

# Machine Learning Approaches for Predicting Grain Size Distribution from Well Logs

## Abstract

The distribution of grain sizes, or the distribution of particle sizes within a reservoir, is an important parameter for the petroleum industry. This can help determine areas that have the best production potential, including optimizing the well completion and production strategies. Other purposes for grain size distribution are formation evaluation and production optimization, which also contributes to further optimizing production and minimizing reservoir damage. This paper presents a study that demonstrates the feasibility of using machine learning to predict grain size distribution from well logs using a Random Forest algorithm. Training of the algorithm was performed using well logs that had a high correlation with the weighted average grain size calculated per depth interval for the whole well intervals. The model achieved an  $R^2$  score of 0.89, which is a promising starting point for future improvements. This paper also highlights the use of statistical methods to improve the accuracy of machine learning algorithms. One of these methods calculates the average value of the adjacent depth interval and reduces the impact of undetected outliers.

**Keywords:** Random Forest algorithm, optimization, reservoir characteristics, grain size distribution, reservoir damage.

## 1. Introduction

Well logs are an essential tool in the analysis of subsurface formations and correlating well log characteristics with grain size distribution (GSD) is important, because integrating GSD analysis with other petrophysical parameters is an effective way to evaluate hydrocarbon reservoirs. Permeability is directly correlated to average grain size in clastic reservoirs, so data on grain size, porosity, sorting, packing, and grain shape can be used to estimate permeability [1]. Several research studies have been conducted on the relationship between well logs and GSD, for example, a case study of a traditional open hole hazardous sand showed that Nuclear Magnetic Resonance (NMR) analysis could predict the pore size distribution [2]. Gamma ray logs are one of the most used logs to evaluate grain size information in subsurface environments and it is believed that gamma ray log values are inversely correlated with grain size in sediments, especially for sandy deposits [3]. In addition, studies have shown that the percentage of fine depositional components have a logarithmic relationship with the average size of the grains in these depositional environments [4]. However, using a linear relationship for interpretation can lead to an incorrect interpretation of siltstones or mudstones. [5]. A more recent finding showed that using a neural

network architecture is better than the multiple regression analysis method for predicting the median grain size of low-permeability sandstone [6]. Other studies have shown that multivariate statistical methods through factor and cluster analysis can indirectly influence sediment grain-size distribution in a basin area in Antarctica [7]. The research deduced that physical properties and lithological units are more clearly and reliably differentiated using cluster analysis, while factor analysis helps identify basic geological controls affecting the measured petrophysical properties. In this regard low permeability zones or clay particles in sandstones can lead to fines migration [8]. It is important to note that density logs will differ for each formation, and results may be inaccurate if wellbore conditions are not considered. For example, dolomitic limestones can cause problems due to having mixed grain densities, requiring careful analysis [9]. Effective machine learning techniques have been successfully used for well logging data prediction, considering the temporal and spatial aspects of the data [10]. Some machine learning models have been developed to predict soil particle size fractions using isometric log ratio transformation [11]. Su et al. (2020) demonstrated that the K-Nearest Neighbors (KNN) algorithm performed better in predicting grain size than other common machine learning algorithms [12]. Different machine learning techniques and statistical methods can help improve the accuracy of predictions. Future research should continue to explore these relationships to better understand hydrocarbon reservoirs. Outliers were identified due to abrupt variations in certain log responses; outliers are data points that deviate significantly from the rest of the data. Several methods for identifying outliers or missing values can be achieved by graphical visualization such as scatter plots and the Z-score methods, therefore, detecting and handling outliers is crucial in data analysis. One of the methods used to detect outliers is Z-score analysis. The Z-score represents the number of standard deviations above or below the mean that each value falls. Aguinis et al. [13] distinguish three types of outliers: low, high, and mixed outliers. A commonly used criterion based on the modified Z-score considers observations with an absolute value greater than 3.0 as potential outliers [14]. However, the use of Z-score-based outlier criteria should be treated with caution when the data are not normally distributed or when the distribution is skewed [15]. Therefore, potential outliers were further assessed using the Interquartile Range (IQR) method and visual inspection through boxplots and scatter plots to verify their significance before inclusion in the analysis. The primary objective of this paper is to develop and evaluate a machine learning model that accurately predicts grain size distribution from well logging data. This paper focuses on a confidential dataset comprising

offshore exploration data and highlights the predictive performance of the Random Forest algorithm. Random Forest algorithm is preferred over KNN due to its ability to oversee high-dimensional data, robustness to noise, and scalability to large datasets. It also provides feature importance insights and does not require feature scaling, making it a more flexible and accurate choice for this machine learning task. Machine learning-based prediction can eliminate the need to take expensive core samples and make better use of the available well logging data.

## ٢. Methodology and data collection

The well log data from ten wells were chosen based on their strong correlation with the weighted average grain size calculated per depth intervals. This means that among multiple well logs available, only those that showed a significant statistical relationship with grain size were considered. The methodology employed in this study follows a structured approach to ensure the reliability and validity of the findings. A quantitative research design was adopted to analyze PSD data obtained from Brunei Shell Petroleum (BSP). Data collection took place over time, adhering to ethical standards set by BSP. Data Extraction. For easy access, the logs are extracted from the main file into a single file. The well logs are 'porosity (POR), deep resistivity (TRDE), shallow resistivity (TRGR), and bulk density (DEN\_E).' For training purposes, statistics such as the mean, logarithmic value, and variance are calculated. These three statistical measures were selected because they are dependable in providing a quantitative approach to characterize the grain size distribution. Mean values are calculated using adjacent values, this method ensures that any potential fluctuations in the data are minimized. This method also provides a more representative mean value, compared to calculating the average of the entire data set **Figure ١**.

```
1 # Load data
2 data = merge_df[log_columns]
3
4 # Calculate statistical measures for each adjacent row
5 features = {}
6 for column in log_columns:
7     features[column] = pd.DataFrame()
8     features[column][column] = data[column]
9     features[column]['mean_' + column] = data[column].rolling(window=2).mean().shift(-1)
10    features[column]['log_' + column] = np.log(data[column])
11    features[column]['variance_' + column] = data[column].rolling(window=2).var().shift(-1)
12
```

**Figure ١**. Python code is used to compute descriptive statistics.

## ۳. Machine Learning – Random Forest Algorithm

Machine Learning (ML) is a branch of artificial intelligence (AI) focused on building systems that can learn from data, identify patterns, and make decisions with minimal human intervention. In other words, machine learning algorithms improve their performance on a task as they are exposed to more data. There are several types and methods in machine learning, each suited to various kinds of problems and tasks. Among them, Random Forest algorithm is a powerful and versatile machine learning technique used for both classification and regression tasks. It is an ensemble learning method, meaning it combines the predictions of several models to improve accuracy and reduce overfitting. Random Forest is built on the concept of decision trees, where multiple decision trees are trained and combined to make more reliable predictions. Random forest classifies each attribute and uses a voting method. This algorithm was implemented using Python scikit-learn library. The Hyper parameter was acquired by searching for a grid with a cross-validation in a training set, where the lowest error is expected. Predictive algorithms, particularly in machine learning, require a dataset to train them.

