and reported for all logs	✓
Caliper log does not show large variations	✓
Resistivity curves do not read less than zero	✓

The risks and uncertainties associated with cut-off analysis using sub-optimal input data can be greatly reduced by starting projects with as much background geological information as possible. This should include information on the expected number of lithologies, the likely proportion of each lithology within the stratigraphic interval of interest, their typical mineralogy (with particular regard to the presence of radiogenic minerals such as glauconite) and the expected range of bed thickness and other parameters of the stratal architecture. The source of this information might be from the well itself in the form of cuttings descriptions, short cored intervals or sidewall cores, or from image logs (optical, acoustic or resistivity). Additional information on the geological formation under consideration might come from regional understanding gained from outcrop descriptions and adjacent boreholes. Finally, a general understanding of the depositional system under investigation can help control and verify the types, stacking order and bed thickness of lithologies that are being generated by the applied cut-offs.

3.3 NORMALISATION OF WELL LOG DATA

Normalisation of well log data is a routine process within petrophysical workflows and is used to correct for inconsistencies between wells in the distribution of values recorded by a log for a particular lithology (Shier, 2004). These variations can arise for many different reasons such as

incorrect tool calibrations, varying tool vintage and changes in downhole environmental conditions between wells.

Some form of normalisation will be required in nearly all cases where cut-off analysis is attempted as a batch process. Exceptions might occur where a cluster of similarly drilled and constructed wells have been logged by one service company who have applied a consistent set of corrections and calibration procedures to the log data.

The normalisation process is a re-scaling or re-calibration procedure so that all logs in the set under consideration are consistent. A defined set of cut-off values in one well should therefore generate the same lithological classification in another. Normalization is commonly applied to gamma-ray logs, but can also be applied to all logs that sample bulk rock properties such as neutron porosity, bulk density, sonic and spontaneous potential logs. Resistivity logs, which are heavily affected by dynamic fluid properties, are generally not normalized unless there is a sufficient reason to do so (Shier, 2004).

There are a number of different approaches to log normalisation –

Normalisation might be carried out by using a simple linear shift of the log curve. This applies particularly where data has been recorded in different units.

$$Curve_{norm} = Curve + Shift$$

In unity-based normalisation curves are rescaled to the minimum and maximum values, bringing all the values within the range [0,1]. The output values are dimensionless and any reference to the original absolute magnitude of the log response is lost.

$$Curve_{norm} = \left(\frac{CurveValue - CurveValue_{min}}{CurveValue_{max} - CurveValue_{min}} \right)$$

An alternative approach involves selecting a reference well (*Ref*) from within the dataset under consideration. Curves values from other wells (*Curve*) are then stretched and squeezed to match the minimum and maximum values of the corresponding curve in the reference well. This approach is frequently applied to gamma-ray data (Shier, 2004).

$$Curve_{norm} = Ref_{min} + (Ref_{max} - Ref_{min}) * \left(\frac{CurveValue - CurveValue_{min}}{CurveValue_{max} - CurveValue_{min}} \right)$$

Anomalous outliers in curve values will adversely impact the outcome of normalisation processes using minimum and maximum values. For this reason it is common practise to use percentiles in the frequency distribution (e.g. 5th and 95th) in place of minimum and maximum values. A quick check on the descriptive statistics of the log data under consideration prior to normalisation will establish whether this is necessary.

In addition to normalising curve data between wells, the types of approach outlined above can also be useful to correct data within a single well. For example, normalising a gamma-ray curve that over some of its length has an attenuated signal due to being run through steel casing (Quatero et al. 2014).

Finally it should be noted that an alternative to normalisation is to simply avoid batch processing and work on a well-by-well basis, generating a bespoke set of cut-off criteria for each well. This may sometimes be the most efficient approach and has the advantage of preserving the integrity (including all of the inherent errors) of the original log data. There is always a risk that normalising log data will remove or alter real geological signal and it should be approached carefully.

3.4 METHOD

Applying cut-off values to a well log is a relatively simple process and most geophysical log handling software has built in functionality to define cut-offs and output a derived lithological log. Figure 15 illustrates the facies calculator in SKUA-GOCAD 2019 as an example. Lithological classifications are achieved in four main steps.

- 1. Well selection.** Wells can be selected and worked on either individually or in batches of any size.
- 2. Well Region.** Where the well length is greater than stratigraphic interval of interest it is generally advantageous to work within a corresponding sub-selection of the log data. Creating and working within pre-defined well 'regions' (in the terminology of SKUA-GOCAD) avoids the need to fragment and subsample the original continuous log dataset.
- 3. Classification.** A pre-defined table is selected which lists the lithologies or sedimentary facies that will be generated by the cut-offs. Each facies is assigned a numerical value and a colour or pattern fill. This table can be edited and augmented as the project proceeds.
- 4. Input properties.** One or more logs are selected that will be used in the analysis. The logs can be either primary measurements or logs that have been derived and modified from these primary measurements (e.g. normalised curves or shale volume V_{shale}).

The screenshot shows the 'Facies Calculator' window. It includes several input fields: 'Object' (WINTERBORNE_KINGSTON), 'Region' (MerciaMudstoneGroup), 'Classification' (Facies), 'Facies property' (Facies), 'Facies log set' (LITH_Facies), and 'Criteria template' (1). Below these is a 'Property selector' table with columns for Property, Minimum, Cutoffs, Maximum, and Classification (Unit). The table contains two rows: GammaRay and Sonic. Below the property selector is a detailed classification table with columns for Lower Limit, Logic, Property, Logic, Upper Limit, and Facies. The table shows various ranges and logic combinations for GammaRay and Sonic properties, resulting in different lithologies like Carbonate, Anhydrite, Halite, DolomiticSiltstone, and Mudstone.

1. Well (single or multiple batch)

2. Defined well interval e.g. formation

3. Applied classification

4. Selected log properties with user-defined single or multiple cut-off values

5. Cut-off values and assigned lithology

Figure 15. Facies calculator in SKUA-GOCAD 2019

5. Cut off values and assigned lithology. In the final step lithologies that have been pre-defined in the classification table are assigned to the available permutations of log type and cut-off value. Clicking 'apply' will generate a continuous lithological log based on these input data for the input wells.

3.5 SELECTING CUT-OFF VALUES

Selecting appropriate cut-off values is clearly the most critical step in this classification approach. There is no singular robust statistical method for determining cut-offs because of the infinite number of ways in which geophysical tools imperfectly sample and respond to the complexities of real-world geology. Cut-offs are often defined both empirically and heuristically to achieve an end-product that is imperfect but deemed geologically plausible. The process

generally involves the determination of clusters within the dataset which are then mapped to specific lithologies using both the absolute magnitude of the responses (e.g. Table 1) but also soft supporting information such as cuttings descriptions and broader geological knowledge. The outcome of a cut-off analysis might also vary with the purpose of the study. For example, in an aquifer study the cut-offs might be deliberately biased toward the recognition of potential flow barriers such as mudstones, even if this overestimates reality. In addition to analysing geophysical logs within conventional downhole tracks (e.g. (Figure 16A) the most commonly used tools for the initial recognition of lithology clusters are frequency distribution histograms and multivariate scatter-plots. There are numerous statistical tools that can be used to enhance the recognition of clusters within these basic types of plot (Ma, 2011; 2019). Cases are considered below of lithology determination based on a single log and using multiple logs.

3.5.1 Case 1: Using a single log type

Where only one lithology log is available (quite often the gamma-ray log) an exploratory analysis of the frequency distribution histogram can help in identifying populations and potential cut-off values (Figure 16B). It is often necessary to experiment with bin sizes to reveal the populations within the histogram. Once the populations are identified a somewhat arbitrary and artificial ‘wall’ or cut-off is selected and used to generate the continuous lithology log. The effectiveness of the cut-off is usually judged against supporting lithological information such as core or cuttings descriptions.

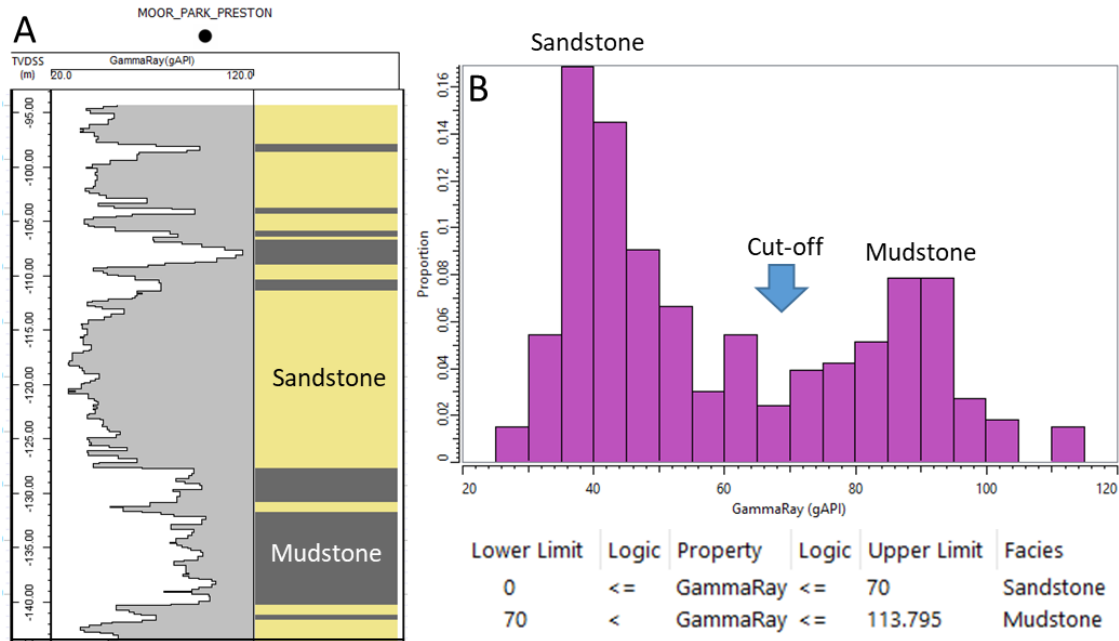


Figure 16. A lithology log (with a simple two-fold classification) created for a short section of Carboniferous Millstone Grit using a gamma-ray log. A histogram highlights the bimodal distribution created by interbedded sandstone and mudstone. The histogram shows distinct modes at 40API and 90API. A cut-off value of 70 API produces the lithology log shown in A. Note that the cut-off is a rather arbitrarily placed ‘wall’ in the lithofacies-component histogram.

