Marine and Petroleum Geology 144 (2022) 105828

Contents lists available at ScienceDirect

Marine and Petroleum Geology

journal homepage: www.elsevier.com/locate/marpetgeo

Clustering of facies in tight carbonates using machine learning

Paul W.J. Glover [a,*], Omar K. Mohammed-Sajed [a,b], Cenk Akyüz [a], Piroska Lorinczi [a], Richard Collier [a]

[a] School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
[b] Department of Geology, College of Science, University of Mosul, Iraq

ARTICLE INFO

Keywords:
Petrophysics
Machine learning
Clustering
Facies recognition
Porosity
Permeability
Butmah formation

ABSTRACT

Machine learning clustering methods offer the potential for recognition and separation of facies based on core or well-log data. This is a particular problem for carbonate rocks because diagenesis produces a wide range of rock microstructures and transport properties. In this work we use a large database of high quality poroperm, electrical, mercury injection capillary pressure and nuclear magnetic resonance spectroscopy measurements (307 core samples), representing 5 stratigraphically defined facies, as well as well log data to examine facies-recognition abilities using 8 different machine learning clustering approaches and a redundancy of 10 to ensure statistically valid results, resulting in a total of over 990 clustering runs. For a 3 cluster problem, we find that the Expectation Maximisation (92.57% success) and two types of Kmeans approaches (89.60% and 91.09%) provide the best methods. Further testing using the best of these shows that the quality of the input parameter (attribute) matters more than the number of attributes used, with the power of attributes in decreasing order being porosity, cementation exponent, permeability, pore throat diameter and free fluid index, implying that some attributes can degrade clustering performance. Further tests show that there should be at least as many attributes as clusters, in which case the machine learning can be left to choose the final number of clusters, providing the best performance in this work (69.35% success for a five cluster problem), otherwise it is best to constrain the cluster number by supervision. Application of the results from the previous testing to a mixed carbonate tight carbonate well from the Butmah formation shows satisfactory determination of 4 petrofacies by clustering (up to 91.65%) when compared to petrofacies determined manually. However, the greater challenge of clustering 9 reservoir quality classes defined using a ternary petrofacies approach did not provide a successful result (<38% success rate).

1. Introduction

In general rocks exhibit a wide range of structures and properties even within formations that seem well-defined stratigraphically (Al Zainaldin et al., 2017). This is especially the case for carbonate rocks because they have almost always been subjected to post-depositional reworking by a range of diagenetic processes (Mohammed Sajed and Glover, 2020; Mohammed Sajed et al., 2021) which can lead to heterogeneity and anisotropy that is not necessarily bed-bound (Sinan et al., 2020). The recognition of groups of core samples or well-logging intervals that share a generic microstructure and transport properties is usually carried out by defining units based on either core scale sedimentological and geological evidence (Jiang et al., 2021), microfacies based on optical or scanning electron micrographic evidence (Liu and Song, 2020; Yarmohammadi et al., 2020) or petrofacies, which are allocated on the basis of petrophysical properties (Kopaska-Merkel et al., 1989; Cao et al., 2020; Silva et al., 2020). Some have used a smaller set of properties which are directly related to reservoir quality to make the classification (Mohammed-Sajed and Glover, 2022). While all of these methods use both qualitative observations and quantitative measurements, they are all classification schemes that do not inform physical processes. Nevertheless, facies allocation is a first and significant step in the analysis of a carbonate rock.

Conventionally, facies definition and distribution is very time-consuming and expensive since it involves the in-depth analysis of well cores, well logs, downhole measurements, analogue studies and is holistic in nature (Bestagini et al., 2017). Such manual analysis of large datasets by a reservoir geologist or petrophysicist must be subjective to some extent, leading to increased uncertainties and the introduction of artefacts (Ren et al., 2019).

* Corresponding author.
E-mail address: p.w.j.glover@leeds.ac.uk (P.W.J. Glover).

https://doi.org/10.1016/j.marpetgeo.2022.105828
Received 18 February 2022; Received in revised form 23 May 2022; Accepted 21 June 2022
Available online 13 July 2022

By contrast, machine learning processes provide the opportunity to make the process of facies recognition more automatic and objective (Witten et al., 2011). The accuracy of the result depends on the relevance and complexity of the method, and critically on the relevance and quality of the attributes that represent the input physical properties (Larose, 2006; Witten et al., 2011).

In machine learning, facies recognition may be approached by using either clustering or classification (Witten et al., 2011). Table 1 shows the differences between these two approaches. Clustering is simpler, does not require foreknowledge of the number or properties of the facies, and does not need training. However, it can only separate facies on the basis of the natural similarity or difference in the properties (which are called attributes in machine learning parlance) and hence it provides more general solutions (Larose, 2006). These aspects of clustering give it advantages in facies recognition where only an attribute dataset exists. However, if facies in one well have already been defined and it is necessary to find to which of these facies a sample or samples in another well belong, classification must be used.

In recent years, initially multivariate statistical methods (Fang and Feng, 2021; Al-Qassab et al., 2000), and more recently machine learning methods have been increasingly used (Al Anazi and Gates, 2010; Alexsandro et al., 2017; Imamverdiyev and Sukhostat, 2019). Approaches looking at microfacies and petrofacies include many of those already cited (Jiang et al., 2021; Liu and Song, 2020; Yarmohammadi et al., 2020; Kopaska-Merkel and Friedman, 1989; Cao et al., 2020; Silva et al., 2020). These approaches are almost always classification schemes which only take account of isolated data for a given core or depth, with no account taken of how the facies will also partially be correlated with depth. Consequently, some of the more recent implementations have attempted to incorporate the constraint imposed by partial depth correlation (e.g., Santos et al., 2022) or seek to improve accuracy by adding methodological complexity (Nanjo and Tanaka, 2020; Tewari and Dwivedi, 2020). While extremely effective, these models are very complex, the Santos model, for example, requiring the implementation of Bidirectional Long-Short-Term Memory Deep Recurrent Neural Networks.

Clustering (Cai et al., 2022) is a much simpler approach that has the potential to provide 'sufficiently good' separation of samples into clusters that represent samples with similar properties with no foreknowledge of the characteristics of each of the clusters, and often does not require knowing the number of clusters (Kanungo et al., 2002; Wu et al., 2008). Clustering can be used as a first step to obtain the number and properties of facies before submitting data to a classification scheme which trains the data to the facies derived from clustering and ultimately then uses the trained process to classify further data (Ren et al., 2022).

The question is what is 'sufficiently good'. At the moment there is little information available on whether clustering can be used on petrophysical data from geologically diverse diagenetically-altered carbonates, which methods perform best, and to understand how to fine-tune the use of such models especially with regard to the number and types of attributes and whether to specify the number of clusters to be produced or to leave that decision to the method.

Table 1
Differences and similarities between Clustering and Classification.

Clustering	Classification
Less complex	More complex
Unsupervised	Supervised or unsupervised
Facies clustered into groups according to similarities in their natural properties (attributes)	Facies classified into known and predetermined classes on the basis of their natural properties (attributes)
Does not need facies (classes) to be already defined.	Needs facies (classes) to be already defined.
Method needs to be tuned for optimum efficacy	Method needs to be tuned for optimum efficacy
No training needed	Method needs to be trained on known facies of the same type.
Validation testing is required.	Validation testing is required.
Can only be used to recognise general groupings such as facies.	Can be used to classify data by finer distinctions in their attributes.

This paper has several goals. The first is to examine and test the accuracy of some of the machine learning clustering methods available, the second is to ascertain which are the most effective for use with petrophysical data. The quality of the results obtained from machine learning methods depends critically on the quality and relevance of its input data. Consequently, it is also the goal of this paper to find which petrophysical input parameters (attributes) make good discriminators for facies separation, and whether it is better to have foreknowledge of the number of groups of data in a given dataset so that the facies definition process can be constrained. Most of the tests have been carried out on a large database of 307 samples containing 5 stratigraphically defined facies, and for which an extensive set of high quality petrophysical measurements are available, including poroperm, electrical, mercury injection capillary pressure (MICP) and nuclear magnetic resonance (NMR) spectroscopy.

1.1. Data sets

In this section, we examine the three main datasets used in this modelling. Table 2 summarises the sample numbers together with which measured and calculated parameters appear in each dataset. All methodologies used to obtain the measured parameters are given in Al Khalifah et al. (2020).

The first dataset contains data from 202 sample from three beds of the Portland limestone, which crops out in quarries on the Isle of Portland (Dorset, UK). All samples were obtained from the Fancy Beach or Jordans quarries (50°33′10″N 02°26′25″W) and are from beds within the Upper Jurassic Portland Freestone, a well-cemented oolitic limestone (Barton et al., 2011) which has traditionally been used in construction.

The Isle of Portland primarily consists of marine strata from the Upper Jurassic overlain by a thin layer of basal Cretaceous Purbeck Formation, which lies on top of the Portland Freestone and which marks the bottom of the Cretaceous. The true Portland Stone, a well-cemented oolitic limestone, lies immediately below the Cretaceous Purbeck Formation. Certain beds of the Portland Freestone have been used both throughout history and currently as high quality architectural building stone. It composes many of the civic and religious buildings in the UK, such as The Cenotaph (1920), St. Paul's Cathedral (1677), The British Museum (1753) and the Parkinson Building (University of Leeds, 1938–1951), but is also used globally, as in the United Nations Building in New York (1952), and Auckland War Memorial Museum, New Zealand (1929). The stone has been designated by the International Union of Geological Sciences as a Global Heritage Stone Resource. Portland Cherty Series lies immediately below the Portland Freestone and above The Portland Sand, which consists predominantly of marls with interbedded sandy layers. Upper Jurassic Kimmeridge Clay lies below the Portland sand and is the lowest formation exposed in the area, occurring beneath Castletown and Portland Harbour. It is exposed at the foot of the high northern cliffs or the Isle of Portland when not covered by talus.

This paper uses data from three beds; the Base Bed (72 samples), Whit Bed (58), and Hard Blue (72). All of these formations are dominated by sparite-cemented oolites (Barton et al., 2011). The data from the first two of these were used previously in initial machine learning permeability prediction studies using genetic algorithms by Al Khalifah et al. (2020), whose results we later compare the results from this work. Al Khalifah et al. (2020) note that the Whitbed contains common shells, usually distributed evenly, but sometimes concentrated in zones. These shells are commonly cemented. The Base bed is less shelly and commonly contains completely cemented shell moulds. The cemented nature of this rock makes it ideal building stone as well as a good, well-studied tight carbonate reservoir analogue. The Hard Blue bed is highly cemented and presents a much lower porosity and permeability

Table 2
Characteristics of the datasets used in this study.