## ۴. Results and discussion

### ۴.۱ Data analysis

Well logging data often include porosity (POR), permeability (PERM), sonic travel time (DT), true resistivity (RTRU), true resistivity conditioned (TRCN), true resistivity deep (TRDE), gamma ray total (TRGR), compressional sonic travel time (DTC\_E), shear sonic travel time (DTS\_E), and density error (DEN\_E). One of the issues that can affect data interpretation is the presence of outliers, which may influence the performance and accuracy of statistical analyses and machine-learning models. Outliers are data points that deviate significantly from the remaining dataset. The Z-score is a useful tool for standardizing and comparing logging data because it can help identify outliers and analyze measurements with different units or scales, as shown in Figure ۳. However, while the Z-score is a powerful statistical method, it may not always be sufficient on its own. Therefore, potential outliers were further assessed using the Interquartile Range (IQR) method and

visual inspection through boxplots and scatter plots to verify their significance before inclusion in the analysis.

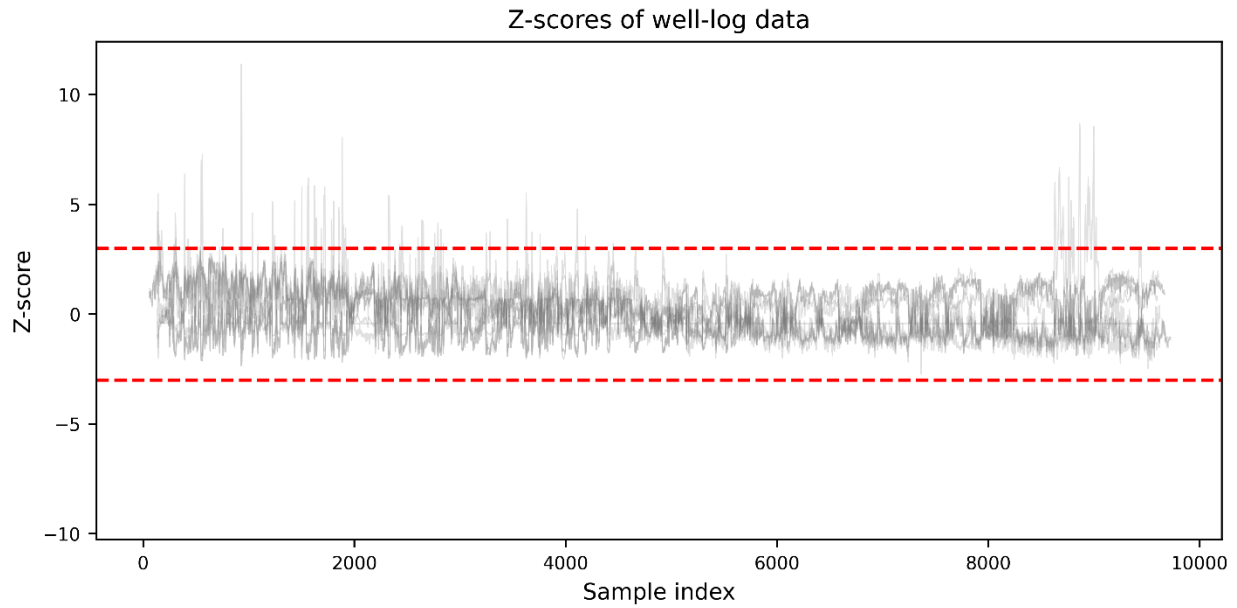


**Figure 3.** Indicates Z-score of well logging data with dotted red line as a threshold for outliers.

In this methodology, the Z-score is used to identify potential outliers in the well logged data. Then these potential outliers are analyzed using other methods to confirm their presence and assess their significance. **Figures 3** and **4** present the well-logging data values represented as a scatter plot. The red dots, in contrast to the blue dots, represent outliers detected.

