Combining high-resolution core data with unsupervised machine learning schemes for the identification of rock types and the prediction of reservoir quality.

Christophe Germay[1,*], *Tanguy* Lhomme[1], and *Paul* Bisset[2]

[1]EPSLOG, Liege, Belgium
[2]Core Technical Services, Aberdeen, UK

Abstract. CoreDNA is an integrated core analysis solution combining transdisciplinary, high resolution, non-destructive measurements on whole cores, for an early yet objective description of cores and the rapid estimation of formation properties. Whole cores, which may still be in their half-open liners, are mounted on one unique table-top equipment and submitted to a battery of tests, all sharing the same depth reference and compatible resolution ranges. A key enabler of CoreDNA is the capability to create a smooth and flat surface along whole cores, which is obtained with a succession of dry cut with a PDC cutter, each removing a sub-millimetre thick layer of rock. On a 4-inches diameter core, removing a 3mm-thick layer of rock suffices to create a 3cm wide flat surface. Such a surface is too small to hinder further sampling and plugging, yet large enough for many different types of sensors, and therefore suitable to take many high-fidelity measurements, which would normally require slabbed cores. Technologies including ultra-high-resolution pictures, elemental composition and the direct measurement of geomechanical properties such as strength and acoustic velocities, are all deployed on the same mini-slab surface along entire cores. Ultra-high resolution panoramic pictures (1.8µm/px) are processed to extract textural and colour features but also continuous grain size distribution from wavelet analysis. Grain size distributions calculated from images are backed-up by analysis of 3D topographical maps created with a laser scan. Results of these fast tests (3ft per hour) are analysed real-time and turned into high resolution, continuous profiles of properties (petrophysical, geomechanical and geochemistry). This knowledge is fed into (unsupervised) machine learning algorithms for the automated identification of lithofacies, the design of fit-for-purpose plug selections and the programming of subsequent steps in core analysis programs, even remotely. Measured core properties are stored under one unique format for all discipline, which eases interdisciplinary work, from the QC of standard tests (Routine Core Analysis, Rock Mechanical Test) to the upscaling of core data and the calibration of robust predictive models from well logs. Such data bases are also formatted for machine learning and can therefore be used to train AI models with reliable data from large numbers of legacy cores, where sedimentological descriptions and plug data are available. The case study of a well drilled in a North Sea field underlines the benefit of this disruptive technology in core analysis when run on intact cores prior to slabbing or taking any sample (plugs and preserved)..

1 Introduction

Because most branches of core analysis rely heavily on plug samples, the competition for intact core material is tough. In practice, not only these samples consume a significant fraction of the material available but also, the selection of sample sites is based on not much *a-priori* information. Indeed, little is known before permanent damage is done to the core by plugging and cutting. In this paper we promote a new approach of core analysis, which is focused on finding practical ways to bypass this issue.

To this end we designed an integrated core analysis concept that combines several non-destructive tests to produce multi-disciplinary data sets continuously along whole cores. The primary purpose of this concept is to equip core specialists with quantitative, high resolution logs of real rock properties as early as possible in their core analysis workflows, without causing any irreparable damage to cores.

* Corresponding author: tanguy.lhomme@epslog.com

2 Methodology

CoreDNA is an integrated core analysis solution combining transdisciplinary, high resolution, non-destructive measurements on whole cores, for an early yet objective description of cores and the rapid estimation of formation properties (Figure 1). Whole cores, which may still be in their half-open liners, are mounted on one unique table-top equipment and submitted to a battery of tests, all sharing the same depth reference and compatible resolution ranges.
The complete CoreDNA test series includes the following tests:

- portable XRF measurements for elemental composition,
- then high resolution photo for panoramic viewing of the cores,

- ultra-high resolution photos for the rock texture and grain properties,
- probe permeability,
- laser scan for grain properties,
- ultrasonic velocity logging,
- strength from the scratch test.
- The MiniSlab is also a suitable surface for sedimentological descriptions.

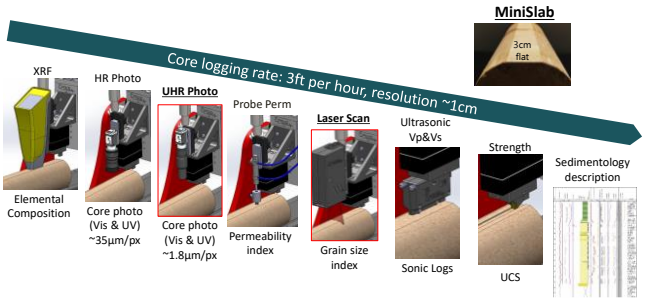

Figure 1: CoreDNA test sequence.

Results of these fast tests (3ft per hour) are analysed real-time and turned into high resolution, continuous profiles of properties (petrophysical, geomechanical and geochemistry). This data is fed into (unsupervised) machine learning algorithms for the automated identification of lithofacies, the design of fit-for-purpose plug selections and the programing of subsequent steps in core analysis programs.

