Politecnico di Torino

Master of Science in Petroleum and mining engineering
A.a. 2021/2022
Sessione di Laurea Novembre 2022.

PETROPHYSICS-DRIVEN 3D DENSITY MODELLING

Relatori:

Dr. Ivan Pires de Vasconcelos
Professor Laura Socco
David Vargas

Candidati:

Dmitry Bublik

ABSTRACT

Subsurface density model building has always been an essential yet challenging topic due to the complexity of compositional and stratigraphic features exhibited in the underground domain, which limits the accuracy of conventional inversion methods in reconstructing 3D mass density maps. Borehole data alone are generally not sufficient for the retrieval of a proper 3D model since in the majority of cases subsurface properties are distributed irregularly. Additionally, most structural features present moderate to large acoustic impedance (the product of density and seismic velocity) contrast that extend laterally. Given the large uncertainty associated with such models, it is common in applied seismology to use a constant or very smooth density model to make the seismic processing computations cheaper and faster at the cost of accuracy. We developed a methodology that introduces several criteria based on borehole and seismic data to retrieve 3D density models. This framework is applied to one of the Norwegian oilfields and aims at becoming a reliable, cost-efficient, and time-saving tool that can be used with little to no experience.

All data used in this study are available and contain the composite well logs, a 3D seismic velocity cube from the open-source Volve data set, and a synthetic 2D velocity slice retrieved by post-stack seismic inversion. Data visualization and analysis are done with the help of dedicated Python libraries, Lasio and Segyio. The first density model was obtained using the standard Brocher and Gardner empirical equations. We then proposed data-driven modified relationships that accommodate well logs to build density models and finally compare against standard Brocher and Gardner-based models. The empirical relationships to be modified are selected based on the applicable P-wave velocity ranges corresponding to the values provided by the 3D velocity model.

An essential part of this work relies on the systematic assessment of well-data outliers. In this regard, the values from the acoustic compression slowness log (AC) are converted to velocity values, that are consequently used to obtain the synthetic density curves utilizing the selected empirical equations. Log curves demonstrated the presence of anomalous values within the data set, proving the importance of quality checks and outlier detection

and removal. Therefore, several methods are applied to identify and eliminate inconsistent values.

Once outliers are identified and removed, the clean data set is used to apply model parameterization for the aforementioned empirical relationships using the Python programming language and the SciPy curve fit method in particular. The procedure was carried out for the entire depth interval and formation tops extracted from the well discovery report and grouped by similar properties into six sections. As a result, we were able to extract new coefficients that were subsequently applied to modify the standard empirical equations, and then used to build new synthetic log curves. Finally, the curves are compared to the real density values by means of an absolute average misfit.

All developed methodologies are evaluated based on the results and the number of proposed criteria to select the best approach. The results are presented as 2D density slices obtained for both the real and sharp synthetic velocity models. The new reconstructed models using our data-driven approach aim to provide higher accuracy, reliability, and efficiency for future scientific research and computations.

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisors Dr Ivan Pires de Vasconcelos and David Vargas at Utrecht University for providing the possibility of cooperation, excellent supervision and constant support.

I would like to acknowledge my supervisor/Politecnico di Torino advisor Professor Laura Socco, who played a significant role in the selection of my primary subjects of interest, and thesis topic and shaped me into the student I am today. In addition, I am grateful to the whole DIATI department and specifically to the lecturers that I met throughout the years of studying at the University during my M.Sc. program in Petroleum and Mining Engineering.

I wish to thank Equinor for publishing the Volve open data in 2018 and creating the possibility for research to be carried out on real data.

Finally, I would like to express my eternal gratitude to my family members that have been supportive for my whole life. I am grateful to my parents and grandparents back in my home country. And I really appreciate all the help, patience and support coming from the person who became my family here.

