PAPER

Quantitative characterisation of Quaternary glaciofluvial aquifer heterogeneity using cluster analysis

Felipe Gallardo Ceron[1] · Landis Jared West[1] · Ian T. Burke[2] · James Graham[3] · Luca Colombera[4]

Abstract

Hydrogeological modelling requires the characterisation of hydrofacies and the representation of their spatial distribution. In this work, a workflow was developed to characterise heterogeneous unconsolidated glaciofluvial sediments in Northwest Cumbria (UK), which involved: (1) field sampling; (2) lithofacies classification; (3) in situ porosity measurements using the sand-replacement method; (4) determination of the particle size distribution (PSD); (5) hydrofacies definition via K-means cluster analysis using PSD data; (6) evaluation of empirical equations for predicting porosity using field measurements and regression analysis; and (7) estimation of hydraulic conductivity (K) using the Kozeny–Carman equation. Nine lithofacies were identified, including glaciofluvial silts, sands and gravels, and local till deposits. Three clusters were defined on the basis of PSD: fine-dominated (cluster-1), sand-dominated (cluster-2) and gravel-dominated (cluster-3). Cluster-1 exhibited the highest porosities (average 44%); cluster-2 showed intermediate to high porosities, with an average porosity of 40%; and cluster-3 had the lowest porosities (average 27%). The logarithm of d_{50} was the parameter with the highest correlation with measured porosities (R^2 of 0.789). K values estimated using the Kozeny–Carman equation ranged between 0.06 and 0.2 m/d for cluster-1, 0.2–11 m/d for cluster 2, and 0.1–62 m/d for cluster 3. Measured porosities were higher than previously reported, while estimated K values were consistent with those from hydraulic tests. Comparison between lithofacies and clustering classification suggests that, for hydrofacies classification, the unsupervised cluster analysis approach is able to generate a classification that captures the hydrogeologically important details without creating an excessive number of categories.

Keywords Hydraulic properties · Heterogeneity · Glaciofluvial · Clustering analysis · Kozeny–Carman

Introduction

Understanding and representing sedimentary and petrophysical heterogeneities of geological media remain vital tasks in subsurface studies. The effects of sedimentary heterogeneity are shown in many studies (e.g. Anderson 1989; Koltermann

✉ Felipe Gallardo Ceron
eefegc@leeds.ac.uk

✉ Landis Jared West
l.j.west@leeds.ac.uk

1   Institute of Applied Geosciences, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK

2   Earth Surface Science Institute, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK

3   National Nuclear Laboratory Ltd., Sellafield CA20 1PG, UK

4   Department of Earth and Environmental Sciences, University of Pavia, Pavia, Italy

and Gorelick 1996; de Marsily et al. 2005; Lee et al. 2007; Ronayne et al. 2010; Savoy et al. 2017; Tahmasebi 2018), including those on transport modelling in groundwater systems (e.g. Zappa et al. 2006; dell'Arciprete et al. 2012; He et al. 2013; Bianchi and Zheng 2016; Montero et al. 2021), hydrocarbon reservoir characterisation (e.g. Sweet et al. 1996; El-Deek et al. 2017; Gray et al. 2022), and geothermal energy exploitation (e.g. Crooijmans et al. 2016; Liu et al. 2019).

In hydrogeological modelling, representing the spatial distribution of hydrogeological units and forecasting fluid flow and solute transport are activities that require good understanding of the geological media being modelled and its inherent heterogeneities (Koltermann and Gorelick 1996; Li et al. 2001; Savoy et al. 2017). Thus, it is paramount that simulations employ appropriate representations of the spatial distribution of hydrogeological units and their hydraulic properties, constrained by data obtained from relevant subsurface or analogue datasets. Model inputs need to be

detailed enough to account for the heterogeneities of interest, but general enough to allow convergence to a solution of flow and transport equations. Particular challenges arise when heterogeneous media exhibit abrupt changes in petrophysical and hydraulic properties, when fractures and faults are present, and/or when processes related to the lithological composition of the aquifer occur, such as sorption and desorption, or reactions between the aquifer minerals and groundwater (Ronayne et al. 2010; Goltz and Huang 2017; Ewis et al. 2022). Failure to adequately represent heterogeneities can result in inaccurate estimations of flow and transport (Zappa et al. 2006; He et al. 2013; Kawo et al. 2023).

Simulations of the spatial distribution of hydrogeological units and their properties commonly use geostatistical—and more generally stochastic—modelling methods. These include object-based algorithms (e.g. Ronayne et al. 2010; Colombera et al. 2019), and methods based on multiple point geostatistics (MPS; e.g. Montero et al. 2021; Kawo et al. 2023) or transition probabilities (e.g. He et al. 2013; Bianchi and Zheng 2016), among others. Thanks to computational advances, it is becoming increasingly easier, faster and more efficient to run stochastic modelling algorithms. However, the required inputs to apply any of those methods need to be geologically consistent and robust to obtain meaningful results.

Given that the permeability heterogeneity of sedimentary aquifers depends on lithofacies distributions (Klingbeil et al. 1999; Zappa et al. 2006; Kawo et al. 2023), defining lithofacies types or operative facies types and then grouping them into hydrofacies (hydrogeologically relevant lithological units) based on their estimated hydraulic conductivity is the first step in the construction of a groundwater model. To define and characterise those units, available geological data and hydraulic tests are regularly used.

Quaternary sedimentary successions, and particularly those containing glaciogenic deposits, represent a case in which sedimentological patterns are difficult to predict, given the high variability in lithology, geometry and topology of sedimentary units (McMillan et al. 2000; Lee 2018; Kurjanski et al. 2020; Smith et al. 2023). Where only borehole and outcrop data are available, other approaches are needed for site characterisation. Recent studies addressing this problem have followed a sedimentological approach founded on detailed facies and architectural-element analysis of outcropping deposits (Smith et al. 2023). This approach results in a conceptual model of aquifer heterogeneity that accounts for the processes involved in the formation of the deposits. However, an excessive level of detail is not ideal for hydrogeological modelling purposes (e.g. Hill 2006).

Modelling a heterogeneous sedimentary aquifer involves representing (1) the spatial distribution of hydrogeological units and (2) their hydraulic properties. To undertake these tasks, one option is to inform the categorization of the units and their petrophysical characterisation using the particle size distribution (PSD) of sedimentary deposits (Alyamani and Şen 1993; Arya et al. 1999; Odong 2008; Rosas et al. 2014; Bianchi and Zheng 2016; Chandel and Shankar 2022). In applications to classification problems, PSD data are especially useful for grouping samples collected for hydrogeological characterisation (Simon et al. 2021; Nichols et al. 2023), given the role of sedimentological properties as controls on hydraulic characteristics (e.g. Anderson et al. 1999). Hydraulic conductivity estimates using the PSD have been derived from empirical equations since the late 19th century (Hazen 1892; Kozeny 1927; Fair et al. 1933; Carman 1937, 1956; Vuković and Soro 1992; Kasenow 2002; MacDonald et al. 2012; Rosas et al. 2014; Wang et al. 2017; Sun et al. 2024).

The relationships between the grain-size distribution and the porosity have been studied by several authors (e.g. Vuković and Soro 1992; Wu and Wang 2006; Wooster et al. 2008; Frings et al. 2011). However, empirical equations may not be readily applicable to sediments with different characteristics and/or from different settings to those for which the relationships were defined. To partly address this problem, the grain-size parameter that best predicts the porosity of the studied glaciofluvial sequences can be identified comparing measured porosities with grain size data. In turn, aiming to obtain a reasonable number of hydrofacies, a comparison between a lithofacies classification, based on a detailed sedimentological criteria, and a classification obtained from an unsupervised clustering approach using the grain-size distribution can be carried out. Unsupervised clustering has the potential to generate groups from large datasets based on input variables, such as the grain size, without needing to label data in advance (Jain et al. 1999), hence generating data-driven labels and reducing subjectivity in the classification.

In that context, the aim of this work is to develop a workflow for objective hydrofacies definition, applied to a heterogeneous unconsolidated glaciofluvial sequence in Northwest Cumbria (Fig. 1). Specific objectives include:

(i) Field characterisation of lithofacies in coastal and quarry outcrops, including field porosity measurements and sampling for laboratory analysis of PSD.

(ii) Identification of an optimum approach to estimate porosity from grain-size parameters where porosity is not measured in the field, as is generally the case given the challenges in undertaking porosity measurements.

(iii) Application of a clustering algorithm to PSD data to identify the optimum number of hydrofacies, and comparison with field observation-based lithofacies classification.

(iv) Definition of statistical hydraulic properties of the defined hydrofacies using estimates of the hydraulic conductivity based on PSD and porosity, and comparison to other data sources.

Fig. 1 Map of the study area, NW Cumbria, UK. Sampled localities are shown. Grid geodetic reference provided in the World Geodetic System 1984 (WGS84) and Universal Transverse Mercator (UTM) coordinate system. Countries labeled following ISO-alpha3 M49 codes
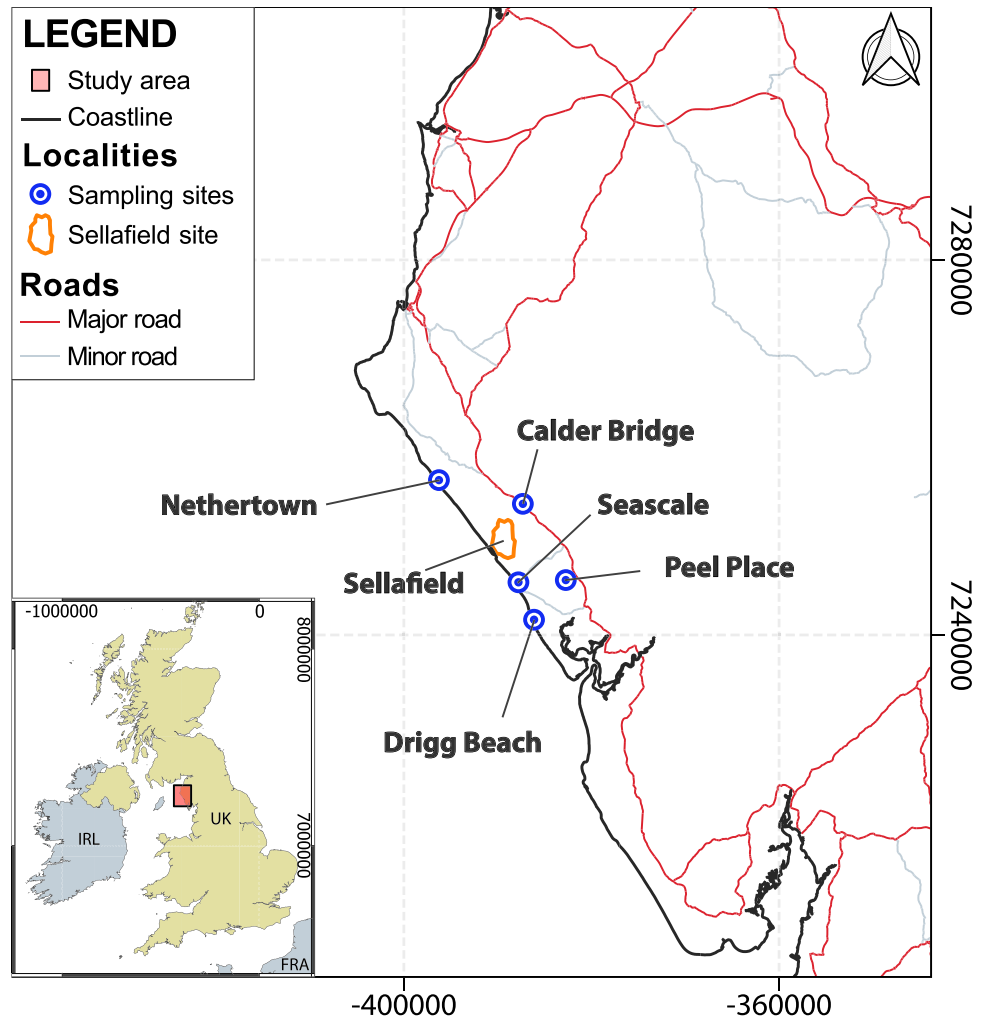
Geological setting

The Quaternary geological history of northwest Cumbria (UK) was characterised by repeated advances and retreats of ice sheets during the Late Devensian (28–13 ka; Merritt and Auton 2000; Smith et al. 2020): main Late Devensian glaciation, Gosforth Oscillation, Main Scottish Readvance and minor readvances periods (Merritt and Auton 2000). This led to the deposition of heterogeneous, glaciofluvial sequences that constitute today's superficial, unconsolidated aquifer in the area. These deposits overlie Permo-Triassic sandstones of the Sherwood Sandstone Group, Ordovician volcaniclastic rocks of the Borrowdale Volcanic Group near the foothills to the east, and, locally, older Paleozoic rocks (Akhurst et al. 1997; McMillan et al. 2000; Merritt and Auton 2000).

Several studies have focused on the geological characterisation of the exposed Quaternary sequences (Akhurst et al. 1997; McMillan et al. 2000, 2011; Merritt and Auton 2000; Coleman et al. 2021; Smith et al. 2023), the past movement of ice sheets (Evans et al. 2005; Livingstone et al. 2012; Chiverrell et al. 2018), and the development of geological

and hydrogeological models (McMillan et al. 2000; Smith et al. 2020, 2023; Coleman et al. 2021). Outwash gravel and sand deposits, of varying thickness, occur along the coast from Nethertown to Drigg Beach (Fig. 1). In the latter area, gravels are a subordinate component, and sands are dominant. Glacial till layers have also been recognised along this coast. Smith et al. (2023) characterised 19 different lithofacies that occur in five different facies associations (FA). The facies associations include sheet-like massive diamicts, gravelly confined flow and unconfined flow deposits, confined flow and lacustrine deposits, sand-dominated unconfined flow deposits and high-energy channelised flow deposits. The FAs map onto 11 types of architectural elements (AE). Glaciotectonic deformation has also been documented in these deposits. Overall, the deposits are heterogeneous in terms of their grain-size and spatial distributions, and their lateral extent is typically limited; as such, traditional approaches of borehole correlation for subsurface characterisation are problematic (McMillan et al. 2000; Smith et al. 2023).