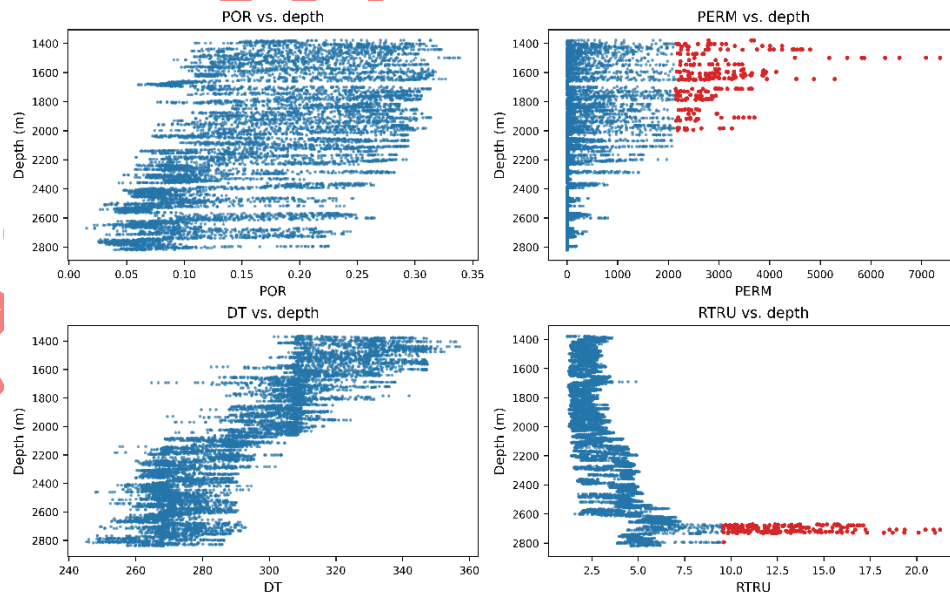


**Figure 4.** Outliers of the well logs POR, PERM, DT and RTRU are shown in red color.

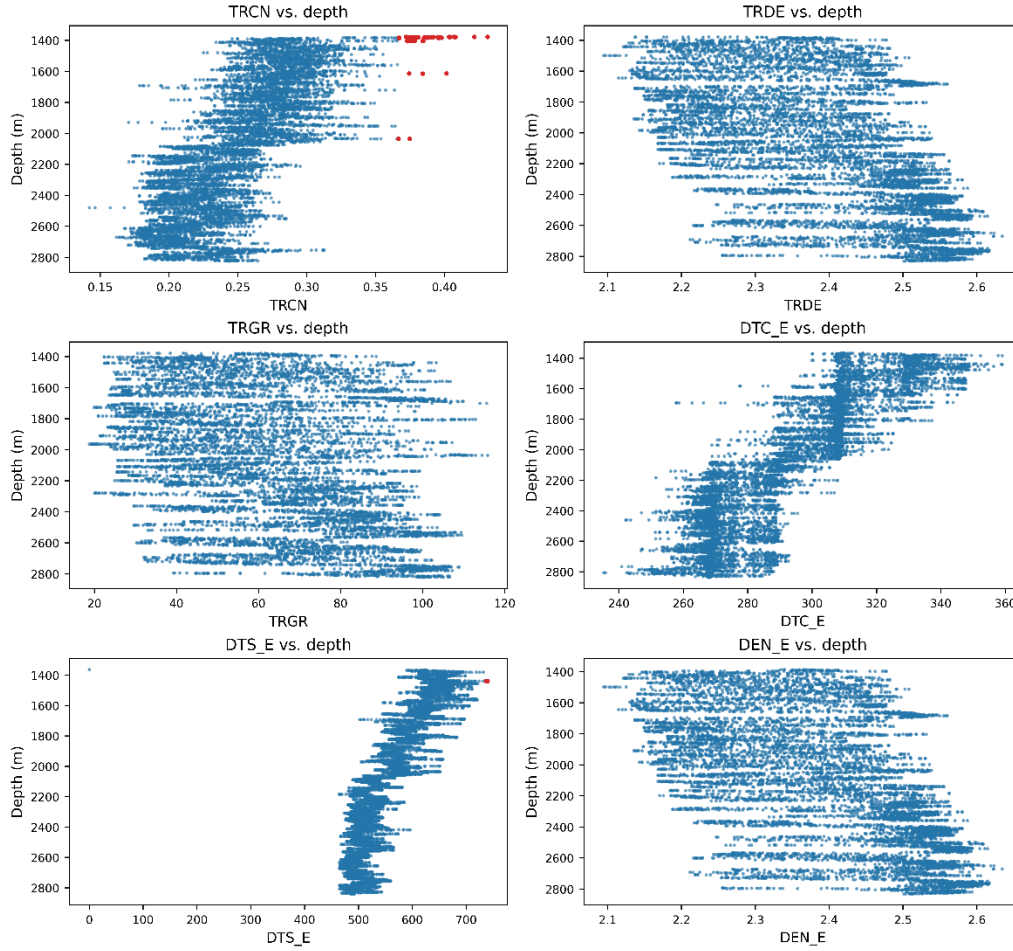


**Figure 4.** Outliers of well logs TRCN, TRDE, TRGR, DTC\_E, DTS\_E and DEN\_E shown in red color.

When an outlier is detected, there are several methods to mitigate its impact. Common methods of dealing with outliers include transforming or removing them entirely. It is important to consider that each outlier can accurately represent the data before ignoring it. To analyze well logging data as shown in **Figure 4**, outliers are managed through statistical modeling. The parameters used are mean, logarithm of the mean and variance. The average is calculated by using adjacent data values from each well log.

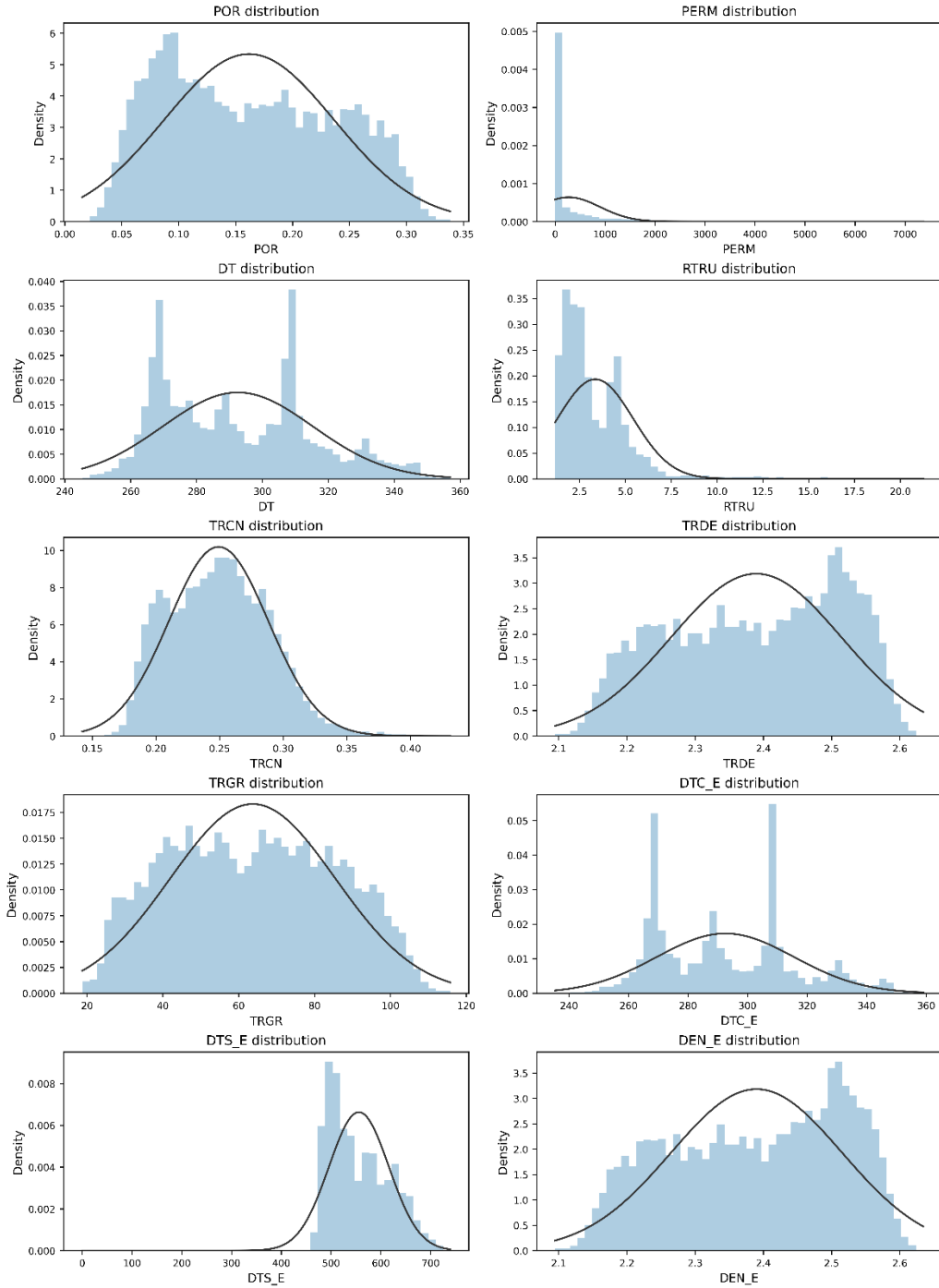


**Figure 9.** Frequency distribution with density curve of logging data from ten wells.

### 9.2 Analyzing Data for Particle Size Distribution

The dataset for the PSD was provided along with the raw dataset from before. The PSD dataset is known to have been obtained using a sieve shaker and the zones are identified solely by their depth. It should be noted that the data only represents a small portion from each well, and to visualize it,