2.1 Mini-Slab

The MiniSlab is created with a succession of dry cuts with a PDC cutter (Figure 2), each removing a sub-millimetre thick layer of rock. On a 4-inches diameter core, removing a 3mm-thick layer of rock suffices to create a 3cm wide flat surface. Such a surface is too small to hinder further sampling and plugging, yet large enough for many different types of sensors, and therefore suitable to take many measurements, which would normally require slabbed cores. Technologies including ultra-high-resolution pictures, elemental composition and the direct measurement of geomechanical properties such as strength and acoustic velocities, are all deployed on the same mini-slab surface along entire cores.

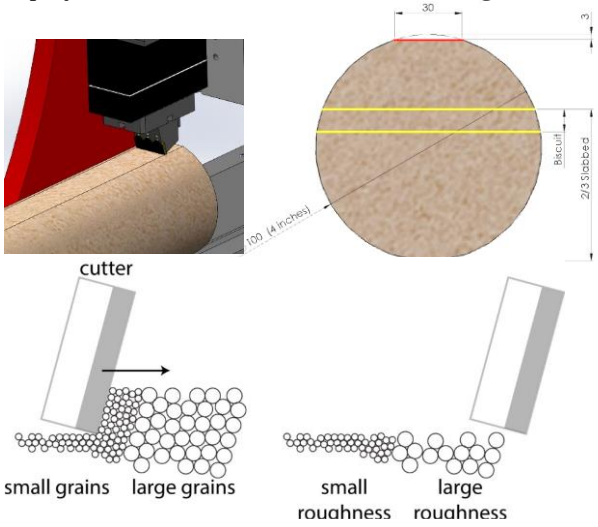

Figure 2: Creation of a MiniSlab surface with a PDC cutter.

2.2 Portable XRF

An Olympus Vanta XRF system is integrated on the CoreDNA bench. The X-Ray detector faces the MiniSlab and reports the elemental composition once the X-Ray source is triggered. Elemental compositions are calculated from the factory-standard Geochem calibration available on portable XRF devices from Olympus. The spot size is 1cm and continuous logs can be created by moving the XRF by steps of 1cm along the core. The measurement time can be chosen or automatically adapted to measure the elemental composition of rock cores with a specified accuracy.

2.3 High Resolution Pictures

High-definition panoramic core pictures (35µm/pixel) of the MiniSlab are taken under white light or UV light. The acquisition of core images is performed while the samples are fastened onto the test bed to guarantee an accurate depth match between the different measurements and core photographs. The camera system is calibrated in order to ensure a consistent picture quality despite varying testing environments and testing setups.

Principal Component Analysis are run on images of the RGB pixels of high-definition core photographs for every centimetre. The two first principal components respectively represent the picture brightness and a log of color differences. They provide useful information for the identification of rock facies.

2.4 Ultra-High Resolution Pictures

Photos of core samples are taken with telecentric lenses giving a resolution of less than 2 microns per pixel. The depth of field these lenses is limited to about 0.3mm, which is too small to fully capture the topographic features related to grain size in most clastic reservoir rocks. We combine stacks of several pictures of the same zone taken at different altitudes in order to increase the sharpness of a single reconstructed picture across the entire depth range (**Error! Reference source not found.**). The actual depth range is measured on the MiniSlab surface by the laser and used to determine the optimal number of pictures per stack.

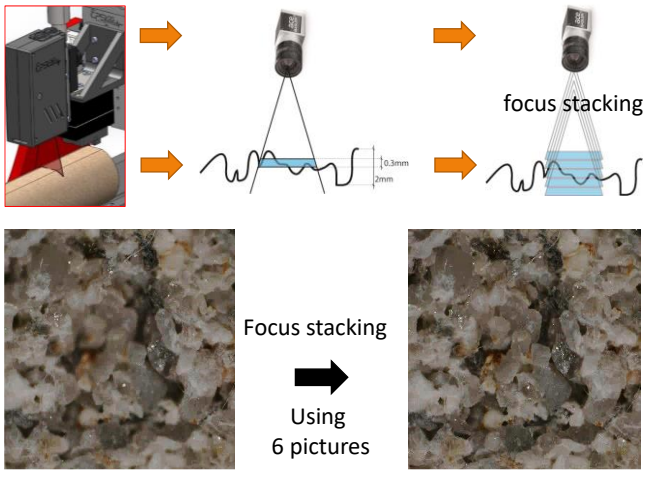

Figure 3: Acquisition and stacking of ultra-high-resolution images.