Understanding sedimentological controls on contaminant migration in this aquifer is desirable, in the light of past

leaks and spills of radionuclides into the subsurface of the Sellafield Nuclear Licenced site (Sellafield Ltd. 2023), and to support the safety case of the nearby UK Low Level Waste Repository (LLWR; LLW Repository Ltd. 2011). Sedimentary heterogeneities in this area determine the occurrence of preferential flow, especially via high-permeability gravel layers, but also spatial variability in sorption processes of contaminants such as ^{90}Sr, which interacts with clays and iron oxides (Barker et al. 2024).

Materials and methods

The identification and characterisation of lithofacies were undertaken through the following activities (the general workflow is shown in Fig. S1 of the electronic supplementary material, ESM): (1) field sampling at localities in the study area where fluvioglacial sequences are known to be exposed (Fig. 1); (2) lithofacies classification based on the visual characterisation of the deposits in the field; (3) in situ porosity measurements using the sand replacement method following British Standard BS1377-9 (British Standards Institution 1990); (4) determination of the particle size distribution (PSD) of the samples by combining mechanical sieving (> 2 mm fraction) and an optical grain sizer equipment (< 2 mm fraction); (5) data processing and data analysis, including an evaluation of empirical equations for predicting porosity (e.g. Vuković and Soro 1992; Wu and Wang 2006; Wooster et al. 2008; Frings et al. 2011) and determination of hydrofacies units via cluster analysis using the sample PSD data; and (6) hydraulic conductivity estimation using Kozeny–Carman equation (Kozeny 1927; Carman 1937, 1956; Bear 1972; Riva et al. 2010). A detailed explanation of each step is described in the following subsections.

Fieldwork

Field sampling was carried out at the coastal outcrops of Drigg Beach, Seascale and Nethertown, and at Peel Place Quarry and the bank of the River Calder upstream of Calder Bridge (Fig. 1) following BS17892-4 (British Standards Institution 2016). Coastal outcrop sampling involved the systematic collection of sediment samples in vertical transects: four cliff sections were sampled at Drigg Beach and Seascale, and three at Nethertown (Table 1). Sampling intervals were identified on the basis of visual changes in the sedimentary composition of the sequence. The extent of vertical sampling was limited by outcrop accessibility. The steep nature of the outcrops precluded porosity measurements at these coastal locations. At Peel Place Quarry site, the stable nature of the slopes allowed sampling along both horizontal and vertical directions, and in situ porosity measurements using the sand replacement method. Three

Table 1 Locations of samples described and obtained from the field (Field), and total number of PSD analyses (Lab). Grid references are provided in the Ordnance Survey National Grid reference system (OSGB)

Location	Grid reference (OSGB)	Number of samples	
		Field	Lab
Borehole 11A	NY 02304 04390[a]	1	4
Borehole 9983	NY 02304 04390[a]	2	8
Calder Bridge	NY 04221 06052	4	9
Drigg Beach	SD 04646 98646	37	88
Nethertown	NX 98733 07488	12	47
Peel Place	NY 06576 01310	18	51
Seascale	NY 03981 00402	15	38
Total	-	89	245

[a]Does not correspond to the exact borehole location but to a reference location of Sellafield site

additional samples were obtained from outcrops of the glaciofluvial sequence along the Calder River, upstream of Calder Bridge. Pre-existing grain-size analyses from samples from two boreholes from the Sellafield site (9983 and 11 A), and Calder Bridge area, and of six further samples taken from Drigg Beach, Nethertown and Seascale Beach in March 2021 were included in the analysed dataset.

Porosity measurements were performed by means of the sand replacement method. The process followed BS 1377–9 (British Standards Institution 1990) with calibration sand characteristics 100% passage through a 600 μm sieve and 100% retention on the 63 μm, large cylinder (200 mm diameter). The calibration sand density was estimated in the laboratory before and after the field study, using a ± 0.05 g precision scale, from which an average value of 1.49 g/cm³ was obtained (difference ± 0.01 between min and max values). A scale with a ± 100 g precision was used in the field to carry out the in situ weight measurements; the resulting error was < 1% porosity. A larger source of error arose from overestimation of the volume of the dug hole, resulting from loss of calibration sand between the cover plate and the surface around the hole (due to irregularities in the surface). The average gap was estimated as 10 mm, and error bars of ± 2% porosity were estimated by assuming minimum and maximum gaps of 5 and 15 mm, respectively. Samples were collected at each porosity measurement location to characterise moisture content and PSD.

Facies analysis

A sedimentological lithofacies classification was carried out on the basis of field observations and photos taken from the sampling points, following the classification schemes by Tucker (2003) and Miall (2006) (Table 2). The facies codes

and sedimentological terminology used are consistent with previous work on the same successions (e.g. Smith et al. 2023). The different types of lithofacies were interpreted in terms of formative depositional processes, and their assemblages were classified into facies associations according to their spatial association and architectural features exhibited in outcrop. The facies associations are interpreted in terms of local depositional environments. Individual samples were linked to their lithofacies and facies association, enabling correlation of sedimentological characteristics and hydraulic properties.

Particle size distribution (PSD) analyses

To obtain the PSD (see detailed workflow in Fig. S2 of the ESM), sieving tests were carried out in accordance with BS 1377–2 (British Standards Institution 2022a). Samples were first oven-dried at 105 °C for 48 h. Coarse-grained samples (coarse gravels or total sample weight > 5 kg) were manually quartered before drying. After drying, the coarse fraction (> 2 mm) of each sample was separated from the fine fraction (< 2 mm) by mechanical sieving, and PSD analyses were performed separately on each fraction. The remaining fine fraction (< 2 mm) was then quartered using a riffle splitter. At least two separate fine-grained subsamples were obtained for each field sample. Given the limited size of the manual riffle, when the weight of the under 2 mm fraction was > 500 g, more pairs of subsamples were obtained. When the finer fraction was evidently poorly sorted and its weight was higher than ~500 g, an additional mechanical sieving step was carried out to obtain the < 1 mm portion. The resultant fine-fraction subsamples were analysed in the CAMSIZER® X2 optical sizer (Technology review 2013; Arora et al. 2022; Microtrac MRB 2024), which is based on the dynamic image analysis principle as presented in BS13322-2 (British Standards Institution 2022b). This equipment allows fast PSD analysis from a teaspoon-sized sample (between 20 and 50 g), for grain sizes < 2 mm. For most samples, the X-FALL module (dry dispersion by gravity - air; 10 μm to 3 mm) was used. For finest-grained samples, particularly those that agglomerated after drying, the X-JET module (dry dispersion by compressed air; 0.8 μm to 1.5 mm) was used as it deagglomerates the samples via a Venturi nozzle. Details on how each module work are explained in the manufacturer's webpage (Technology review 2013; Microtrac MRB 2024). For each analysed sample, the CAMSIZER gives two PSD outputs: one considering the minimum diameter of each grain, and another one using the grain area. The former was used here as it is more comparable to mechanical sieving results.

Data processing and analysis

Compiled PSD data was used to obtain the percentage of fine (< 0.0625 mm), sand (0.0625–2 mm) and gravel (> 2 mm) of each sample. These thresholds are widely recognised as key thresholds (Miall 2006; British Standards Institution 2018) and have been used for the sedimentological lithofacies classification of this work. Given that the same sample was analysed more than once, average values were obtained. The three grain-size fractions were plotted on a 2D ternary plot by reducing them to two independent variables using Eqs. [1] and [2],

$$X = \frac{1}{2} \bullet \frac{2 \bullet \%\text{Gravel} + \%\text{Fine}}{\%\text{Gravel} + \%\text{Sand} + \%\text{Fine}} \qquad (1)$$

$$Y = \frac{\sqrt{3}}{2} \bullet \frac{\%\text{Fine}}{\%\text{Gravel} + \%\text{Sand} + \%\text{Fine}} \qquad (2)$$

The resultant X and Y variables were used to classify the samples using the K-means algorithm. K-means is an unsupervised iteration-based clustering algorithm that stands out for its simplicity and ability to classify data points into clusters based on similarity (Jain et al. 1999; Pedregosa et al. 2011). It has been used, for example, to obtain lithofacies groups using seismic data (Troccoli et al. 2022) or geophysical logging (Newell et al. 2021) as input variables. It aims to minimise within-cluster distances and maximise between-cluster distances. It is initialised by setting the number of clusters then generating randomly positioned centroids to those clusters. Then, samples are assigned to their closest cluster (Euclidean distance), and centroids are recalculated. The algorithm iterates by reassigning samples their closest cluster considering the new centroids obtained. Finally, when no more samples are reassigned or centroids do not change anymore, the algorithm stops. To determine the optimum number of clusters, four methods were used. The elbow method plots the resultant sum squared error (SSE) against the number of clusters (the

Table 2 Lithofacies codes for siliciclastic sediments (based on Miall 2006 and Tucker 2003)

Parameter	Code
Lithology	G: gravel S: sand F: fines D: diamicton
Depositional structure	m: massive p: planar cross-bedded t: trough cross-bedded r: ripple cross-laminated h: horizontal bedding l: laminated
Fabric	c: clast supported m: matrix supported

SSE is zero where the number of clusters equals the number of data points). The optimum number of clusters is that where the rate of reduction in SSE with increasing cluster number reduces (*elbow* in the plot). The silhouette method analyses the distance between clusters and number of samples contained in each (Rousseeuw 1987; Pedregosa et al. 2011). Plots are generated for different values of *n* (number of clusters) and a silhouette score is calculated. The optimum number of clusters is that when the average silhouette score is the highest, i.e. the clusters maximum silhouette coefficients are all above the average score, and cluster sizes are similar. The Davies–Bouldin index measures the similarity between clusters as the ratio between average within-cluster and between-cluster distances (Davies and Bouldin 1979; Pedregosa et al. 2011). A lower number means clusters are more separate from one another and less dispersed internally, which indicates the number of clusters selected is better. Finally, the Calinski–Harabasz criteria is defined as the ratio between the overall between-cluster variance and the overall within-cluster variance (Calinski and Harabasz 1974; Pedregosa et al. 2011). The result is higher when clusters have a lower within-cluster variance, and individual clusters are more separate between each other. The value will usually increase as the number of clusters increases, so algorithms usually look for the first local maximum to decide for an optimum number of clusters. Here, the Scikit-Learn Python package (Pedregosa et al. 2011) was used to run the K-means algorithm and obtain the elbow, silhouette, Davies–Bouldin and Calinski–Harabasz scores, with the optimum cluster number being found manually by comparison of efficacy of these four approaches.

Following the analysis of the ternary plot and the cluster analysis results, the grain-size variables required for subsequent porosity and hydraulic conductivity estimates were extracted and included in a database; these variables include the d_{10}, d_{25}, d_{50} and d_{75} values; the uniformity coefficient ($U = d_{60}/d_{10}$) and the geometric standard deviation (GSD) of Φ: ($\Phi = -\log_2(\text{grain size})$).

Porosity and hydraulic conductivity estimation

Relationships were evaluated between porosity and d_{50} (following Wu and Wang 2006), the uniformity coefficient (*U*; following Vuković and Soro (1992)) and the geometric standard deviation of Φ (σ_Φ or GSD(*Φ*), following Frings et al. 2011; based on the equation by Wooster et al. 2008). A regression analysis was carried out and a site-specific calibration was found for each variable from which the approach showing the narrowest confidence intervals (least uncertainty) was selected. Porosity values were then estimated for the samples from outcrops where no direct porosity measurements were available, using that approach. These statistical porosity

characteristics for each cluster were then defined using measured porosities (where available) and estimated values.

Hydraulic conductivity was then estimated from porosity and PSD data using empirical approaches. One of the most widely used is the Kozeny–Carman equation (K-C; Eq. 3), which relies on the porosity and on the representative grain size, and which has proven as a good fit for sediments similar to the ones being studied (Vuković and Soro 1992; Odong 2008; Chandel and Shankar 2022):

$$K = 8.3 \cdot 10^{-3} \cdot \frac{\mathbf{g}}{\nu} \cdot \frac{\theta^3}{(1-\theta)^2} \cdot d_{\mathrm{e}}^2 \tag{3}$$

where **g** is the gravitational acceleration (9.81 m/s^2), ν the viscosity coefficient of water (1.002 m^2/s at 20 °C), θ the fractional porosity and d_{e} is the effective grain diameter (m). For this application, the characteristic grain size has been assumed to be d_{10} (following Vuković and Soro 1992; Kasenow 2002; Riva et al. 2010; Chandel and Shankar 2022).

Results

Samples and geological characterisation

A total of 81 samples were collected in the field and 12 different lithofacies have been identified (Table 3). A summary of the number of samples, locations and lithofacies is presented in Table 4.

Sands and gravels are overall dominant (28 and 26 samples classified as such, respectively), while fines are slightly less abundant, having been recognised mostly at Drigg Beach (Table 4). Diamicton-type deposits (Dmm; Fig. 2b and d) have been observed at Drigg Beach, Seascale and Nethertown areas.

Gravel facies are pebbly, poorly sorted, clast supported and polymictic (Gcm, Gh; Fig. 2b, e and g). Clasts sizes are mostly in the 0.2 cm to 20 cm range, locally reaching > 1 m in diameter (e.g. Nethertown) and are dominantly red sandstone (presumably Sherwood Sandstone) and igneous (volcanic and occasionally granitic). Sand facies are mostly very fine to medium grained, well-sorted, occasionally containing sparse gravel-sized clasts (Sm, Sp, Sh; Fig. 2a and c). Fine-grained facies are represented by silts, clayey silts and sandy-silts, occasionally presenting clasts of up to 0.5 cm diameter (Fm, Fl; Fig. 2a, e, f and g).

Most of the studied deposits demonstrate a massive structure (Gcm, Sm, Fm; Fig. 2a and f), although horizontal stratification and cross-stratification are observed (Gh, Gp, Sh, Sp, Fl; Fig. 2b, c and e). Some gravel deposits show imbrication, with both a(t)b(i) (Fig. 2b) and a(p) fabrics (Fig. 2e).