Formation	No. Samples	He Porosity	Formation Factor	Cementation exponent	Characteristic pore throat size	Characteristic pore size	Characteristic grain size	Gas Permeability	T_2	FFI	BVI
Portland limestone	202	✓	✓	✓	✓	✓	✓	✓		✓	✓
Base Bed	72	✓	✓	✓	✓	✓	✓	✓		✓	✓
Whit Bed	58	✓	✓	✓	✓	✓	✓	✓		✓	✓
Hard Blue	72	✓	✓	✓	✓	✓	✓	✓		✓	✓
Purbeck limestone	61	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Button Bed	61	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Solnhofen limestone	44	✓	✓	✓	✓	✓	✓	✓		✓	✓
Source of parameter		i	iii	i&iii	iv	iv	i,iii&iv	ii	v	v	v

Sources of measurement: (i) pycnometry, (ii) permeametry, (iii) electrical, (iv) MICP, and (v) NMR. T_2 = nuclear magnetic resonance T_2 time, FFI = Free fluid index, BVI=Bulk volume index.

than the other two beds.

Each bed has its own characteristic range of porosities and permeabilities, which overlap each other, with the Whit Bed having the highest porosity and permeability and the Hard Blue having the least porosity and permeability when taken as a group. The petrophysical characteristics of the beds will be described in detail later in the paper.

The second dataset contains 61 samples of Purbeck limestone from the Swanage Quarry (now California Quarry) at (50°35′49″N 01°58′30″W). This Upper Jurassic to Lower Cretaceous limestone occurs stratigraphically just above the Portland Freestone. In this work all samples were taken from the Button Bed of the Purbeck limestone with an age of about 155 Ma, which has a lower porosity, permeability and pore throat size than other Purbeck limestone horizons (Baud et al., 2021; Brantut et al., 2018).

The third dataset contains 44 samples of Solnhofen limestone from the Solnhofen Stone Group Quarry near Solnhofen in Germany at (48°53′02″N 10°58′44″W). Solnhofen stone is a Late Jurassic limestone from the Jura mountains which has long been famed for its building and sculpting qualities as well as its exceptional preservation of a range of palaeobiota. Its small porosity, permeability and pore throat diameter distribution result from an homogeneous microstructure which arises from the sedimentation of micritic particles into an extremely calm environment (Koch, 2007; Baud et al., 2000).

1.2. Experimental methods

The measurements comprising the datasets arise from five different independent measurement methods. These are (i) pycnometry, (ii) permeametry, (iii) electrical measurements, (iv) mercury injection capillary pressure measurements (MICP), and (v) NMR spectroscopy. Consequently, any of the basic parameters shown in Table 1 from any of these independent approaches can be compared with each other according to the codes given in the bottom line of the table.

Pycnometry was carried out to obtain porosity on 1.5" diameter cores with an apparatus designed and built in-house with an accuracy of ±0.001 (i.e., 0.1% porosity). Porosity was also obtained by fluid saturation/Archimedes bulk volume and by MICP, but these data are not used as it was found that the difficulty in fully saturating these tight limestones with water or mercury led to systematic underestimations of porosity.

Permeametry was carried out using two methods depending on the permeability of the samples. A Klinkenberg-corrected steady-state helium flow approach was used for the higher permeability samples. The permeability of lower permeability samples was measured using helium pulse-decay permeametry (Jones, 1997). No systematic difference between the approaches was noted.

Electrical measurements of resistivity were made on fully saturated samples of the fluid saturating the rock using a Quadtech LCR meter at a frequency where the quadrature component was smallest (near 1 kHz). Formation factor F, connectedness G, connectivity χ, and cementation exponent m were all calculated from the electrical measurements and porosity according to the methodologies set out in Glover (2015).

Mercury injection capillary pressure (MICP) measurements were carried out with a Micromeritics Autopore V up to a maximum applied pressure of 60,000 psi. The measured capillary pressure/incremental mercury intrusion data were inverted to provide the pore throat diameter distribution of each sample. This data was used to calculate a pore diameter distribution using the method of (Glover and Déry, 2010), and the grain size distribution was subsequently calculated by combining the pore size distribution, porosity and electrical measurements according to the method of Glover and Walker (2009).

Nuclear Magnetic Spectroscopy was carried out using a 2 MHz MARAN Ultra spectrometer from Oxford Instruments, inverting the T_2 spectra to obtain Free Fluid Index (FFI) and Bulk Volume Irreducible (BVI) using a variable cut-off informed by the capillary pressure curve.

More details of experimental measurements are given in Al Khalifah

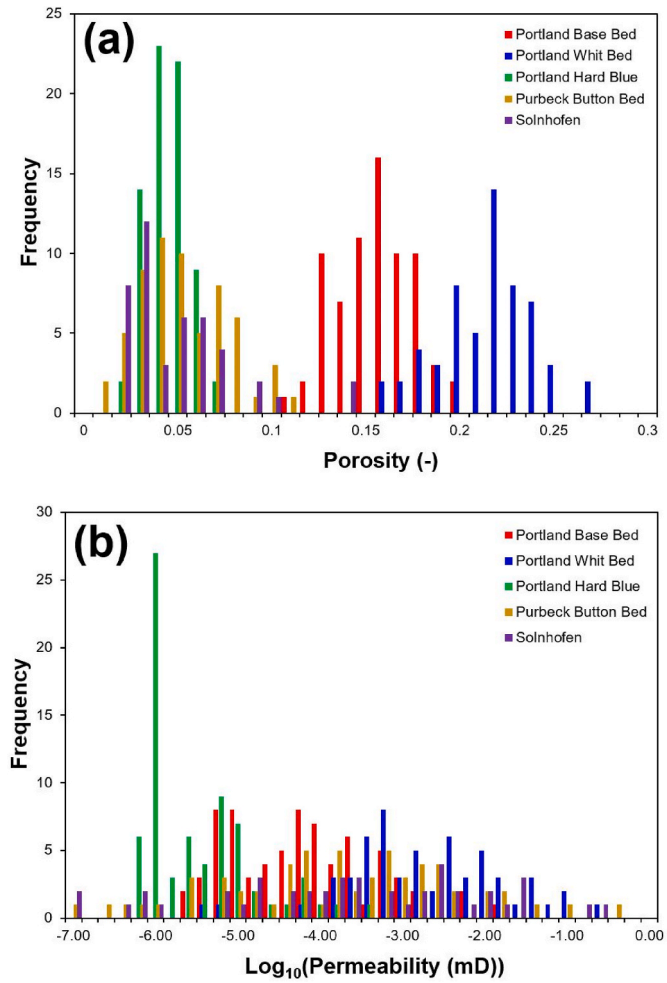

Fig. 1. Distributions of (a) porosity and (b) the logarithm of permeability to base 10 for the three types of Portland limestone, Purbeck limestone, and Solnhofen limestone.

et al. (2020).

2. Data summary

Figure 1 shows the distributions of porosity and permeability for each of the datasets. For the Portland limestone, the three component beds can be distinguished reasonably clearly on the basis of porosity, with the Hard Blue bed occupying low values with no overlap with the values of the other two beds. The Base Bed and Whit Bed exhibit higher porosities, with the Whit Bed having the higher porosities, but with a significant overlap. Table 3 shows the descriptive statistics for each bed and each dataset. However, when the permeability is considered, there is a significant overlap of all three beds even though the order of the modal values for the data from each bed is the same as that for porosity. There are even several Whit Bed samples with permeabilities lower than most of the Hard Blue samples. The porosities exhibited by the Purbeck limestone are roughly coincident with those of the Portland Hard Blue, while the Solnhofen limestone has porosities covering the same range, but extending higher, to about 14%. The permeabilities exhibited by both the Purbeck and Solnhofen limestones are distributed over the whole range that was occupied by all three Portland limestone beds.

Clearly, there are other parameters than porosity controlling permeability. These are partially known, and discussed in the next section. They include characteristic grain size and cementation exponent (Rashid et al., 2015a; 2015b; Glover et al., 2006). Fig. 2 shows cross-plots between porosity and formation factor, cementation exponent, characteristic pore throat and pore diameters and characteristic grain size as well as between characteristic pore throat size and free fluid index (FFI) from nuclear magnetic resonance (NMR) spectroscopy measurements and with characteristic grain size, all for all datasets. They show that the Hard Blue bed ought to be easy to separate from the other two Portland limestone beds, but would probably be difficult to separate from a mixture of data from all the datasets.

Part (a) of Fig. 3 shows a Poroperm diagram of all of the datasets. This figure summarises the relationship of hydraulic flow with the parameters controlling it (porosity, characteristic grain size, η, and cementation exponent).

We have imposed dashed theoretical curves from the RGPZ-carbonate model (Rashid et al., 2015b) with constant $m = 2.1$ (being the arithmetic mean of the combined Portland dataset), constant $\eta = 1.73$ (from Rashid et al., 2015b) and using four characteristic grain sizes, which vary from 3×10^{-8} m to 3×10^{-5} m. We have also added a further set of RGPZ-carbonate model curves holding characteristic grain size constant at $d_{grain} = 1.58 \times 10^{-6}$ m, which is the geometric mean of the characteristic grain size for the combined Portland dataset, with a constant $\eta = 1.73$ (from Rashid et al., 2015b) and using four values of cementation exponent varying from 2.0 to 3.5 on increments of 0.5.

It is clear that each control on permeability has a significant role to play. Permeability generally increases with increasing porosity and grain size but decreases with increasing cementation exponent or η. The clouds of data points can be associated with a particular range of each of the parameters. However, the spread of the data with respect to different RGPZ-carbonate curves for different characteristic grain size, η and cementation exponent illustrates the difficulty in predicting permeability or cementation exponent from such overlapping data, where permeability is controlled by multiple parameters.

Part (b) of Fig. 3 shows a Poroconn diagram of all of the datasets. By contrast with the Poroperm plot, this plot summarises the relationship of electrical flow with the parameters controlling it (only porosity and cementation exponent). As before, we have imposed theoretical lines on the plot in accordance with the equation $G = \varphi^m$ (Glover, 2015) for values of cementation exponent from 1.5 to 3.0 in increments of 0.5.

Once again this figure intimates that the Base Bed and Whit Bed samples might be separated from each other and from the rest of the datasets on the basis of these data, but that the Hard Blue, Purbeck Button Bed and Solnhofen limestone data is likely to be inseparable.

The next two sections consider the machine learning methods that may be used to cluster a mixed dataset and to implement the methods using the combined Portland limestone dataset. Subsequently, all of the datasets will be used to test the limitations on the efficacy of prediction of permeability using 12 machine learning datasets, 7 conventional approaches and using Genetic Algorithms.

2.1. Clustering methods

We have used input data that are based on stratigraphically-defined beds. It would be easier to define the test machine learning clustering in terms of petrofacies-defined classes than stratigraphically-defined classes. This is because the process of defining petrofacies-based classes automatically 'cleans-up' the natural variability of a given stratigraphically delineated bed by not including samples which do not fit the strict criteria upon which the petrofacies are defined. This would be the case even if the natural variability led to samples of a given type having petrophysical characteristics falling outside the petrofacies definition. Such a process would be an artificial intervention and partially invalidate any attempt to cluster by using machine learning. In addition, if machine learning techniques are to be useful in the real world, they must be able to handle data as it comes to us, and that means data that is not previously classified.

A total of 8 data clustering methods have been implemented using the artificial intelligence platform Weka® (ver. 3.8.5) (Frank et al., 2016). All of the methods used in this paper are shown in Table 4, noting

Table 3
Descriptive statistics for porosity and permeability of the datasets used in this study.