In many cases, frequency distribution histograms of log values will not display distinct populations but are single mode histograms with a skewed long tail. This can occur even where the geology consists of two well-defined rock types, such as the Triassic St Bees Sandstone Formation illustrated below which is predominantly sandstone but also contains subordinate mudstone (Figure 17).

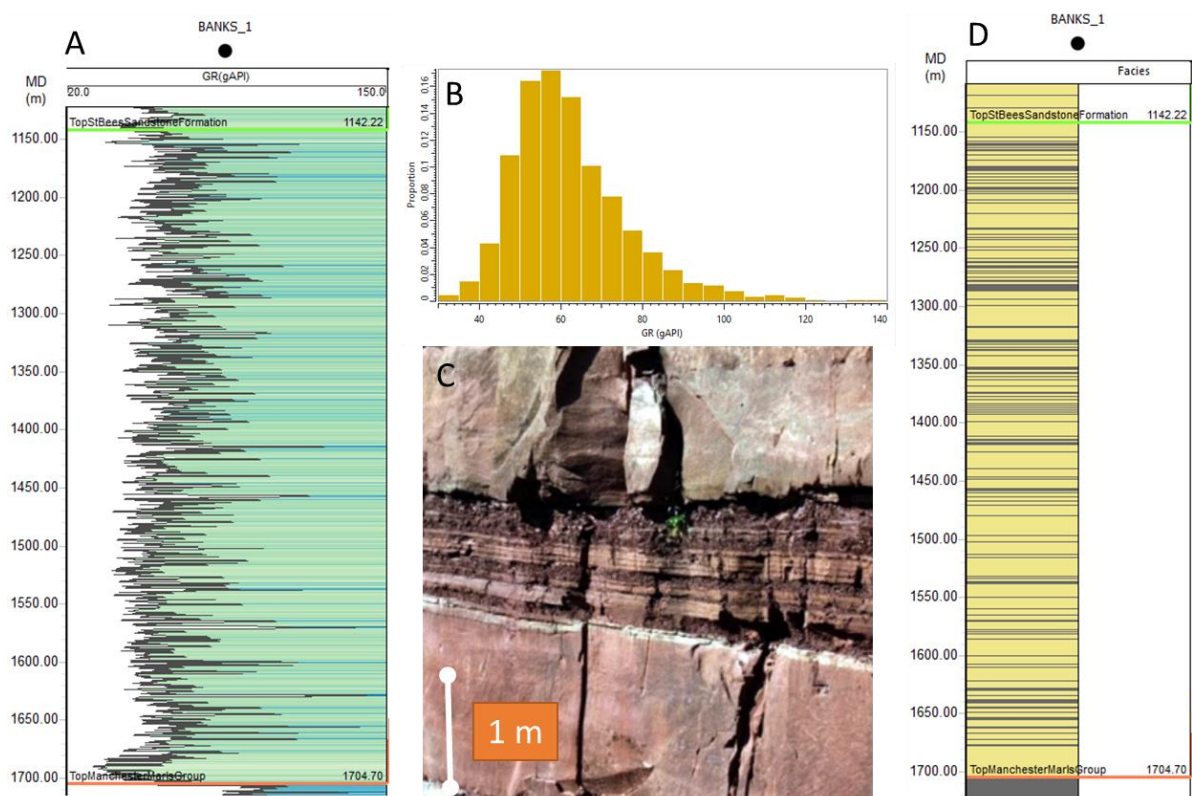


Figure 17. (A) Gamma-ray log of the Triassic St Bees Sandstone in NW England (B) histogram of the gamma-ray values (C) Outcrop of the St Bees Sandstones showing thinly interbedded sandstones and mudstones (D) gamma-derived lithology log.

Gamma-ray spikes related to thin interbedded mudstones are clearly seen on the gamma-ray log (Figure 17A) but they do not form a distinct population on the unimodal skewed histogram (Figure 17B). This is a common feature of geological units that contain thin beds because the logging tool will record a signal that is a mixture of the formation properties within the tool's vertical range of investigation. For example the zone of investigation of gamma-ray tool records signals from a hemispherical zone that is typically around 0.9 m across vertically. Coupled with continuous downhole logging-tool travel, thinly interbedded mudstones such as those illustrated in Figure 17C will always produce a convoluted or 'mixed-lithology' signal mid-range between sandstone and mudstone end members. This drowns the presence of a distinct peak in the histogram for thinly bedded mudstones. The vertical resolution will vary according to the type of tool (Figure 18) as well as other factors, in particular logging speed.

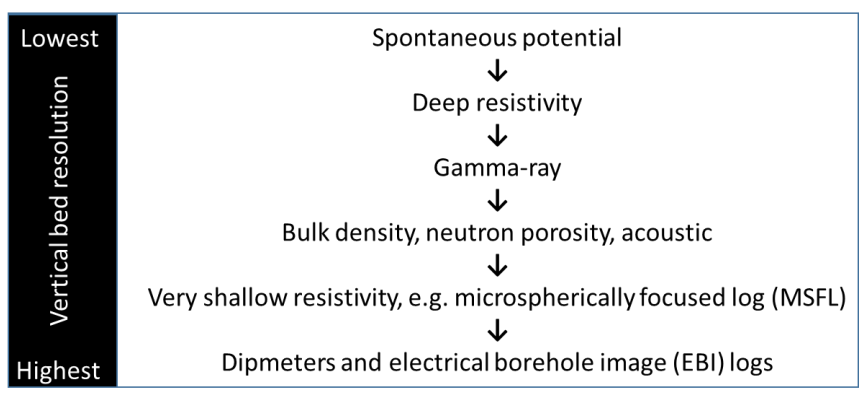


Figure 18. Common logs ranked from lowest to highest vertical resolution as follows. The absolute resolution of a given log varies depending on the tool, the sampling rate, logging speed, and processing methods.

While efforts can be made to deconvolve the long tail of the histogram into additional populations (Ma 2019) the selection of a cut-off value is generally based on supporting information from core and cuttings and on a regional understanding of the typical bulk proportion and bed thickness of mudstone within the St Bees Sandstone Formation.

3.5.2 Case 2: Using multiple log types

Using two or more log types will often produce a clearer separation of lithofacies from geophysical log datasets. There are no fixed rules on what logs can be combined but some advocate against combining logs that are highly correlated because they do not bring significant additional information into the lithofacies classification process (Ma, 2019). The main limiting factor in many projects undertaken in BGS using data archives assembled over many decades and derived from many heterogeneous sources are the low number of wells with comparable datasets of similar curve types and log quality..

Two logs can be readily plotted on a 2D scatter-plot. Figure 19 illustrates a cross-plot of sonic and gamma-ray for the Mercia Mudstone and Penarth Group in the Winterborne Kingston borehole in South Dorset (see Figure 22). While there are substantial overlaps, a number of high-density points clusters points toward the existence of at least four distinct lithologies in the stratigraphic interval.

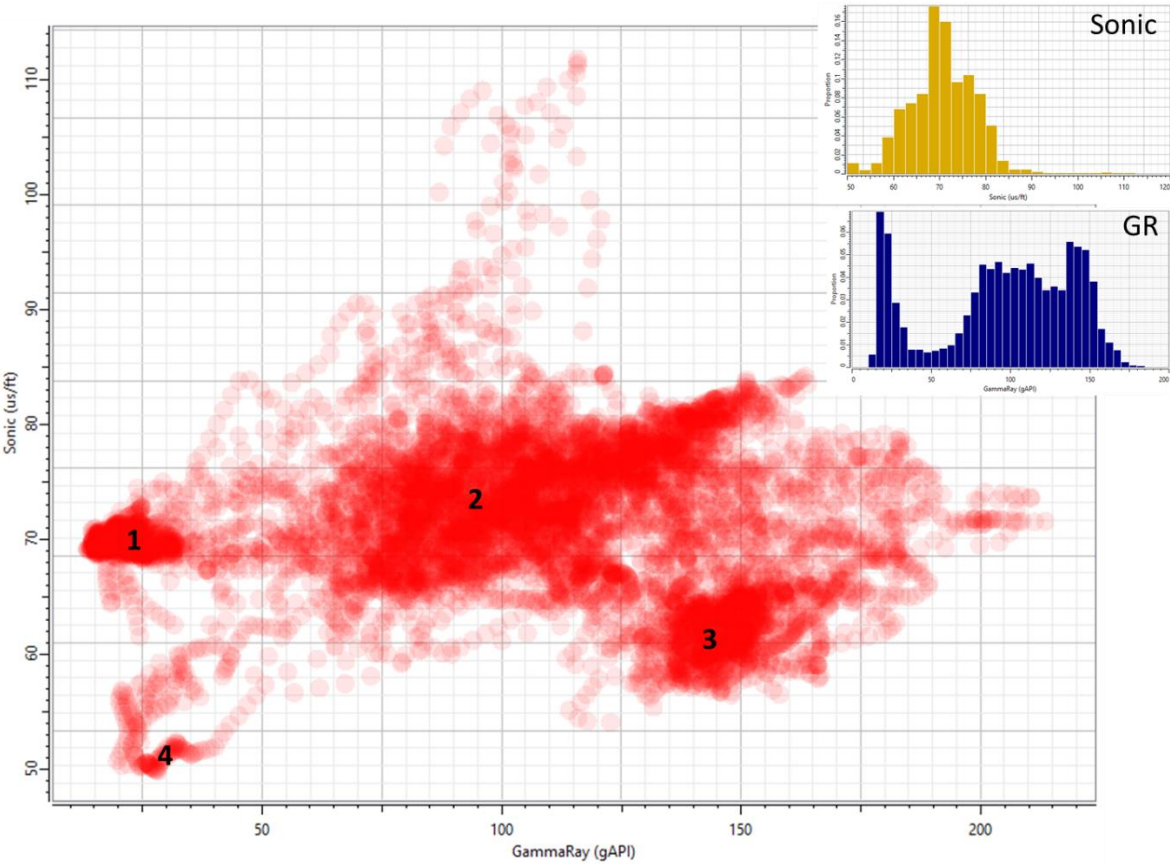


Figure 19. Cross-plot of gamma-ray and sonic logs for the Mercia Mudstone Group of the Winterborne Kingston Borehole (see Fig. 7) reveals four high density (darker red) clusters more effectively than logs plotted as individual distributions.