Alternation of finer and coarser grained sized beds are observed at a scale of decimetres to metres (e.g. Fig. 2g);

Table 3 Sedimentological lithofacies characteristics

Facies code	Description	Interpretation (cf. Smith et al. 2023)
Diamicton Dmm (Fig. 2b, d)	Clayey-silt to gravelly silt diamicton. Polymictic, with subrounded to subangular clasts. Massive depositional structure. Brown colour. Very high stiffness	Subglacial till deposit (Evans et al. 2006; Smith et al. 2023)
Laminated silts Fl (Fig. 2e)	Sandy silt. Orange-brown, horizontally laminated. Bed thickness up to 0.2 m. Minor presence (<1%) of subrounded clasts of up to 0.5 cm	Very low or inexistent flow, slack water deposit. Deposition from suspension (Bennett et al. 2002; Smith et al. 2023)
Massive silts Fm (Fig. 2a, f)	Silts and sandy silts. Occasional clasts (<4%) up to 0.5 cm in size. Massive structure. Grey-brown to red-brown colour	Loess, overbank or abandoned channel fill deposits (Bennett et al. 2002; Miall 2006; Smith et al. 2023)
Massive sand Sm (Fig. 2a, b, e)	Very fine to coarse sands, well sorted. Massive, often normal graded. Occasionally with abrupt contacts between fine- and coarse-grained sands. Light brown to orange-brown colour. Occurs in units with thickness from 0.2 to 5 m	Deposition of sands out of suspension by energetic turbulent flows infilling channels (Smith et al. 2023)
Horizontally bedded sand Sh (Fig. 2c)	Fine to medium grained sand, well-sorted. Centimetric lamination. It is locally folded. It occurs in beds that have thickness from 0.5 to 2 m. Light brown to orange-brown colour	Unconfined flow events. Streamflow deposit under upper-flow regime conditions (Smith et al. 2023)
Planar cross-bedded sand Sp (Fig. 2c)	Very fine to medium sands. Planar cross-stratified. It shows intercalations of finer grained red-brown silty-sands. Mud drapes along the cross-stratified lamination. It may display centimetric scale folding. It occurs in units with thickness from 0.5 to 1 m. Orange-brown to light brown colour	Migration of straight-crested mesoforms, such as alluvial dunes, or linguoid and transverse bars (Miall 2006; Smith et al. 2023)
Massive clast-supported gravel Gcm	Sandy fine to coarse gravels. Poorly sorted, polymictic. Clasts size vary from 0.2 to 20 cm. Locally, clasts of up to 2 m size (Nethertown)	High energy mass flow and gravelly barforms (Miall 2006; Smith et al. 2023)
Planar cross-bedded gravel Gp	Sandy medium to coarse gravels with clasts that are up to 10 cm in diameter. Clast supported, poorly sorted, polymictic. Planar cross-bedded	Migration of gravelly barforms (Miall 2006; Smith et al. 2023)
Horizontally bedded gravel Gh (Fig. 2b, e)	Gravel and sandy gravels, with gravel-clast sizes varying mostly from 0.2 to 10 cm, but locally up to 50 cm. Poorly sorted, with polymictic subrounded clasts. Horizontally graded, locally imbricated	Fluvial sheet gravel or channel floor sediments in a fluvial environment (Miall 2006; Smith et al. 2023)

fining- and coarsening-upwards trends within beds are observed, although the former are dominant (Fig. 2). At Peel Place, local centimetric to metric scale faulting and folding are observed in sand-prone deposits (Fig. 2c). Larger-scale folds, of the order of metres to decametres, are evident in all coastal areas. Smith et al. (2023) attributed deformation features as such to be of glaciotectonic origin.

At Drigg Beach, the studied vertical transects conform to two main fining-upwards successions. Laterally, the succession varies in terms of the abundance and relative position of the different facies (Figs. 3a and b, and 4). The main facies types recognised correspond to massive clast-supported gravels (Gcm), horizontal plane-bedded and massive sands (Sh, Sm) and massive silts (Fm). A laminated sandy silt (Fl) is observed at the top and bottom of the exposed sequence, and an up to 1 m thick diamicton layer (Dmm) is observed

along the sequence exposed in the south (Figs. 3a and 4). In the north exposure, the bottom fining-upwards succession is represented by an intercalation of silts and sands. In the base of the facies sequence, a laminated silt is overlain by a thin (<50 cm) gravel bed (Figs. 3b and 4), which then shows a continuous, non-erosional, normal-graded sequence (Fig. 3c). The observations are consistent with the facies sequence being deposited in a glaciofluvial context, suggesting an origin related to stream-flow processes. The massive fine-grained deposits (Fm) with a predominance of silt and fine sand suggest the development of either loess, overbank or abandoned channel fill deposits. Smith et al. (2023) have characterised architectural elements in this area as sand-prone unconfined flow deposits (facies Sh, Sm), intercalated with unconfined flow gravels. Their fine-grained sequences

Table 4 Number of identified sedimentological facies per location

Location	Identified lithofacies									Total
	Dm	F		G			S			
	m	l	m	h	m	p	h	m	p	
Calder Bridge	-	-	1	-	2	-	-	-	-	3
Nethertown	1	-	2	-	4	-	1	-	1	9
Peel Place	-	-	1	-	6	-	4	4	3	18
Seascale	1	1	3	3	3	-	-	5	-	16
Drigg Beach	2	2	13	2	4	2	4	6	-	35
TOTAL	4	3	20	5	19	2	9	15	4	81

in the area were attributed to subaqueous deposition in a glaciolacustrine environment, which could be consistent with the laminated silts and sandy silts observed. Diamictons observed by Smith et al. (2023) were interpreted as being deposited directly from glacier ice.

In Peel Place Quarry, the succession consists of coarse-grained, poorly channelized sand and gravelly deposits (Figs. 3d and 4). Facies recognised in this area include massive and horizontally laminated gravels (Gcm, Gh), and massive, horizontally bedded and planar cross-bedded sands (Sm, Sh, Sp). Smith et al. (2023) have characterised these facies as unconfined flow elements, interbedded with channel bar confined flow deposits, and solitary sand/silt fills. In general, they interpreted this facies sequence as gravel and sand dominated outwash deposits that might have accumulated under confined to unconfined flow conditions. At Seascale, the succession is dominated by gravels and sands, forming a facies association similar to that observed at Peel Place Quarry. However, massive and layered silts are also found along the facies sequence at Seascale. Layered silts are observed at the bottom of one of the vertical sections on this site; a massive silt overlies a clast-supported coarse gravel; and a massive fine-grained bed tops the succession. As there are no further exposures of the sediments on this site, it is unclear whether the overlying and underlying facies associations are similar to the one observed at Drigg Beach. Smith et al. (2023) interpret this sequence as being stratigraphically below the sequence observed in Peel Place Quarry.

Particle size distribution

The PSDs of 89 samples were obtained from a total of 245 analyses (Table 1). The frequency distributions of the summary statistics (d_{10}, d_{50}, GSD(Φ) and U) are shown in Fig. 5a to d, after log-transformation (base 10) for ease of visualisation. Distributions are all positively skewed; some grade of bimodality is observed in the U, d_{10} and d_{50} distributions, while GSD(Φ) shows a wider and more uniform dispersion of values. The uniformity coefficient (U) shows a high positive skewness with a median value of 0.90 for log U (Fig. 5a; $U \sim 8$), consistent with samples being predominantly poorly sorted. D_{10} is mostly distributed in the 0.01–0.1 mm range (log d_{10} between −2.0 and −1.0; Fig. 5b), showing a lower positive skewness. The mean and median values of log d_{10} are similar (−1.473 and −1.467, respectively) and the distribution can be approximated by a normal distribution, although a tendency to being bimodal is observed with a cluster of log d_{10} values between −2.5 and −2.0. In contrast, d_{50} shows a clear bimodal distribution (Fig. 5c), which also suggests that the uniformity coefficient distribution is more dependent on d_{50} than d_{10}. The high skewness and kurtosis values are consistent with a wider distribution of this parameter. GSD(Φ) is also more evenly distributed (Fig. 5d), although its distribution is much more uniform, concentrated between 1.2 and 2.4. A major peak is observed at GSD(Φ) ~ 2.0.

Analyses

Cluster analysis

The resultant proportions of fines, sand and gravel are presented in the ternary plot of Fig. 6a. Samples with limited gravel fractions (< 10%) collectively cover the entire spectrum of sand- to fine-dominated deposits. Samples with a higher gravel content form a separate group, visually distinguishable in the plot, in which the sand content is < 60% and the content of fines is < 20%. Only a small number of samples are seen to plot in intermediate ranges between those two major groups.

The four approaches described in the previous section were applied and their results manually compared to determine to optimum number of clusters. The elbow method
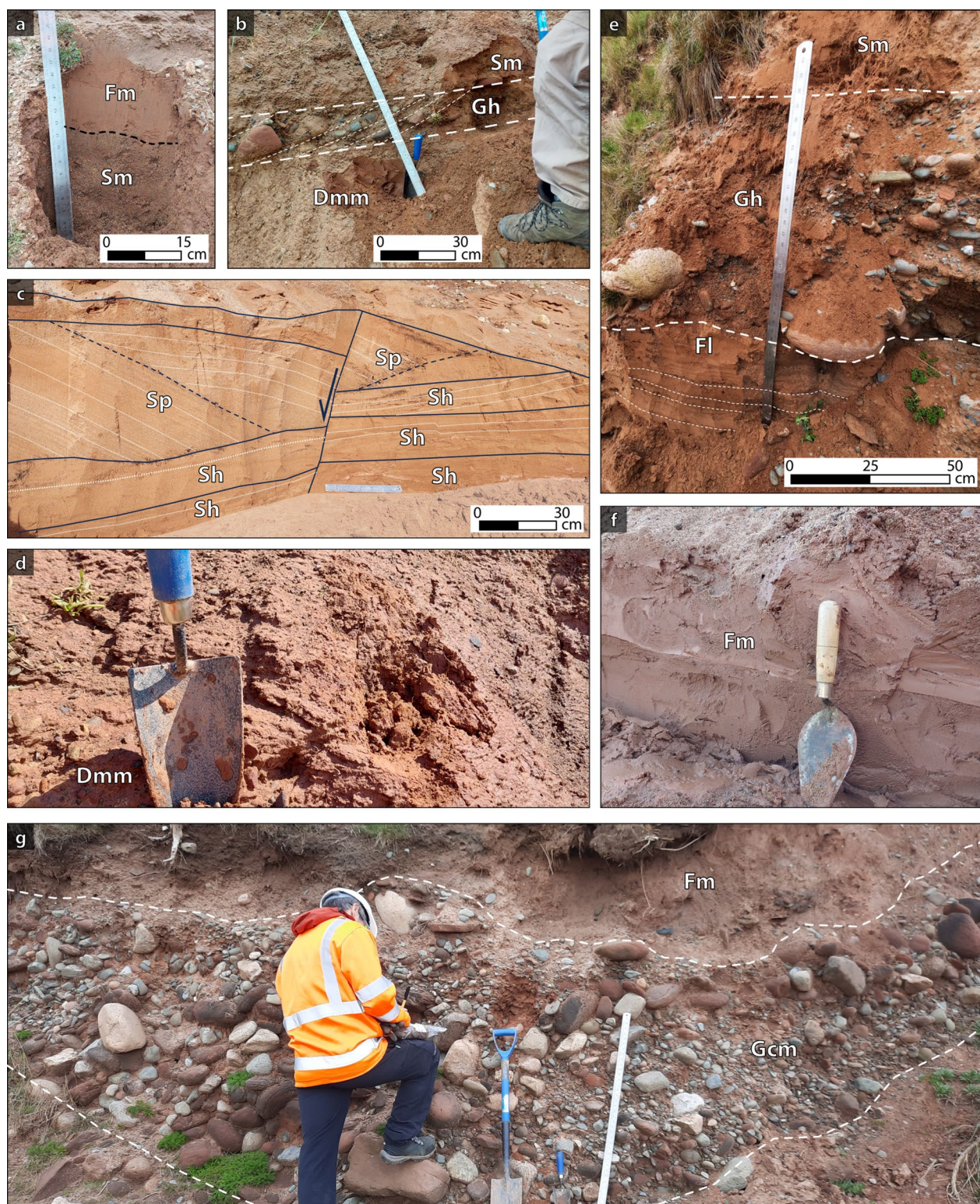
Fig. 2 Outcrop photographs of facies characterised on the field. (**a**) Drigg beach, silt layer (Fm) overlying a massive sand (Sm). (**b**) Drigg beach, diamicton (Dmm) underlying a gravel layer with imbricated clasts (Gh), and a massive sand on top (Sm). (**c**) Peel Place quarry, horizontal and planar cross bedded sands (Sh and Sp), and a normal fault interrupting the continuity of the deposits. (**d**) Diamicton (Dmm). (**e**) Seascale, layered silt (Fl) overlain by an ~80 cm, fining-upwards, horizontally bedded gravel (Gh). A massive sand overlies the whole sequence. (**f**) Peel Place quarry, clayey silt (Fm). (**g**) Seascale, massive clast supported gravel (Gcm), folded, overlain by a massive silt (Fm). Photograph accreditation: Felipe Gallardo, Jared West