2.5 Wavelet Analysis

These pictures are turned into grey-scale images and analysed using wavelet transforms according to the following sequence (**Error! Reference source not found.**):

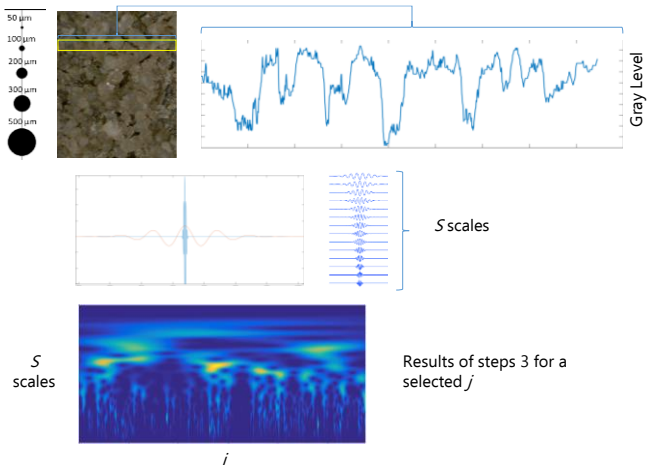

Figure 4: Wavelet analysis of Ultra-High Resolution Images for grain size distribution mapping.

1. First we extract lines of pixel intensity, and then
2. we calculate convolution products of these intensity vectors by wavelets of different scales ***s***.
3. Then we calculate the norm of these convolution products. At this point, for every line in the working window we have created such a map where the horizontal axis is the position of the pixel on the line and the vertical axis is the scale of the wavelet used in the convolution product. A high value at the coordinate (***i,s***) , shown with a green colour in **(**Figure 4**),** shows a good match between local features of the intensity profile centered on the ***i***th pixel and a wavelet of scale ***s***.
4. We average the norm of the convolution products over the entire line
5. And over the number of lines in the working window
6. We normalize the product to create a cumulative distributed function (cdf) in scale and
7. Transform this cdf into another one in terms of grain diameter
8. Finally we use a Cahn-Fullman transform to turn this results into a cdf in grain volume [1].

There are a few limitations in the technique using ultra high resolution images for Grain size distributions mapping. First the wavelet transform often ends up with non-zero low frequency residues, which are interpreted as large particles. Although in small numbers, these large ghost particles amount for a significant fraction of the cumulative grain volume. We therefore need a low-frequency filter, or a maximum grain size cut-off value to remove these ghost particles from the distributions.
These is also the matter of bedding planes and fractures which sometimes come with high contrasts and are interpreted as large particles. Strong colour contrasts may also amplify the detectability of some grains with respect to others. Windows of investigation are 1cm wide, which means that particle larger than 5mm will be hardly visible with this method.

For these reasons, the best option to obtain reliable grain sizes is two combine two independent data sources, which are the laser scans and the ultra-high resolution pictures.
Practically, maximum grain size cut-off values can be determined from laser scans and the QA/QC of results can be greatly improved such that spotting outliers and invalidating dubious results becomes much easier.

2.6 Laser Topography

The laser scan technology uses a laser beam to map the topography of the MiniSlab surface continuously along whole cores, with a vertical accuracy of 1 micron and a horizontal resolution of 20*20 microns **(**Figure 5**)**.

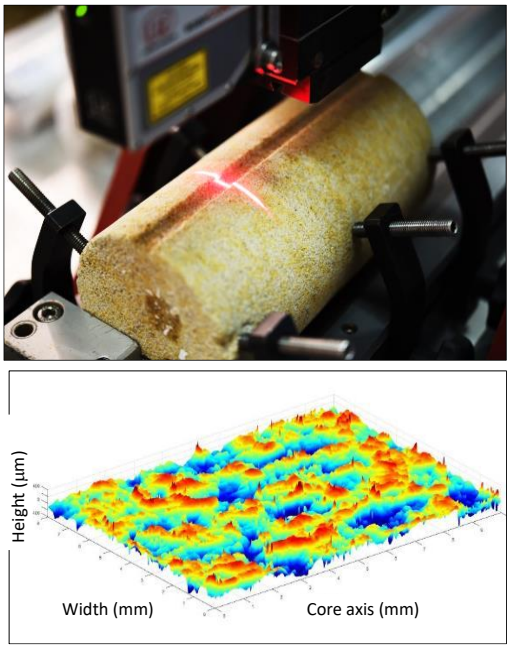

Figure 5: - Laser topography: acquisition.

The fundamental assumption here is that the topographic features of the Mini-Slab surface are related to the size of constitutive grains.
This is mostly the case in clastic rocks, unless large grains embedded in a strong matrix are sheared by the PDC cutter during rock cutting, which blurs the correlation of the measured surface roughness with the size of large grains.

2.7 Strength and Ultrasonic Velocities

Continuous high resolution profiles of rock strength profile and ultrasonic compressional velocity (Vp) profile are acquired on fresh cores. Details on the scratch tests can be found in [2] and [3].