Additional lithological resolution can be achieved by colouring the points with a third property such as density. Figure 20 shows how this can be used to distinguish anhydrite (which has a relatively high density) within the scatter plot.

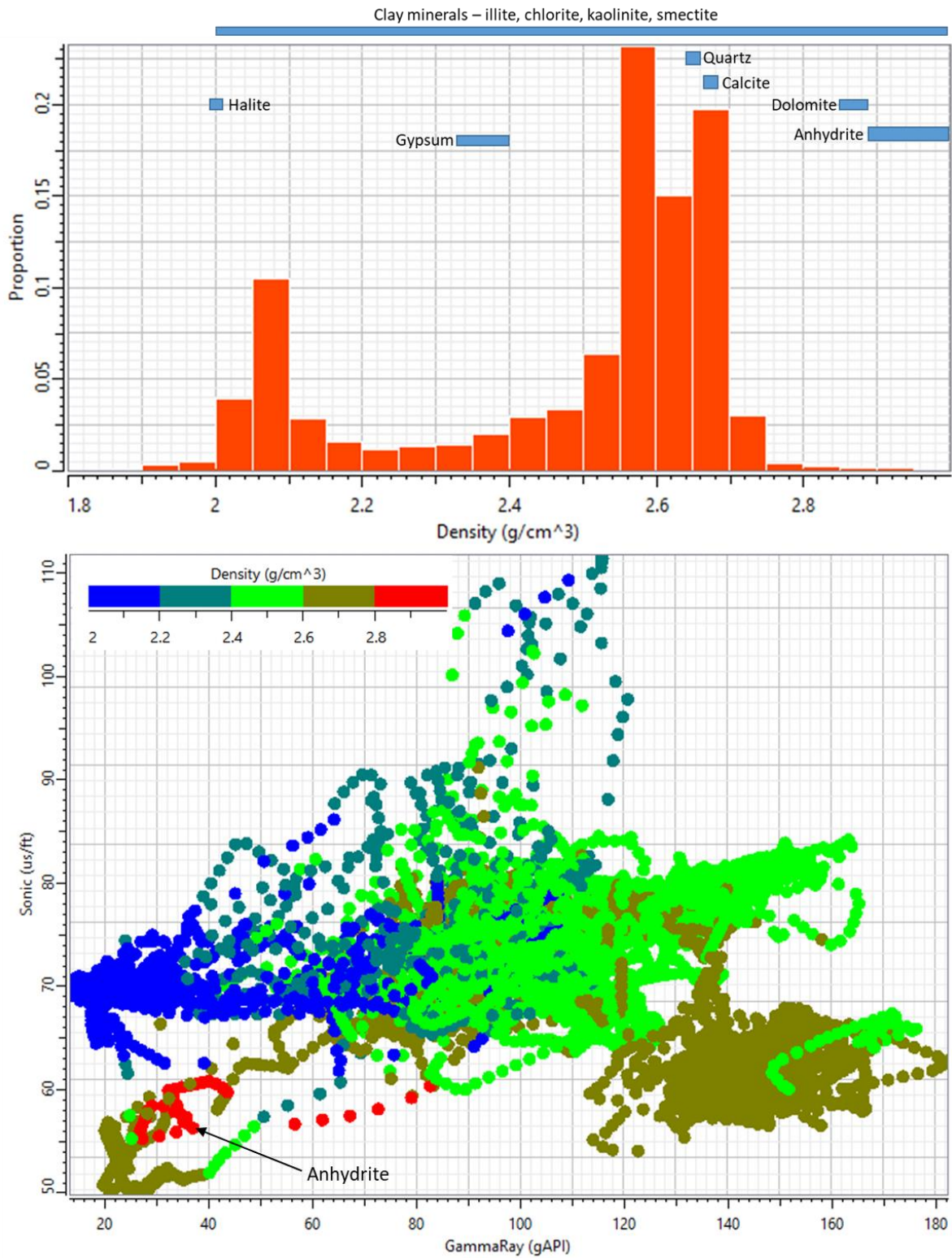


Figure 20. (A) Histogram of density log data for the Mercia Mudstone and Penarth Groups of the Winterborne Kingston borehole. The typical densities for a range of rock materials are shown on the plot. Cross-plot of sonic and gamma-ray logs with points coloured up by the density log. This reveals a cluster of high density anhydrite which was not apparent before including the additional log information on the 2D cross-plot.

An alternative method of cross-plotting three properties is to use a 3D scatter plot (Figure 21). A fourth property can be used to colour up the points, or as shown in Figure 21 this can be used to highlight the derived lithological clusters. Such 3D plots can however become difficult to interpret. Any addition of further properties requires the application of techniques such as

principal component analysis which reduces dimensional data into a low dimensional sub-space that can be visualized in 2–3 dimensions.

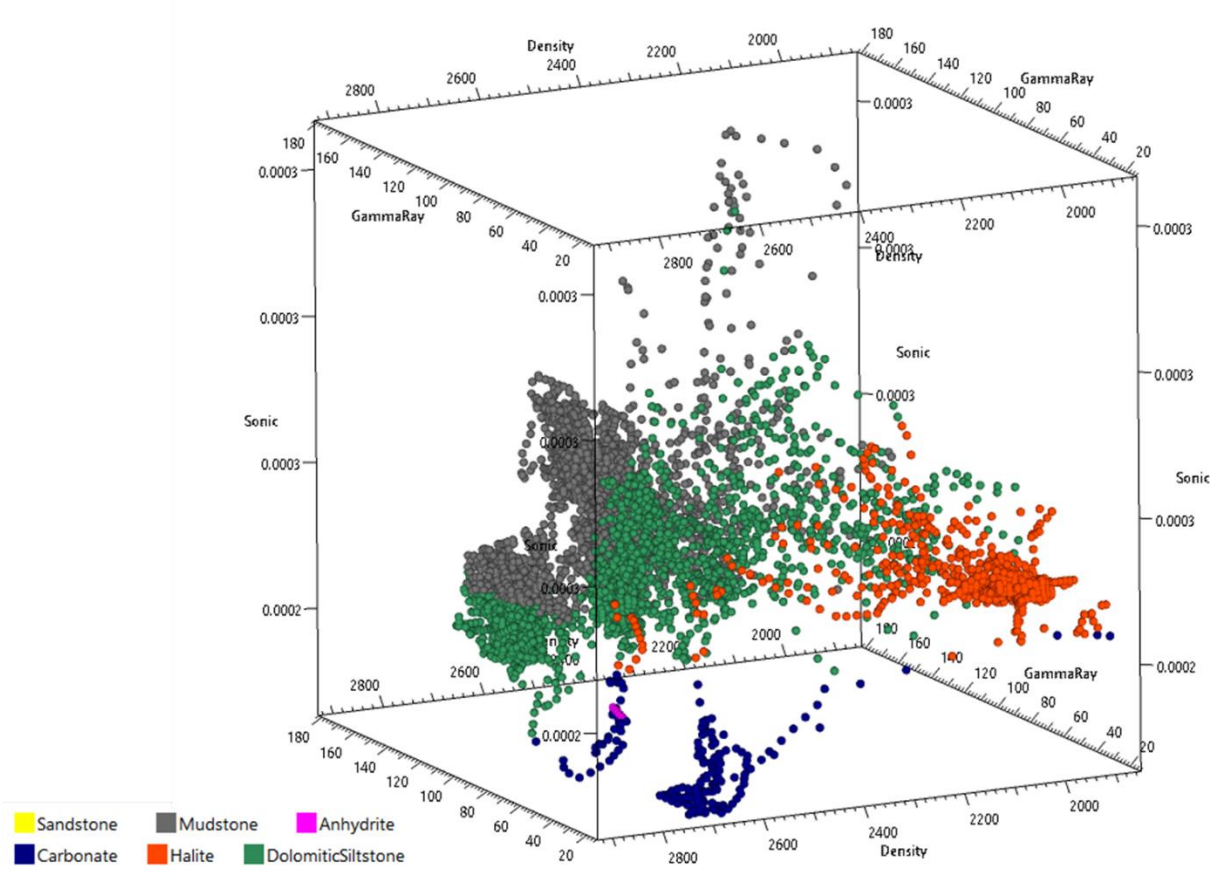


Figure 21. 3D scatter plot of density, sonic and gamma-ray for the Mercia Mudstone Group of the Winterborne Kingston borehole with the points coloured by derived lithology information.

Once the optimal combination of logs has been established for the stratigraphic interval under consideration a table of cut-off values can be created and the analysis run (Figure 22). As in most cases of cut-off analysis the approach is highly iterative. Cut-off values can be adjusted and the lithological log re-created in a highly interactive way until there is a close match between the lithology profile and what should be expected from additional evidence such as cuttings and general stratigraphic understanding.