first we need to plot a cumulative distribution curve. **Figure 7** presents the plot for all grain sizes available (gravel, very coarse sand, coarse sand, medium sand, fine sand, very fine sand, coarse silt, medium silt, fine silt, very fine silt, and clay). The frequency of various grain sizes in the samples is displayed in the cumulative grain size distribution plot. A well-sorted distribution is shown by the steep slopes, but a mix of grain sizes is suggested by the gradual curves. The curves' fluctuation indicates that sediment deposition is heterogeneous, due to shifting depositional or environmental conditions.

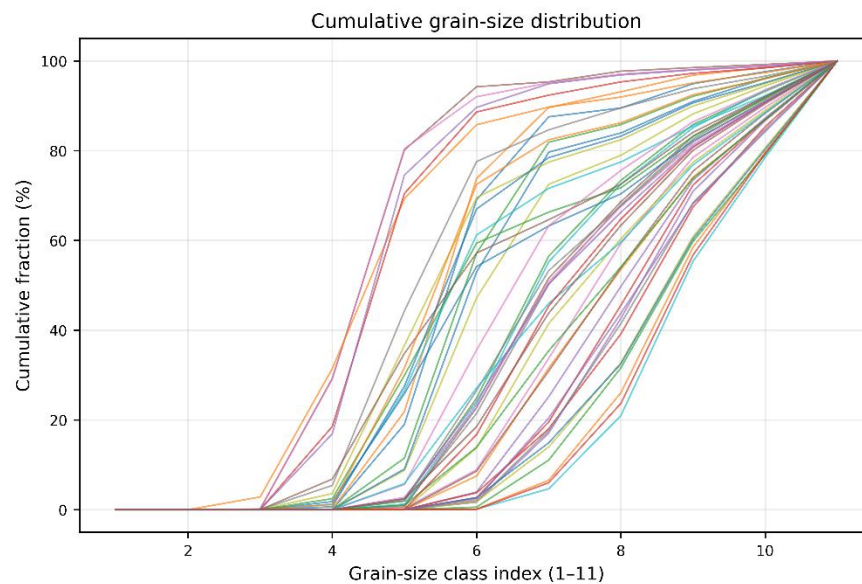


**Figure 7:** Grain size distribution against the cumulative frequency.

**Figure 8** below represents the symmetric distribution of the upper grains and shows the depth according to the ratio of particle size. The colors represent each grain size, and the graph also shows the availability of each grain size at each depth. Depth values are taken from the center of each sample and are displaced in ascending order. The grain size distribution chart shows the variation in sediment composition with depth. It indicates a mix of sand, silt, and clay fractions, with sand (yellow, brown, and red shades) dominating at certain depths and finer sediments (pink, purple, and blue shades) being more prevalent at others. This suggests changes in depositional environments, due to fluctuating energy conditions over time.

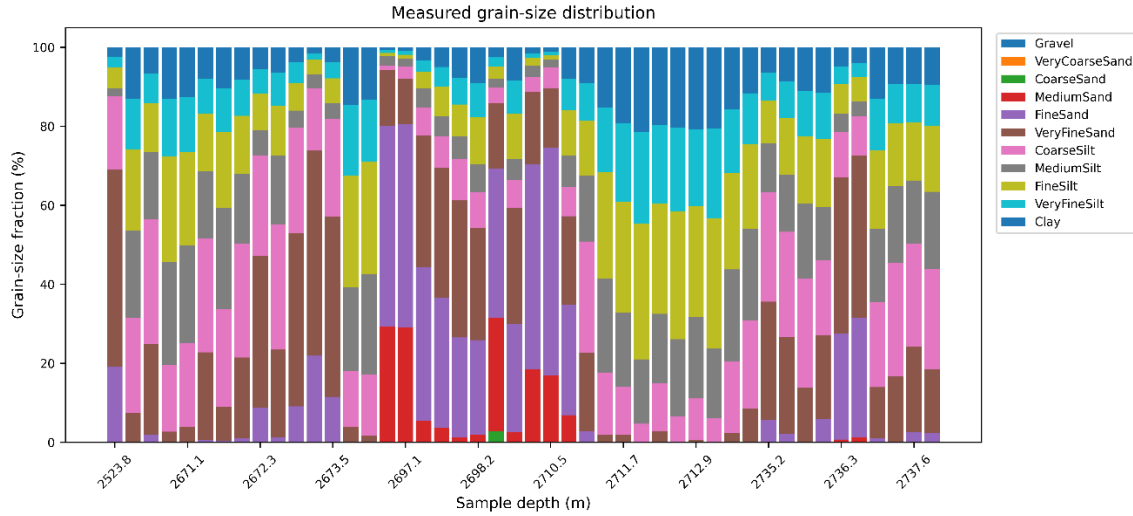


**Figure 9:** A histogram of the particle size distribution of the sample.

### 4.3 Correlation Analysis

There are a total of nine well logs, and each well was calculated using the weighted average grain size equation. The correlations used are: 'POR' log vs weighted average grain size calculated per depth interval, 'DT' log vs weighted average grain size calculated per depth interval, 'RTRU' log vs weighted average grain size calculated per depth interval, 'TRCN' log vs weighted average grain size calculated per depth interval, 'TRDE' log vs weighted average grain size calculated per depth interval, 'TRGR' log vs weighted average grain size calculated per depth interval, 'DTC\_E' log vs weighted average grain size calculated per depth interval, 'DTS\_E' log vs weighted average grain size calculated per depth interval, and 'DEN\_E' log vs weighted average grain size calculated per depth interval. Pearson correlation coefficient is used to find possible linear correlation, while Spearman correlation is used to find non-linear correlation. Variables are used to find the correlation between well logs and weighted average of the grain sizes. Plots for both methods are shown in **Figures 8** and **9**. The scatter plots in the picture show how several geophysical characteristics relate to grain size, with trend lines showing patterns of correlation. Certain measures, such as acoustic transit time (DT) and porosity (POR), exhibit a negative relationship with grain size, indicating that these characteristics decline with increasing grain size. On the other hand, there is a positive correlation between other factors like TRDE and DTS, suggesting that greater levels of these characteristics are linked to larger grain sizes. A few plots show erratic or weak trends, suggesting that there is little relationship between some parameters and grain size. Numerous factors, including mineral composition, compaction, and diagenetic processes, may