Fig. 3 Facies associations observed in Drigg Beach (**a**, **b** and **c**) and Peel Place Quarry (**d**). Outcrops (**a**) and (**b**) are separated 100 m one from the other (south and north, respectively; Fig. 4). Highlighted rectangle in (**b**) is the area zoomed in (**c**). Sampling points are represented by yellow circles. Photograph accreditation: Felipe Gallardo
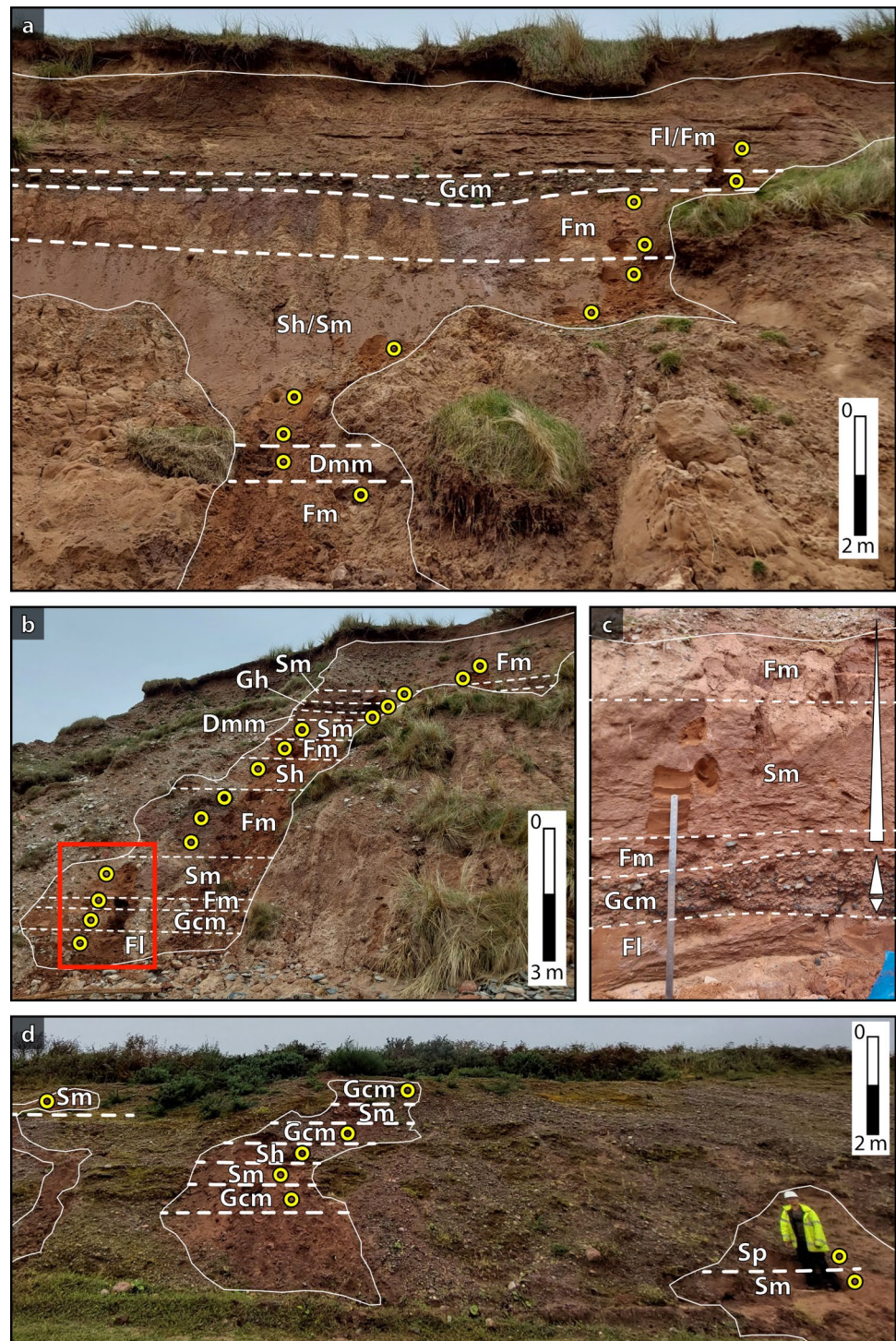
suggested that three clusters are optimum (Fig. 6b). An optimum of three clusters is also supported by the silhouette and Davies–Bouldin maximum and minimum values, respectively (Fig. 7a and b). The Calinski–Harabasz score suggests five clusters as an optimum number, although three or four would also be acceptable. Given that three out of four methods strongly suggest three being the optimum number of clusters, three clusters were defined as follows: fine-dominated (32 samples; cluster 1), sand-dominated (31 samples; cluster 2) and gravel-dominated (26 samples; cluster 3) clusters (Fig. 6a). A summary of key statistics of each cluster and sample is included in the Tables S1 and S2 of the ESM).

Fig. 4 Schematic stratigraphic columns representing the deposit distribution in Drigg Beach and Peel Place Quarry
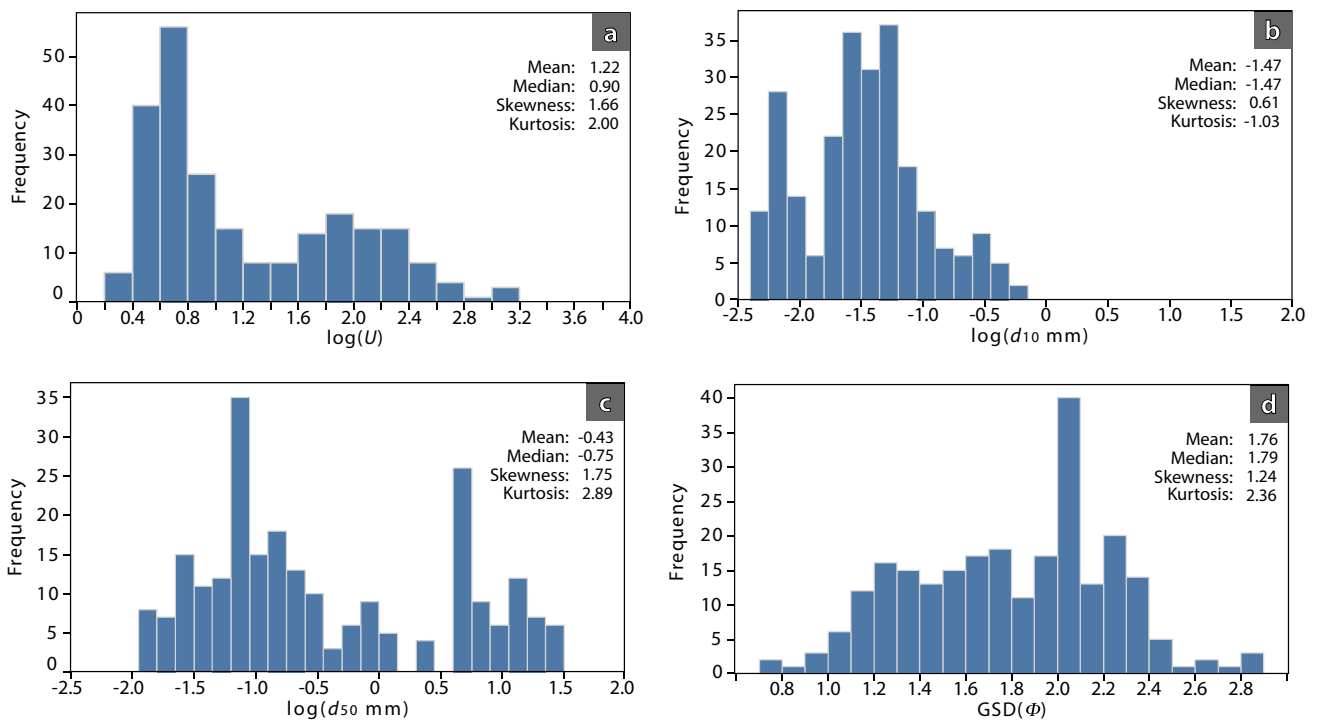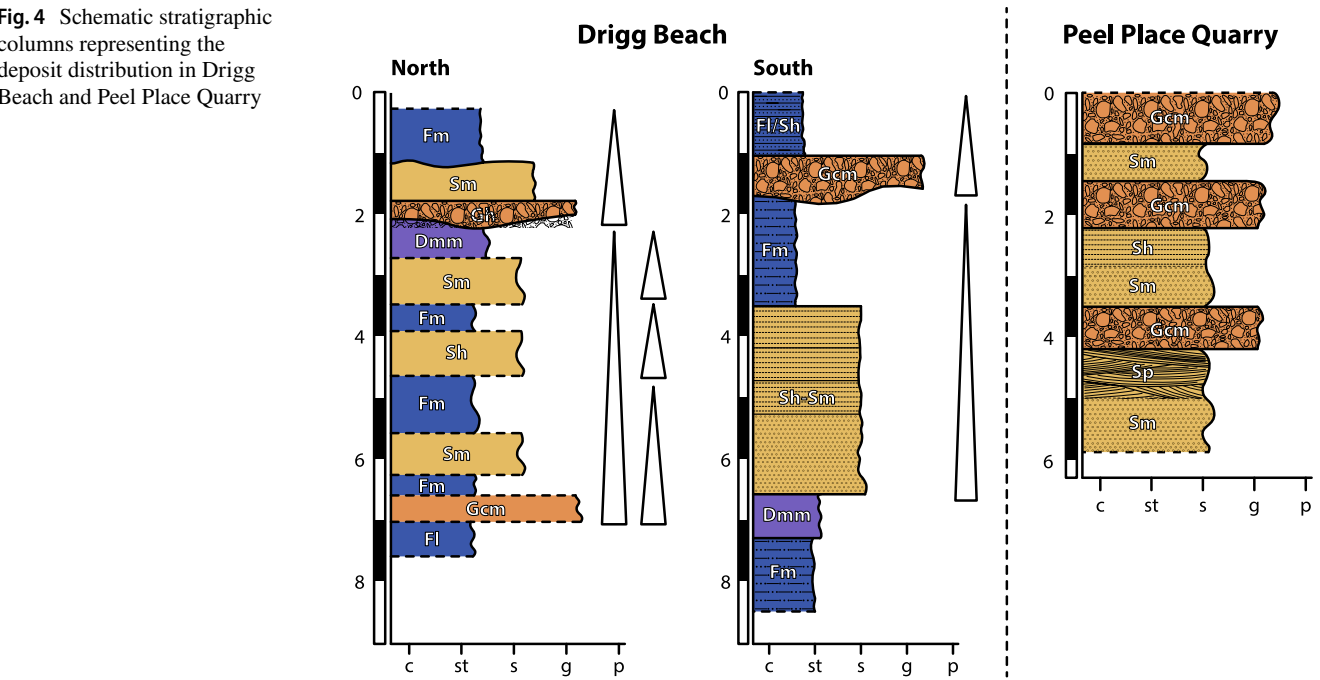
Fig. 5 Frequency histograms: (**a**) log U (uniformity coefficient), (**b**) log d_{10}, (**c**) log d_{50}, (**d**) GSD(Φ), for the 245 analysed samples (from 89 sample locations)

Fine- and sand-dominated clusters are characterised by a low content of gravel-sized grains: 3% and 2% on average, respectively. These clusters have the lowest d_{10} (Fig. 8a) and d_{50} values (Fig. 8b), with a group of values exceeding 1.5 times the interquartile range (IQR) observed in the fine-dominated (d_{10}), and in the sand-dominated cluster (d_{50}). They

also have the lowest uniformity coefficients, with an average of 5.6 in cluster 2 and 6.3 in cluster 1 (Fig. 8c). The gravel-dominated cluster comprises the samples that were visually identified as a separate group (Fig. 6a). Fine-sized grains represent, on average, 10% of the samples belonging to the gravel-dominated cluster. In turn, the uniformity coefficient
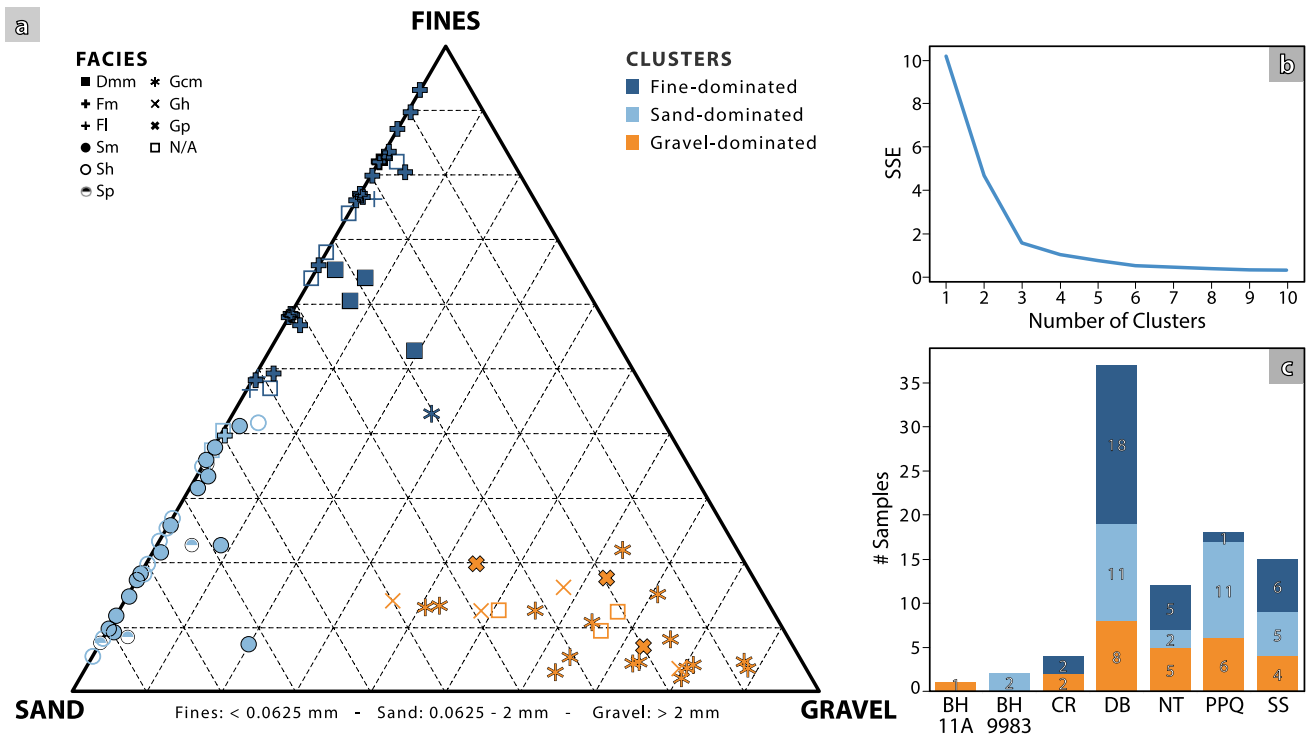
Fig. 6 (**a**) Ternary plot showing the percentage of fines (< 0.0625 mm), sand (0.0625–2 mm) and gravel (> 2 mm) for each sampled location. Field lithofacies classification has been included as symbols, and colours represent the cluster in which samples have been classified. (**b**) Elbow method graph. (**c**) Number of samples per location (*BH* Borehole; *CR* Calder river; *DB* Drigg Beach; *NT* Nethertown; *PPQ* Peel Place Quarry; *SS* Seascale)

of this group is considerably higher than the other two, with an average of 175 and extreme values above 1000 (Fig. 8c). This is consistent with these samples being extremely poorly sorted. The GSD*(Φ)* shows a distribution more similar among the three clusters (Fig. 8d), although the sand-dominated cluster has lower values (average 1.4), and the gravel-dominated cluster has the highest ones (average 2.1).