Formation	He Porosity (–)	Formation Factor (–)	Cementation exponent (–)	Characteristic pore throat size (m)	Characteristic pore size (m)	Characteristic grain size (m)	Gas Permeability (mD)	T_2 (ms)	FFI (–)	BVI (–)
Portland limestone	0.129 ± 0.0736	152 ± 188	2.10 ± 0.431	*8.53×10^{-9} ±8.47×10^{-8}	*1.41×10^{-8} ±5.75×10^{-8}	*1.58×10^{-6} ±6.82×10^{-6}	*4.60×10^{-5} ±1.64×10^{-2}	–	1.98×10^{-2} ±4.51×10^{-2}	9.8×10^{-1} ±4.51×10^{-2}
Base Bed	0.153 ±0.020	61.5 ±24.5	2.145 ±0.201	*8.40×10^{-9} ±2.55×10^{-8}	*1.39×10^{-8} ±4.23×10^{-8}	*9.76×10^{-7} ±2.38×10^{-6}	*4.92×10^{-5} ±1.42×10^{-3}	–	1.00×10^{-2} ±1.40×10^{-2}	9.9×10^{-1} ±1.40×10^{-2}
Whit Bed	0.212 ±0.025	63.2 ±35.1	2.580 ±0.253	*2.83×10^{-9} ±4.90×10^{-8}	*4.69×10^{-8} ±8.11×10^{-8}	*3.88×10^{-6} ±1.04×10^{-5}	*1.53×10^{-3} ±2.92×10^{-2}	–	1.60×10^{-2} ±2.00×10^{-2}	9.84×10^{-1} ±2.00×10^{-2}
Hard Blue	0.039 ±0.011	314 ±237	1.677 ±0.248	*3.29×10^{-9} ±6.79×10^{-9}	*5.45×10^{-9} ±1.12×10^{-8}	*1.24×10^{-6} ±4.79×10^{-6}	*2.56×10^{-6} ±4.03×10^{-5}	–	3.30×10^{-2} ±7.00×10^{-2}	9.67×10^{-1} ±7.00×10^{-2}
Purbeck limestone	0.047 ±0.024	582 ±994	1.83 ±0.223	*3.22×10^{-8} ±6.52×10^{-8}	*5.32×10^{-8} ±1.08×10^{-7}	*1.82×10^{-5} ±8.16×10^{-5}	*1.32×10^{-4} ±3.72×10^{-2}	–	5.00×10^{-2} ±4.30×10^{-2}	9.50×10^{-1} ±4.30×10^{-2}
Solnhofen limestone	0.044 ±0.029	810 ±882	1.768 ±0.297	*4.93×10^{-8} ±1.42×10^{-7}	*8.16×10^{-8} ±2.35×10^{-7}	*2.59×10^{-5} ±1.08×10^{-5}	*1.68×10^{-4} ±3.48×10^{-2}	114 ±179	1.45×10^{-1} ±1.56×10^{-1}	8.55×10^{-1} ±1.56×10^{-1}

All averages are arithmetic means except those labelled with a * which are geometric means of log-normally distributed data. All scatter represented as standard deviations.

that the KMeans and Hierarchical clustering algorithms were implemented using both the Euclidean and Manhattan distance functions. We have tuned each method to optimise its efficiency.

Some methods allow for optional approaches which we have implemented whenever it was pragmatic to do so. For example, the KMeans and Hierarchical clustering algorithms allow a number of distance functions to be explored. We have used the Euclidean, Minkowski or Manhattan distance functions in this work. All distance-based methods used in this work include a normalisation of the input values to ensure that the results can be compared. Initial testing has implemented eight methods on the Portland dataset with the number of clusters set to 3 *a priori* if allowed by the algorithm (all cases except the Cobweb method) in order to examine the efficacy of each method.

It is not known in advance which attributes are powerful in aiding clustering, which are helpful in distinguishing between two particular clusters, or even which attributes are so insensitive to cluster membership that they degrade the efficacy of clustering if included in the attributes for a given test. Consequently, we have used a 3-cluster implementation of the EM method to examine the results from all 31 combinations of the five independent parameters in this work. These tests include using all attributes, using all combinations of 4, 3 and 2 attributes as well as each attribute alone.

Most methods allow the number of clusters to be set in advance, but this is not always required. In the case of the Portland limestone, there are three clusters, each associated with a bed within the formation. We have implemented clustering with unconstrained number of clusters, where possible, as well as for 2, 3, 4 and 5 defined clusters using the combined Portland-Purbeck-Solnhofen dataset. We note that it is extremely challenging to separate 5 clusters with overlapping properties using 5 attributes.

Finally, all implementations described above have been run ten times with different stochastic seeds such that all results represent statistically valid implementations, and allow a measure of spread to be given in all of our results plots.

In summary, there were a total of 5 facies in 3 databases that were clustered with 8 clustering approaches using all combinations of 5 attributes, 11 pre-defined cluster targets and one instance of an undefined cluster target number, resulting in a total of 990 machine learning implementations.

The accuracy of clustering has been judged on the basis that the algorithms provide not only the correct number of clusters, but that the number of samples falling into each cluster are in accord with the real partitioning of the Base Bed, Whit Bed and Hard Blue formations, and most critically, that each sample is correctly clustered. These are judgements of success based on the known number of original classes and the known stratigraphical classification of each sample.

2.1.1. *Clustering results*

Machine learning clustering algorithms have been applied initially to the Portland dataset and later to the combined Portland-Purbeck-Solnhofen dataset. Machine learning processes generally perform better when used with normally-distributed attributes (Witten et al., 2011). In all cases we use the untransformed values of porosity, cementation exponent and FFI because they are distributed quasi-normally. The pore throat size and permeability are distributed approximately log-normally. Hence we use the logarithm to the base 10 of these parameters to obtain a quasi-normal distribution for them as well.

Hereafter, we use the term ‘clustering success’ rather than ‘clustering accuracy’. The word ‘accuracy’ would be a good one to use for the machine learning process of classification because this process places samples in any one of a number of pre-defined classes for which a degree of ‘accuracy’ can be calculated. However, the process of clustering does not have predefined clusters for which such an accuracy could be calculated. Instead we consider the success with which samples of the same type are gathered together in a cluster. The difference may appear trivial, but arises from fundamental differences between the processes of

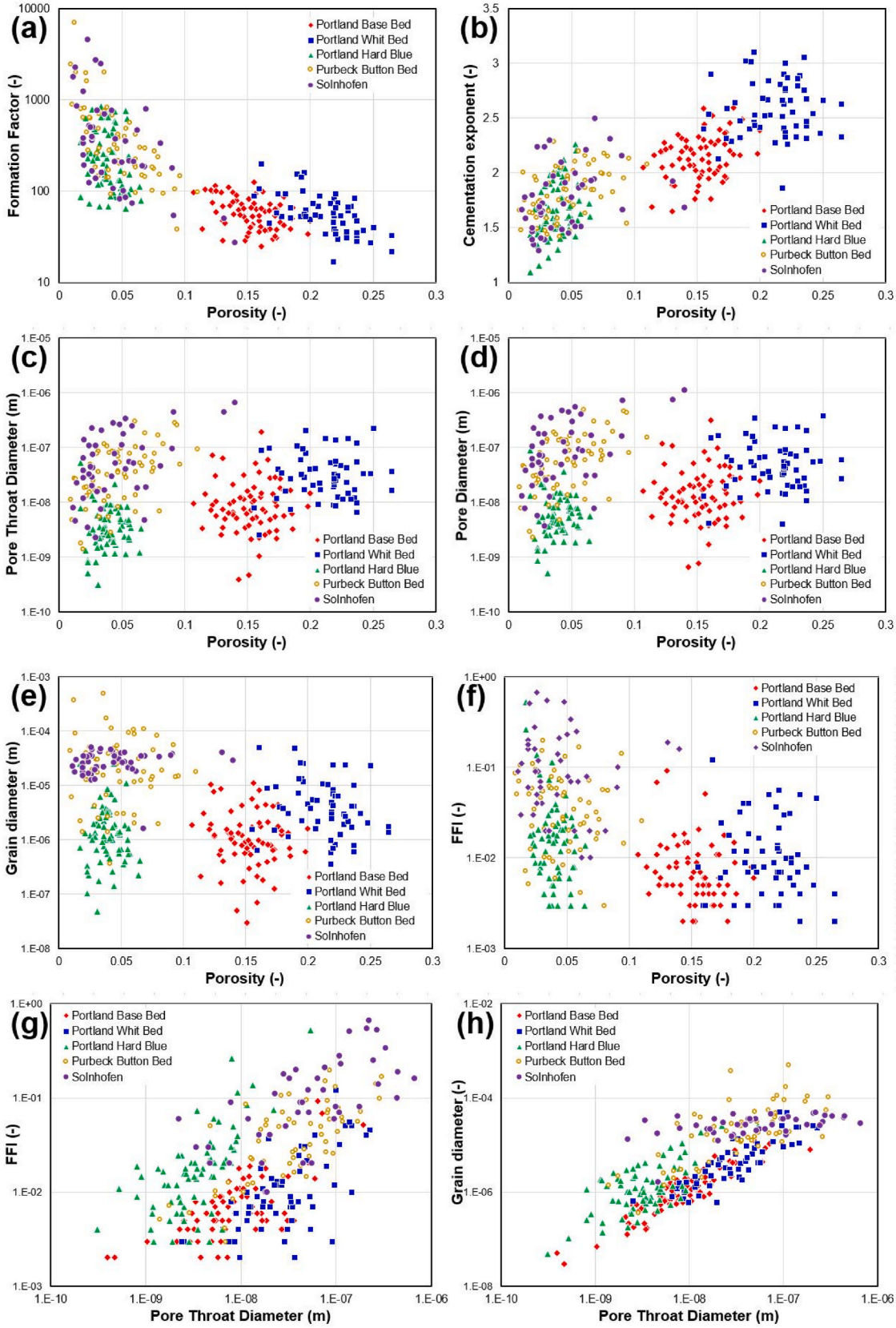

Fig. 2. Cross-plots between porosity and (a) formation factor, (b) cementation exponent, (c) characteristic pore throat diameter, (d) characteristic pore diameter, (e) characteristic grain diameter, and (f) free fluid index (FFI) size for all datasets, together with cross-plots between characteristic pore diameter and (g) FFI, and (h) characteristic grain diameter.