influence these correlations, as indicated by the general variety in correlation strength and direction. The Pearson and Spearman correlation coefficients between the well-log parameters and weighted average grain size are presented in **Table 1**. The results indicate varying degrees of correlation, with porosity (POR) exhibiting a negative correlation, while deep resistivity (TRDE), gamma ray (TRGR), and bulk density (DEN\_E) show positive correlations with grain size. Based on these relationships, POR, TRDE, TRGR, and DEN\_E were selected as input variables for the Random Forest prediction model. The remaining well-log parameters exhibited comparatively weaker correlations and were therefore excluded from the predictive analysis.

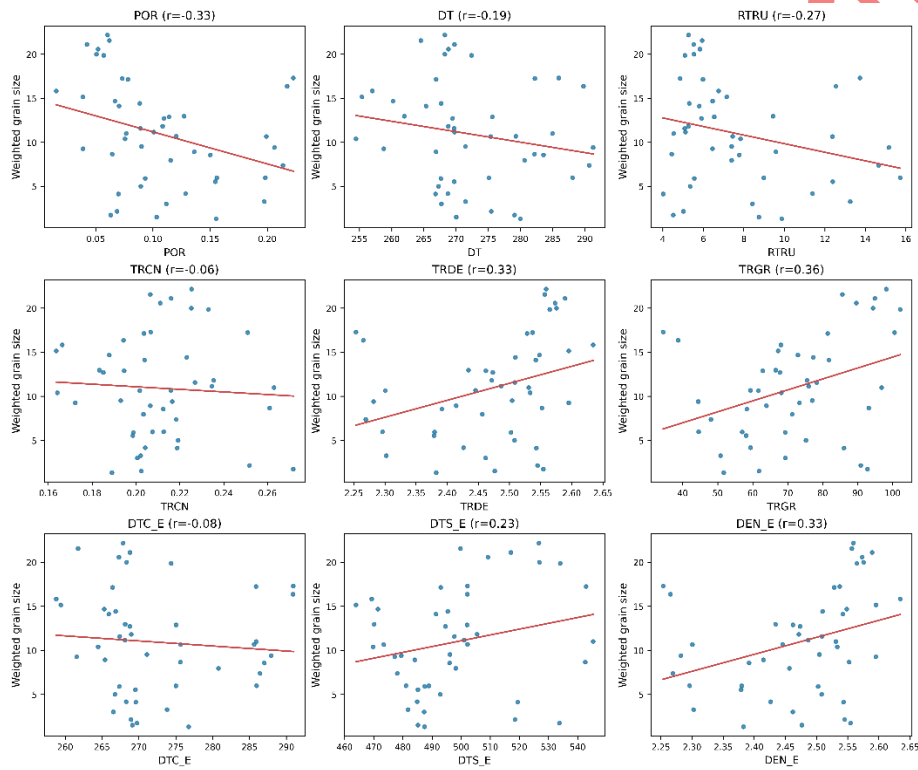


**Figure 1:** Scatter plots showing the relationship between grain size and various geophysical parameters, with trend lines indicating overall Pearson correlations.

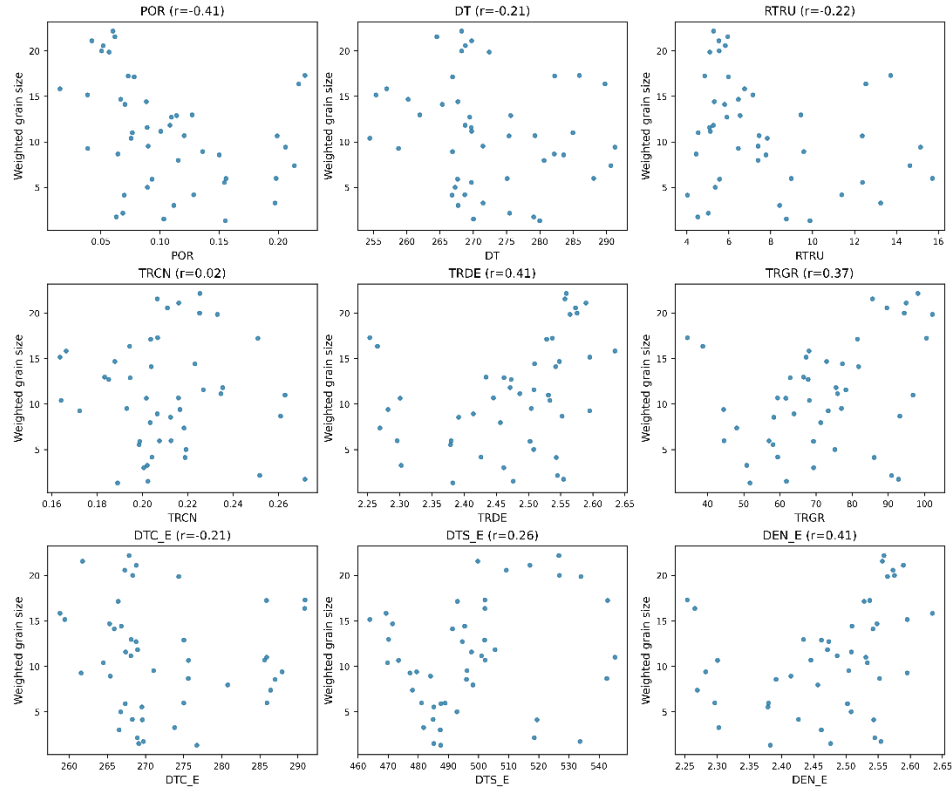


**Figure 9:** Scatter plots showing nine individual relationships between grain size and the corresponding geophysical log parameters, with units included for all axes (grain size in mm; gamma ray in API; deep and shallow resistivity in ohm·m; porosity in %; bulk density in g/cm<sup>3</sup>; and transit time in μs/ft).