Porosity measurements and estimates

A total of 14 porosity measurements were obtained from deposits found at Peel Place Quarry site. These represent total porosities, which are equivalent to effective porosities when studying unconsolidated sediments (Woessner and Poeter 2020). Porosity from field measurements show a 35% average porosity, with a minimum value of 20% and a maximum value of 44.5%. Porosity measurements were plotted against the uniformity coefficient (*U*), the logarithm (base 10) of d_{10} (representative grain size used in hydraulic conductivity estimators); the logarithm of d_{50} and GSD*(Φ)*. The regression analysis shows that log d_{50} has the highest correlation with measured porosity values (Fig. 9), which is supported by log d_{50} having the highest Pearson's correlation coefficient and coefficient of determination R^2, and the lowest mean squared error and sum of squared errors (MSE

and SSE; Table 5). In contrast, log d_{10} and GSD*(Φ)* show lower correlation with porosity. Overall, log d_{50} is identified as the variable that can be most confidently used to estimate the porosity of the remaining samples using equation [4]:

$$n = -0.0773 \cdot \log(d_{50}) + 0.3282 \tag{4}$$

Porosity estimates using this relationship (Fig. 10) show that the fine-dominated cluster has the highest porosity range (average 44%, range 41–48%), consistent with it corresponding to well-sorted fine-grained sediments. The range is slightly lower in the sand-dominated cluster (average 40%, range 35–42%), whereas the gravel-dominated cluster shows the lowest porosities (average 27%, range 22–36%). This is also consistent with the latter being poorly sorted (high *U*).

Hydraulic conductivity estimates

The Kozeny–Carman equation was used to estimate the hydraulic conductivity (*K*) of the 89 samples. Figure 11 shows the distribution of log(*K*) for each cluster. The gravel-dominated cluster shows the highest median and geometric mean values of hydraulic conductivity (2.2 and 2.8 m/d, respectively) and the largest ranges of values (from 0.1 to 62 m/d). The sand-dominated cluster has a lower median
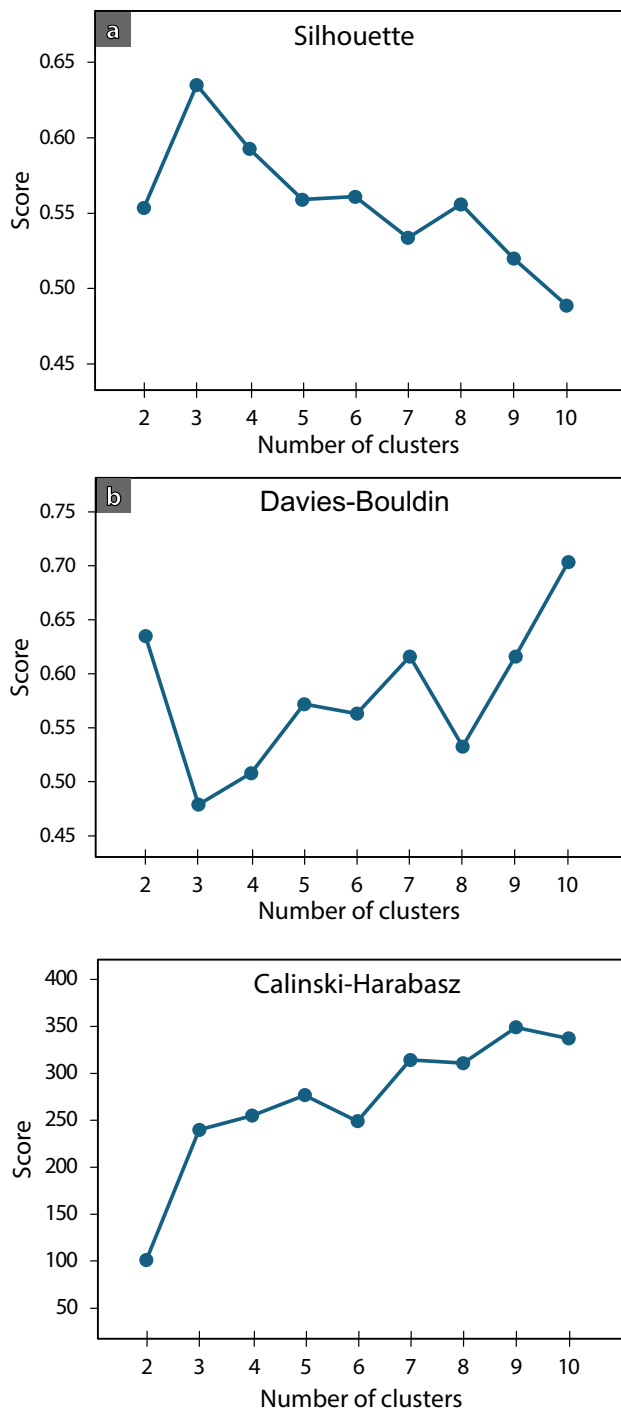
Fig. 7 (**a**) Silhouette, (**b**) Davies–Bouldin and (**c**) Calinski–Harabasz scores for deciding the optimum number of clusters to use in the K-means algorithm

and geometric mean than the gravel-dominated cluster (1.1 and 1.4 m/d, respectively), and a narrower range, from 0.2 to 11 m/d. The fine-dominated cluster shows a hydraulic conductivity range that is one order of magnitude lower, although six outlier samples have a K range overlapping

with the second quartile of cluster 2 (sands). These outlier samples have relatively high d_{10} values (see outliers for fine-dominated cluster in Fig. 8a), which explains their higher K estimates. This is consistent with these samples having been analysed using the X-FALL module instead of the X-JET, as was done with all the other fine-grained samples. Hence, they should be discounted in any characterisation of clusters' K-ranges. Treating these samples as outliers, K values for the fine-grained cluster are in the 10^{-2}–10^{-1} m/d range (0.06–0.2 m/d; geometric mean of 0.1 m/d), while both sand- and gravel-dominated clusters have a range between 10^{-1} and 10^1 m/d, with the latter being closer to the 10^2 m/d limit. All the log(K) distributions show positive skewness.

Discussion

Cluster analysis and lithofacies

Clusters obtained from the PSD are consistent with field characterisation of lithofacies (Figs. 6 and 12). For instance, all the 'G' coded lithofacies fall into the gravel-dominated cluster, except for a single sample, identified as Gmm, which falls within cluster 1 (dark blue asterisk in the middle of the ternary plot; Fig. 6). Similarly, all of the 'S' coded lithofacies fall into cluster 2 (circles; Fig. 6) and all of 'F' coded and 'D' coded lithofacies fall into cluster 1 (dark blue crosses and squares; Fig. 6), with the exception of a single sample—coded F – light blue cross—which is close to the sand-fines midpoint on the ternary plot. This is an important finding because it highlights how the approach followed for lithofacies characterisation is based on objective observations made in the field, but that the statistical analysis based on laboratory PSD alone replicates the key lithotextural classes of fines/diamicton, sands and gravels without expert knowledge. Note that cluster analysis was unable to distinguish 'D' – diamicton, from 'F' – fines coded samples, as these plot relatively close together on the ternary plot. However, the d_{10} and porosity values for 'F' and 'D' samples are relatively similar, as are their predicted hydraulic conductivities (Fig. 13). Hence, while the samples are geologically distinct, it is appropriate to treat them as a single hydrofacies unit. In this sense, the unsupervised cluster analysis proves itself useful and efficient to classify lithofacies using the PSD and focusing on hydrogeologically important details, without creating an excessive number of categories.

Additional measurements, such as geochemical analyses (Simon et al. 2021; Nichols et al. 2023) or geophysical logging (Al-Mudhafar 2017; Kumar et al. 2022) could also be used for scopes of cluster analysis together with the PSD. Such data can help differentiate deposits. However, given the difficulty in obtaining these data from subsurface studies, using available grain size data in the manner explained
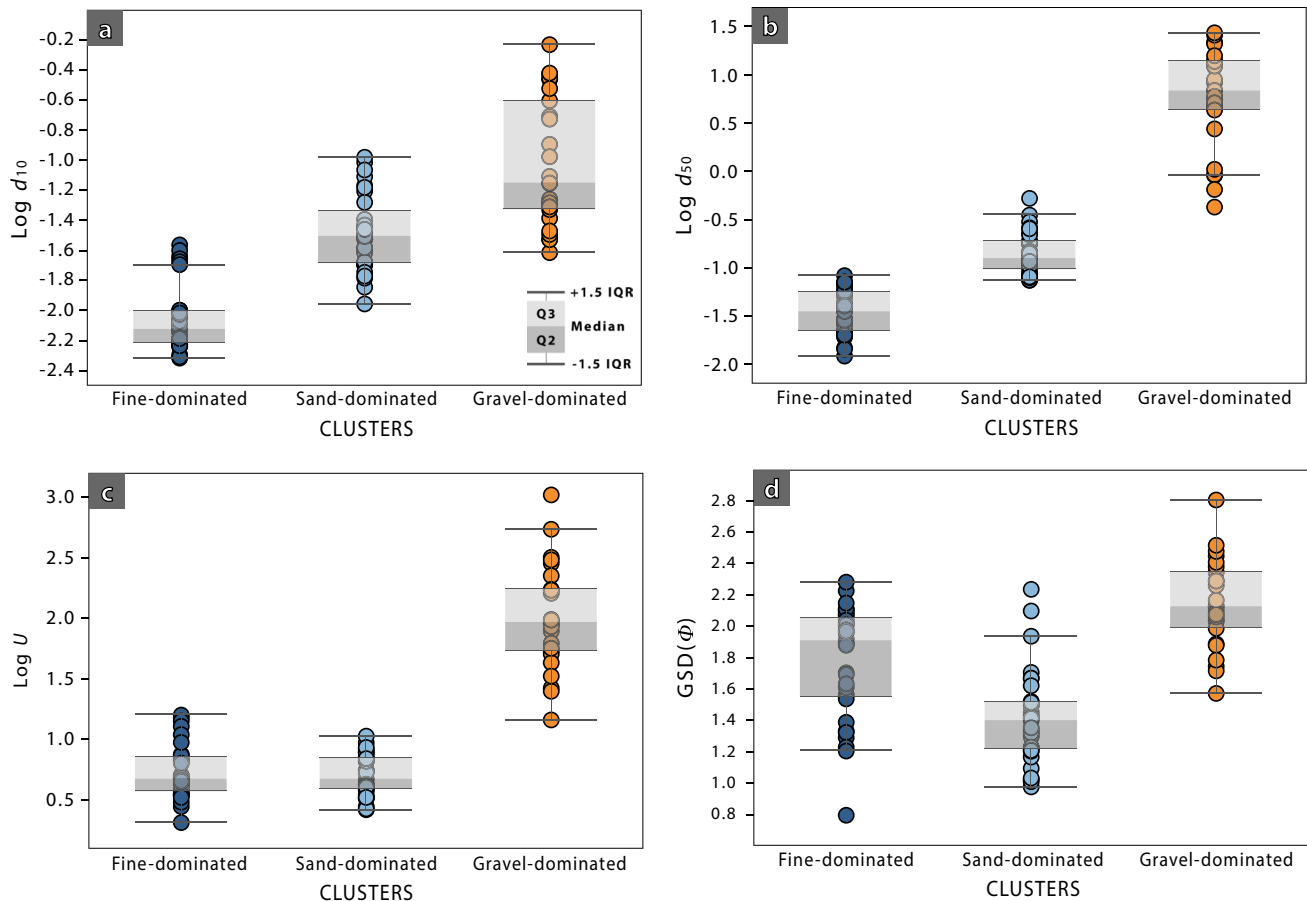
Fig. 8 Distribution of (**a**) d_{10}, (**b**) d_{50}, (**c**) log U and (**d**) GSD(Φ) for each cluster

in this work remains a more practical and broadly applicable approach to the characterisation of hydrogeological units.

The number of clusters used in the algorithm must align with the aim of the project rather than solely adhering to a specific criterion. For instance, the Calinski–Harabasz criterion (Fig. 7a; Calinski and Harabasz 1974) suggests an optimum of five clusters from the dataset, instead of three as suggested by the elbow method, silhouette analysis and Davies–Bouldin score. However, increasing the number of clusters reduces the correspondence with lithofacies, and does not provide additional hydrogeological or sedimentological distinction (the bottom-level lithofacies classes are not visually clustered, Fig. 6a). Notably, including log d_{50} as an additional third variable in the cluster analysis makes the number of clusters recommended on the basis of the Calinski–Harabasz criterion decrease to three, aligning with the number selected.