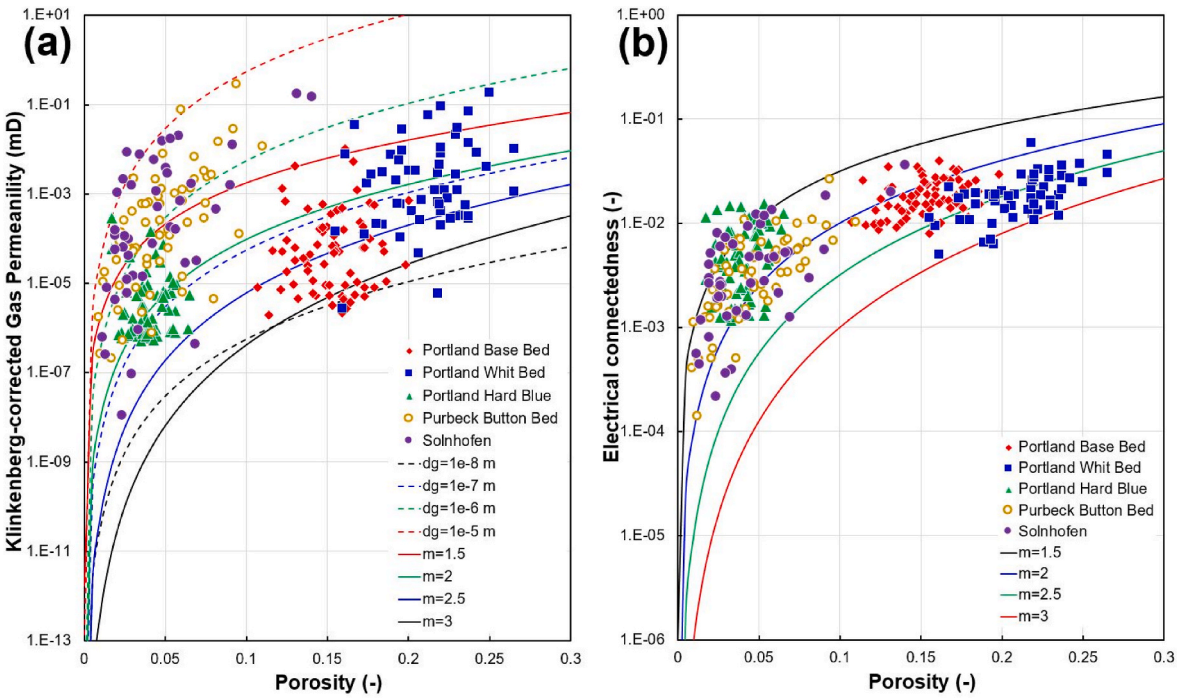

Fig. 3. (a) Poroperm diagram for all data, with RGPZ-carbonate models (Rashid et al., 2015b) for varying characteristic grain size (with $m = 2.1$, $\eta = 1.73$) shown as dashed lines, and with RGPZ-carbonate models varying cementation exponent (with $d_g = d_{grain} = 3 \times 10^{-7}$ m, $\eta = 1.73$). (b) Poroconn diagram (Electrical connectedness vs. Porosity) for all data, with theoretical lines for connectedness (Glover, 2015) for varying cementation exponent.

Table 4
Machine learning permeability prediction methods used in this work.

Method	Parameters & Notes
1. Canopy	Clustering data using the canopy clustering algorithm (McCallum et al., 2000) with the ability to define the number of searched clusters.
2. Cobweb	Clustering based in the Cobweb and ClassIt algorithms (Fisher, 1987).
3. EM (Expectation Maximisation)	EM assigns a probability distribution to each instance which indicates the probability of it belonging to each of the clusters, deciding how many clusters to create by cross validation, or specified in advance.
4. Farthest First	Defines a first cluster centre randomly and subsequent centres as farthest progressively from chosen centres followed by a nearest first population of the clusters (Kumar, 2013; Hochbaum and Shmoys, 1985).
5. KMeans	Clusters data using the k means algorithm (Arthur and Vassilvitskii, 2007), using various distance functions, including Euclidean, Minkowski or Manhattan distance functions.
6. Hierarchical Clusterer	Clustering based on hierarchical decomposition of the dataset, using various distance functions, including Euclidean, Minkowski or Manhattan distance functions.

machine learning clustering and machine learning classification (Witten et al., 2011).

3. Efficacy of different clustering methods

As we have already seen, the Portland dataset contains data from 202 samples from three beds representing three facies with different but overlapping properties. We examine how effective each of the clustering algorithms are in recognising data from the three beds.

Figure 4 shows a comparison of 8 machine learning clustering methods, constraining the cluster number to 3 in all cases except for the Cobweb method where such a constraint is not available. In this figure the clustering has been carried out with the maximum available independent data. Accordingly, five input attributes are used, i.e., porosity, cementation exponent, the logarithm to the base 10 of the characteristic pore throat size, the logarithm to the base 10 of the permeability, FFI. Other parameters are derived from one or from a number of the other parameters. Clustering success rate is calculated as the percentage of samples clustered successfully.

It is clear that some methods are much better than others. The EM and two Kmeans methods (Euclidean and Manhattan distance models) perform best, with clustering success rates of over 90% (92.58%, 89.60% and 91.09%, respectively). The five other methods performed significantly worse. The hierarchical models were the worst of all (36.14% and 36.63% success rates for the Euclidean and Manhattan distance models, respectively), which is only slightly higher than the success rate expected from a random choice of one of the 3 facies (i.e., 33.3̇3%).

Table 5 shows the confusion matrices for the 8 clustering methods examined in this paper. In each case the horizontal columns reference the actual labels for the Base Bed (BB), Whit bed (WB) and Hard Blue (HB), while the vertical labels represent the clustering results. Success is measured by the closeness of the prime diagonal values to those of the 'Control' which represents 100% success. Off-diagonal values show errors where data from a certain formation has been included in another cluster. The values shown are the rounded means of the event numbers from 10 implementations of clustering (as in Fig. 4).

The clear distinction between the Hard Blue formation and the other two would lead to the expectation that the off-diagonal components related to the Hard Blue formation would be at or close to zero. This is indeed the case for the EM, Farthest First and both Kmeans methods. The remaining methods were not capable of distinguishing between the HB

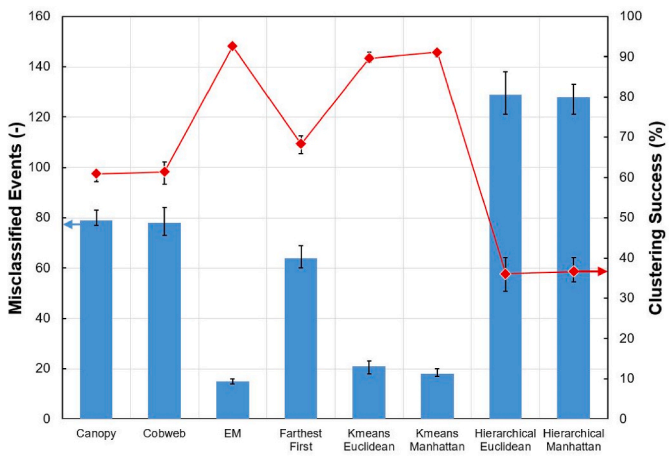

Fig. 4. Clustering results for the three facies of the Portland limestone using 8 machine learning clustering methods, constrained to 3 clusters. Five input attributes are used (porosity, cementation exponent, the logarithm to the base 10 of the characteristic pore throat size, the logarithm to the base 10 of the permeability, FFI). Arithmetic mean of misclassified events as blue bars (left-hand *y*-axis) and Arithmetic mean of percentage success as red symbols (right-hand *y*-axis). Error bars represent the minimum and maximum of 10 tests; Total samples $N = 202$; Tests = 80. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 5
Confusion matrices for the rounded mean of 10 clustering tests for each clustering method.

Control	**Actual ↓**	**BB**	**WB**	**HB**	**← Clustered as**
	BB	72	0	0	
	WB	0	58	0	
	HB	0	0	72	
Canopy	**Actual ↓**	**BB**	**WB**	**HB**	**← Clustered as**
	BB	62	10	0	
	WB	4	54	0	
	HB	65	0	7	
Cobweb	**Actual ↓**	**BB**	**WB**	**HB**	**← Clustered as**
	BB	1	38	33	
	WB	0	56	2	
	HB	0	5	67	
EM	**Actual ↓**	**BB**	**WB**	**HB**	**← Clustered as**
	BB	65	7	0	
	WB	8	50	0	
	HB	0	0	72	
Farthest First	**Actual ↓**	**BB**	**WB**	**HB**	**← Clustered as**
	BB	69	0	3	
	WB	57	1	0	
	HB	2	2	68	
Kmeans/Euclidean	**Actual ↓**	**BB**	**WB**	**HB**	**← Clustered as**
	BB	58	1	13	
	WB	0	52	6	
	HB	0	1	71	
Kmeans/Manhattan	**Actual ↓**	**BB**	**WB**	**HB**	**← Clustered as**
	BB	61	1	10	
	WB	0	52	6	
	HB	0	1	71	
Hierarchical/Euclidean	**Actual ↓**	**BB**	**WB**	**HB**	**←Clustered as**
	BB	72	0	0	
	WB	58	0	0	
	HB	70	1	1	
Hierarchical/Manhattan	**Actual ↓**	**BB**	**WB**	**HB**	**←Clustered as**
	BB	72	0	0	
	WB	58	1	0	
	HB	68	2	1	

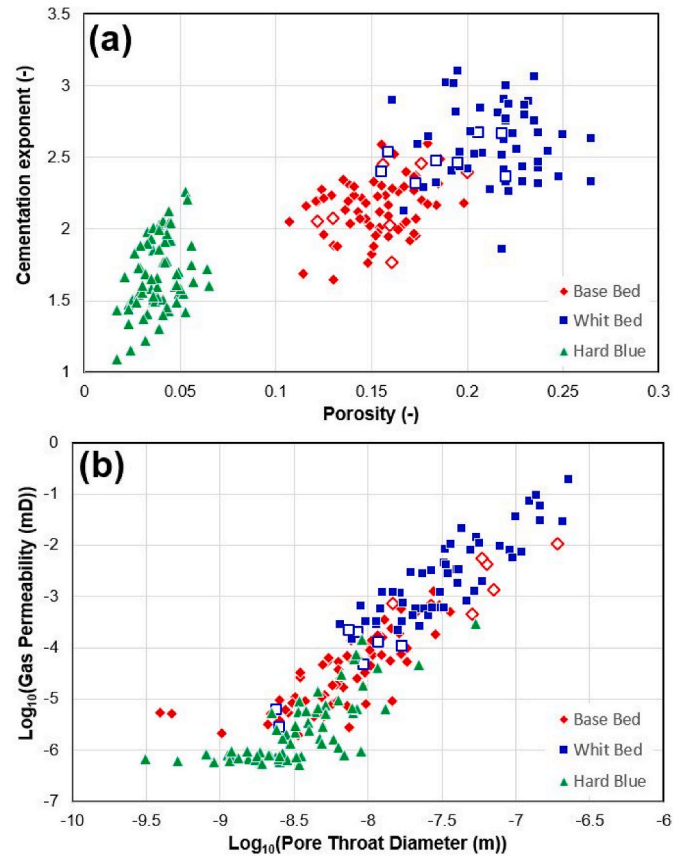

Fig. 5. An example of two attribute cross-plots coloured for each of the three facies with solid symbols showing successfully clustered samples and open samples for incorrectly clustered samples (line coloured for correct cluster) for the EM approach with the same 5 attributes as in Fig. 4 and for (a) the porosity-cementation cross-plot space, and (b) for the pore throat diameter-permeability space. (Result of part of a single test.) Total samples $N = 202$.

and the other formations. By contrast, it would be expected that the Base Bed and Whit Bed data would be more difficult to distinguish, and this is shown to be the case, with the Canopy, Cobweb, Farthest First and both Hierarchical methods effectively recognising the majority of BB and WB data to be of the same cluster. The likelihood is that this confusion is driven by overlap in more than one of the main petrophysical characteristics.