**Table 1:** Pearson and Spearman correlation coefficients for various well logs.

| Well log | Pearson correlation coefficient | Spearman correlation coefficient |
|----------|---------------------------------|----------------------------------|
| POR      | -0.334                          | -0.406                           |
| DT       | -0.180                          | -0.207                           |
| RTRU     | -0.266                          | -0.223                           |
| TRCN     | -0.061                          | 0.018                            |
| TRDE     | 0.332                           | 0.410                            |
| TRGR     | 0.309                           | 0.367                            |
| DTC_E    | -0.084                          | -0.206                           |

|       |       |       |
|-------|-------|-------|
| DTS_E | 0.234 | 0.206 |
| DEN_E | 0.332 | 0.410 |

### 4.4 Random Forest application

Random forest classifies each attribute and uses a voting method. This algorithm was implemented using Python scikit-learn library. The Hyper parameter was acquired by searching for a grid with a cross-validation in a training set, where the lowest error is expected. Predictive algorithms, particularly in machine learning, require a dataset to train them. These data are extracted and inputted into the training algorithm. The model achieved an  $R^2$  score of 0.89, indicating that approximately 89% of the variability in the weighted average grain size is explained by the selected well-log inputs. It also implies that the model can capture the underlying patterns in the training data to a considerable extent. The well log data is extracted from the “UTB\_1” well logs at different depths. As new well log data is generated, the trained algorithm is used to predict grain size values. The predicted data is printed in a tabular format for better understanding. Due to the way its statistics are calculated which involves ‘shift (-1)’ method, the last row has no value and is therefore ignored. Another approach is to calculate a weighted average and compare the statistical measurements to the trained values. The expected result is that these measurements are close to the range of the trained data. **Figure 10** shows two stacked bar graphs of the original and predicted particle sizes. The trends of both distributions are similar, indicating that the prediction model

successfully captures the principal grain-size distribution patterns and provides reliable estimates of particle-size proportions.

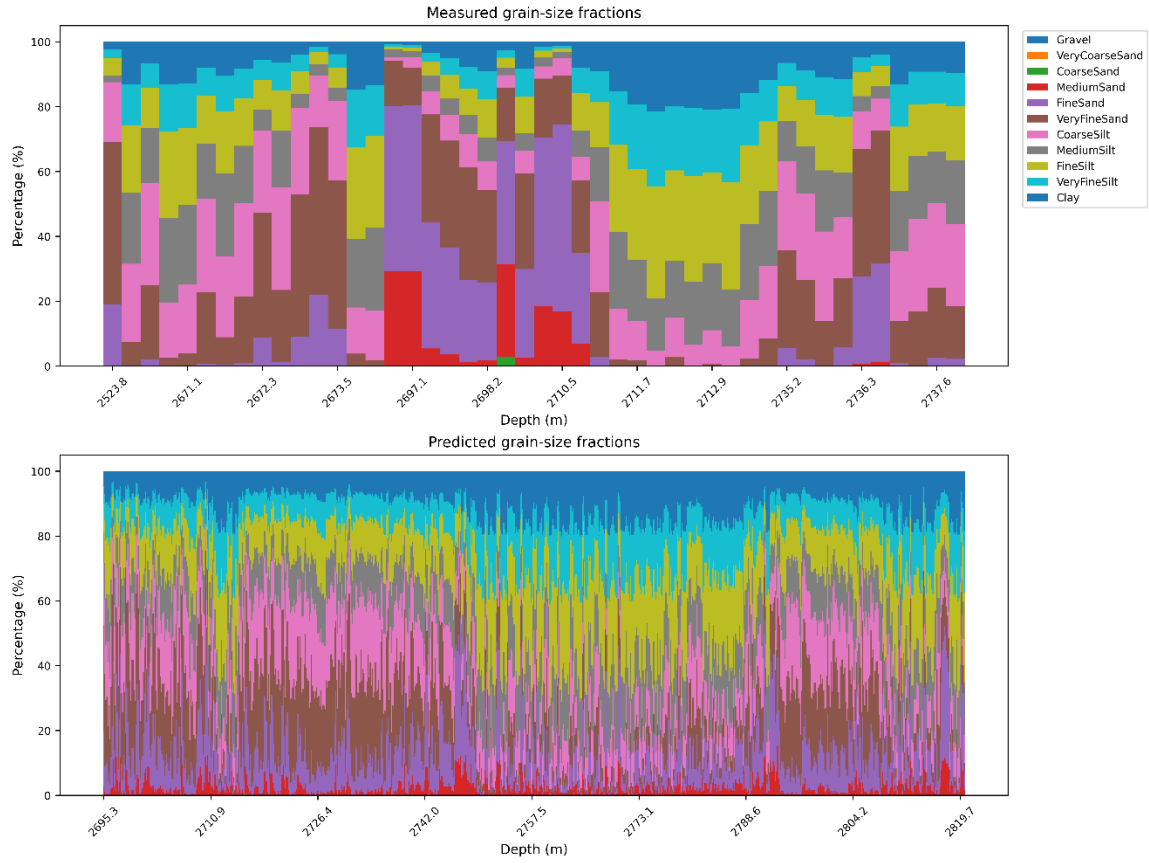


**Figure 10:** trained (top) and predicted (bottom) grain size values in a stacked bar chart.

A more accurate way to evaluate the reliability of the predicted grain-size values is to examine their cumulative distribution behavior. This confirms that the predicted grain sizes follow a systematic trend rather than random scatter, producing the expected cumulative response as shown in **Figure 11**. The figure presents the cumulative grain-size distribution (CSD) curves for multiple depth intervals, where each curve reflects the progressive increase in cumulative frequency (%) with grain size. The spread and shape of the curves highlight the diversity of sediment characteristics across the formation. Intervals with a sharp rise in cumulative percentage indicate dominant grain fractions and better sorting, while more gradual transitions imply mixed or poorly sorted sediment populations. Differences among the curves reveal variations in sediment composition, depositional processes, and potential changes in sediment source or transport mechanisms.

Overall, the cumulative distribution trends demonstrate that the predicted grain-size values exhibit geologically meaningful patterns consistent with natural sediment variability. These outputs, generated through machine-learning modeling, were exported into an Excel file (“Predicted\_Grain\_Size.xlsx”) to improve accessibility and user convenience, with representative values presented in Appendix B.