2.8 Grain Size From Autocorrelation of Laser Profiles

Topographic maps are analysed for the purpose of deriving continuous grain size distributions along tested cores.
First we compute 1D autocorrelation functions along lines perpendicular to the core length, which are then averaged over a 1cm window. A proprietary transform, which has been calibrated on more than 10 outcrop rock samples, is then used

to turn the average autocorrelation function into a cumulative distributed function in volume of the grain size (Figure 6).

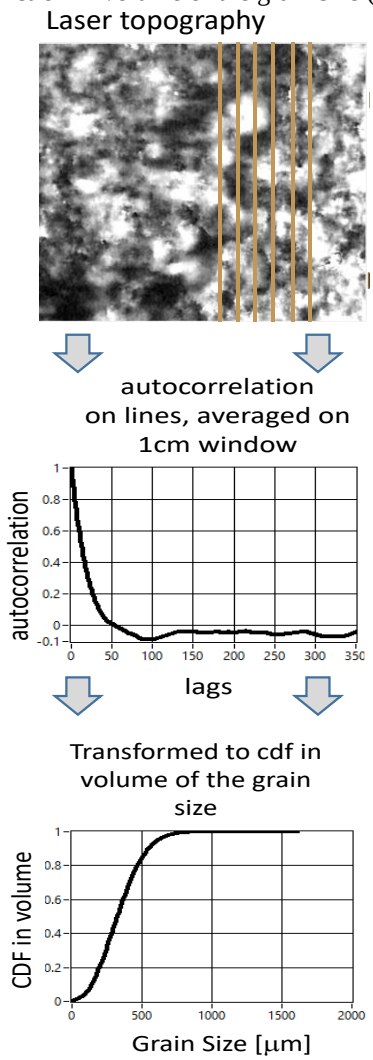

Figure 6: - Grain size from autocorrelation of laser profiles.

2.9 Estimated Permeability from Grain Size

An estimate of permeability is obtained with an empirical transform of grain size data from laser scans. This strong link between CoreDNA and conventional core data had already been demonstrated in other case studies involving other formations, as seen for example in the two charts in Figure **7**: the blue curves show CoreDNA data (strength and grain size), while the red dots show porosity and permeability values measured on plug samples much later [4].

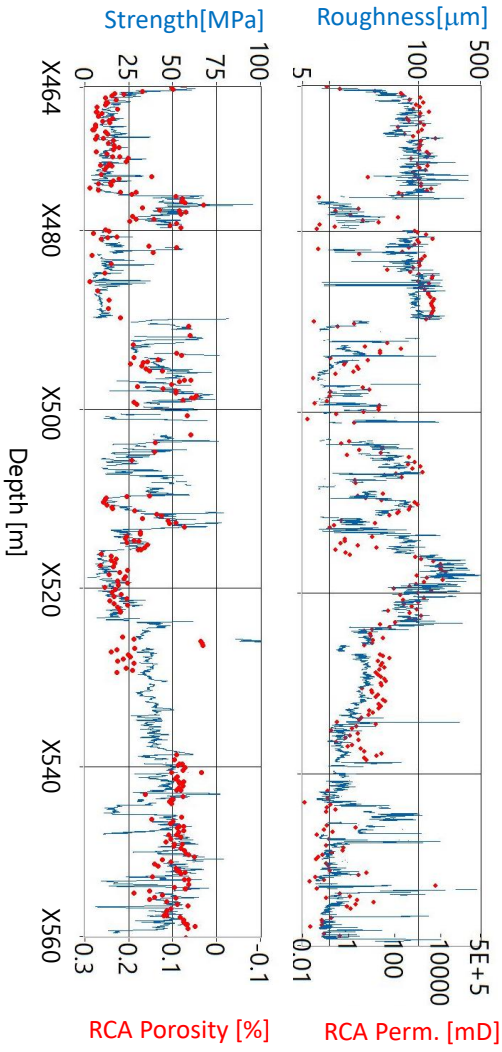

Figure 7: - Reservoir quality indexes derived from CoreDNA data

3 Applications

CoreDNA data was acquired on two cores from an observation well drilled by Neptune Energy in a North Sea prospect [5].

3.1 Data Recovery

The two cores had about similar lengths of 37m each, although Core1 had been subject to intensive plugging, including seal peels, prior to being tested with CoreDNA, while Core 2, which was tested immediately after opening its barrel, had not been plugged and was therefore in much better condition overall. CoreDNA data recovery was significantly lower in Core1 than in Core2, as seen in Figure 8, where coloured intervals correspond to sections where CoreDNA data could be retrieved, while blank intervals are seen where core material was too sparse even for CoreDNA sensors.