Figure 5 shows the results for the best performance (EM model) on a sample-by-sample basis for two attribute spaces, where correctly clustered points are shown as solid points and incorrectly clustered points are shown as open points with a line colour representing the cluster they should have been associated with. The perfect clustering of the Hard Blue facies is mainly due to the complete separation of this facies in terms of porosity as shown clearly in Fig. 5a. However, the 5-fold attribute dimension is important to be able to attain the high quality clustering between facies with overlapping properties, as between the Base Bed and Whit Bed in both parts of Fig. 5. It is instructive that the mis-clustered samples are not always in the zone of overlap between any two attributes.

4. Effect of different attributes

The importance of each attribute in the clustering process depends on the extent to which each attribute shows an ability to separate samples into their correct groups. Clearly, this will vary to some extent from dataset to dataset, but will also be related to what each individual attribute physically represents. Consequently, we have chosen to work

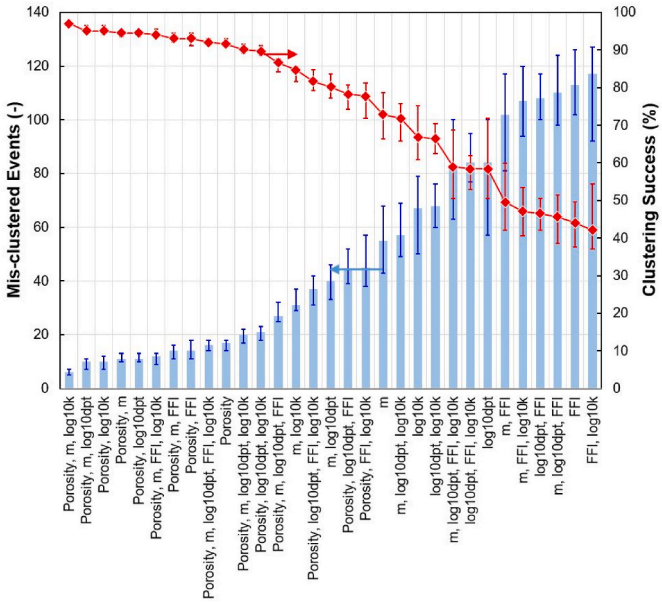

Fig. 6. Arithmetic mean of mis-clustered events (left-hand y-axis) and arithmetic mean of percentage clustering success rate (right-hand y-axis) for all combinations of 5 attributes using the EM method with 3 clusters set in advance, ranked in decreasing efficacy left to right. Error bars represent the total spread (minimum to maximum of the 10 tests per data point). Total samples $N = 202$. Total tests = 310).

with the best method we found in the previous section, the EM method, and have run it for all combinations of the 5 possible attributes, logging how each combination performs on the tight carbonate data represented by the Portland limestone group. In this process, we have started with the results of the 5-attribute, 3-cluster result already reported, then examined the clustering efficacy using the 5 combinations of 4 attributes, the 10 combinations of 3 attributes, then 10 combinations of 2 attributes and finally the 5 combinations of 1 attribute. The summary of all of this data is shown in Fig. 6. The data shown in this figure have been arranged by rank on the basis of success from the best on the left-hand of the figure to the worst on the right-hand side.

It is clear that the adage that more data is always better does not work. The implementations using all five attributes is ranked 9th in clustering quality. This implies that inclusion of some of the attributes actively degrades the clustering ability. Indeed, the top 5 attribute combinations are of only 2 or 3 attributes.

Porosity occurs as an attribute in all of the top 13 combinations and in 16 of the top 18 combinations. Indeed use of porosity as a single attribute provides the 10th best performance (91.58%). This is an indicator that porosity is the most powerful attribute for clustering in this dataset. However, it will not necessarily be the case in all petrophysical datasets.

Eight of the 9 worst performers contain FFI in the attribute set, indicating that this attribute does not contribute to improvement in clustering, and probably degrades clustering. The addition of FFI to any combination of other attributes degrades performance by as much as 13 rank places, except, oddly when it is added to the single attribute porosity. Indeed, there is a small group of highly performing combinations which contain both porosity and FFI. There is something about this combination that seems to work well together for this carbonate dataset.

Analysis of the strength of attributes on the basis of Fig. 6 results in the order first porosity, followed by cementation exponent and the logarithm to the base 10 of permeability, then the logarithm to the base 10 of pore throat size, and finally FFI.

The spread in values (calculated as the range of data over 10 iterations) are not symmetrical and grow as the clustering worsens. This is to be expected because for those tests with good success rates there are fewer possible combinations of samples in each cluster, while poorly performing implementations struggle to allocate samples to a cluster, mis-clustering many samples which vary greatly between implementations with different seeds.

5. Effect of different cluster numbers

The results described in this section use the combined Portland-Purbeck-Solnhofen dataset. Figures 1 and 2 show that there is significant overlap in the attributes of the Base Bed and Whit Bed in the Portland samples as well as between the Portland Hard Blue samples and samples of the Purbeck and Solnhofen limestones.

In this section we carry out a number of clustering tests using (i) all 5 attributes, (ii) the 4 attributes [porosity, m, $\log_{10}k$, $\log_{10}d_{pt}$, FFI], (iii) the triad [porosity, m, $\log_{10}k$], (v) the dyad [porosity, m], and (v) porosity alone. Each set of attributes has been tested for 11 different specified numbers of clusters as well as the case where the method is unconstrained by a number of clusters. Each combination has been implemented 10 times, as throughout this paper, with the plotted results being the arithmetic mean of the 10 trials with error bars representing the full spread of results (i.e., minimum to maximum). This results in 600 trials of the EM method in this sub-section.

Overall, all the results are worse than for the Portland dataset, with the best clustering success rate at only 69.38% (using all five attributes and leaving the number of clusters unconstrained). This arises because the clustering problem we have provided in the combined dataset is more difficult than the Portland dataset, with all clusters significantly overlapping with either one or two other clusters.

For the case where the number of clusters is undefined, the use of all attributes lead to a correct outcome of five clusters and a reasonable success rate (69.38%), which was the best of all of the tests. In other words, the method recognises that there are five clusters in the data even though their properties overlap significantly.

Repeating the unconstrained tests of the EM method for the 4 attributes excluding FFI results in a significantly worse performance (58.96%) and larger spread of results as shown by the larger error bars. Both are due separate trials of the method recognising different numbers of clusters, in one case choosing 10 clusters.

The triad [porosity, m, $\log_{10}k$] performed best when applied to the Portland dataset as described in the previous subsection and shown in Fig. 6. It was not as successful with the more complex combined dataset. When asked to recognise clusters, the EM method recognised only 3 for all 10 implementations, failing to recognise both the Purbeck Button Bed and Solnhofen limestone as clusters in their own right. However, it recognised the other facies well, with a success rate of 63.52%.

The dyad porosity and cementation exponent was the best of the 2-attribute combinations when applied to the Portland dataset. Here in the more complex combined dataset, it performs fairly creditably with a success rate of 65.8% compared to most other combinations, but completely fails to recognise the Solnhofen limestone as a cluster in its own right, returning only 4 clusters.

As discussed in the previous subsection, porosity was the most powerful attribute in separating clusters for the Portland dataset. In the more complex combined dataset, it fails with a success rate of 46.90% because only 2 clusters were recognised in all 10 trials. In all cases the Base Bed was combined with the Whit Bed for one cluster, while the remaining 3 facies were combined into the second cluster. Difference in the detailed results between trials was restricted only to small variations on this theme.

Consequently, trusting the clustering method to recognise the right number of clusters is difficult, but can produce the best results if there are enough attributes to separate the clusters reasonably.

For the results where the number of clusters was defined by the operator, some clear trends are apparent from Fig. 7. First, the number of mis-clustered samples increases with decreasing percentage

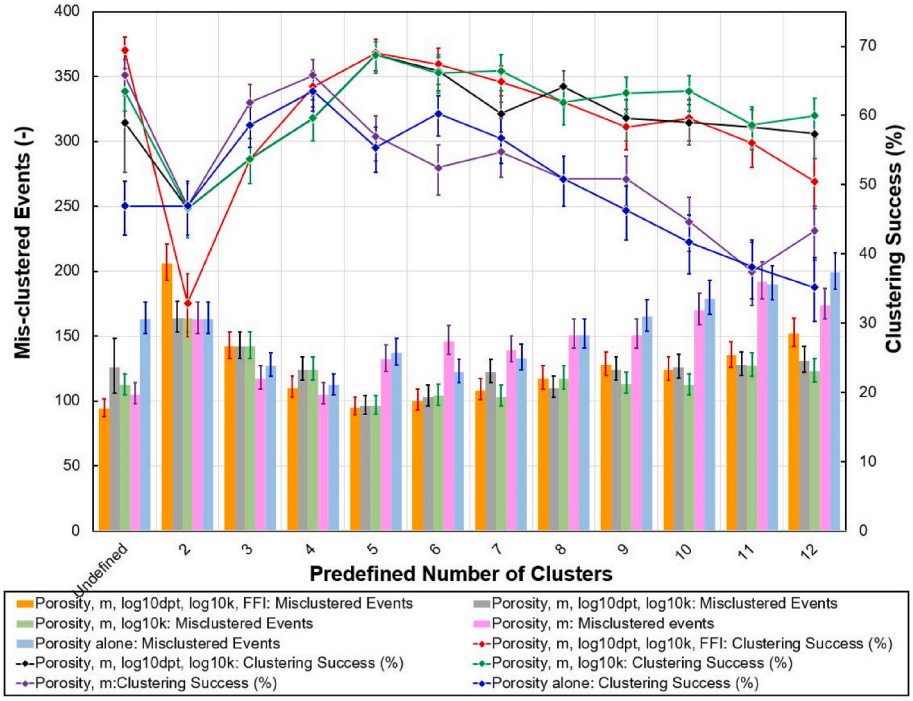

Fig. 7. Arithmetic mean of mis-clustered events (bars, left-hand *y*-axis) and arithmetic mean percentage clustering success rate (symbols/lines, right-hand *y*-axis) for 12 different pre-defined or undefined cluster numbers using the EM method using (i) all 5 attributes, (ii) the 4 attributes [porosity, m, $\log_{10}k$, $\log_{10}d_{pt}$, FFI], (iii) the triad [porosity, m, $\log_{10}k$], (iv) the dyad [porosity, m], and (v) porosity alone. Error bars represent max-min extent of 10 tests (Samples $N = 307$, total tests = 600).

Table 6
The wireline log attributes used in this work.