SUMMARY

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS

1) **AC** Acoustic compression slowness (sonic log)
2) **CALI** Caliper log
3) **CMP** Common midpoint
4) **CSP** Common shot point
5) **DEN** Density log (RHOB – bulk density)
6) **bcm** Billion cubic meters of gas
7) **GP** Gaussian process
8) **GR** Gamma ray log
9) **IF** Isolation Forest
10) **LOF** Local outlier factor
11) **LWD** Logging while drilling
12) **MD** Measured depth
13) **ML** Machine learning
14) **mln bbl** Million barrels of oil
15) **MWD** Measurement-while-drilling
16) **NEU** Neutron porosity log
17) **RDEP** Deep resistivity log
18) **RMED** Medium resistivity log
19) **SVM** Support vector machines
20) **TVD ss** True vertical depth subsea
21) **UTM** Universal Transverse Mercator

1 INTRODUCTION

The main goal of the project is to develop a density modelling workflow applicable to different fields located both on- and offshore. The available data used for the current project are the velocity model and well logs provided by the Volve open data source (Volve data village 2020). In addition, there is a 2D high-resolution velocity slice obtained by means of post stack seismic inversion (Ravasi 2022).

Bulk density is the total mass of material including rock matrix and pore fluid within a defined volume. It is an essential parameter in subsurface characterization that represents a challenge for the majority of geophysical methods that aim at reconstructing detailed 3D model with high fidelity. Despite being an active subject of research in the area of geoscience, bulk density estimation methods are still a matter of debate since such structural parameters can only be poorly quantified and most of the models lead to misinterpretations due to the high number of compositional and stratigraphic uncertainties (Mcknight et al., 2020).

Densities of pure, dry, geologic materials range from 0.88 g/cm^3 for ice up to 8 g/cm^3 for some rare minerals. Rock density commonly lies within the range of 1.6 g/cm^3 for sediments to 3.5 g/cm^3 for gabbro (Sharma, 1997). Figure 1.1 shows an estimated set of density ranges for various rocks. Igneous rocks represent the heaviest, while unconsolidated sediments like alluvium or soil are the lightest.

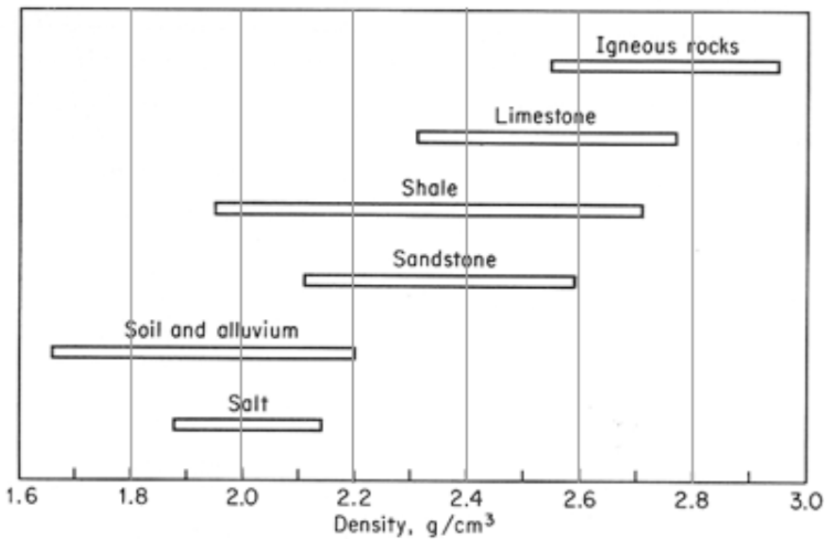

Fig. 1.1: Realistic ranges of rock densities (Sharma, 1997)

Accurate knowledge of the underground bulk density represents an essential aspect for the determination of lithologic and stratigraphic features of the subsurface as well as for the estimation of economically important rock properties in the petroleum industry (Bulhões et al., 2015).

Once density values are determined, additional rock's mechanical properties can be inferred as is the case in the following examples:

1. The bulk density is a required parameter for the conversion of acoustic velocities to dynamic elastic moduli.
2. A sufficient estimation of the vertical stress can be derived from the use of the density logs integrated over the vertical depth of the well, especially in the areas of low tectonic activities (Fjœr et al., 2008).