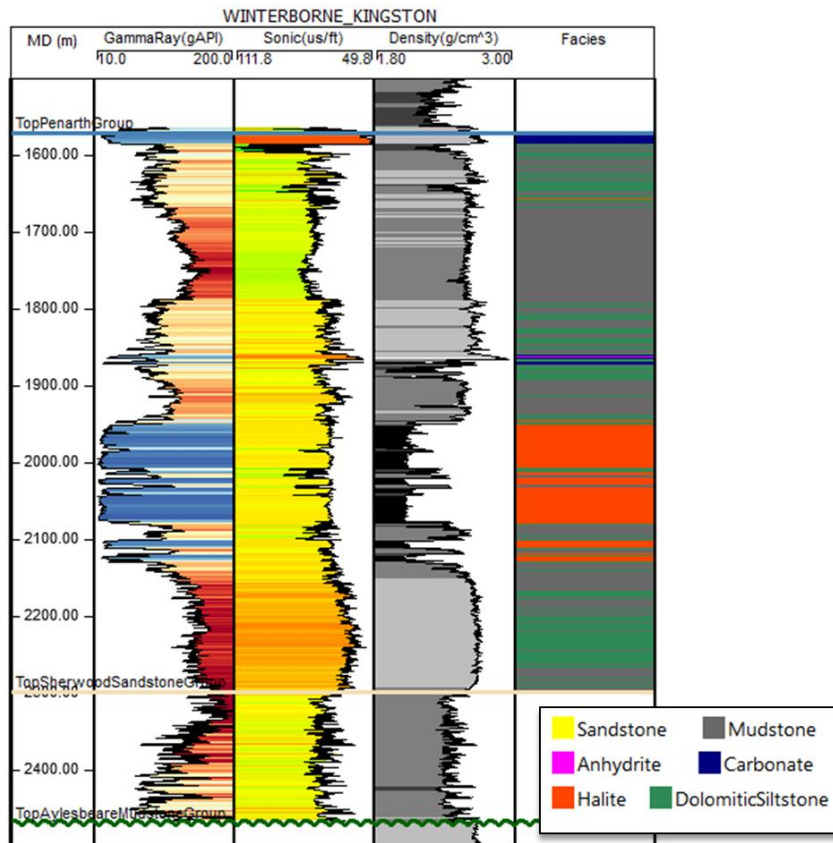


Figure 22. Resultant lithology log (central track, see A for key to colours).

3.6 THE IMPORTANCE OF WORKING WITHIN WELL REGIONS

The main difficulty of using simple cut-off analysis is discriminating lithologies when there is large overlap in their log properties. Figure 23 illustrates an example from the Winterborne Kingston borehole in South Dorset where the dolomitic siltstones and mudstones of the Blue Anchor Formation show a remarkably similar log response to the Sherwood Sandstone at greater burial depth. While effort could be expended to discriminate these rock types by introducing additional logs or by applying more sophisticated clustering techniques a simple solution is to restrict the analysis to a shorter well region that excludes the overlapping rock types.

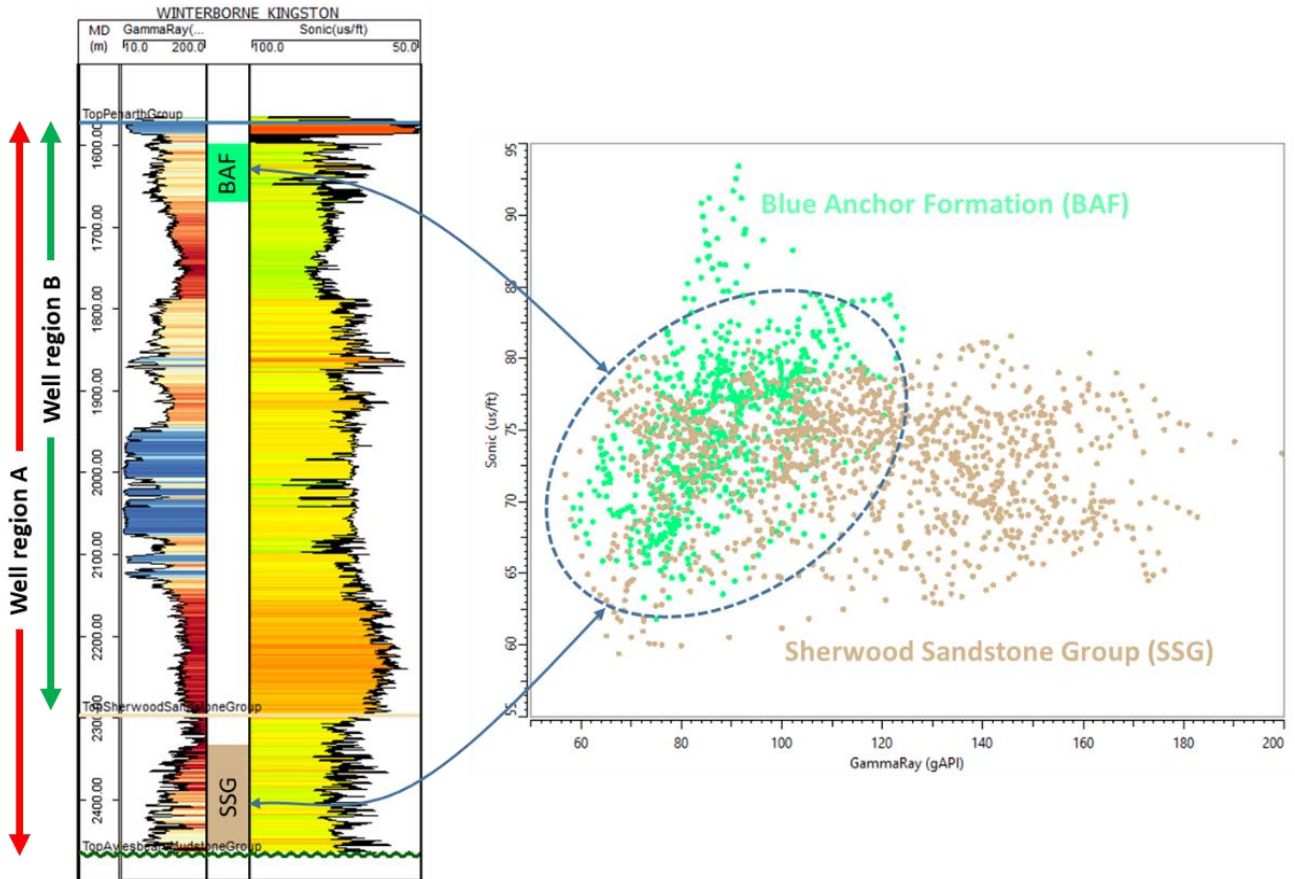


Figure 24. Gamma-ray and sonic log from Triassic strata in the Winterborne Kingston Borehole in South Dorset. The interbedded mudstones and dolomitic siltstones of the Blue Anchor Formation show a large overlap in their gamma-ray and sonic log response with the interbedded sandstones and mudstones of the Sherwood Sandstone at greater depth. Rather than attempting to discriminate lithofacies by introducing more log data it is often simpler and more efficient to work in smaller well regions e.g. B rather than A.

4 Mineral composition analysis by linear inversion of log data

4.1 BACKGROUND

The volumetric determination of mineral composition from petrophysical logs has a long history originating in efforts to estimate reliable porosity values from rocks with variable mineralogy and associated log responses (Doveton, 2018). By (mathematically) generating mixtures of mineralogical components with known petrophysical properties (Table 3) and comparing their calculated log response against the observed values it was possible to inverse model the likely composition of the rock by solving a system of linear equations.

Table 3. Table showing a section of the library of petrophysical properties used in the inverse modelling process. (DENS density, DT Compressional Delta Time (Slowness), DTS Shear Delta Time, VP P-Wave velocity), VS S-Wave velocity, PE Photoelectric Index (barns/electron)

Mineral	DENS(g/cc)	DT(μ sec/ft)	DTS(μ sec/ft)	VP(m/s)	VS(m/s)	PE
Quartz	2.65	43.90	88.80	6037.62	4120.82	1.82
Shale	2.60	62.50	150.00	2559.92	1129.87	3.42
Calcite	2.71	47.20	89.90	2559.92	3436.29	5.09
Clay	2.65	64.30	98.90	5966.29	3079.51	3.03
Dolomite	2.87	43.90	74.80	7346.57	3959.73	3.13
Anhydrite	2.95	50.00	85.00	6105.55	3366.50	5.08
Gypsum	2.35	52.40	85.40	6513.39	3603.75	4.04
Muscovite	2.83	47.20	91.10	5773.50	3342.19	2.40
Biotite	3.20	55.50	100.60	5374.84	3027.65	8.59
Kaolinite	2.64	64.30	101.70	5637.62	2995.90	1.47
Glauconite	2.83	55.50	157.40	3257.57	1935.28	4.77
Illite	2.77	64.30	98.90	5966.29	3079.51	3.03
Chlorite	2.87	55.50	61.30	9268.86	4969.11	4.77
Orthoclase	2.54	68.90	84.90	4926.98	3586.24	2.87
Siderite	3.91	43.90	84.90	6957.66	3588.70	14.30
Pyrite	5.00	39.60	55.90	8429.07	5448.05	16.40
Halite	2.03	66.70	114.50	4594.68	2661.45	4.00
FreshWater	1.00	205.00	-999.25	1482.00	-999.25	0.36
Brine	1.10	188.00	-999.25	1522.00	-999.25	0.81
Oil	0.85	238.00	-999.25	1280.00	-999.25	0.12

Since the log response is a product of both the rock matrix and the fluid fill of the pore space, the fluid composition and their variable petrophysical properties must be taken into account (Table 3).