Attributes	Source	Petrofacies Tests					Reservoir Class Tests		
		(i)	(ii)	(iii)	(iv)	(v)	(i)	(ii)	(iii)
Depth	Wireline log measurements	✓					✓		
Caliper Log [CAL]		✓					✓	✓	
Gamma Ray Log [GR]		✓					✓	✓	
Density Log [RHOB]		✓					✓	✓	
Neutron Log [NPHI]		✓	✓				✓	✓	
Sonic Log [DT]		✓			✓		✓	✓	
Shale volume [VSH]	Derived from wireline log measurements by analysis	✓					✓		
Corrected Neutron Porosity [CNPHI]				✓					
Bulk porosity [PHIB]		✓					✓		
Effective porosity [PHIE]		✓					✓		
$\log_{10}$(Permeability) [$\log_{10}k$]		✓	✓	✓	✓	✓	✓		✓

The data have already been analysed to define 4 petrofacies, 9 reservoir classes and 19 more restrictive reservoir zones (Mohammed Sajed et al., 2021). Fig. 8 shows the analysed well together with the interpreted lithologies, permeabilities, petrofacies and reservoir classes from Mohammed Sajed et al. (2021).

clustering success as the number of clusters is progressively less than the real number of clusters. This is primarily because the clustering method is forced to place samples into too few categories, partitioning clusters between other clusters. Second, the number of mis-clustered samples also increases, again with decreasing percentage clustering success as the number of clusters is progressively more than the real number of clusters. In this case, the decrease in clustering efficiency is due to there being more categories than necessary, which leads to samples being placed in a cluster that is not required.

Examination of Fig. 7 also shows generic differences between the number of attributes and the efficacy of classification which occur irrespective of the pre-defined number of clusters. The tests with 3 or more attributes all share one behaviour that leads to higher clustering success and lower mis-clustered samples compared with the tests with one or two attributes. These 3+ attributes show peak clustering success for 5 clusters as ought to be the case, whereas the 1- and 2-attribute trends peak at 4 clusters with the Solnhofen limestone samples generally being distributed between the Portland Hard Blue and Purbeck Button Bed clusters. Here is another example of needing to have enough attributes to provide efficient clustering.

Taking all of the results in this subsection into account, we propose a rule-of-thumb that the number of independent attributes should be at least the same as the number of expected clusters, and preferably more.

6. Wireline log data

A set of data was used from well Bm-15 of the Butmah field, which is a tight carbonate field in the north-western Iraq (Mohammed Sajed and Glover, 2020; Mohammed Sajed et al., 2021). This data includes 11 possible attributes shown in Table 6 over a depth range from 394.8 m continuous depth interval, representing 1317 depths.

Petrofacies are classes of rock based upon petrophysical characteristics. There are 4 of these in the dataset used in this work as defined by and fully described in Mohammed Sajed et al. (2021). They have been given the labels A, B, C and D. Petrofacies A represents the best reservoir quality rock, followed by the others in alphabetic order. It is important

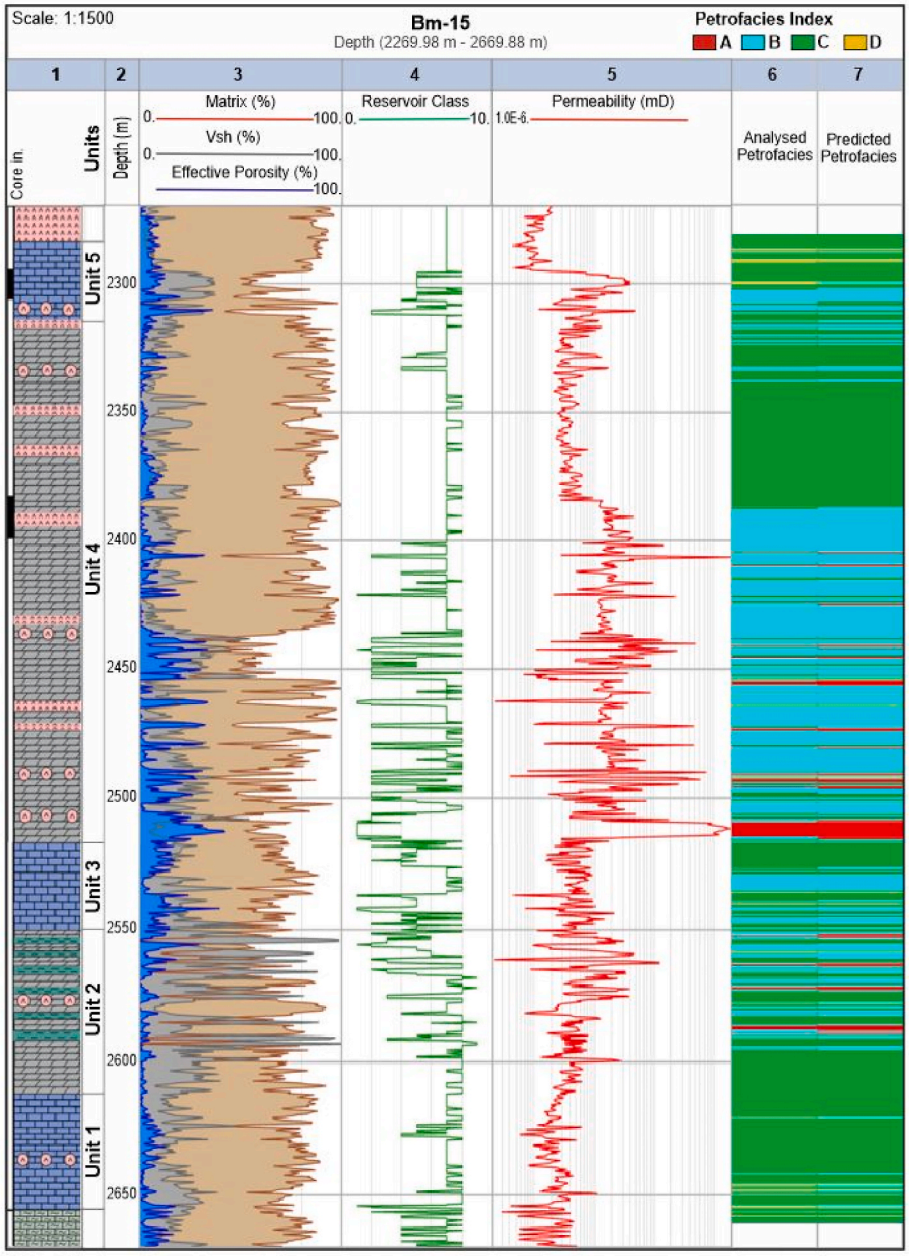

Fig. 8. The Bm-15 well analysed for lithology, matrix, porosity, shale volume, and permeability, together with the interpreted petrofacies (Col 6) and reservoir quality classes (Col 4) and the machine learning predicted petrofacies (Col 7) (based on Fig. 9 from Mohammed Sajed et al. (2021)).

that the clustering can distinguish between the best rock (Petrofacies A and B) from the lower quality rock, represented by Petrofacies C and especially Petrofacies D. Track 6 shows the petrofacies as defined by detailed petrophysical analysis, while Track 7 shows the results of clustering.

Track 4 of Fig. 8 also contains a more detailed classification of reservoir rocks called Reservoir Classes. There are 9 reservoir classes, labelled numerically from 1 to 9, which are also described fully in Mohammed Sajed et al. (2021). These classes may be viewed as a more subtle classification of the rocks within the test well, and hence likely to pose a greater challenge to the clustering algorithm.

Figure 9 shows cross-plots of five of the attributes, coloured for the recognised petrofacies. It is immediately clear that there is considerable overlap between the petrofacies in all of the attributes except permeability. Consequently, we hypothesise that permeability will be the most powerful attribute in clustering, and that the other attributes may contribute little to the overall clustering success or even actively degrade it. We have not shown a diagram coloured for the 9 reservoir classes as these are even more interlinked. Consequently, the challenge to separate the nine reservoir classes is a very difficult one.

We have applied the Expectation Maximisation clustering method to ascertain if it can reasonably recognise the petrofacies and the reservoir classes by comparing the results with the manually analysed petrofacies and reservoir classes.

6.1. Petrofacies clustering

Figure 10 shows the results of 450 implementations of the Expectation Maximisation algorithm for 9 scenarios, including letting the method chose the number of clusters and for a predefined number of clusters from 1 to 8. Each of these has been carried out for 5 combinations of attributes, which have been chosen strategically (i.e., the use of

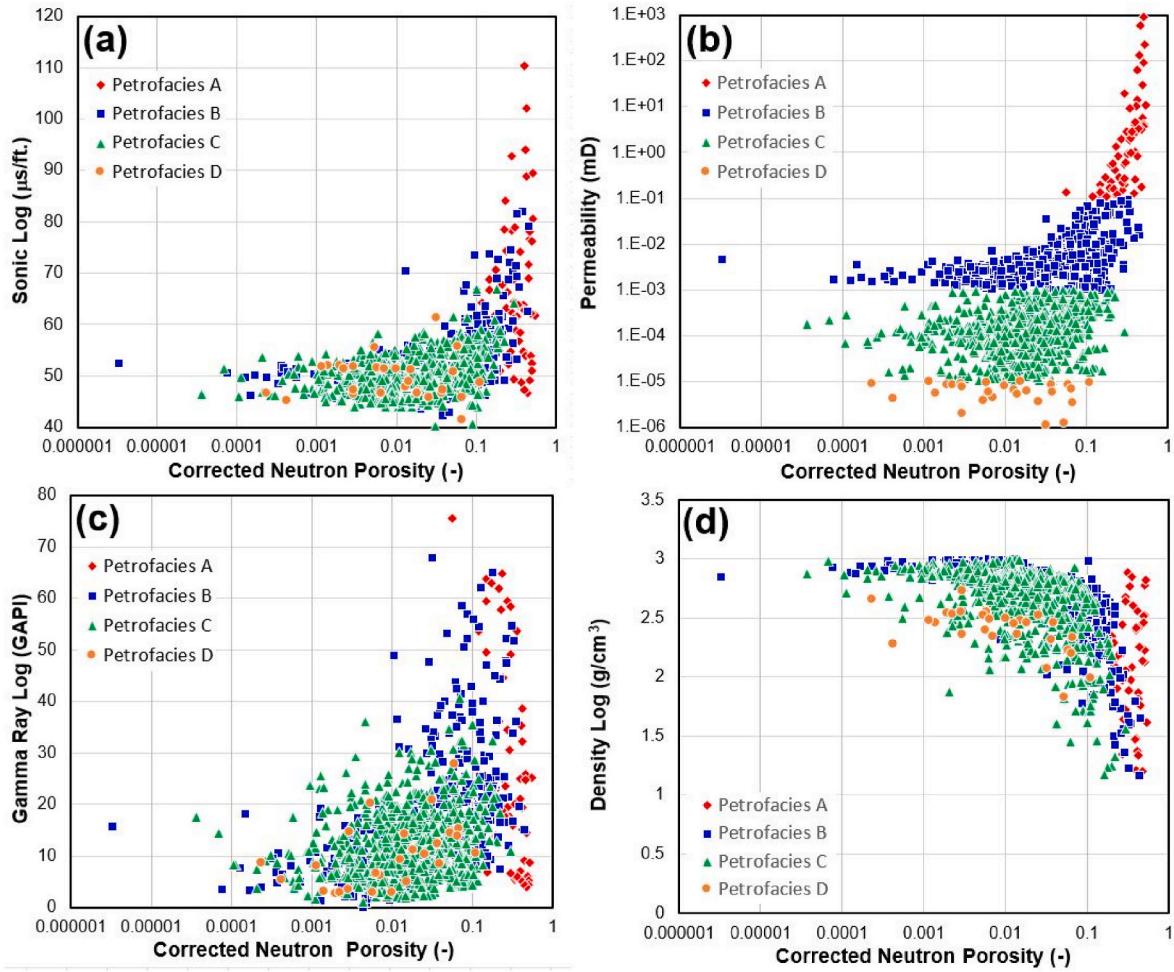

Fig. 9. Cross-plots between five of the main attributes in the Butmah Bm-15 dataset. In each case the *x*-axis represents the neutron porosity corrected for lithology ($N = 1317$).

all attributes) or based on efficacy (i.e., which is best). The combinations are: (i) all 11 attributes except CNPHI, (ii) the $\log_{10}k$ and NPHI attributes, (iii) the $\log_{10}k$ and CNPHI attributes, (iv) the $\log_{10}k$ and DT attributes, and (v) $\log_{10}k$ alone. As throughout this work, all implementations have been carried out 10 times with different seeds and the results shown are the arithmetic mean values with the error bars representing the total range of responses.