In the case of low tectonic activity, vertical stress is also considered to be the principal stress. Especially important is the realization that, provided density data is available, the determination of the full in situ stress field only depends on the magnitude and orientation of the horizontal stresses (Fjœr et al., 2008).

In some situations, prior knowledge of density is specifically important for seismic acquisition, gravimetric studies, and imaging. Through profiling and core samples of wells, density information is sampled spatially.

The density parameter is of utmost importance to plan the positioning of new production and injection wells. In addition, it is required to evaluate the mechanical properties of rock, such as rock matrix compressibility, Bulk modulus, Shear modulus, and Young’s modulus (Chang et al., 2006). To provide comprehensive information about lithology and pore fluid, formation density log data should be applied in combination with the neutron porosity log. Moreover, density is a necessary parameter to define overburden pressure and porosity (Yusuf et al., 2019). Integration of density logs from the surface to the depth of interest provides the possibility to estimate the magnitude of overburden pressure, pore pressure, and origin of subsurface overpressure conditions (Feng et al., 2019).

A density log is one of the logs frequently acquired in a well, along with gamma-ray, resistivity, and neutron logs. Despite the possibility for the logging tool to fail or provide inconsistent data, the values obtained reveal rock parameters within the volume very close to the well. Therefore, there are situations in which the prediction of formation bulk density is required for further estimation of parameter distribution.

In addition, due to the difficulty of retrieving such logs in large diameter boreholes, density logs are not commonly run in the top holes (Zoback, 2010). In top-hole sections, excessive washouts usually occur in the loose sediments, limiting the acquisition of density logs. Based on the need for bulk density to determine overburden pressure, synthetic logs can be predicted from additional log measurements through empirical relationships. In such a case, it is expected to include density values defined for the whole predefined depth interval. Resolving such a strategy allows compensating for the absence of values in certain regions while at the same time revealing the true log pattern. Constructing synthetic well logs carry polluted measurement that leak into the proposed data series and ultimately motivates the call for systematic outlier detection and removal. Even in cases where there are no missing values and obvious outliers can be spotted, a comparison of measured values to estimated formation density might be applied as a quality check tool. In the regions with no log data, available density can be predicted by use of the interpolation techniques (e.g., Kriging) or from P-wave velocity through the application of available empirical relationships. Kriging is a mathematical interpolation method. The technique is based on the stationary Gaussian process (GP). In terms of stationarity, the GP has a constant expected value, variance, and covariances that vary according to the gaps between points in a k-dimensional space ($k \leq 3$). In addition, these Gaussian processes were used in the context of machine learning (Kleijnen, 2009).

In addition, as was stated before bulk density can be directly estimated using empirical correlations applied in non-fractured formations, which are a rare realistic occasion (Lobkovsky et al., 1996).

The main issues related to existing empirical equations are:

1. Specified lithologies, depth, and velocity ranges for empirical relationships application. The equations are less reliable for the lithological columns that consist of several stratigraphic units.

2. The empirical relationships do not consider the presence of microcracks and fractures in the rock.

The main goal of the current project is to develop a robust, yet efficient workflow, that substitutes the practice of using constant density values, commonly used in most of the reservoir properties computations, with a high-resolution density model based on both velocity models obtained from seismic data and well log data.

We visualized both the smooth velocity model obtained through the processing of the seismic data and high-resolution 2D velocity slice retrieved by means of post stack seismic inversion. The obtained velocity range is used to select the applicable empirical relationships and provide the first models for further comparison. The well logs are preprocessed and used for modifying the empirical coefficients in the selected equations. The innovative strategy followed in the project led to the development of a data-driven framework that includes artificial intelligence algorithms to identify well-log outliers and proposes new empirical relationships to build accurate density models. The workflow is implemented using the python programing language and follows a modular design that facilitates the integration with available libraries. The methodology includes multiple steps and techniques composing different combinations of methods to be compared and evaluated by several criteria.

The final results are provided as a singular 2D density slice for the high-resolution velocity model obtained by post stack seismic inversion (Ravasi 2022) and a smooth 2D density slice developed from the velocity model retrieved from the Volve data set. Nevertheless, a 3D density model can be formed by applying the approach to each of the slices within the provided data set.