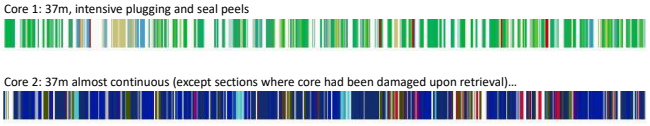

Figure 8: Data recovery on Cores 1 and 2

This difference in recovery advocates for continuous, non-destructive core measurements such as CoreDNA to be taken as a first step in core data acquisition programs, before any core material is spent in more destructive tests.

3.2 Missing data points

If the presence of specific elements in the cores is such that their concentrations cannot be detected wherever it is lower than the limit of detection (LOD) of the XRF device, the core data analyst has the two following options:

- the dataset only covers sections where all selected vectors are defined, or;
- the data set is completed with values equal to 10% of the device LOD wherever measured element concentrations where found below the LOD.

3.3 Data Coherence

The internal coherence of CoreDNA data is visible in the 2 cross-plots seen in Figure 9Figure *9*.

Brightness and color index values are normalized over the entire core depth range. High values of the brightness index means brighter formations under UV light.

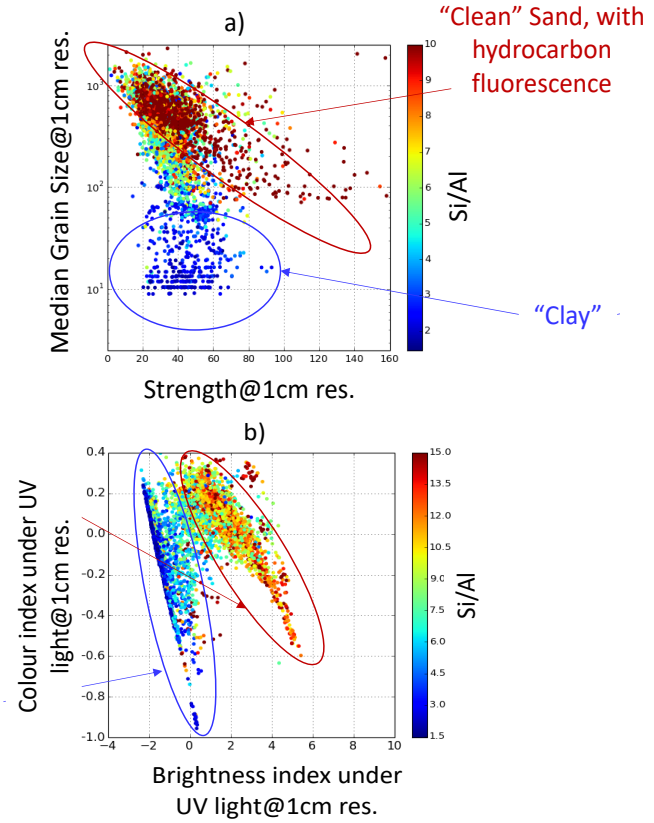

Figure 9: a) P50 of grain size distribution vs Rock Strength@1cm resolution vs. b) Colour index under UV light vs Brightness index under UV light. Colours indexed on Si/Al concentration concentrations ratio given by the portable XRF measurements.

Sections with low Si/Al concentration ratios have strength limited below 80MPa and smaller median grain sizes. They are also darker, although their colour index under UV light span across the entire range. These attributes are all compatible with the properties of clays.

Conversely, sections with higher Si/Al concentration ratios have strengths varying across the whole range but always have large grain sizes. They also appear brighter under UV light, with a slightly tighter range of colour index under UV lights than "clay" sections. These sections therefore have attributes compatible with those of clean, hydrocarbon-bearing sands.

The data organisation visible in Figure 9 prompts for a data analysis scheme based on the following data vectors:

- Strength,
- Grain Size P50 from Laser Scan,
- Elemental Compositions from XRF ,
- Brightness and colour index from high resolution core photos.

3.4 Identifying Lithofacies

A K-mean clustering scheme was run on the CoreDNA data set consisting of selected vectors, all at a 1cm resolution. The clustering analysis is performed in the multidimensional space defined by the principal components of the original data vectors. The number of selected components is adjusted to capture a meaningful fraction of the total variance. The results of the clustering algorithm are visualized with a colour scheme specifically built to represent the degree of proximity of clusters centroids. To this end, the RGB bits of each cluster colours are related to the position of cluster centroids in a 3D space defined by the three first principal components of the data space analysed by clustering (Figure 10).

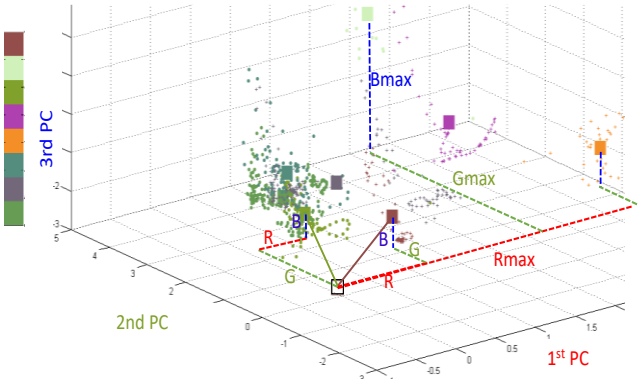

Figure 10: RGB colour scheme indicative of cluster centroid proximity.