Leaving the number of target clusters undefined led to a large range of results. Using 10 attributes gave the worst response with an average of 990.3 depths being mis-clustered into 20 clusters instead of the 4 that were actually present, giving a success rate of only 24.8%. By contrast the use of $\log_{10}k$ and the dyad of $\log_{10}k$ with the corrected neutron porosity were best (249 and 373 depths mis-clustered, 81.09% and 79.27% success rates, both respectively, and both predicting the correct number of clusters. These results indicate strongly that the inclusion of attributes which do not effectively separate data actively degrades the performance of the clustering process.

The trivial case where only 1 cluster is forced for this method counter-intuitively produce values less than 100% for all attribute combinations because the method recognises the density of data points as a cluster, automatically excluding outliers. This is the best that using 10 attributes can do, with all greater predefined cluster numbers giving lower efficacy, and only a 35.53% success rate for the imposition of the actual number of clusters present. For all other predefined cases the success rate increases to a peak near the actual number of clusters (3 or 4), sometimes significantly, for other attribute combinations, followed by a decrease as the predefined number of clusters increases further.

Considering only those results for the correct number of clusters, the combination of $\log_{10}k$ and NPHI performs only moderately (458 depths mis-clustered and 66.56% success rate), but the performance increases substantially when the value of NPHI is corrected to take account that this log is only accurate for limestone and needs to be corrected for use in anhydrite, dolomite and shales. Once correction has been carried out, the combination of $\log_{10}k$ and CNPHI performs best overall (110 depths mis-clustered and 91.65% success rate). Other reasonable performances include using $\log_{10}k$ alone and the combination of $\log_{10}k$ and the sonic log, giving (249 and 302 depths mis-clustered, 81.09% and 77.07% success rates, both respectively). The results of the case for using log10 k alone are also shown in Column 7 of Fig. 8, showing how close the prediction is to the petrofacies defined from analysis of the petrophysical properties and logs from the well (Column 6).

Consequently, we infer that the clustering process using the Expectation Maximisation method can work very well even if the number of attributes is less than the number of clusters, but only if the attributes separate the data well. This is a general feature of even the best machine learning codes, if the input data is garbage, the output will also be garbage (GIGO).

6.1.1. *Reservoir class clustering*

The best performance for predicting the 4 petrofacies was 91.65%. The question now arises how effectively the 9 reservoir classes may be clustered. Fig. 11 shows the results of 390 implementations of the Expectation Maximisation algorithm for 13 scenarios, including letting the method chose the number of clusters and for a predefined number of

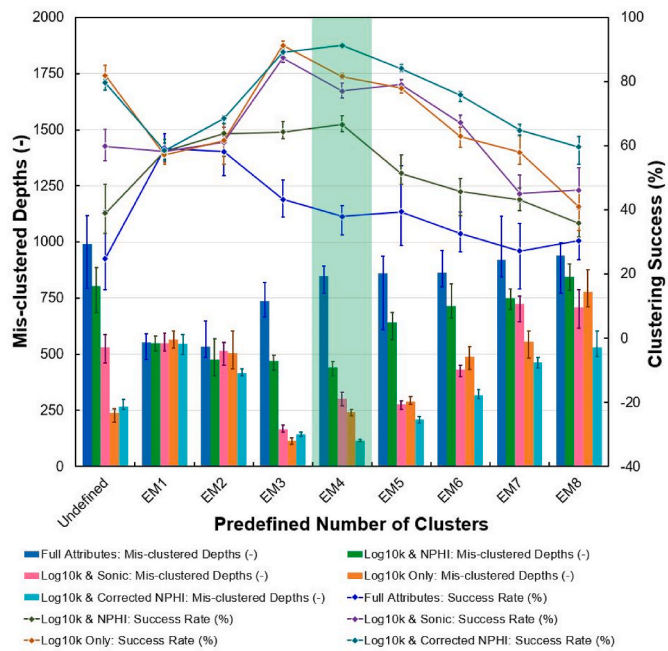

Fig. 10. Arithmetic mean of mis-clustered events (bars, left-hand *y*-axis) and arithmetic mean percentage clustering success rate (symbols/lines, right-hand *y*-axis) for 9 different pre-defined or undefined cluster numbers using the EM method using (i) all 11 attributes except CNPHI, (ii) the $\log_{10}k$ and NPHI attributes, (iii) the $\log_{10}k$ and CNPHI attributes, (iv) the $\log_{10}k$ and DT attributes, and (v) $\log_{10}k$ alone. Error bars represent max-min extent of 10 tests (Samples $N = 1317$, total tests = 450). The green box represents the actual number of clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

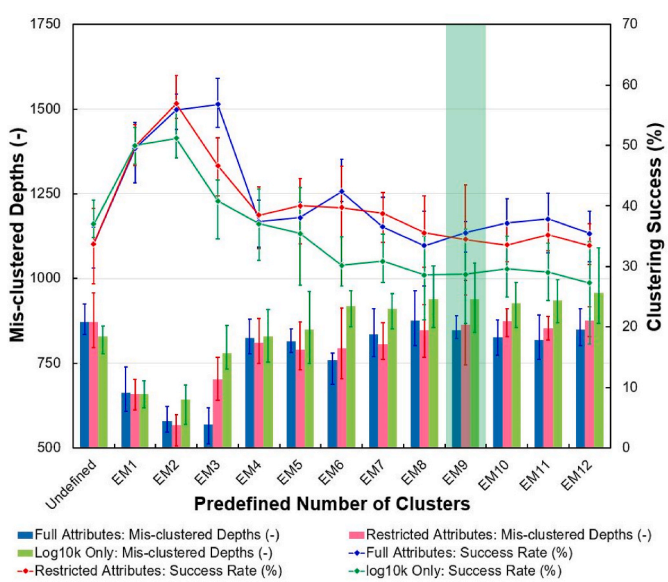

Fig. 11. Arithmetic mean of mis-clustered depths (bars, left-hand *y*-axis) and arithmetic mean percentage clustering success rate (symbols/lines, right-hand y-axis) for 13 different pre-defined or undefined cluster numbers using the EM method using (i) all 10 attributes except CNPHI, (ii) a restricted attribute set of 5 attributes [caliper, gamma ray, density, neutron and sonic logs], and (iii) $\log_{10}k$ alone. Error bars represent max-min extent of 10 tests (total tests = 390). The green box represents the actual number of clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

clusters from 1 to 12. Each of these has been carried out for the 3 combinations of attributes: (i) all 11 attributes except CNPHI, (ii) a restricted range of attributes defined in Table 5, and (iii) log10 k only. As throughout this work, all implementations have been carried out 10 times with different seeds and the results shown are the arithmetic mean values with the error bars representing the total range of responses.

As we have already noted, separating 9 clusters from a set of attributes which overlap so much is extremely challenging. The results shown in Fig. 11 are typical for most combinations of attributes. Leaving the number of clusters undefined resulted in 12, 14 and 4 predicted clusters for the three sets of attributes, respectively, with low success rates (33.82%, 33.74%, 36.90%, respectively). All of these can be considered to be failures as they provide clearly wrong sets of clusters, but success rates are still higher than a random allocation of depths to clusters, which would have a success rate of 11.11%.

For the predefined cluster tests, there was a peak for all sets of attributes caused by groups of clusters being well-recognised, but a low success rate for most pre-defined cluster results from 4 to 12 clusters, including the 9-cluster result which represents reality. For the 9-cluster solutions, the worst result was for the use of $\log_{10}k$ alone (938.9 mis-clustered events, 28.71% success rate), which is the opposite than we found for the prediction of the 4 petrofacies earlier. The results from the full 10 attributes and the restricted attribute dataset were similar (848.4 and 863.1 mis-clustered depths and 35.58% and 34.46% success rates, respectively) with the full dataset providing marginally the better results. Once again, all of these results, while better than random allocation, are insufficiently good to be used in practice. Consequently, we conclude that the complexity of the wireline dataset used is too great a challenge for the attributes that were available to us.

6.2. Advantages and limitations of clustering

Clustering has a number of clear advantages over classification schemes. The prime advantage is that it requires no training and does not depend on a previously created classification scheme. This makes clustering an inherently more general and robust method for grouping data. The lack of any assumptions and training to those assumptions, implies that the clustering scheme has, in principle, a less constrained ability to form clusters.

Nevertheless, it is certainly the case that clustering has not been nearly as popular as classification schemes. This is almost certainly due to the nervousness of researchers to let the clustering algorithms loose without supervision. Those requiring facies to be separated into like clusters would rather it be done into classes they define for whatever anthropic predisposition.

So, fundamentally, clustering does not provide recognition of any particular set of facies but a recognition of similarity, to which a human may later assign a facies label. By contrast, classification starts with the definition of particular facies, to which a machine learning algorithm might assign data, having been trained to do so. The clustering algorithm is to the classification algorithm as an artist is to an accountant. The later might be more reliable, but is less likely to provide as much insight.

Clustering does have a number of clear disadvantages. The first is that this work makes it clear that clustering works better if there is prior knowledge of the number of clusters that would best describe the dataset. If this is unknown, some algorithms exist which can optimise for the number of clusters, and this work has shown them to work well. Alternatively, the search for the best number of clusters could be done by batch application with a human choice of best cluster number. This has also been done in this work, and found to work well, but is onerous if not automated.

The second important disadvantage is that the clustering algorithm can become overwhelmed if there are a large number of clusters. In this work, clustering performed well when the data formed 4 clusters (4 petrofacies), but badly when the data formed 9 clusters (reservoir

classes). In general there needs to be an increased number of data if more clusters are expected, and we define a rule of thumb whereby there should be at least as many independent characteristics being rested than expected clusters.

7. Conclusions

This paper presents a study on the clustering efficiency of 8 machine learning clustering methods with a large database ($N = 307$) of five carbonate facies. The facies were the Base Bed, Whit Bed and Hard Blue bed of the Portland limestone, the Purbeck Button Bed and the Solnhofen limestone. A wide range of petrophysical tests were carried out and have been analysed in detail prior to machine learning testing.