The method works as an iterative search procedure. The input logs and the output mineralogical components (and pore-space filling fluid) are first defined and an initial composition is estimated. A series of intermediate solutions is calculated each comparing the input log responses with those predicted from the computed mineral proportions. The process terminates when convergence has been reached and there is no appreciable difference between successive solutions.

The appropriate logs for this method are those that discriminate well between the various mineralogical components under consideration. The method traditionally uses those logs which are indicative of porosity (density, neutron porosity and sonic) but today can include resistivity, spectral gamma-ray and geochemical logs. The number of available logs has a direct limiting effect on the number of mineralogical components that can be distinguished.

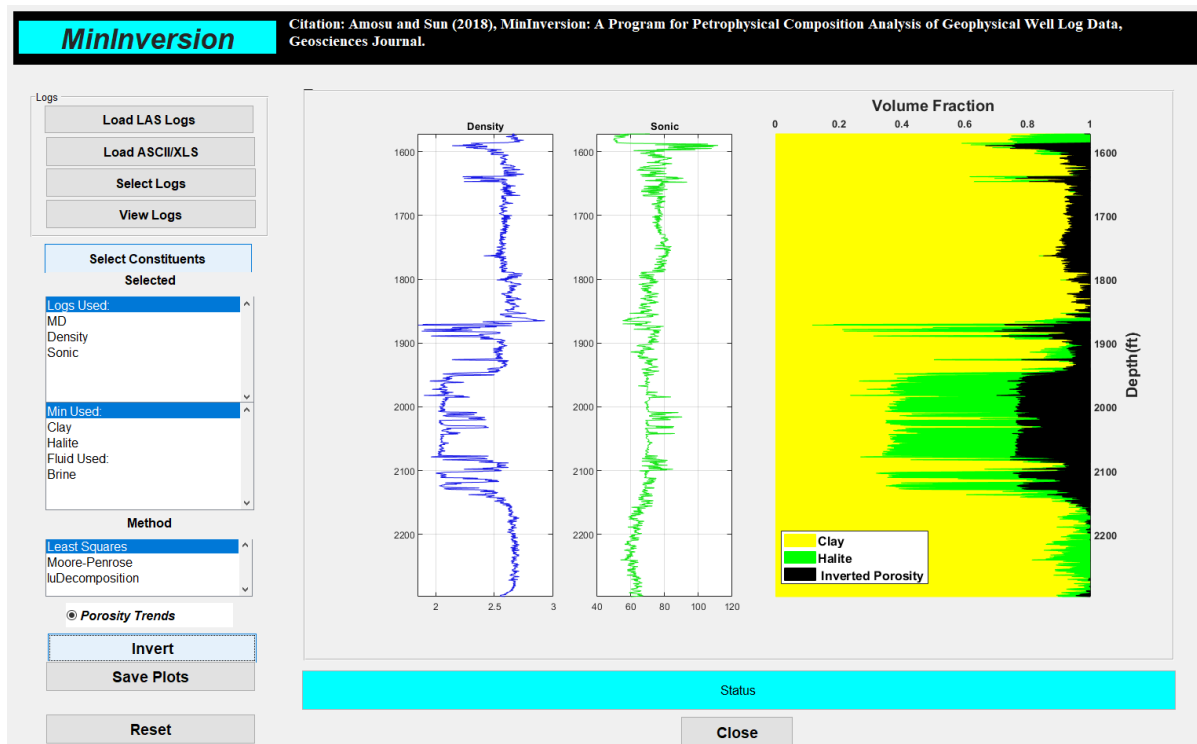
This method is useful for carbonate rocks of mixed calcite-dolomite composition and a wide range of other lithologies, although shales, mudstones and clay minerals can present problems because of their compositional variability. The method can generate reproducible solutions that are entirely erroneous if the initial mineral component model is incorrectly specified. Meaningful results are thus best obtained with as much a priori knowledge of the rock formation under consideration as possible. Careful geological evaluation of a range of possible solutions is a good approach together with pre-model calibration and post-model validation against cored intervals and cuttings where possible.

4.2 PRACTICAL APPLICATION

Functionality to estimate mineral composition from petrophysical logs is a standard feature in many log analysis software packages and has also been implemented within Excel spreadsheets. Amosu and Sun (2018) provide an openly available MATLAB program (MinInversion) for petrophysical composition analysis of geophysical well log data that features an easy-to-use graphical user interface, a selection of inputs that include LAS and a choice of

three inversions methods. The underpinning table of mineral/fluid components and their petrophysical values is user editable.

Table 4. Graphical user interface of MinInversion with input data and results from the Mercia Mudstone Group of the Winterborne Kingston Borehole. The availability of only two input logs (Density and Sonic) restricts the analysis to two specified mineralogical components (halite and clay) and inverted porosity.



5 Unsupervised cluster analysis

5.1 BACKGROUND

Clustering (or cluster analysis) is a technique that finds groups of similar objects that are more related to each other than to objects in other groups. In the case of geophysical logs the clusters might relate to different lithofacies types. There are many different clustering algorithms but one of the more widely used as a method for deriving lithologies from geophysical logs is k-means clustering (Cerqueira et al. 2019). K-means clustering can be applied to two or more logs and does not use any training data to guide the formation of clusters. Training data might include predetermined knowledge of how geophysical response varies with lithofacies gained from examining borehole core or image logs. K-means clustering thus falls into the category of an unsupervised technique.

In summary k-means clustering proceeds in four automated steps:

1. From the sample points, a pre-defined number of cluster centroids, k , are randomly picked as initial cluster centres
2. Each sample is assigned to the nearest centroid
3. Centroids are then relocated to the centre of the samples that were assigned to it. The similarity between points is based on the squared Euclidean distance between two points in m -dimensional space.
4. Repeat steps 2 and 3 until the cluster assignments do not change, or a user-defined tolerance or maximum number of iterations is reached.

A possible drawback of k-means is that the number of clusters (or lithofacies) must be specified before running the analysis. While the number of clusters may be obvious in datasets of only two log types or where the formations comprise only a few easily distinguishable rock types it may be less obvious when a larger number of logs are brought into the analysis. In such cases an 'elbow plot' showing sum squares of distances between each sample to the centre of its cluster group for a range of cluster numbers can be useful (Figure 25). The optimal number of clusters is generally thought to be at the point of maximum curvature. In poorly-known rock formations the elbow plot may give a useful initial indication of how many lithologies are present (or resolvable using the available logs) that is independent of any preconceived (and possibly erroneous) ideas of the geologist.

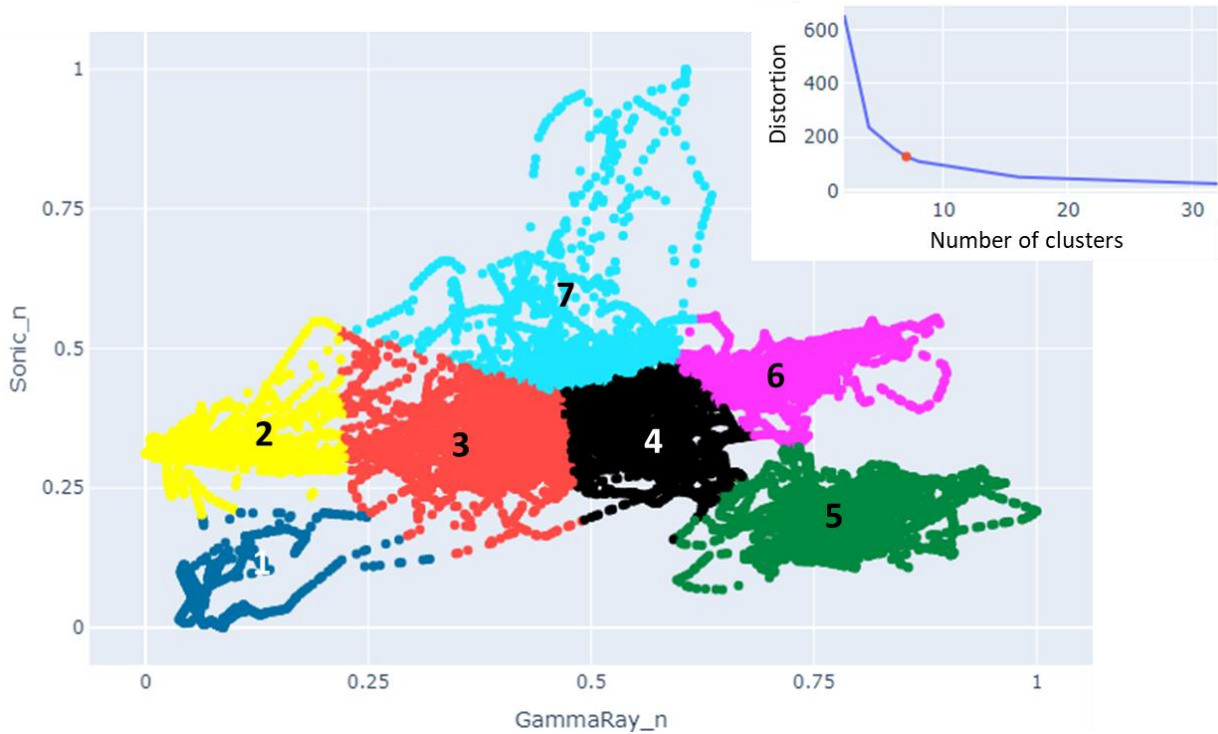


Figure 25. K-means cluster analysis of normalised gamma-ray and sonic log data for the Mercia Mudstone Group of the Winterborne Kingston borehole. Seven centroids have been pre-selected corresponding to the inflection point on the elbow plot.