Porosity and hydraulic conductivity estimates

To determine the most accurate porosity estimation equation, a pairwise bivariate analysis was followed using measured porosities and separate parameters (log d_{10}, log d_{50}, log U and GSD(Φ)). The best-fit equation was obtained using d_{50} (Eq. 4). As expected, a multiple linear regression (MLR) results in an even better fit than a bivariate one, which is achieved using log U, log d_{50} and GSD(Φ) together (equation [5]; $R^2 = 0.899$, MSE $= 0.0006$, SSE $= 0.008$):

$$n = 0.53 - 0.04 \cdot \left(\log(U) + \log d_{50} + 2 \cdot \mathrm{GSD}(\Phi) \right) \qquad (5)$$

However, the improvement in predictive power over the use of log d_{50} alone is marginal. When applying the MLR equation to the samples where porosity was not directly measured, some porosity estimates were significantly lower than any of the measured values. This happened in samples with high uniformity coefficient and d_{50} (exceeding 80 and 4 mm, respectively). Furthermore, for a few samples predicted porosities exceeded 50%, which is physically

Fig. 9 Scatter plots showing relationships between measured porosity (n) and particle size distribution parameters: (**a**) log U; (**b**) log d_{10}; (**c**) log d_{50}; (**d**) GSD(Φ). *Blue lines* indicate linear regressions (equation shown in each graph), whereas the *light blue fields* indicate the 95% confidence intervals
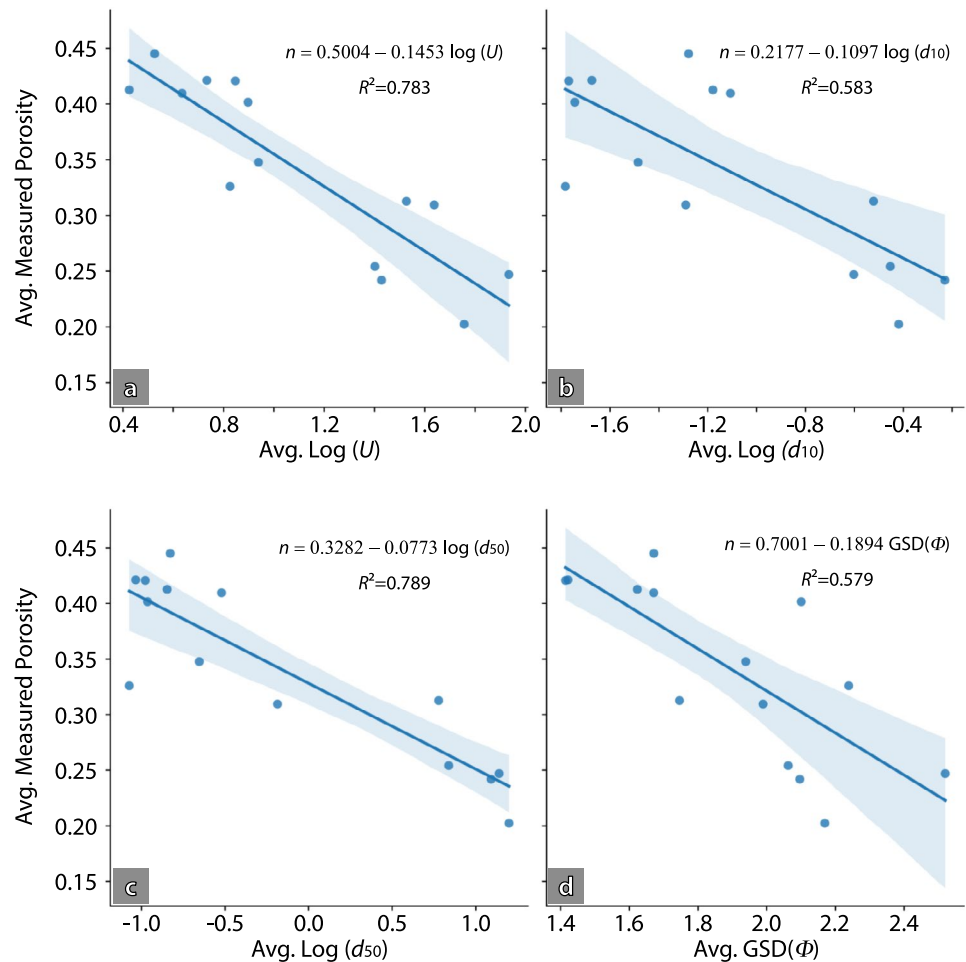

Table 5 Correlation coefficients between measured porosity and log U, log d_{10}, log d_{50} and GSD(Φ). σ_Φ: geometric standard deviation of Φ

Statistic	log U	log d_{10}	log d_{50}	GSD(Φ)
Pearsons	−0.885	−0.764	−0.888	−0.758
Spearmans	−0.859	−0.644	−0.798	−0.767
Best fit equation	$n = -0.1453\log(U) + 0.5004$	$n = -0.1097\log(d_{10}) + 0.2177$	$n = -0.0773\log(d_{50}) + 0.3282$	$n = 0.975e^{-0.562\sigma_\Phi}$
SSE	0.0183	0.0352	0.0178	0.0356
MSE	0.0015	0.0025	0.0014	0.0025
R^2	0.783	0.583	0.789	0.579

unrealistic for these sediment types. Given these considerations, the relationship between porosity and log d_{50} (Fig. 9c) has been used for estimating the porosity where it was not measured.

Porosity estimates obtained using empirical equations proposed in earlier studies (Vuković and Soro 1992; Wu and Wang 2006; Wooster et al. 2008; Frings et al. 2011) reflect known relationships between porosity and the considered PSD variables (d_{10}, d_{50}, GSD(Φ) or U). However, the

application of those equations did not accurately match the measured values in this study. This highlights how empirical best-fit equations derived from other areas may not be universally applicable.

Using empirical equations to estimate the permeability of hydrogeological units is commonly done in practice when data are limited. McMillan et al. (2000) used the equation by Fair and Hatch (Fair et al. 1933) to estimate K values on glaciofluvial and glaciolacustrine sediments from the Sellafield area, whereas Bianchi and Zheng
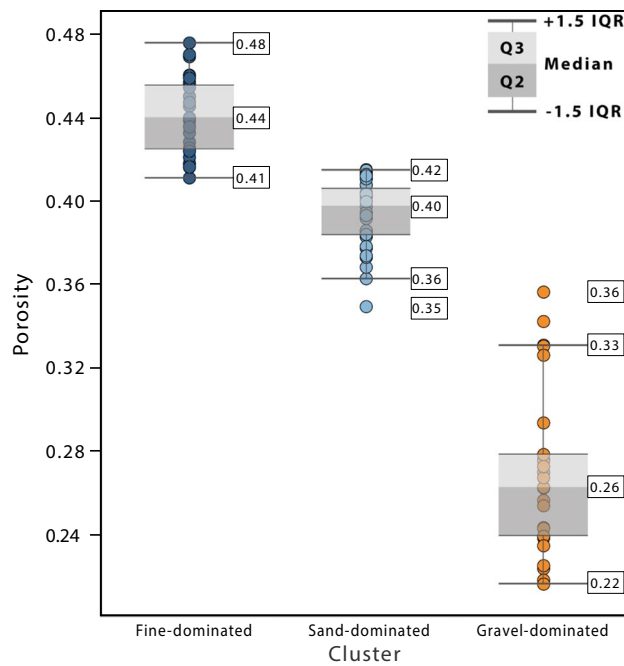
Fig. 10 Distribution of porosity values per cluster using porosity d_{50} equation

Fig. 12 Frequency histogram of samples by lithofacies, per cluster. Gcm* and Fm* classified as cluster 1 and cluster 2, correspond to two samples located close to the boundaries of the corresponding clusters
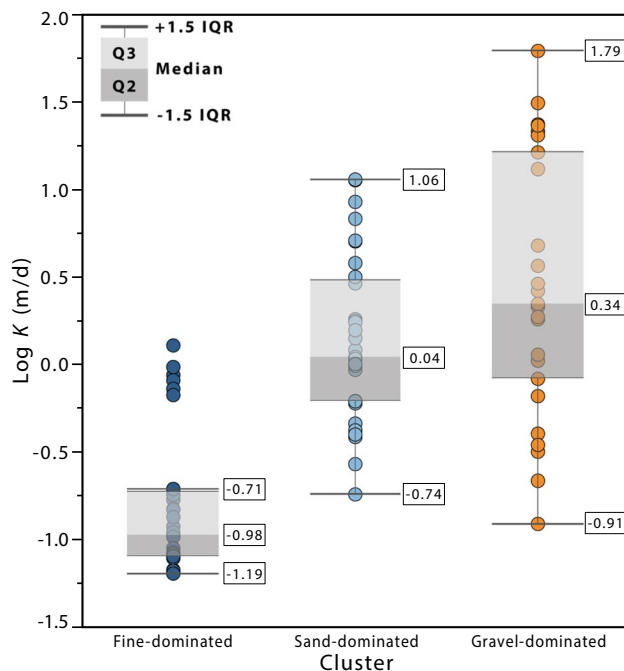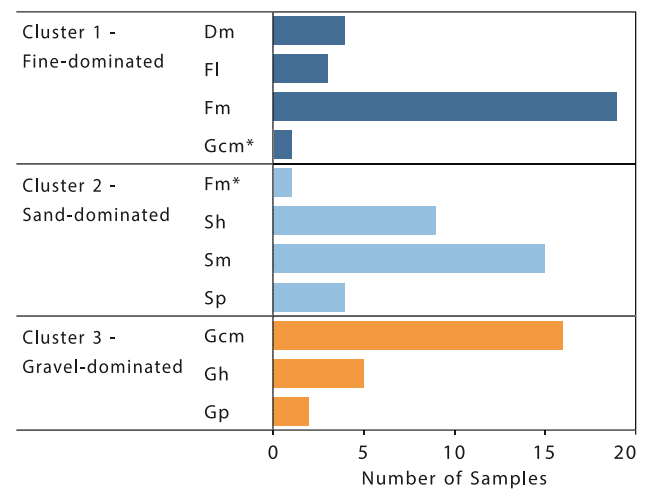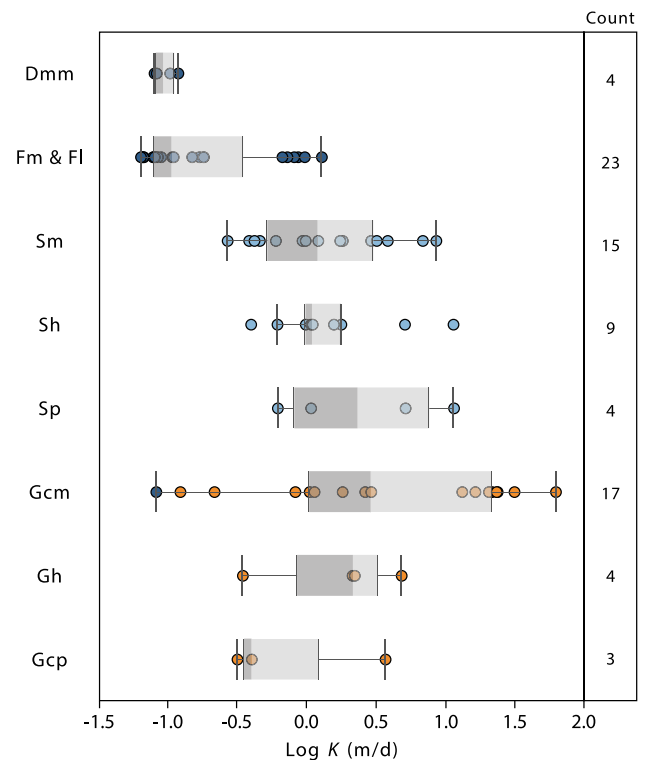

Fig. 11 Distribution of hydraulic conductivity estimates, per cluster, using the Kozeny–Carman equation. Porosity estimates using the d_{50} equation were used as an input

Fig. 13 Distribution of log of hydraulic condictivity, estimated using the Kozeny–Carman equation, for each lithofacies

(2016) used the Kozeny–Carman equation to estimate the hydraulic conductivity of their lithofacies and for modelling a *K* field in glaciofluvial sediments from North America (MADE site, Mississippi, USA). As empirical approaches do not always result in a good match with measured values, it becomes important to know the applicability of different equations to different depositional environments (Rosas et al. 2014). In that sense, the Kozeny–Carman equation has proven to generate better estimates than other equations when used in sediments

similar to the ones being studied (Odong 2008; Rosas et al. 2014; Chandel and Shankar 2022). To further support its usage, the applicability of the Kozeny–Carman equation was checked via comparison with hydraulic conductivities derived by other approaches and field data.

Several studies have reported K values for the study site (Nirex 1997a; McMillan et al. 2000; Smith et al. 2023). The ranges of these K values are summarised in Fig. 14 and tabulated in Table S3 of the ESM. The K estimates for the sand-dominated and gravel-dominated clusters using the Kozeny–Carman equation show a good correlation with K measurements made by Nirex (1997a; reported in McMillan et al. 2000) on sand and gravel deposits using cone penetrometer, slug and pumping tests. The fine-dominated cluster has a geometric mean K value of 0.1 m/d, which is less than one order of magnitude higher than the estimated geometric mean of Nirex (1997a) obtained through a pumping test on a silt channel facies (0.019 m/d). Estimates reported in McMillan et al. (2000) using Fair and Hatch equation (Fair et al. 1933) are 0.5 to 3 orders of magnitude lower than the estimates presented in this work, which are considered unrepresentative.

Smith et al. (2023) reports a range of K values for each lithofacies (following LEwis et al. 2006 rock-types ranges). The values are useful for comparisons between different units; notably, it is observed that the predictions from this study are contained in the ranges reported. However, these ranges are for a very broad variety of geological materials and therefore do not provide useful site-specific information. Purkis et al. (2023) have recently measured the permeability of gravel and sand samples from the Peel Place Quarry site using the constant head method. Their results are one to two orders of magnitude higher than the estimates in this study. This is likely because their samples were re-compacted from loose sediment and hence had much higher porosities than in situ sediments, as well as no sedimentary structures. K ranges can also be determined during tasks of conceptual modelling and hydrogeological modelling calibration, when it is useful to have a reference range of values supported by appropriate analyses. Extensive hydrogeological flow modelling of the superficial aquifer from the Sellafield site produces overall hydraulic conductivities in the range 0.2 to 95 m/d (Nirex 1997b), with a geometric mean of 1.9 m/d. This range is comparable to the values predicted in this work using the PSD and Kozeny–Carman equation, although closer to the ones estimated for the sand-dominated and gravel-dominated clusters. This finding likely arises from the fact that modelled values represent the hydraulic conductivity of the sequence, which are dominated by the more permeable units (arithmetic weighted average).