2 METHOD AND DATA

In the current chapter, we provided an overview of the general information about the Volve field, available data and the related theoretical concepts. We outline the main elements of the workflow depicted in Figure 2.1.

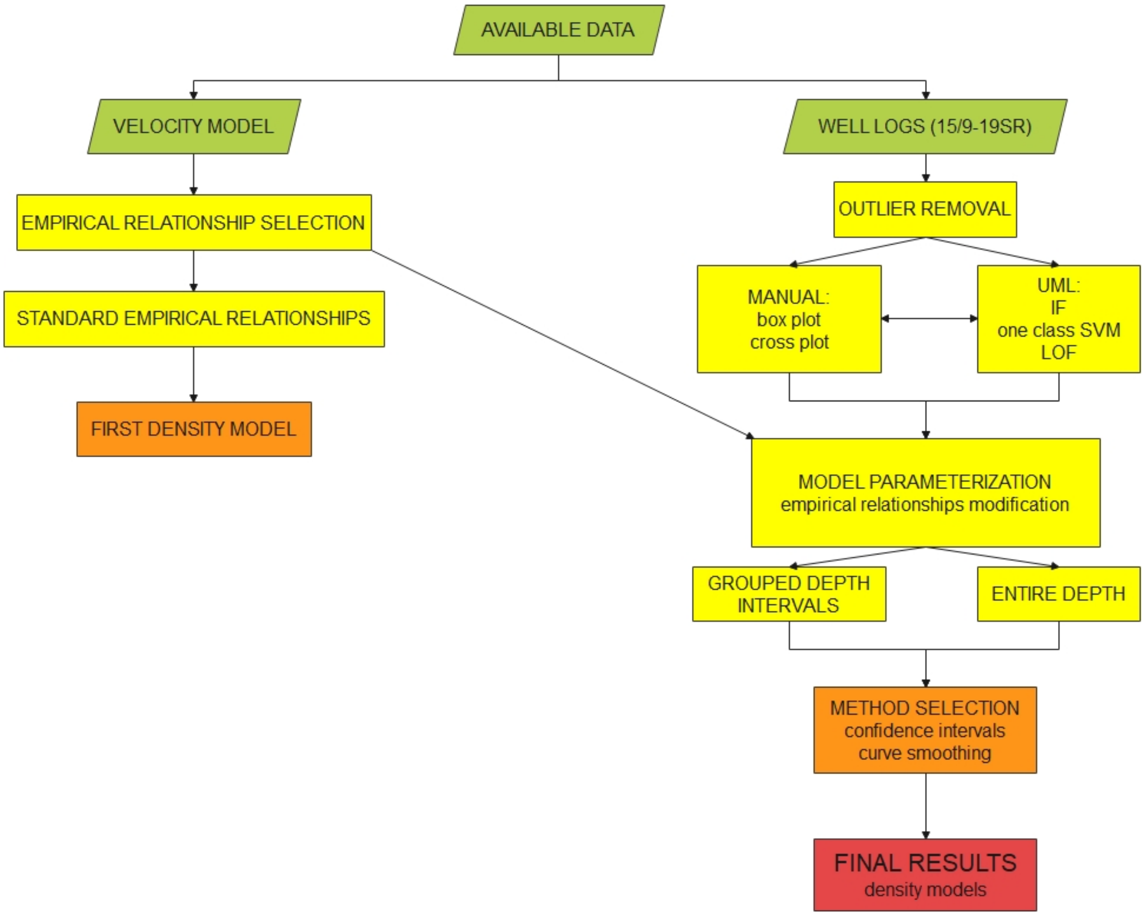

Fig. 2.1: Project workflow

We first start by collecting the **available data** from the provided repository; then the data are reviewed and grouped. On one hand, the available **velocity models** are used together with **selected empirical relationships** to build a **first density model**. On the other hand, well logs undergo a series of pre-processing steps to identify and **remove outliers**. Then, **model parameterization** is carried out by fitting the selected equations to the real data and modifying the empirical coefficients. Afterwards, based on a number of criteria the

most beneficial **combination of methods is selected** and applied to reconstruct the optimal high-resolution **density model** that represents the **final results**. In the following sections, we provide detailed descriptions of each module of the workflow.
The application of the workflow to the available data will be shown in Chapter 3.