5.2 PRACTICAL IMPLEMENTATION OF K-MEANS CLUSTERING

K-means cluster analysis can be undertaken easily and rapidly on large datasets using open-source python tools such as those provided by scikit-learn (<https://scikit-learn.org/stable/index.html>). The geoapps project (<https://pypi.org/project/geoapps/>) created by Mira Geoscience has been used here. It includes the scikit-learn k-means clustering algorithm within a Jupyter-Notebook application that includes a range of [Plotly](#) visualisation tools to assess the results using histogram, box, scatter, inertia and cross-correlation plots (Figure 26).

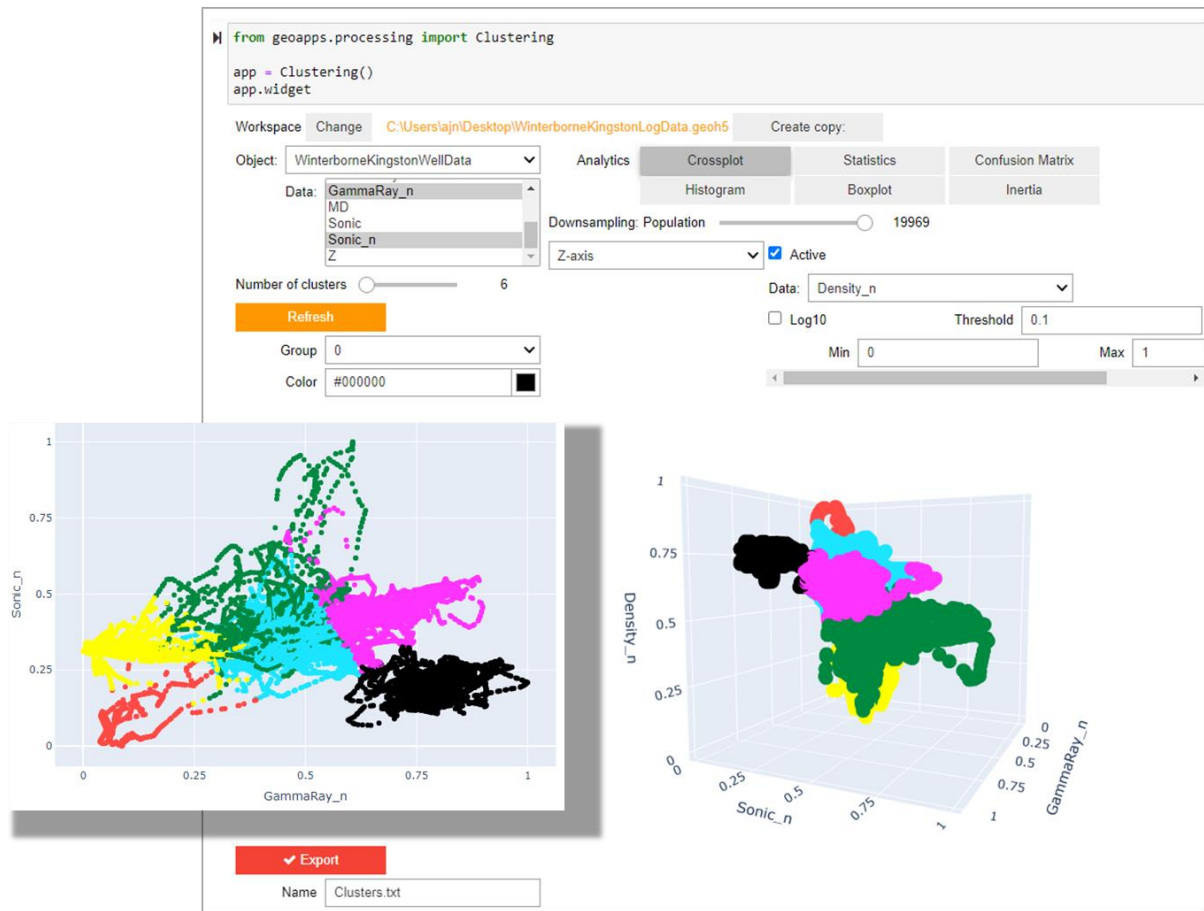


Figure 26. K-mean clustering as implemented in Mira Geoscience geoapps Jupyter-Notebook. Inbuilt Plotly visualisation tools provide a highly interactive environment. 2D and 3D cross-plots show normalised gamma-ray, sonic and density curve data for the Mercia Mudstone Group of the Winterborne Kingston Borehole.

Results can be saved and displayed within the free Mira Geoscience ANALYST 3D viewer or exported as text files to use in other log handling applications.

5.3 EXAMPLE OF RESULTS

Figure 27 shows the results of a k-mean cluster analysis for the Mercia Mudstone Group of the Winterborne Kingston borehole and highlights some of the similarities and differences with the previous cut-off analysis. While unsupervised cluster analysis of this type is unlikely to provide a definitive lithological classification of a borehole based on log data it is nonetheless a rapid and powerful method to derive insight into the dataset and could be a guide for additional supervised work or manual intervention and adjustment.

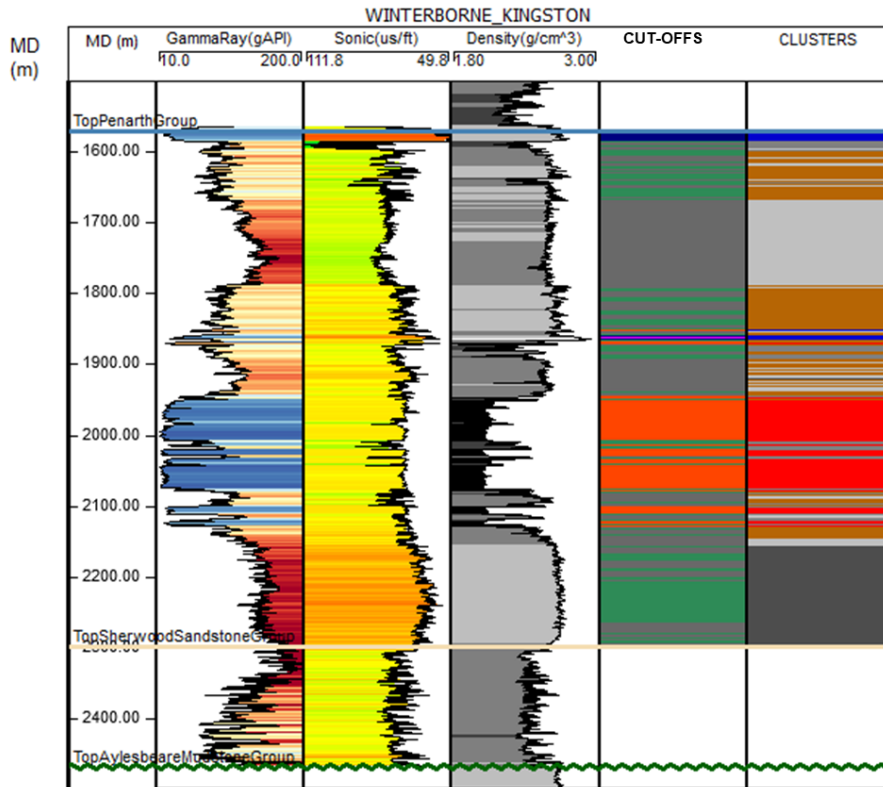


Figure 27. Results of k-means cluster analysis (right-hand track) performed in geo-apps and imported to SKUA-GOCAD. Note both similarities and differences between the cut-off analysis and cluster analysis.

6 Facies classification using supervised machine learning

6.1 BACKGROUND

Machine learning is essentially a set of data-analysis methods that includes classification, clustering, and regression. Machine learning algorithms can be used to discover similarities and trends within large complex datasets without being explicitly programmed, in essence learning from the data itself.

Machine learning methods would appear well-suited to the task of deriving lithology categories from geophysical log data. In borehole datasets there are often boreholes where the lithology or lithofacies is known from expert-descriptions of core and other boreholes where only geophysical logs are available. Boreholes (or parts of boreholes) where core and geophysical logs coexist can be used as training data for the machine learning algorithm and the relationships that are established can then be applied to unknown borehole sections.

Growth in the use of machine learning methods for facies classification of log data has occurred in parallel with the availability of many open-source packages, much of which used to be only available in proprietary software platforms. The best known general example is scikit-learn (<http://scikit-learn.org/>) a collection of machine learning tools coded in Python. These tools form the core of many Jupyter notebooks which have been specifically compiled for the purpose of classifying geophysical log data (e.g. https://github.com/brendonhall/facies_classification). This notebook (which employs a support vector machine or SVM) has been trialled for the purpose of this report and forms the basis for the discussion below. Many other notebooks using alternative supervised-learning algorithms such as random forest classification are available but have not yet been trialled (e.g. <https://github.com/seg/2016-ml-contest>).

6.2 INPUT DATA

The preparation of input data is straightforward and is a simple table comprising a series of log measurements with a known lithology or lithofacies class (Table 5). Here the training values represent short intervals of the borehole where the determination of lithology from the log response is clear from log response and cuttings. Ideally of course the training dataset should be based on core (or high-quality image logs) and might be assembled from multiple boreholes which prove the same formation within a basin.

Table 5. Example format of training data. In this case a subset of the geophysical measurements (INPUT_FACIES) is used to classify the entire stratigraphic interval (SVM_CLASS). The training data would often be based on expert descriptions of core.