Nevertheless, incorporation of the less permeable units represented by cluster 1 (fine-dominated) may be necessary to correctly model solute transport. Finally, MacDonald et al. (2012) measured the permeability of glacigenic deposits in Northern Scotland using a Guelph permeameter. Considering only the deposits they classified as glaciofluvial and glaciolacustrine, their estimated K ranges and median values for their sands (median~2.5 m/d) and gravels (median~5.5 m/d) are in the same

Fig. 14 Ranges of values of log K (m/d) for different sediment sizes, according to different authors: (**a**) This work; (**b**) Nirex (1997b); (**c**) McMillan et al. (2011), including field measurements and PSD-based estimates as reported in Nirex (1997a); (**d**) MacDonald et al. (2012), including only measurements on glaciofluvial and glaciolacustrine deposits; (**e**) Purkis et al. (2023). Bars represent the whole range (min to max), while dots represent single measurements
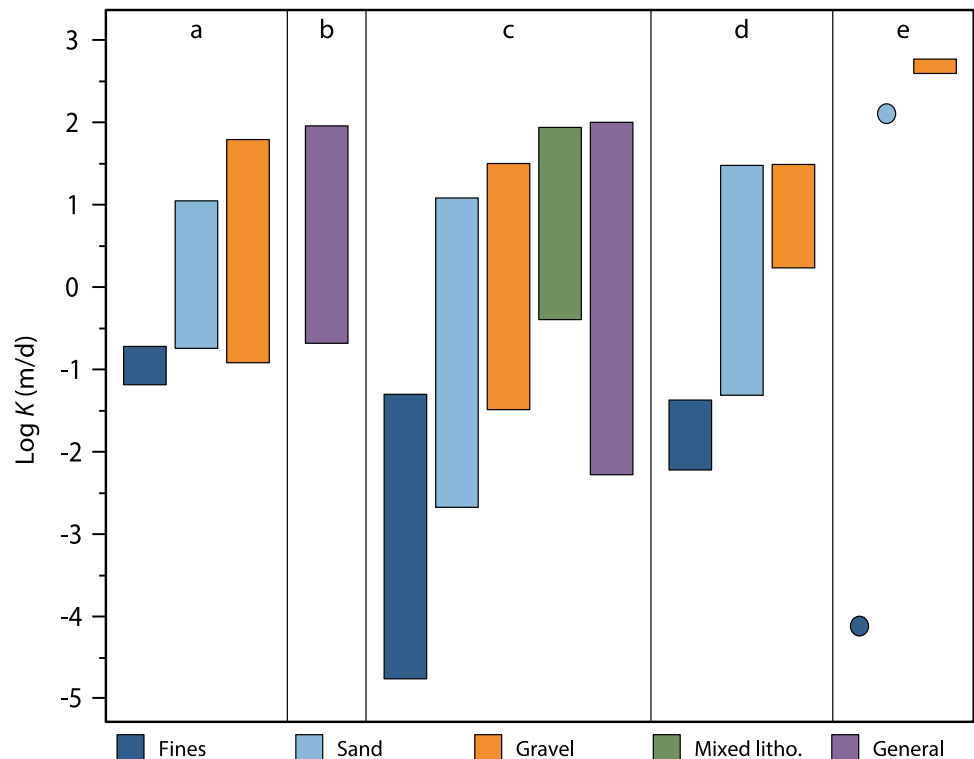
order of magnitude as those obtained in this work (Fig. 14). Their silt deposits show a median K one order of magnitude lower (median 0.02 m/d) than this work. However, only two samples, one glaciolacustrine and one glacial till, have been included in this estimate, and their sand and silt classification might not exactly fit the one obtained through cluster analysis.

Using the Fair and Hatch equation (Fair et al. 1933; Rosas et al. 2014) instead of Kozeny–Carman, with the dataset here presented, leads to lower K estimates. Sand and gravel hydraulic conductivities are one to three orders of magnitude lower (Fair and Hatch ranges 2×10^{-3} to 3×10^{-1} m/d and 4×10^{-4} to 4×10^{-1} m/d, respectively). Estimates for the fine-dominated cluster are also lower than Kozeny–Carman estimates and better match other authors' ranges, with values in the 10^{-3} to 10^{-2} m/d range. However, the method relies on the sum of the ratio between the percentage of grains retained between adjacent sieves and the geometric mean of the sieve's sizes, and changing the number of sieves used can affect the hydraulic conductivity by up to one order of magnitude.

Implications for hydrogeological modelling in heterogeneous porous media

Hydrogeological modelling requires definition of hydrofacies on a scale appropriate for the problem. For all but the smallest site-scale problems, a lithofacies classification that is too detailed might not be optimal for definition of hydrofacies, and some generalisation might be preferable whereby lithofacies units that have similar hydrogeological characteristics are grouped together into a single hydrofacies. A methodology that uses PSD data only is presented here to define hydrofacies using an unsupervised cluster-analysis approach. This methodology was tested against conventional classification based entirely on field outcrop observations. The agreement between cluster analysis and top-level lithofacies characteristics (F, S, G, D) was necessarily only partial, since the cluster analysis grouped D (diamict) together with F (fines). However, given that the predicted hydraulic characteristics of the latter two categories are similar, it is entirely appropriate to group them as a single hydrofacies for the purpose of hydrogeological modelling. Using the PSD data thus allowed solving two problems simultaneously: defining hydrofacies units and characterising their hydraulic properties. Such an approach has the advantage that it could be based entirely on borehole sampling, where, as is usually the case, no outcrop evidence is available. In this case study, no significant grouping pattern was observed between hydrogeological parameters and bottom level lithofacies indicators groups (m, c, p; i.e. those based on sedimentary structure within the top-level—particle size based—groupings F, S, G, D). However, it is possible that this may be an effect of the approach used to predict hydraulic conductivity, which used only d_{10} and porosity. Where there are additional variations in sediment texture and structure that may influence the hydraulic conductivity, additional variables needed to capture such features should be included in the cluster analysis. Using PSD data remains, however, a critical step, in view of the strong relationships between this and other petrophysical properties, sedimentological facies and depositional context.

The results presented in this work can be used as in dell'Arciprete et al. (2012) and Bianchi and Zheng (2016). They conducted a lithofacies characterisation of a heterogeneous porous media. Then, they used it as an input to generate realisations of the spatial distribution of the lithofacies and their hydraulic conductivity using the Markov chain/transition probability method (MC/TP; TPROGS code from Carle 1999). That same approach is suitable to the Sellafield site given the heterogeneous nature of the sediments at this site. Following this approach, several realisations can be generated to represent the spatial distribution of the lithofacies while keeping their connectivity as well as their heterogeneous distribution. PSD data from Sellafield site boreholes can be used to assign samples to the clusters identified in this work, which represent the appropriate hydrofacies to be used as an input for such geostatistical methods. Additionally, given that K ranges obtained for cluster 2 (sand-dominated) are contained within the range of those for cluster 3 (gravel-dominated), a potential bimodal hydrofacies definition could be tested on a hydrogeological model, as has been done by Bianchi and Zheng (2016).

Analogously, the workflow presented in this study is widely applicable to other sites in which sedimentary heterogeneities impact groundwater flow and contaminant transport, where inputs for hydrogeological modelling are needed, and where particle size distribution data are available. Heterogeneous fluvial sediments are widespread in terms of resource geology contexts, and simulations for hydrocarbon extraction, carbon capture and storage, hydrothermal potential, and general groundwater contaminant transport would potentially benefit from the approach described. Using the Kozeny–Carman equation to obtain hydraulic conductivity ideally requires either porosity data, or a site-specific evaluation of the relationship between PSD and porosity so that the latter variable can be inferred. Other empirical equations such as Hazen's rule, the Terzaghi or Beyer equations, which do not need porosity, may be applied instead of Kozeny–Carman where this can be justified in terms of the grain-size characteristics of the sediment (Rosas et al. 2014).

Conclusion

An unsupervised hydrofacies identification approach was evaluated on a heterogeneous unconsolidated glaciofluvial sequence. The approach used the particle size distribution (PSD) with K-means unsupervised clustering algorithm.

Sedimentological facies were defined from field observations and compared against the identified hydrofacies from cluster classification. Three clusters—fines/diamicts, sands and gravels—were defined using the K-means algorithm. Results showed excellent correlation between sedimentological lithofacies and cluster analysis. Although each cluster typically contained multiple sedimentary facies, these had essentially similar predicted hydrogeological parameters. For instance, both fines-only and diamicton facies fell within the fine-dominated cluster, but these lithofacies had similar d_{10} grain size and, therefore, similar predicted hydraulic conductivities. For hydrogeological classification, this unsupervised cluster analysis approach is simpler to apply than lithofacies classification; it captures hydrogeologically important details without creating an excessive number of categories, does not require the same level of expert knowledge to apply and reduces observational bias while remaining consistent with a sedimentological lithofacies classification.

The workflow developed here used a limited number of field porosity measurements (using the sand replacement method) to correlate those porosities to various grain-size characteristics (d_{10}, d_{50}, U, geometric standard deviation of Φ). From these, log d_{50} showed the best correlation with the measured porosities. Porosities are higher than previously reported by Nirex (1997a) and used in subsequent works (e.g. McMillan et al. 2000), with the gravel cluster having the lowest porosity (average 27%; range 22–36%), sand cluster intermediate (average 40%; range 35–42%) and fines/diamicts cluster highest (average 44%; range 41–48%).

Hydraulic conductivity ranges for each cluster were estimated using the Kozeny–Carman equation. Gravel and sand clusters have higher K values (geometric means of 2.8 and 1.4 m/d, respectively) than fines/diamicts (geometric mean of 0.1 m/d). These results are consistent with previously measured values from pumping and slug tests. As the sand cluster K range is contained in the gravel cluster range, a potential bimodal hydrofacies definition may be appropriate when generating realizations of hydrofacies distributions for groundwater modelling, as has been suggested in other similar studies (e.g. Bianchi and Zheng 2016).

The methodology developed here is useful for classifying heterogeneous sediments for hydrogeological purposes when the particle size distribution can be measured, and facilitates consistent hydrogeological parameter assignment compared to a more traditional approach. Specifically, the clustering approach can be used to optimise hydrofacies definitions for the purposes of generating realisations for hydrogeological modelling using geostatistical methods. Site-specific conditions should be considered to adequately decide on the number of clusters and empirical relationships needed to infer hydraulic conductivity. Further studies should focus on (1) the comparison between different empirical equations and other methods by which K can be estimated, either in situ or laboratory based; and (2) the use of other easily measurable parameters, such as those obtained through geophysical measurements, to complement hydrofacies classifications and hydraulic conductivity estimates.

Declarations

Conflicts of interest James Graham of NNL contributed to investigation; resources, writing-review and editing. Landis Jared West, Ian T. Burke, Luca Colombera and Felipe Gallardo Ceron 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

Akhurst MC, Chadwick RA, Holliday DW et al (1997) Geology of the west Cumbria district. Memoir of the British Geological Survey, Sheets 28, 37 and 47 (England and Wales). Memoir of the British Geological Survey

Al-Mudhafar WJ (2017) Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. J Pet Explor Prod Technol 7:1023–1033. https://doi.org/10.1007/s13202-017-0360-0

Alyamani MS, Şen Z (1993) Determination of hydraulic conductivity from complete grain-size distribution curves. Groundwater 31:551–555. https://doi.org/10.1111/j.1745-6584.1993.tb00587.x

Anderson MP (1989) Hydrogeologic facies models to delineate large-scale spatial trends in glacial and glaciofluvial sediments. Geol Soc Am Bull 101:501–511.

Anderson MP, Aiken JS, Webb EK, Mickelson DM (1999) Sedimentology and hydrogeology of two braided stream deposits. Sediment Geol 129:187–199. https://doi.org/10.1016/S0037-0738(99)00015-9

Arora N, Kumar A, Singal SK (2022) Measurement of sediment particle size distribution and shape using dynamic imaging analysis. In: IOP Conference Series: Earth and Environmental Science. Institute of Physics

Arya LM, Leij FJ, Shouse PJ, van Genuchten MTh (1999) Relationship between the hydraulic conductivity function and the particle-size distribution. Soil Sci Soc Am J 63:1063–1070. https://doi.org/10.2136/sssaj1999.6351063x

Barker GR, West LJ, Graham JT et al (2024) Effect of grain size variation on strontium sorption to heterogeneous aquifer sediments. J Environ Radioact 277. https://doi.org/10.1016/j.jenvrad.2024.107451

Bear J (1972) Dynamics of fluids in porous media. American Elsevier, New York

Bennett MR, Huddart D, Thomas GSP (2002) Facies architecture within a regional glaciolacustrine basin: Copper River, Alaska. Q Sci Rev 21:2237–2279. https://doi.org/10.1016/S0277-3791(02)00027-6

Bianchi M, Zheng C (2016) A lithofacies approach for modeling non-Fickian solute transport in a heterogeneous alluvial aquifer. Water Resour Res 52:552–565. https://doi.org/10.1002/2015WR018186

British Standards Institution (1990) BS 1377–9: methods for test for soils for civil engineering purposes - In-situ tests. British Standards Institution

British Standards Institution (2016) BS 17892–4: Geotechnical investigation and testing. Laboratory testing of soil - Determination of particle size distribution. British Standards Institution

British Standards Institution (2022a) BS 1377–2: Methods of test for soils for civil engineering purposes - Classification tests and determination of geotechnical properties. British Standards Institution

British Standards Institution (2022b) BS13322–2: article size analysis. Image analysis methods - Dynamic image analysis methods

British Standards Institution (2018) BS EN ISO 14688–1:Geotechnical investigation and testing. Identification and classification of soil - Identification and description. British Standards Institution

Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. Commun Stat Theory Methods 3:1–27. https://doi.org/10.1080/03610927408827101

Carle SF (1999) T-PROGS: transition probability geostatistical software. University of California, Davis, p 78

Carman PC (1937) Fluid flow through granular beds. Trans Inst Chem Eng London 15:150–156. https://doi.org/10.1016/S0263-8762(97)80003-2

Carman PC (1956) Flow of gases through porous media. Butterworths Scientific, London

Chandel A, Shankar V (2022) Evaluation of empirical relationships to estimate the hydraulic conductivity of borehole soil samples. ISH J Hydraul Eng 28:368–377. https://doi.org/10.1080/09715010.2021.1902872

Chiverrell RC, Smedley RK, Small D et al (2018) Ice margin oscillations during deglaciation of the northern Irish Sea Basin. J Quat Sci 33:739–762. https://doi.org/10.1002/jqs.3057