2.1 Available data set

2.1.1 General information about the Volve field

In 1993, the Volve field was discovered in the central part of the North Sea, through the successful prospection of well 15/9-19SR (Fig. 2.4). It was the first well to be drilled into the Theta Vest structure. The exact location of Theta Vest is shown in Fig 2.3. The location of the Volve field (Fig 2.2) is 200 km west of Stavanger and 8 km from the Sleipner Ost field. The water depth in the region is about 80 m, and it has been established that the field was formed by the collapse of adjacent salt ridges (Volve data village 2020).

Fig. 2.2: Location of the Volve field (Wang et al., 2021)

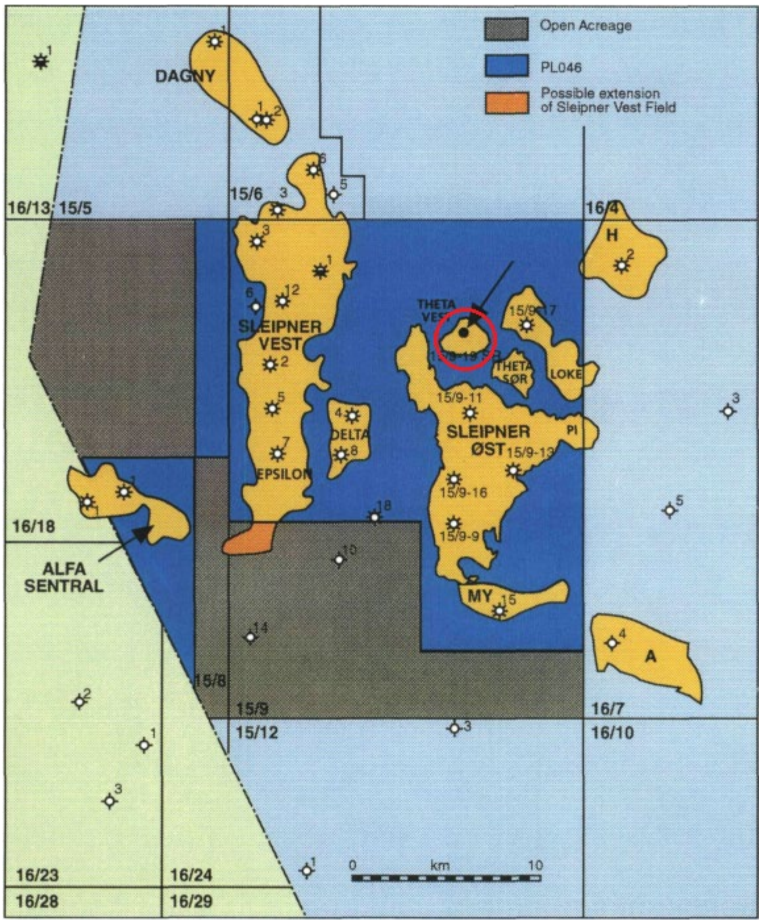

Fig. 2.3: Location of the Theta Vest (Volve data village, 2020)

Drilling was put into operation using the world's largest jack-up rig, at the moment, called "Maersk Inspirer", which is equipped with its wellhead and process module for production. It was supposed to be used and reopened for exploitation of the Respol-operated Yme field, following the production shutdown of the Volve field. As the main oil storage vessel, FSU Navion Saga was used (Volve data village 2020).