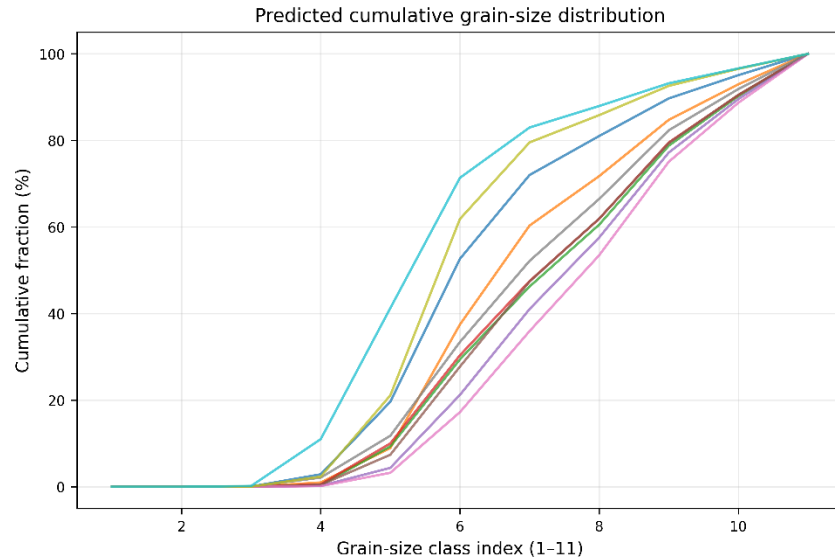


**Figure 11:** Cumulative Distribution Function (CDF) of Grain Size (%) for Selected Depth Intervals.

#### 9. Conclusion

In conclusion, predicting particle size distribution from well logs using machine learning has proven to be a promising method for application in the oil and gas industry, despite the challenges posed by limited data availability due to confidentiality restrictions and time constraint. The forecast was applied using a random forest algorithm as a type of machine learning. Several data analysis techniques were used to ensure the accuracy of the forecast. These methods include identification and handling outliers of well logs to minimize the impact on finding the correlation between well log and grain size distribution. Another technique employed was to use a weighted grain size instead of the mean or median while training the algorithm. Additionally, statistics were used on the trained well log values to predict new grain size values, improving model accuracy by up to 89%. To ensure the legitimacy of the results, the project team has employed data visualization techniques such as using graphs to represent the data in a clear and concise manner.

## Funding

This research received no specific grant from funding agencies in the public, commercial, or non-profit sectors.

## Declaration of Competing Interest

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

## Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) solely for English language editing, grammatical corrections, and improving the clarity and readability of the text. The research methodology, data analysis, interpretation of results, and scientific conclusions were independently developed by the authors. All AI-assisted edits were carefully reviewed and verified by the authors, who take full responsibility for the content and scientific integrity of the manuscript.

## References

- [1] Masumura, R., Mikada, H., & Takekawa, J. (2019). The relation between permeability and grain size distribution. *\*Proceedings of the 11st EAGE Conference & Exhibition\**, 1–5. <https://doi.org/10.3997/2302-8265.2019.0242>.
- [2] Anifowose, F. A., Mezghani, M. M., & Shahrani, S. S. (2020). Indirect estimation of elastic reservoir rock grain size from wireline logs using a supervised nearest neighbor algorithm. *\*SPE Middle East Oil and Gas Show and Conference\**. <https://doi.org/10.2118/20106-MS>.
- [3] Clarke, S. M., Pettigrew, R., Priddy, C., & Clarke, S. M. (2024). Outcrop spectral gamma-ray signals of distal arid continental basins: Facies identification and controls. *\*Marine and Petroleum Geology\**, 157, 107047. <https://doi.org/10.1016/j.marpetgeo.2024.107047>.
- [4] Chen, L., Ji, H., Qiu, N., & Zhang, L. (2021). A conversion for average grain size based on gamma ray well log: Application in the Second Member of Xujiahe Formation of Anyue Area, Sichuan Basin. *\*Frontiers in Physics\**, 9, 807303. <https://doi.org/10.3389/fphy.2021.807303>.
- [5] Fan, H. (2022). Quantitative prediction of low-permeability sandstone grain size based on conventional logging data by deep neural network-based BP algorithm. *\*Geofluids\**, 2022, 7498449. <https://doi.org/10.1155/2022/7498449>.

- [٦] Sazali, Y., Gödeke, S., Sazali, W. L., Ibrahim, J. M., Graham, G. M., Kidd, S. L., & Ohen, H. A. (٢٠١٨). Fines migration during CO<sub>2</sub>-saturated brine flow in carbonate reservoirs with migratory clay minerals: Malaysian formations. *\*SPE International Symposium on Formation Damage Control\**. <https://doi.org/10.2118/189069-MS>.
- [٧] Bücker, C. J. (٢٠٠١). Statistical analysis of wireline logging data of CRP-٣ drillhole. *\*Terra Antarctica\**, ٨(٤), ٤٩١–٥٠٥. <https://epic.awi.de/id/eprint/٢٧٤٠٠/>.
- [٨] Edwards, J. M. (١٩٥٩). Quantitative evaluation of the density log in the Rocky Mountain area. *\*Journal of Petroleum Technology\**, ١١(٤), ٢٩–٣٤. <https://doi.org/10.2118/1229-G>.
- [٩] Wu, L., Dong, Z., Li, W., Jing, C., & Qu, B. (٢٠٢١). Well-logging prediction based on hybrid neural network model. *\*Energies\**, ١٤(٢٤), ٨٥٨٣. <https://doi.org/10.3390/en14248083>.
- [١٠] Martinelli, G., & Gasser, M.-O. (٢٠٢١). Machine learning models for predicting soil particle size fractions from routine soil analyses in Quebec. *\*Soil Science Society of America Journal\**, ٨٥(٣), ٩٦٢–٩٧٦.
- [١١] Su, Q. (٢٠٢٠). Sedimentary environment prediction of grain-size data based on a machine learning approach. *\*Interpretation\**, ٨(٣), ١–٦. <https://doi.org/10.1190/INT-2019-0103.1>.
- [١٢] Aguinis, H., Gottfredson, R. K., & Joo, H. (٢٠١٣). Best practice recommendations for defining, identifying, and handling outliers. *\*Organizational Research Methods\**, ١٦(٢), ٢٧٠–٣٠١. <https://doi.org/10.1177/1049124112470848>.
- [١٣] Engineering Statistics Handbook. (n.d.). Detection of outliers. NIST. <https://www.itl.nist.gov/div898/handbook/eda/section3/eda30h.htm>.
- [١٤] Verkoeijen, P. P. J. L., Polak, M. G., Bouwmeester, S., Vazire, S., & Savalei, V. (٢٠٢١). A practical illustration of methods to deal with potential outliers: A multiverse outlier analysis. *\*Collabra: Psychology\**, ٧(١), ٨. <https://doi.org/10.1020/collabra.11.8>.