The adequate number of clusters required to describe the spatial distribution of lithofacies detected along the tested core is determined with the use of a stabilisation diagram, synthesizing colour-based clustering results obtained for the same data set with an increasing number of clusters. The optimal number of cluster is set equal to the minimum number of cluster above which the colour-based representation of facies does not change anymore.

Statistical properties of lithofacies identified as clusters are shown using the logic depicted in Figure 11.

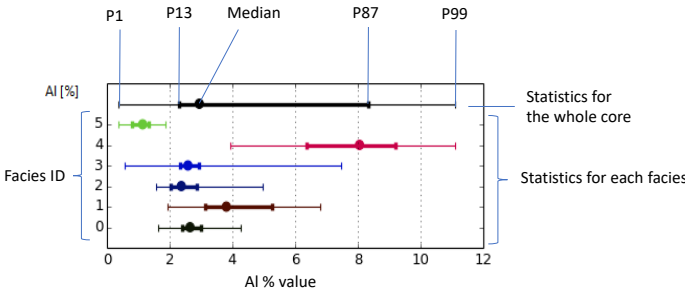

Figure 11: Statistic box example

The vertical axis of the statistic box corresponds to the facies number, while the horizontal axis corresponds to the range for one of the physical properties used in the clustering scheme. The topmost line shows the statistics for the physical property displayed by the statistic box for the entire core. Each facies is shown on a different line below, with its specific colour, extending from the 1st to the 99th percentiles for the physical property being displayed. The dot on this bar corresponds to the median value of the physical property for this facies. The left hand side extremity of the thick bar shows the 13th percentile while the right hand side of the thick bar shows the 87th percentile. In a normal distribution, the range covered by the thick bar would correspond to the median plus or minus the standard deviation.

3.5 Results

Clustering analysis for lithofacies identification was run on the two cores separately.

3.5.1 Core1

The first core was analysed using the following CoreDNA vectors:

1. Strength;
2. Brightness Index;
3. P50 of GSD from Laser;
4. Al, Ca, Fe, S & Si concentration

A total of seven different lithofacies were identified in Core1. Statistic boxes are shown in Figure 12 for each of the physical properties used in the clustering scheme. The main characteristics of the 7 lithofacies identified in Core1 are described in Table 1.

Table 1: Lithofacies identified in Core1

Fa-cies	Core cover-age	Strength	Grain size	Mineralogy
0	13%	Weakest	Coarse to very coarse	Low Al
1	2%	Strong	Small	High Ca, low Al, low Si
2	13%	Interme-diate	Medium	High Si, low Ca, low Al, low S
3	34%	Weak	Medium to coarse	High Si/Al ratio, low Ca, low S, low Fe
4	9%	Weak	Medium to coarse	High Al, High Fe
5	10%	Strong	Medium to coarse	High Si, low Ca,
6	13%	Strong	Small	High Al, high Fe, High S