Up to five independent attributes were selected for clustering tests. A total of 990 tests were carried out to examine (i) the efficacy of 8 clustering methods, (ii) the effect of using each of the 31 different combinations of attribute available, and (iii) the effect of leaving the target number of clusters undefined or each of 11 defined cluster numbers.

We find that the Expectation Maximisation (EM) performs best on our petrophysical data, with the Kmeans approach with both Euclidean and Manhattan distance models coming a creditable second and third. The other methods tested in this work did not provide results of sufficient quality to allow us to recommend them for use with petrophysical data.

Good clustering requires the attribute set to be chosen carefully, with pre-implementation tuning to ensure good performance. It is clear that the adage that more data is always better does not work. We found that often the use of a limited number of good attributes was better than using a larger set of attributes. We associate this observation with the idea that some attributes (porosity in this work) are more powerful than others in separating data into clusters, while other attributes (FFI in this work) can degrade clustering efficiency if included in the attribute set. However, the efficacy of any given attribute will probably vary from dataset to dataset.

Clustering performance of a set of attributes in one dataset will be different than that in another. Here we see that clustering performance is different in the larger combined dataset than the Portland dataset.

It can be effective to let the clustering method automatically recognise the number of clusters required for the data. Indeed it provided the best clustering in this study. However, it is only effective when the number of independent attributes is equal to or greater than the number of clusters existing in the dataset. It is probably better to pre-define the number of target clusters if it is possible to do so.

Finally, we propose a rule-of-thumb that the number of independent attributes should be at least the same as the number of expected clusters, and preferably more.

The conclusions drawn from earlier in the work have been applied to the problem of facies recognition by clustering using a set of wireline logs. We find that clustering is capable of recognising major facies boundaries providing the number of independent wireline logs used as attributes exceeds the number of facies in the well.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Al Anazi, A., Gates, I.D., 2010. On the capability of support vector machines to classify lithology from well logs. Natl. Resour. Res. 19 (2), 125–139. https://doi.org/10.1007/s11053-010-9118-9.

Al Khalifah, H., Glover, P.W.J., Lorinczi, P., 2020. Permeability prediction and diagenesis in tight carbonates using machine learning techniques. Mar. Petrol. Geol. 112.

Al Qassab, H.M., Fitzmaurice, J., Al-Ali, Z.A., Al-Khalifa, A., Aktas, G.A., Glover, P.W.J., 2000. Cross-discipline integration in reservoir modeling: the impact on fluid flow simulation and reservoir management. Proc. - SPE Ann. Tech. Conf. Exhibit. 2000, 27–38.

Alexsandro, G.C., da P Carlos, A.C., Geraldo, G.N., 2017. Facies classification in well logs of the Namorado oilfield using support vector algorithm. Proc. Intl. Congr. Braz. Geophys. Soc. 1853–1858.

Al-Zainaldin, S., Glover, P.W.J., Lorinczi, P., 2017. Synthetic fractal modelling of heterogeneous and anisotropic reservoirs for use in simulation studies: implications on their hydrocarbon recovery prediction. Transport Porous Media 116 (1), 181–212.

Arthur, D, Vassilvitskii, S., 2007. k-means++: the advantages of carefull seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp. 1027–1035.

Barton, C.M., Woods, M.A., Bristow, C.R., Newell, A.J., Westhead, R.K., Evans, D.J., Kirby, G.A., Warrington, G., 2011. Geology of South Dorset and South-East Devon and its World Heritage Coast, Special Memoir. British Geological Survey, p. 161.

Baud, P., Schubnel, A., Wong, T.-F., 2000. Dilatancy, compaction, and failure mode in Solnhofen limestone. J. Geophys. Res. Solid Earth 105 (B8), 19289–19303.

Baud, P., Hall, S., Heap, M.J., Ji, Y., Wong, T.-F., 2021. The brittle-ductile transition in porous limestone: failure mode, constitutive modeling of inelastic deformation and strain localization. J. Geophys. Res. Solid Earth 126 (5).

Bestagini, P., Lipari, V., Tubaro, S., 2017. A Machine Learning Approach to Facies Classification Using Well Logs. Society of Exploration Geophysicists *Technical Program Expanded Abstracts*, pp. 2137–2142. https://doi.org/10.1190/segam2017-17729805.1.

Brantut, N., Baker, M., Hansen, L.N., Baud, P., 2018. Microstructural control of physical properties during deformation of porous limestone. J. Geophys. Res. Solid Earth 123 (6), 4751–4764.

Cao, B., Luo, X., Zhang, L., Lei, Y., Zhou, J., 2020. Petrofacies prediction and 3-D geological model in tight gas sandstone reservoirs by integration of well logs and geostatistical modeling. Mar. Petrol. Geol. 114.

Cai, L., Wang, H., Jiang, F., Zhang, Y., Peng, Y., 2022. A New Clustering Mining Algorithm for Multi-Source Imbalanced Location Data, vol. 584. Information Sciences, pp. 50–64.

Fisher, D., 1987. Knowledge acquisition via incremental conceptual clustering. Mach. Learn. 2 (2), 139–172.

Frank, E., Hall, M.A., Witten, I.H., 2016. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques, fourth ed. Morgan Kaufmann. 2016.

Fang, X., Feng, H., 2021. Study on discriminant method of rock type for porous carbonate reservoirs based on Bayesian theory. Sci. Rep. 11 (1).

Glover, P.W.J., 2015. Geophysical properties of the near surface earth: electrical properties. Treat. Geophys. 89–137 second ed.

Glover, P.W.J., Déry, N., 2010. Streaming potential coupling coefficient of quartz glass bead packs: dependence on grain diameter, pore size, and pore throat radius. Geophysics 75 (6), F225–F241.

Glover, P.W.J., Walker, E., 2009. Grain-size to effective pore-size transformation derived from electrokinetic theory. Geophysics 74 (1), E17–E29.

Glover, P.W.J., Zadjali, I.I., Frew, K.A., 2006. Permeability prediction from MICP and NMR data using an electrokinetic approach. Geophysics 71 (4), F49–F60. https://doi.org/10.1190/1.2216930.

Hochbaum, D.S., Shmoys, D.B., 1985. A best possible heuristic for the k-center problem. Math. Oper. Res. 10 (2), 180–184.

Imamverdiyev, Y., Sukhostat, L., 2019. Lithological facies classification using deep convolutional neural network. J. Pet. Sci. Eng. 174, 216–228. https://doi.org/10.1016/j.petrol.2018.11.023.

Jiang, J., Xu, R., James, S.C., Xu, C., 2021. Deep-learning-based vuggy facies identification from borehole images. SPE Reservoir Eval. Eng. 24 (1), 250–261.

Jones, S., 1997. A technique for faster pulse-decay permeability measurements in tight rocks. SPE Form. Eval. 12, 19–26. https://doi.org/10.2118/28450-PA.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithms: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. https://doi.org/10.1109/TPAMI.2002.1017616.

Koch, R., 2007. Sedimentological and petrophysical characteristics of Solnhofen monument stones - lithographic limestone: a key to diagenesis and fossil preservation. Neues Jahrb. Geol. Palaontol. - Abhand. 245 (1), 103–115.

Kopaska-Merkel, D.C., Friedman, G.M., 1989. Petrofacies analysis of carbonate rocks: example from lower paleozoic hunton group of Oklahoma and Texas. AAPG (Am. Assoc. Pet. Geol.) Bull. 73 (11), 1289–1306.

Kumar, M., 2013. An Optimized Farthest First Clustering Algorithm. IEEE, 978-1-4799-0727-4/vol. 13.

Larose, D.T., 2006. Data Mining Methods and Models. Data Mining Methods and Models, pp. 1–322.

Liu, X., Song, H., 2020. Automatic identification of fossils and abiotic grains during carbonate microfacies analysis using deep convolutional neural networks. Sediment. Geol. 410.

McCallum, A., Nigam, K., Ungar, L.H., 2000. Efficient Clustering of High Dimensional Data Sets with Application to Reference Matching. Proceedings of the sixth ACM SIGKDD internation conference on knowledge discovery and data mining ACM-SIAM symposium on Discrete algorithms 169–178.

Mohammed Sajed, O.K., Glover, P.W.J., 2020. Dolomitisation, cementation and reservoir quality in three Jurassic and Cretaceous carbonate reservoirs in north-western Iraq. Mar. Petrol. Geol. 115.

Mohammed Sajed, O.K., Glover, P.W.J., Collier, R.E.L., 2021. Reservoir quality estimation using a new ternary diagram approach applied to carbonate formations in north-western Iraq. J. Petrol. Sci. Eng. 196.

Mohammed-Sajed, O.K., Glover, P.W.J., 2022. Influence of an hydritisation on the reservoir quality of the Butmah Formation in north-western Iraq. Mar. Petrol. Geol. 135.

Nanjo, T., Tanaka, S., 2020. Carbonate lithology identification with generative adversarial networks. Int. Petrol. Technol. Conf. 2020. IPTC 2020.

Rashid, F., Glover, P.W.J., Lorinczi, P., Collier, R., Lawrence, J., 2015a. Porosity and permeability of tight carbonate reservoir rocks in the north of Iraq. J. Petrol. Sci. Eng. 133, 147–161.

Rashid, F., Glover, P.W.J., Lorinczi, P., Hussein, D., Collier, R., Lawrence, J., 2015b. Permeability prediction in tight carbonate rocks using capillary pressure measurements. Mar. Petrol. Geol. 68, 536–550.

Ren, X., Hou, J., Song, S., Liu, Y., Chen, D., Wang, X., Dou, L., 2019. Lithology identification using well logs: a method by integrating artificial neural networks and sedimentary patterns. J. Petrol. Sci. Eng. 182, 106336.

Ren, Q., Zhang, H., Zhang, D., Zhao, X., Yan, L., Rui, J., 2022. A novel hybrid method of lithology identification based on k-means++ algorithm and fuzzy decision tree. J. Pet. Sci. Eng. 208, 109681 https://doi.org/10.1016/j.petrol.2021.109681.

Santos, D.T.D., Roisenberg, M., Nascimento, M.D.S., 2022. Deep recurrent neural networks approach to sedimentary facies classification using well logs. IEEE Geosci. Rem. Sens. Lett. 19 https://doi.org/10.1109/LGRS.2021.3053383.

Silva, A.A., Tavares, M.W., Carrasquilla, A., Misságia, R., Ceia, M., 2020. Petrofacies classification using machine learning algorithms. Geophysics 85 (4), WA101–WA113.

Sinan, S., Glover, P.W.J., Lorinczi, P., 2020. Modelling the impact of anisotropy on hydrocarbon production in heterogeneous reservoirs. Transport Porous Media 133 (3), 413–436.

Tewari, S., Dwivedi, U.D., 2020. A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies. J. Pet. Explor. Prod. Technol. 10 (5), 1849–1868.

Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques, third ed. Morgan Kaufmann Publishers, Burlington.

Wu, X., Kumar, V., Ross, Q.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. Knowl. Inf. Syst. https://doi.org/10.1007/s10115-007-0114-2.

Yarmohammadi, S., Wood, D.A., Kadkhodaie, A., 2020. Reservoir microfacies analysis exploiting microscopic image processing and classification algorithms applied to carbonate and sandstone reservoirs. Mar. Petrol. Geol. 121.