Coleman CG, Grimoldi E, Woollard H et al (2021) Developing 3D geological and hydrogeological models for the low level waste repository site, west Cumbria, UK. Quart J Eng Geol Hydrogeol 54. https://doi.org/10.1144/qjegh2020-026

Colombera L, Mountney NP, Medici G, West LJ (2019) The geometry of fluvial channel bodies: Empirical characterization and implications for object-based models of the subsurface. Am Assoc Pet Geol Bull 103:905–929. https://doi.org/10.1306/10031817417

Crooijmans RA, Willems CJL, Nick HM, Bruhn DF (2016) The influence of facies heterogeneity on the doublet performance in low-enthalpy geothermal sedimentary reservoirs. Geothermics 64:209–219. https://doi.org/10.1016/j.geothermics.2016.06.004

Davies DL, Bouldin DW (1979) A Cluster Separation Measure. IEEE Trans Pattern Anal Mach Intell PAMI-1:224–227. https://doi.org/10.1109/TPAMI.1979.4766909

de Marsily G, Delay F, Gonçalvès J et al (2005) Dealing with spatial heterogeneity. Hydrogeol J 13:161–183. https://doi.org/10.1007/s10040-004-0432-3

dell'Arciprete D, Bersezio R, Felletti F et al (2012) Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments. Hydrogeol J 20:299–311. https://doi.org/10.1007/s10040-011-0808-0

El-Deek I, Abdullatif O, Korvin G (2017) Heterogeneity analysis of reservoir porosity and permeability in the Late Ordovician glacio-fluvial Sarah Formation paleovalleys, central Saudi Arabia. Arabian J Geosci 10. https://doi.org/10.1007/s12517-017-3146-2

Evans DJA, Clark CD, Mitchell WA (2005) The last British ice sheet: a review of the evidence utilised in the compilation of the Glacial Map of Britain. Earth Sci Rev 70:253–312. https://doi.org/10.1016/j.earscirev.2005.01.001

Evans DJA, Phillips ER, Hiemstra JF, Auton CA (2006) Subglacial till: formation, sedimentary characteristics and classification. Earth Sci Rev 78:115–176. https://doi.org/10.1016/j.earscirev.2006.04.001

Ewis D, Ba-Abbad MM, Benamor A, El-Naas MH (2022) Adsorption of organic water pollutants by clays and clay minerals composites: a comprehensive review. Appl Clay Sci 229. https://doi.org/10.1016/j.clay.2022.106686

Fair GM, Hatch LP, Hudson HE (1933) Fundamental factors governing the streamline flow of water through sand [with discussion]. J Am Water Works Assoc 25:1551–1565

Frings RM, Schüttrumpf H, Vollmer S (2011) Verification of porosity predictors for fluvial sand-gravel deposits. Water Resour Res 47. https://doi.org/10.1029/2010WR009690

Goltz M, Huang J (2017) Analytical modeling of solute transport in groundwater: using models to understand the effect of natural processes on contaminant fate and transport. 235–239. https://doi.org/10.1002/9781119300281.index

Gray E, Hartley A, Howell J (2022) The influence of stratigraphy and facies distribution on reservoir quality and production performance in the Triassic Skagerrak Formation of the UK and Norwegian Central North Sea. In: Geological Society Special Publication. Geological Society of London, pp 379–409

Hazen A (1892) Some physical properties of sands and gravels, with special reference to their use in filtration. 24th Annual Report, Massachusetts State Board of Health, pp 539–556

He X, Sonnenborg TO, Jørgensen F et al (2013) Analyzing the effects of geological and parameter uncertainty on prediction of groundwater head and travel time. Hydrol Earth Syst Sci 17:3245–3260. https://doi.org/10.5194/hess-17-3245-2013

Hill MC (2006) The practical use of simplicity in developing ground water models. Groundwater 44:775–781. https://doi.org/10.1111/j.1745-6584.2006.00227.x

Jain AK, Murty MN, Flynn PJ (1999) Data clustering. ACM Comput Surv 31:264–323. https://doi.org/10.1145/331499.331504

Kasenow M (2002) Determination of hydraulic conductivity from grain size analysis. Water Resources Publication. https://www.wrpllc.com/books/gsadhc.html

Kawo NS, Korus J, Gulbrandsen ML (2023) Multiple-point statistical modeling of three-dimensional glacial aquifer heterogeneity for improved groundwater management. Hydrogeol J 31:1525–1546. https://doi.org/10.1007/s10040-023-02658-x

Klingbeil R, Kleineidam S, Asprion U et al (1999) Relating lithofacies to hydrofacies: outcrop-based hydrogeological characterisation of Quaternary gravel deposits. Sediment Geol 129:299–310. https://doi.org/10.1016/S0037-0738(99)00067-6

Koltermann CE, Gorelick SM (1996) Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches

Kozeny J (1927) Über kapillare leitung des wassers im boden. Royal Acad Sci Vienna Proc Class I 136:271–306

Kumar T, Seelam NK, Rao GS (2022) Lithology prediction from well log data using machine learning techniques: a case study from Talcher coalfield, Eastern India. J Appl Geophy 199. https://doi.org/10.1016/j.jappgeo.2022.104605

Kurjanski B, Rea BR, Spagnolo M et al (2020) A conceptual model for glaciogenic reservoirs: from landsystems to reservoir architecture. Mar Pet Geol 115. https://doi.org/10.1016/j.marpetgeo.2019.104205

Lee J (2018) Glacial lithofacies and stratigraphy. In: Menzies J, van der Meer JJM (eds) Past glacial environments, second edition. Elsevier, Amsterdam, pp 377–429

Lee S-Y, Carle SF, Fogg GE (2007) Geologic heterogeneity and a comparison of two geostatistical models: sequential Gaussian and transition probability-based geostatistical simulation. Adv Water Resour 30:1914–1932. https://doi.org/10.1016/j.advwatres.2007.03.005

Lewis MA, Cheney CS, Ódochartaigh BÉ (2006) Guide to Permeability Indices: British Geological Survey Report CR/06/160N

Li L, Eykholt GR, Benson CH (2001) Groundwater modeling: semi-analytical approaches for heterogeneity and reaction networks. https://www.wri.wisc.edu/research/groundwater-modeling-semi-analytical-approaches-for-heterogeneity-and-reaction-networks/

Liu G, Pu H, Zhao Z, Liu Y (2019) Coupled thermo-hydro-mechanical modeling on well pairs in heterogeneous porous geothermal reservoirs. Energy 171:631–653. https://doi.org/10.1016/j.energy.2019.01.022

Livingstone SJ, Evans DJA, Ó Cofaigh C et al (2012) Glaciodynamics of the central sector of the last British–Irish Ice Sheet in Northern England. Earth Sci Rev 111:25–55. https://doi.org/10.1016/j.earscirev.2011.12.006

LLW Repository Ltd. (2011) Environmental safety case - main report. Available at: https://assets.publishing.service.gov.uk/media/5a9eb747ed915d07a3b5dc63/LLWR_ESC_Main_Report_-_May_2011.pdf. Holmrook

MacDonald AM, Maurice L, Dobbs MR et al (2012) Relating in situ hydraulic conductivity, particle size and relative density of superficial deposits in a heterogeneous catchment. J Hydrol (Amst) 434–435:130–141. https://doi.org/10.1016/j.jhydrol.2012.01.018

McMillan AA, Hamblin RJO, Merrit JW (2011) A lithostratigraphical framework for onshore Quaternary and Neogene (Tertiary) superficial deposits of Great Britain and the Isle of Man. British Geological Survey Research Report, RR/10/03

McMillan AA, Heathcote JA, Klinck BA et al (2000) Hydrogeological characterization of the onshore Quaternary sediments at Sellafield using the concept of domains. Q J Eng GeolHydrogeol 33:301–323. https://doi.org/10.1144/qjegh.33.4.301

Merritt JW, Auton CA (2000) An outline of the lithostratigraphy and depositional history of Quaternary deposits in the Sellafield district, west Cumbria. Proc Yorks Geol Soc 53:129–154. https://doi.org/10.1144/pygs.53.2.129

Miall AD (2006) The geology of fluvial deposits. Sedimentary facies, basin analysis, and petroleum geology. Springer, Berlin

Microtrac MRB (2024) Camsizer X2. https://www.microtrac.com/products/particle-size-shape-analysis/dynamic-image-analysis/camsizer-x2/. Accessed 10 Jun 2024

Montero JM, Colombera L, Yan N, Mountney NP (2021) A workflow for modelling fluvial meander-belt successions: combining forward stratigraphic modelling and multi-point geostatistics. J Pet Sci Eng 201:108411. https://doi.org/10.1016/j.petrol.2021.108411

Newell AJ, Woods MA, Graham RL et al (2021) Derivation of lithofacies from geophysical logs: a review of methods from manual picking to machine learning. British Geol Survey Open Report, OR/21/006.

Nichols TE, Worden RH, Houghton JE et al (2023) Sediment texture and geochemistry as predictors of sub-depositional environment in a modern estuary using machine learning: a framework for investigating clay-coated sand grains. Sediment Geol 458:106530. https://doi.org/10.1016/j.sedgeo.2023.106530

Nirex (1997a) Synthesis of the hydrogeological characteristics of Quaternary sequences in the Sellafield area. UK Nirex Ltd. Report No. SA/97/001

Nirex (1997b) Sellafield surface hydrological modelling: near-surface groundwater model. UK Nirex Ltd. Report No. SA/97/009

Odong J (2008) Evaluation of empirical formulae for determination of hydraulic conductivity based on grain-size analysis. J Am Sci 4

Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830

Purkis J, Robinson C, Shaw S et al (2023) Characterisation data for sediments collected from Peel Place Quarry, Cumbria, UK – A Sellafield nuclear site analogue. The University of Manchester version 3. https://doi.org/10.17632/xn63xpnkbx.1

Riva M, Guadagnini L, Guadagnini A (2010) Effects of uncertainty of lithofacies, conductivity and porosity distributions on stochastic interpretations of a field scale tracer test. Stoch Env Res Risk Assess 24:955–970. https://doi.org/10.1007/s00477-010-0399-7

Ronayne MJ, Gorelick SM, Zheng C (2010) Geological modeling of submeter scale heterogeneity and its influence on tracer transport in a fluvial aquifer. Water Resour Res 46. https://doi.org/10.1029/2010WR009348

Rosas J, Lopez O, Missimer TM et al (2014) Determination of hydraulic conductivity from grain-size distribution for different depositional environments. Groundwater 52:399–413. https://doi.org/10.1111/gwat.12078

Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Savoy H, Kalbacher T, Dietrich P, Rubin Y (2017) Geological heterogeneity: goal-oriented simplification of structure and characterization needs. Adv Water Resour 109:1–13. https://doi.org/10.1016/j.advwatres.2017.08.017

Sellafield Ltd. (2023) Sellafield Ltd annual review of environmental performance 2022/23. Available at: https://www.gov.uk/government/publications/sellafield-ltd-annual-review-of-environmental-performance-202223/sellafield-ltd-annual-review-of-environmental-performance-202223.

Simon N, Worden RH, Muhammed DD et al (2021) Sediment textural characteristics of the Ravenglass Estuary; development of a method to predict palaeo sub-depositional environments from estuary core samples. Sediment Geol 418:105906. https://doi.org/10.1016/j.sedgeo.2021.105906

Smith NT, Merritt JW, Phillips ER (2023) High-resolution 3D geological modelling of heterogeneity in poorly exposed glacial deposits using sedimentary and glaciotectonic architectural element analysis: a case example from Sellafield in west Cumbria, UK. Quart J Eng Geol Hydrogeol 56. https://doi.org/10.1144/qjegh2022-022

Smith NT, Shreeve J, Kuras O (2020) Multi-sensor core logging (MSCL) and X-ray computed tomography imaging of borehole core to aid 3D geological modelling of poorly exposed unconsolidated superficial sediments underlying complex industrial sites: an example from Sellafield nuclear site. UK J Appl Geophy 178:104084. https://doi.org/10.1016/j.jappgeo.2020.104084

Sun D, Luo N, Vandenhoff A et al (2024) Evaluation of hydraulic conductivity estimates from various approaches with groundwater flow models. Groundwater 62:384–404. https://doi.org/10.1111/gwat.13348

Sweet ML, Blewden CJ, Carter AM, Mills CA (1996) Modeling heterogeneity in a low-permeability gas reservoir using geostatistical techniques, hyde field, southern North Sea. Am Assoc Pet Geol Bull 80:1719–1735. https://doi.org/10.1306/64eda14a-1724-11d7-8645000102c1865d

Tahmasebi P (2018) Multiple point statistics: a review. In: Handbook of mathematical geosciences: fifty years of IAMG. Springer, Cham, pp 613–643

Technology review (2013) Optimising the measurement of fine particles. Metal Powder Report 68:34–37. https://doi.org/10.1016/S0026-0657(13)70128-0

Troccoli EB, Cerqueira AG, Lemos JB, Holz M (2022) K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis. J Appl Geophy 198:104555. https://doi.org/10.1016/j.jappgeo.2022.104555

Tucker ME (2003) Sedimentary Rocks in the Field, 3rd edn. Wiley, Chichester

Vuković M, Soro A (1992) Determination of hydraulic conductivity of porous media from grain-size composition. Water Resources Publications, Littleton, Colorado

Wang J, François B, Lambert P (2017) Equations for hydraulic conductivity estimation from particle size distribution: a dimensional analysis. Water Resour Res 53:8127–8134. https://doi.org/10.1002/2017WR020888

Woessner WW, Poeter EP (2020) Hydrogeologic properties of earth materials and principles of groundwater flow. The Groundwater Project, Guelph, Ontario, Canada

Wooster JK, Dusterhoff SR, Cui Y et al (2008) Sediment supply and relative size distribution effects on fine sediment infiltration into immobile gravels. Water Resour Res 44:. https://doi.org/10.1029/2006WR005815

Wu W, Wang SSY (2006) Formulas for sediment porosity and settling velocity. J Hydraul Eng 132:858–862. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:8(858)

Zappa G, Bersezio R, Felletti F, Giudici M (2006) Modeling heterogeneity of gravel-sand, braided stream, alluvial aquifers at the facies scale. J Hydrol (Amst) 325:134–153. https://doi.org/10.1016/j.jhydrol.2005.10.016

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.