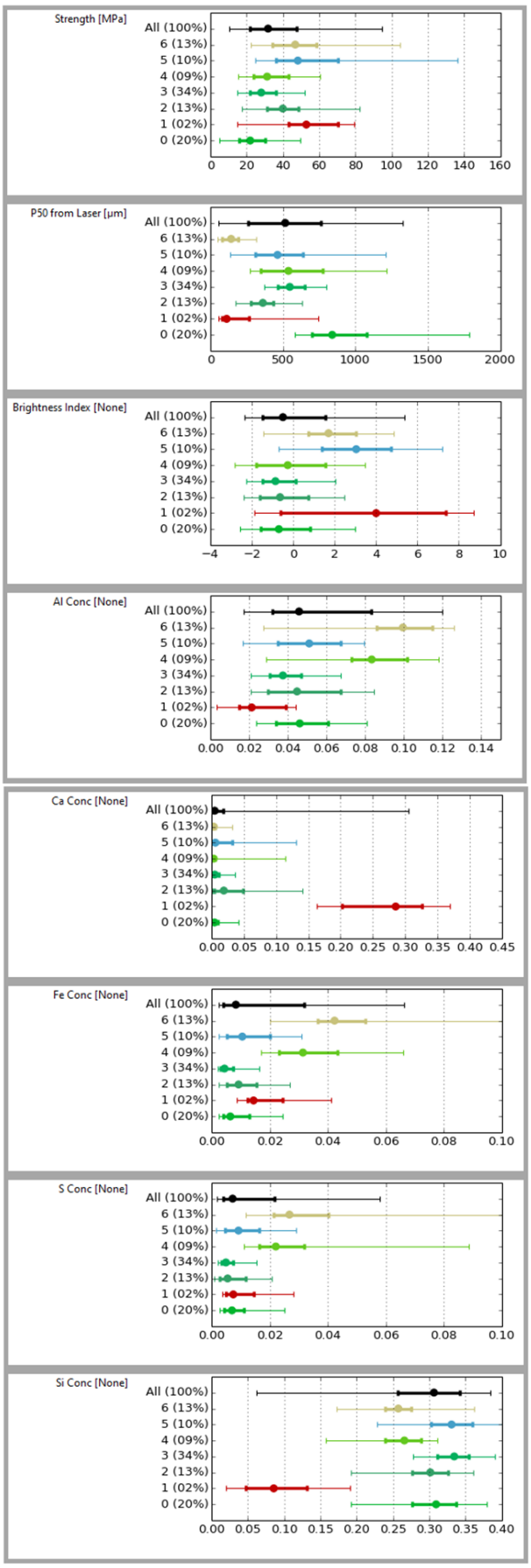

Figure 12: Lithofacies identified by clustering; Core1.

3.5.2 Core2

The second core was analysed using the following CoreDNA vectors:

1. Strength;
2. P50 of GSD from Laser;
3. Al, Ca, Fe, S & Si concentration.

A total of six different lithofacies were identified in Core2. Statistic boxes are shown in **Error! Reference source not found.** for each of the physical properties used in the clustering scheme.

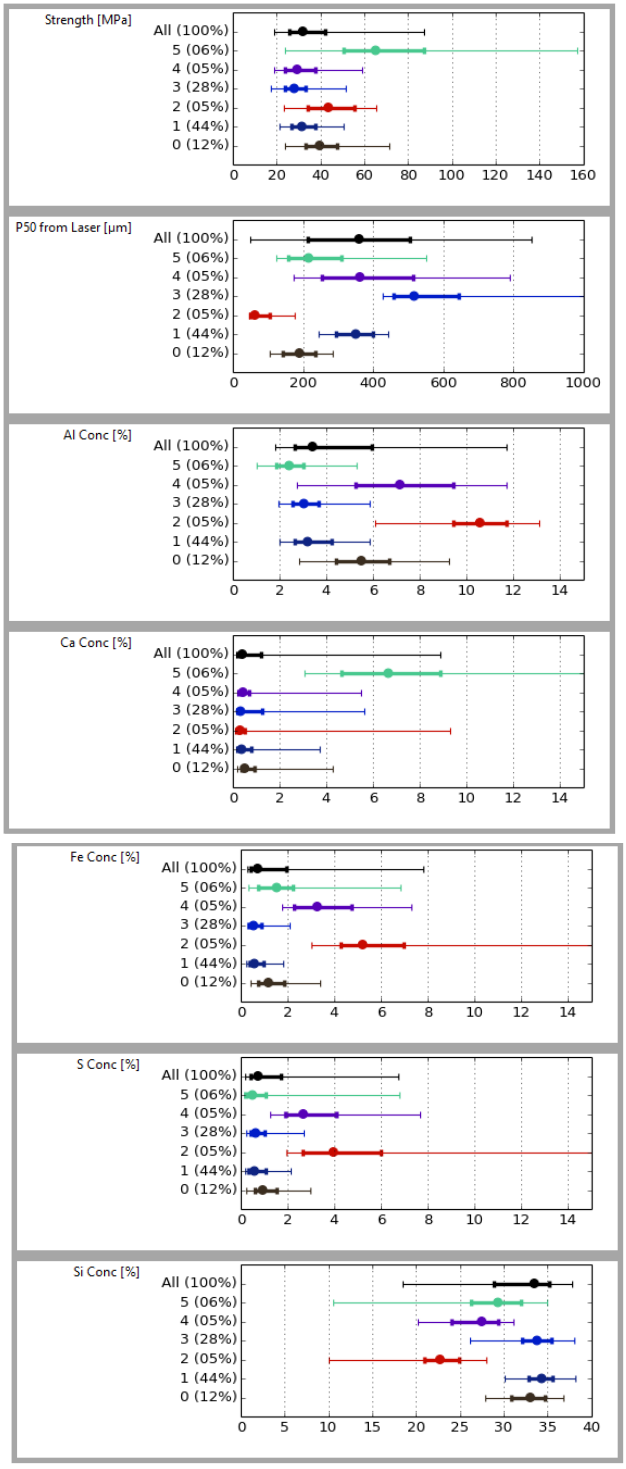

Figure 13: Lithofacies identified by clustering: Core2.