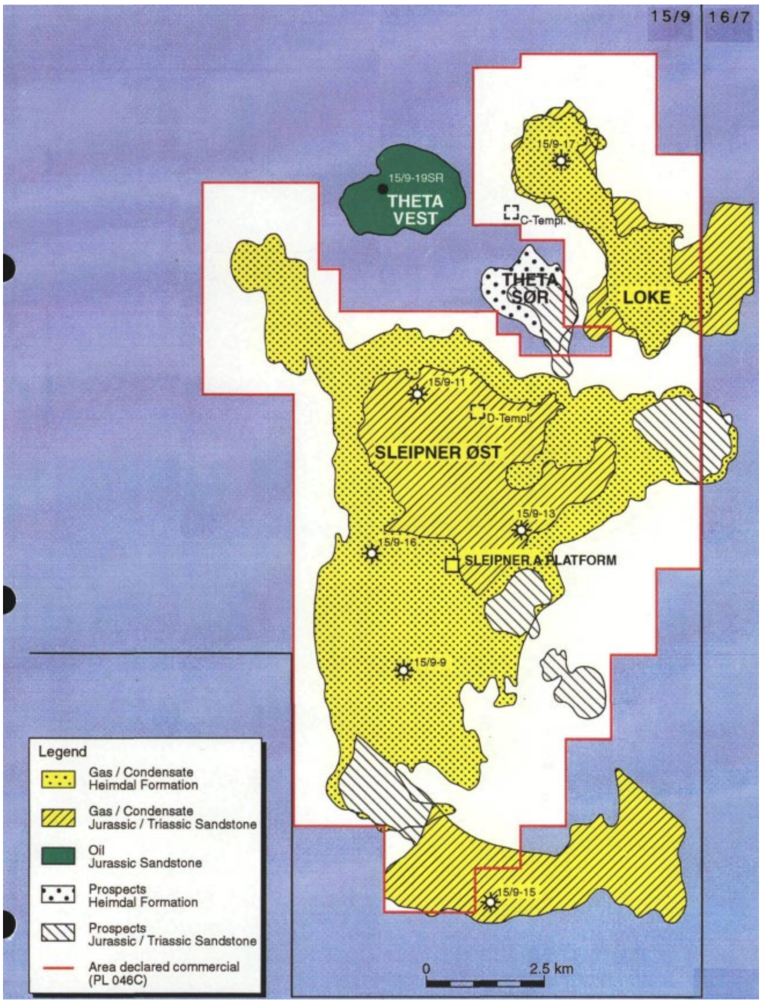

Fig. 2.4: Location of the well 15/9-19SR (Volve data village, 2020)

The Volve field was produced using water injection to support the pressure in the reservoir, which consists of Jurassic sandstones at 2750 - 3120 m TVD ss. Likewise, drilling operations were initiated in May 2007, followed by production starting the next year with an expected life cycle of 3-5 years.

A significant increase in both recovery factor and field life expectancy was achieved thanks to a new set of wells put into exploitation in the period up to 2013. However, the drop in oil prices in recent years made clear that new wells were no longer profitable due to the limited remaining resources.

Closed in 2016, the Volve oil and gas field operated for three years longer than the estimated project life cycle. During the production period, high production results were achieved by exploring different technical solutions that lead to a prolonged exploitation extension period.

The estimated recoverable resources projected were 78.6 mln bbl of oil and 1.5 bcm of natural gas, with daily production of 56000 bbl of oil a day. A sustained increment in production was achieved. In a period spanning over 8 years, it increased by around 9.5 mln bbl of oil, which by far exceeded the expectations stated in the development and operation plans. Overall, the Volve field recovery rate reached 54%, and the complete removal of subsea equipment was carried out in the summer of 2017 (The Volve data village, 2020).

2.1.2 Data set information

The approximate number of files available in the Volve field dataset repository is around 40,000. The whole dataset is open-source in order to provide scientists and students with all the realistic information available for future research and studies. The data available is shown in Fig. 2.5.

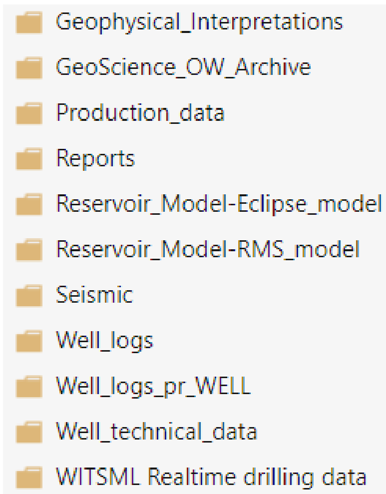

Fig. 2.5 – Volve data set (Volve data village, 2020)

Therein, one finds all geophysical interpretations as datasets containing fault polygons, faults, horizons, and well picks. Monthly production data is available per well, and the reports folder includes discovery and PUD (Plan for Utbygging og Drift) reports. Likewise, two types of reservoir models were added to the data set: the dynamic fluid simulation models produced in the Eclipse software, and the geological reservoir models made created using the RMS software.