Facies	Formation	Well Name	Depth	GR	DT	DEN
3	MMG	WINTERBORNE_KINGSTON	1576.00	21.39	1.71E-04	2620.00
3	MMG	WINTERBORNE_KINGSTON	1576.15	21.57	1.72E-04	2620.00
3	MMG	WINTERBORNE_KINGSTON	1576.30	21.67	1.72E-04	2617.91
3	MMG	WINTERBORNE_KINGSTON	1576.45	21.86	1.73E-04	2613.77
3	MMG	WINTERBORNE_KINGSTON	1576.60	22.04	1.73E-04	2610.00
4	MMG	WINTERBORNE_KINGSTON	1576.75	22.04	1.73E-04	2610.00
4	MMG	WINTERBORNE_KINGSTON	1576.90	22.01	1.75E-04	2610.00
4	MMG	WINTERBORNE_KINGSTON	1577.05	22.00	1.75E-04	2610.00
4	MMG	WINTERBORNE_KINGSTON	1577.20	21.99	1.75E-04	2610.00

6.3 RUNNING THE CLASSIFICATION

After loading the input data file the Jupyter notebook takes the user through the SVM classification process in a step-by-step way with blocks of runnable Python code, text providing instructions and commentary and interactive tabular and graphical outputs (Figure 28).

The screenshot shows a Jupyter notebook interface with a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations and execution. The code cell contains the following Python code:

```
facies_labels = ['MUD', 'ANH', 'CARB', 'HAL', 'DSST']
#facies_color_map is a dictionary that maps facies labels
#to their respective colors
facies_color_map = {}
for ind, label in enumerate(facies_labels):
    facies_color_map[label] = facies_colors[ind]

def label_facies(row, labels):
    return labels[ row['Facies'] -1]

training_data.loc[:, 'FaciesLabels'] = training_data.apply(lambda row: label_facies(row, facies_labels), axis=1)
training_data.describe()
```

The output cell shows a statistical summary for the 'Facies' variable:

```
Out[101]:
```

	Facies	Depth	GR	DT	DEN
count	2159.000000	2159.000000	2159.000000	2159.000000	2159.000000
mean	2.063918	1973.010439	101.406997	0.000226	2500.541213
std	1.445516	234.146303	57.277732	0.000028	248.665906
min	1.000000	1576.062256	13.360000	0.000164	2020.000000
25%	1.000000	1753.009522	26.060647	0.000208	2520.000000
50%	1.000000	2053.699951	136.725937	0.000225	2600.000000
75%	4.000000	2177.356079	148.129089	0.000250	2668.092895
max	5.000000	2285.847656	170.059998	0.000276	2930.000000

Below the output is a text box with the following text:

This is a quick view of the statistical distribution of the input variables. Looking at the 'count' values, most values have 4149 valid values except for 'PE', which has 3232. In this tutorial we will drop the feature vectors that don't have a valid 'PE' entry.

The final code cell shows the following code:

```
In [4]: PE_mask = training_data['PE'].notnull().values
training_data = training_data[PE_mask]
```

Figure 28. Part of the Jupyter notebook for facies classification (https://github.com/brendonhall/facies_classification) showing the typical mixture of runnable Python code, text providing instructions and commentary and interactive tabular output (or graphical plots).

The process breaks down into three stages.

1. Examining the input data

After the text file of training data are loaded some descriptive statistics are generated and there are options to plot and view the log data in conventional tracks or as cross-plots between different types of log measurements coloured according to lithofacies.

2. Splitting the dataset

A standard practice when training supervised-learning algorithms is to separate some data from the training set to evaluate the accuracy of the classifier. For example, one or more wells could be removed in a multi-borehole training dataset to act as test data (Figure 29). These test data play no part in the training or cross-validation of the SVM classifier.

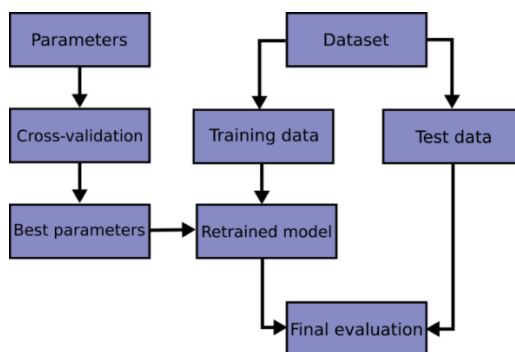


Figure 29. Splitting the training dataset into training data and test data

3. Standardise the dataset

Geophysical log measurements record a range of properties in different units of widely varying magnitude (e.g. see the summary statistics table in Figure 28). Many machine-learning algorithms however assume that all features are centred around zero and have variance in the same order. If one log type has a variance that is orders of magnitude larger than the others it will dominate the function and impede learning from the other features. For this reason all log data (including both training set and later input data) must be standardised to similar scale and deviation. This can be undertaken using the [StandardScaler](#) function in Scikit-learn.

4. Training the support vector classifier

Training the support vector classifier is an optimization process. The SVM classification learns from the training dataset the projection into a higher dimensional space where classes can be separated by a hyperplane (or set of hyperplanes) that maximizes the margin separating the classes. Hyperplanes are higher-dimensional generalisations of a plane. In 2D the hyperplane or decision boundary corresponds to a line (Figure 30). New uncategorised samples are classified according to the side of the hyperplane on which they fall when projected into the same space. “Soft margin” classification can accommodate some classification errors on the training data, in the case where data is not perfectly separable.

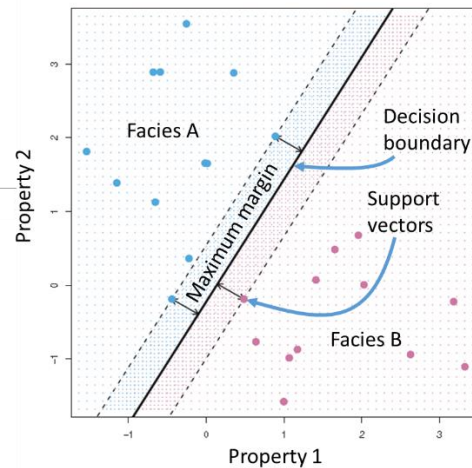


Figure 30. 2D example of a linear boundary that maximises the margin between the closest pair of data points belonging to two classes. The support vectors are the points on the dashed lines. Modified from [Wilimitis \(2018\)](#).

A cross-validation dataset is used to tune the parameters of the training model and is created by randomly splitting the training data into subsets (Figure 29).

The SVM implementation in scikit-learn takes a number of parameters which control the learning rate and the specifics of the kernel functions which map the original non-linear inseparable observations into a higher-dimensional space in which they become separable. A succession of models is created with different parameter values and the combination with the lowest cross-validation error is used for the classifier.

5. Evaluating the classifier

To evaluate the accuracy of the classifier the borehole that was set aside at the beginning of the process can be used to compare the predicted facies with the actual ones which may have been determined from expert description of core. A range of accuracy metrics are calculated, or the observed and predicted facies can simply be visually compared in adjacent log tracks. As demonstrated by periodic machine learning competitions (e.g. <https://github.com/seg/2016-ml-contest>) using hidden control boreholes, SVM (and related supervised ML approaches) while sophisticated in their approach do not generate completely accurate results with **F-1 values** ranging from 0.4 to 0.6.

6. Applying the classifier

Once created the SVM classifier can be applied to other boreholes with a comparable range of lithofacies (e.g. they are part of the same formation) and have a similar array of geophysical logs. These will need to be rescaled using the same parameters used to rescale the training set. The results can be saved as text files and loaded into other log handling or geological modelling software (Figure 31).

6.4 EXAMPLE OF RESULTS

An example of the output is shown in Figure 31 for the Winterborne Kingston borehole. Here the input data were simply short extracts from three log types attributed with a lithofacies (shown in the input facies column). SVM is here used simply to upscale these inputs to the entire log and here (with well-defined lithologies) appears to produce meaningful results.

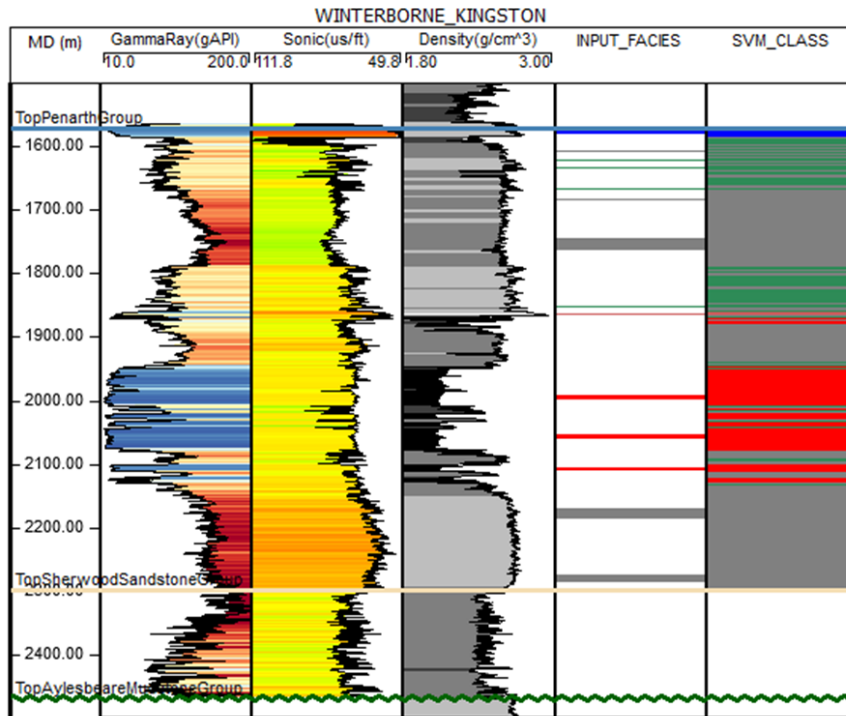


Figure 31. Example of SVM classifier

7 Conclusions

The aim of this report has been to evaluate a range of practical and accessible methods to derive lithology (or lithofacies) information from geophysical logs. All of the methods have their advantages and disadvantages (Table 6) and none will produce perfect results because of the inherent fuzziness of rock classifications and the convolved sampling signal of geophysical logs. Future work on fuzzy methods (such as [Fuzzy-C-Means](#) or [Fuzzy-SVM](#)) may be useful for this reason. Thus instead of predicting just one class as output they predict a fuzzy output i.e. a rock could belong to 40% of class A and 60% of class B etc.