The main characteristics of the 6 lithofacies identified in Core2 are described below:

Table 2: Lithofacies identified in Core2

Facies	Core coverage	Strength	Grain size	Mineralogy
0	12%	Intermediate	Very fine	Low Si/Al ratio
1	44%	Weak	Medium	High Si/Al ratio, low Ca
2	5%	Intermediate	Very fine	Very low Si/Al ratio
3	28%	Weak	Coarse to very coarse	High Si/Al ratio, low Ca, low Al
4	5%	Weak	Medium to coarse	Low Si/Al ratio, large S
5	6%	Strong	fine	Large Ca, High Si/Al ratio

3.5.3 Facies groups

Facies identified above are grouped by types of lithology, using the information synthetized in Table 2. For instance, the aluminium concentration detected by the XRF measurements is indicative of the presence of clay, and therefore of the sand cleanliness. Facies can therefore be grouped according to their position with respect to a mid-range threshold (Figure 14).

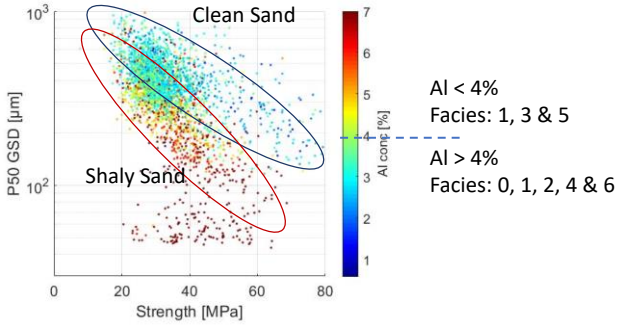

Figure 14: Distinction between shaly sand and clean sand based on the average Al concentration in each facies.

Using this classification, we group lithofacies as seen in Figure 15.

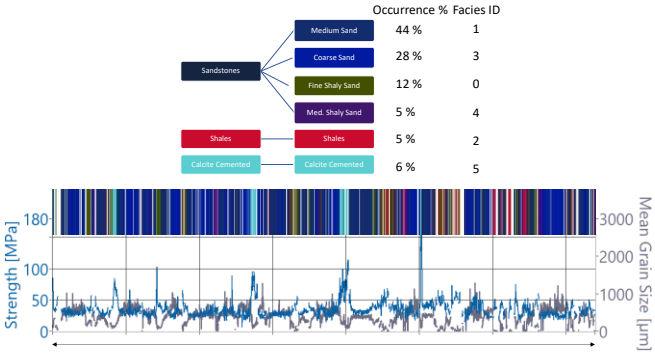

Figure 15: Lithofacies identified and grouped along Core2.

4 Conclusions

The complete CoreDNA array of transdisciplinary, high resolution, non-destructive tests was deployed as an integrated core analysis package on cores from one observation well drilled by Neptune Energy in a North Sea prospect.

One of the two cores from this well was tested with CoreDNA prior to any other testing program and was therefore virgin

from any sample hole, which allowed for maximum CoreDNA data recovery along its 37 meter of length.
A comprehensive data set ranging from textural and colour features of the rock to grain size distribution statistics, elemental concentrations, elastic wave velocities and rock strength was fed in an unsupervised machine learning algorithm for the early yet objective identification of lithofacies.
Six different lithofacies were identified in the 37m core interval, ranging from coarse clean sand to fine grain shaly sands, including limited thin calcite cemented layers.
In less than 24 hours following the testing campaign, a detailed sequence of rock facies with associated properties including reservoir quality index was provided to sedimentologists and core analysis specialists.
Such a detailed and comprehensive knowledge of the distributions of core properties, available under one unique format for all discipline, eases interdisciplinary work and significantly improves existing core analysis standards. It also provided a sound basis to train AI rock reservoir property predictors linking well-log data to well established core-based lithofacies signatures.

References

[1] J. W. Cahn and R. L. Fullman, “On the use of Lineal Analysis for Obtaining Particle Size Distribution Functions in Opaue Samples,” *Trans. AIME 206, 610,* 1956.

[2] C. Germay, T. Richard, C. Lindsay and H. T. Woo, “SPE-174086-PA The Continuous Scratch Profile: A High Resolution Strength Log for Geo-Mechanical and Petro-Physical Characterization of Rocks,” *SPE Reservoir Evaluation & Engineering-Formation Evaluation*, publication pending.

[3] T. Richard, F. Dagrain, E. Poyol and E. Detournay, “Rock Strength determination from scratch tests,” *Engineering Geology*, pp. 91-100, 2012.

[4] C. Germay, L. Perneder and T. Lhomme, “Combining the analysis of ultra-high resolution images with continuous direct measurements to identify rock types,” in *IRIS Rock Imaging SUmmitt*, Aberdeen, 2020.

[5] C. Germay, “Technical Talk: Epslog Engineering: Early Trans-Disciplinary High Resolution Core Logs To Steer The Core Analysis Workflow,” 2nd December 2020. [Online]. Available: https://www.afes.org.uk/wp-content/uploads/2020/12/EPSLOG_GENERAL-CORE-DNA-AFES-2020.pdf.