## Appendix:

### Appendix A: Statistical Measures

| POR    | mean_POR | log_POR   | variance_POR | TRDE   | mean_TRDE | log_TRDE | variance_TRDE | TRGR    | mean_TRGR | log_TRGR | variance_TRGR | DEN_E  | mean_ |
|--------|----------|-----------|--------------|--------|-----------|----------|---------------|---------|-----------|----------|---------------|--------|-------|
| 0.0697 | 0.06820  | -2.663555 | 4.500000e-06 | 2.5432 | 2.54570   | 0.933423 | 1.250000e-05  | 86.1105 | 79.51650  | 4.455631 | 86.961672     | 2.5432 |       |
| 0.0667 | 0.05270  | -2.707550 | 3.920000e-04 | 2.5482 | 2.57170   | 0.935387 | 1.104500e-03  | 72.9225 | 73.14645  | 4.289397 | 0.100307      | 2.5482 |       |
| 0.0387 | 0.02700  | -3.251916 | 2.737800e-04 | 2.5952 | 2.61475   | 0.953664 | 7.644050e-04  | 73.3704 | 70.70930  | 4.295521 | 14.162906     | 2.5952 |       |
| 0.0153 | 0.02695  | -4.179902 | 2.714450e-04 | 2.6343 | 2.61480   | 0.968617 | 7.605000e-04  | 68.0482 | 67.65830  | 4.220216 | 0.304044      | 2.6343 |       |
| 0.0386 | 0.05705  | -3.254503 | 6.808050e-04 | 2.5953 | 2.56425   | 0.953702 | 1.928205e-03  | 67.2684 | 67.72225  | 4.208691 | 0.411960      | 2.5953 |       |
| 0.0755 | 0.10135  | -2.583623 | 1.336445e-03 | 2.5332 | 2.48365   | 0.929483 | 4.910405e-03  | 68.1761 | 67.36285  | 4.222094 | 1.322751      | 2.5332 |       |
| 0.1272 | 0.16315  | -2.061995 | 2.584805e-03 | 2.4341 | 2.36735   | 0.889577 | 8.911125e-03  | 66.5496 | 64.09395  | 4.197948 | 12.060434     | 2.4341 |       |
| 0.1991 | 0.20640  | -1.613948 | 1.065800e-04 | 2.3006 | 2.28490   | 0.833170 | 4.929800e-04  | 61.6383 | 54.84405  | 4.121283 | 92.323666     | 2.3006 |       |
| 0.2137 | 0.20995  | -1.543182 | 2.812500e-05 | 2.2692 | 2.27560   | 0.819427 | 8.192000e-05  | 48.0498 | 46.23775  | 3.872238 | 6.567050      | 2.2692 |       |
| 0.2062 | 0.20220  | -1.578909 | 3.200000e-05 | 2.2820 | 2.28910   | 0.825052 | 1.008200e-04  | 44.4257 | 44.47175  | 3.793818 | 0.004241      | 2.2820 |       |
| 0.1982 | 0.19780  | -1.618479 | 3.200000e-07 | 2.2962 | 2.29930   | 0.831256 | 1.922000e-05  | 44.5178 | 47.67305  | 3.795889 | 19.911205     | 2.2962 |       |
| 0.1974 | 0.17595  | -1.622523 | 9.202050e-04 | 2.3024 | 2.34065   | 0.833952 | 2.926125e-03  | 50.8283 | 54.45965  | 3.928453 | 26.373406     | 2.3024 |       |
| 0.1545 | 0.18860  | -1.867561 | 2.325620e-03 | 2.3789 | 2.31630   | 0.866638 | 7.837520e-03  | 58.0910 | 46.24090  | 4.062011 | 280.849740    | 2.3789 |       |
| 0.2227 | 0.22000  | -1.501930 | 1.458000e-05 | 2.2537 | 2.25955   | 0.812573 | 6.844500e-05  | 34.3908 | 36.54455  | 3.537789 | 9.277278      | 2.2537 |       |

### Appendix B: Predicted Grain Size

| Depth     | Gravel | VeryCoarseSand | CoarseSand | MediumSand | FineSand  | VeryFineSand | CoarseSilt | MediumSilt | FineSilt  | VeryFineSilt | Clay      |
|-----------|--------|----------------|------------|------------|-----------|--------------|------------|------------|-----------|--------------|-----------|
| 2695.3464 | 0.0    | 0.0            | 0.064634   | 2.813592   | 16.757426 | 33.017219    | 19.356051  | 9.004611   | 8.686298  | 5.376334     | 4.923835  |
| 2695.4988 | 0.0    | 0.0            | 0.031502   | 0.969043   | 7.928912  | 28.549869    | 22.846994  | 11.422809  | 13.000021 | 8.218100     | 7.032750  |
| 2695.6512 | 0.0    | 0.0            | 0.000064   | 0.383512   | 8.932625  | 20.100051    | 16.828736  | 14.291084  | 18.244265 | 11.458767    | 9.760897  |
| 2695.8036 | 0.0    | 0.0            | 0.000064   | 0.564315   | 9.451282  | 20.354636    | 17.134509  | 14.396857  | 17.625008 | 10.953151    | 9.520179  |
| 2695.9560 | 0.0    | 0.0            | 0.000132   | 0.217981   | 4.151240  | 16.851102    | 19.819982  | 16.534458  | 19.623691 | 12.352748    | 10.448666 |
| ...       | ...    | ...            | ...        | ...        | ...       | ...          | ...        | ...        | ...       | ...          | ...       |
| 2819.7048 | 0.0    | 0.0            | 0.001165   | 2.582739   | 8.480003  | 3.627283     | 9.621153   | 15.975351  | 25.319551 | 17.473578    | 16.919177 |
| 2819.8572 | 0.0    | 0.0            | 0.005575   | 11.054567  | 38.133420 | 13.525039    | 7.791775   | 7.510797   | 9.379498  | 6.286616     | 6.312714  |
| 2820.0096 | 0.0    | 0.0            | 0.004969   | 7.874803   | 26.452277 | 11.694827    | 9.512404   | 11.162589  | 14.406751 | 9.542930     | 9.348451  |
| 2820.1620 | 0.0    | 0.0            | 0.004036   | 7.328803   | 24.187772 | 8.552579     | 8.195653   | 11.483575  | 17.111565 | 11.707560    | 11.428457 |
| 2820.3144 | 0.0    | 0.0            | 0.000577   | 1.531751   | 5.506033  | 2.866071     | 8.619132   | 16.254994  | 27.781686 | 19.077251    | 18.362506 |

821 rows × 11 columns

### Appendix C: Grain-Size Fraction Consistency Check

The predicted grain-size fractions were checked to confirm that their percentages sum to 100% for each depth interval. This check assesses the compositional consistency of the predicted grain-size distributions.