The method chosen is likely to depend on the quality and range of the available geophysical logs, user preferences and the scale of the project that is being considered (e.g. one borehole versus many hundreds). The best solution might be to combine a number of methods e.g. k-means to gain some initial insight into the data, classification by machine learning or cut-offs and finally some manual editing if required. An encouraging finding of the review is the large volume of high quality, open source software (mostly written in Python) that is now available for log handling, display and analysis. The widespread use of Jupyter notebooks makes this software accessible and easy to run.

Table 6. Table summarising some of the advantages and disadvantages of the different methods trialled.

Method	Advantage	Disadvantage
Manual classification	<ul style="list-style-type: none"> - Uses all the background knowledge and expertise of the geologist - Tools for manual digitising are built into most log handling and 3D geomodelling software - Doesn't require high-quality log datasets or normalisation of data for multi-well projects - Can be undertaken on raster log images if no digital version is available 	<ul style="list-style-type: none"> - Slow manual process - Exact criteria for each classification decision can vary and may not be clear or reproducible by other users
Cut-offs	<ul style="list-style-type: none"> - Rapid, semi-automated method - Reproducible by any user applying the same cut-off criteria - Tools are built into most log handling and 3D geomodelling software - User has a strong degree of control on the outcome by adjusting cut-offs in different borehole regions - Can work effectively on a single log such as gamma-ray in simple siliciclastic formations 	<ul style="list-style-type: none"> - Often requires a degree of trial and error to find ideal cut-offs which slows the process - Becomes complex as the number of log types increase above two - Requires normalisation of log data if criteria are applied across multiple boreholes
Mineral composition analysis by linear inversion	<ul style="list-style-type: none"> - Rapid automated technique - Provides insight into the mineralogical composition of the formation - Useful in limestone-dolomite formations - Realistic in the sense that geophysical log measurements (with wide vertical sampling) are often 'averages' across different lithologies and mineralogies - Included in many log analysis packages or can be undertaken in spreadsheets and open-source software 	<ul style="list-style-type: none"> - Results must be post-processed into more conventional discrete lithofacies classifications - Requires high-quality geophysical log data, in particular a diverse suite of (relatively uncommon) porosity logs - Automated technique with little opportunity to include the users general geological knowledge of the formation in the outcome
Unsupervised cluster analysis (K-means clustering)	<ul style="list-style-type: none"> - Rapid automated technique - Reproducible by any user using the same input criteria - Deployable using open-source Python tools - Elbow plots provide unbiased insight into the likely number of facies classes that can be resolved from a particular set of log measurements 	<ul style="list-style-type: none"> - Requires a priori knowledge on the number of expected facies - No opportunity to include general user geological expertise or machine-learned support from cored intervals
Supervised cluster analysis (Support Vector Machines)	<ul style="list-style-type: none"> - Rapid automated technique, particularly once the SVM training process is complete - Easy to deploy using open-source Python tools - Incorporates geological expertise by using training examples of what constitutes the "correct" clustering of a dataset - Computes a range of metrics for the performance of the classifier based on the modelled versus the observed for the test data 	<ul style="list-style-type: none"> - Some unfamiliar terms and concepts for new starters in machine learning - Sophisticated but not infallible and (as shown by competitions against standard test datasets) will still generate lithology logs with many errors because of the inherent fuzziness of rock classifications and the convolved sampling signal of geophysical logs.

References

British Geological Survey holds most of the references listed below, and copies may be obtained via the library service subject to copyright legislation (contact libuser@bgs.ac.uk for details). The library catalogue is available at: <https://envirolib.apps.nerc.ac.uk/olibcgi>.

AMOSU, A., & SUN, Y. (2018). MinInversion: A program for petrophysical composition analysis of geophysical well log data. *Geosciences*, 8(2), 65.

CERQUEIRA, A. G., DE LIMA, O. A. L., & RIOS, R. A. (2019). A nonparametric approach using clustering analysis to estimate shaliness in shaly-sand formations. *Journal of Applied Geophysics*, 164, 11-18.

DOVETON J.H. (2018) Mathematical Minerals: A History of Petrophysical Petrography. In: Daya Sagar B., Cheng Q., Agterberg F. (eds) *Handbook of Mathematical Geosciences*. Springer, Cham. https://doi.org/10.1007/978-3-319-78999-6_24

EVANS, D J, HOPSON, P. 2000. The seismic expression of synsedimentary channel features within the Chalk of southern England. *Proceedings of the Geologists' Association*, Vol. 111, 219 – 230.

GALE, A S, YOUNG, J R, SHACKLETON, N J, CROWHURST, S J, WRAY, D S. 1999. Orbital Tuning of Cenomanian Marly Chalk Successions: Towards a Milankovitch Time-Scale for the Late Cretaceous. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, Vol. 357, No. 1757, 1815 – 1829.

GRADSTEIN, F M, WATERS, C N. 2018. Summary of the new stratigraphic guide to the Chalk Group in the UK and Norwegian sectors of the North Sea. *Proceedings of the Yorkshire Geological Society*, Vol. 62, 187 – 194.

JOHNSON, H. & LOTT, G.K. 1993. 2. Cretaceous of the Central and Northern North Sea. In: Knox, R.W.O'B & Cordey, W.G. (eds.) *Lithostratigraphic nomenclature of the UK North Sea*. British Geological Survey, Nottingham.

LOTT, G.K. & KNOX R.W.O'B. 1994. Post Triassic of the Southern North Sea. Volume 7. In: Knox, R.W.O'B. & Cordey, W.G. (eds), *Lithostratigraphic nomenclature of the UK North Sea*. British Geological Survey, Nottingham.

MA, Y.Z. (2019) Facies and Lithofacies Classifications from Well Logs. In: *Quantitative Geosciences: Data Analytics, Geostatistics, Reservoir Characterization and Modeling*. Springer, Cham. https://doi.org/10.1007/978-3-030-17860-4_10

MA, Y.Z. (2011). Lithofacies Clustering Using Principal Component Analysis and Neural Network: Applications to Wireline Logs. *Mathematical Geosciences - MATH GEOSCI.* 43. 401-419. [10.1007/s11004-011-9335-8](https://doi.org/10.1007/s11004-011-9335-8).

MORTIMORE, R N, GALLAGHER, L T, GELDER, J T, MOORE, I R, BROOKS, R, FARRANT, A R. 2017. Stonehenge—a unique Late Cretaceous phosphatic Chalk geology: implications for sea-level, climate and tectonics and impact on engineering and archaeology. *Proceedings of the Geologists' Association*, Vol. 128, 564 – 598.

MORTIMORE, R N, WOOD, C J. 1986. The distribution of flint in the English Chalk, with particular reference to the 'Brandon Flint Series' and the high Turonian flint maximum. In: *The Scientific Study of Flint and Chert*, Sieveking, G de G (ed), Cambridge University Press: Cambridge, 7 – 20.

MORTIMORE, R N. 1986. Stratigraphy of the Upper Cretaceous White Chalk of Sussex. *Proceedings of the Geologists' Association*, Vol. 97, 97 – 139.

- MORTIMORE, R N. 2014. *Logging the Chalk*. Whittles Publishing: Caithness, pp. 357 pp.
- NEWELL, A J, WOODS, M A, FARRANT, A R, SMITH, H, HASLAM, R B. 2018. Chalk thickness trends and the role of tectonic processes in the Upper Cretaceous of southern England. *Proceedings of the Geologists' Association*, Vol. 129, 610 – 628.
- QUARTERO, ERIK & BECHTEL, D. & LEIER, ANDREW & BENTLEY, LAURENCE. (2014). Gamma-ray normalization of shallow well-log data with applications to the Paleocene Paskapoo Formation, Alberta. *Canadian Journal of Earth Sciences*. 51. 327-340. 10.1139/cjes-2013-0148.
- RAWSON, P F, ALLEN, P, GALE, A.S. 2001. The Chalk Group—a revised lithostratigraphy. *Geoscientist*, Vol. 11, 21.
- RIDER, M., AND KENNEDY, M. 2011. *The Geological Interpretation of well logs*, 3rd edition. Rider and French Consulting Ltd., pp. 1–4, pp. 117–146.
- SHIER, D.E. 2004. Well Log Normalization: Methods and Guidelines. *Petrophysics*, 45.3: 268–280.
- WHITHAM, F. 1991. The stratigraphy of the Upper Cretaceous Ferriby, Welton and Burnham formations north of the Humber, north-east England. *Proceedings of the Yorkshire Geological Society*, Vol. 48, 227–254.
- WHITHAM, F. 1993. The stratigraphy of the Upper Cretaceous Flamborough Chalk Formation north of the Humber, north-east England. *Proceedings of the Yorkshire Geological Society*, Vol. 49, 235–258
- WHITTAKER, A, HOLLIDAY, D, PENN, I E. 1985. Geophysical logs in British Stratigraphy. *Special Report of the Geological Society, London*, No. 18, 74 pp.
- WOOD, C J, SMITH, E G. 1978. Lithostratigraphical classification of the Chalk in North Yorkshire, Humberside and Lincolnshire. *Proceedings of the Yorkshire Geological Society*, Vol. 42, 263 – 287.
- WOODS, M A, NEWELL, A, HASLAM, R, FARRANT, A, SMITH, H. 2016. Predicting the Subsurface. *Geoscientist*, Vol. 26, 10 – 13.
- WOODS, M.A. 2018. Geophysical log characterization of the Flamborough Chalk Formation (Late Cretaceous, Santonian – Early Campanian), East Yorkshire, UK: implications for understanding the onshore and offshore occurrence of the Late Campanian Rowe Chalk Formation. *Proceedings of the Yorkshire Geological Society*, Vol. 62, 201 – 209.