2.2 Velocity model

As per the seismic data models, it contains a P-wave 3D velocity model (Fig 2.6). The Volve velocity model was developed by the specialists at Equinor by the processing of the available seismic data and the velocity analysis.

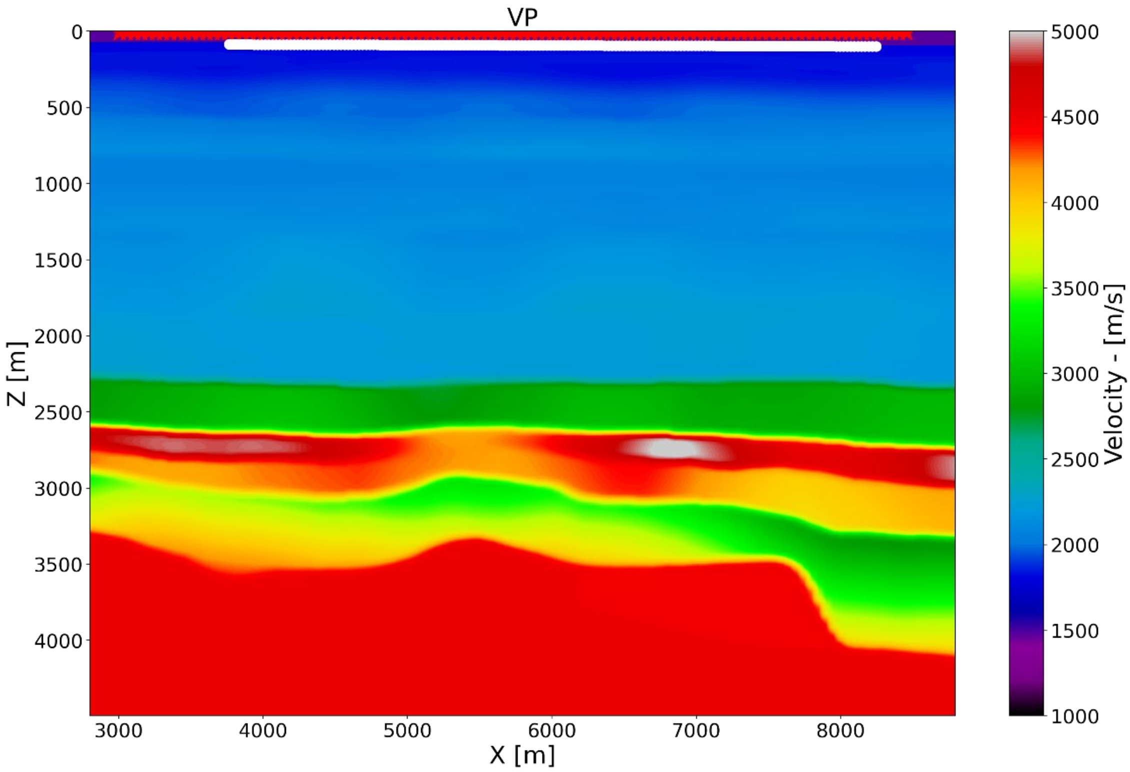

Fig. 2.6: 2D velocity slice №1 (The Volve data village, 2020)

We used the estimated 2D velocity slice retrieved by post-stack seismic inversion (Ravasi, 2022) (Fig. 2.7) to obtain the sharp density model with higher resolution. This model presents a sharper and more realistic look in terms of interface visualization and better explains the open-source data from the Volve field when used for numerical modelling. It is stated that the selected 2D slice is the closest to a 2D receiver cable that was used for modelling purposes (Ravasi, 2022). Further calibration of the model was carried out using the well log data.

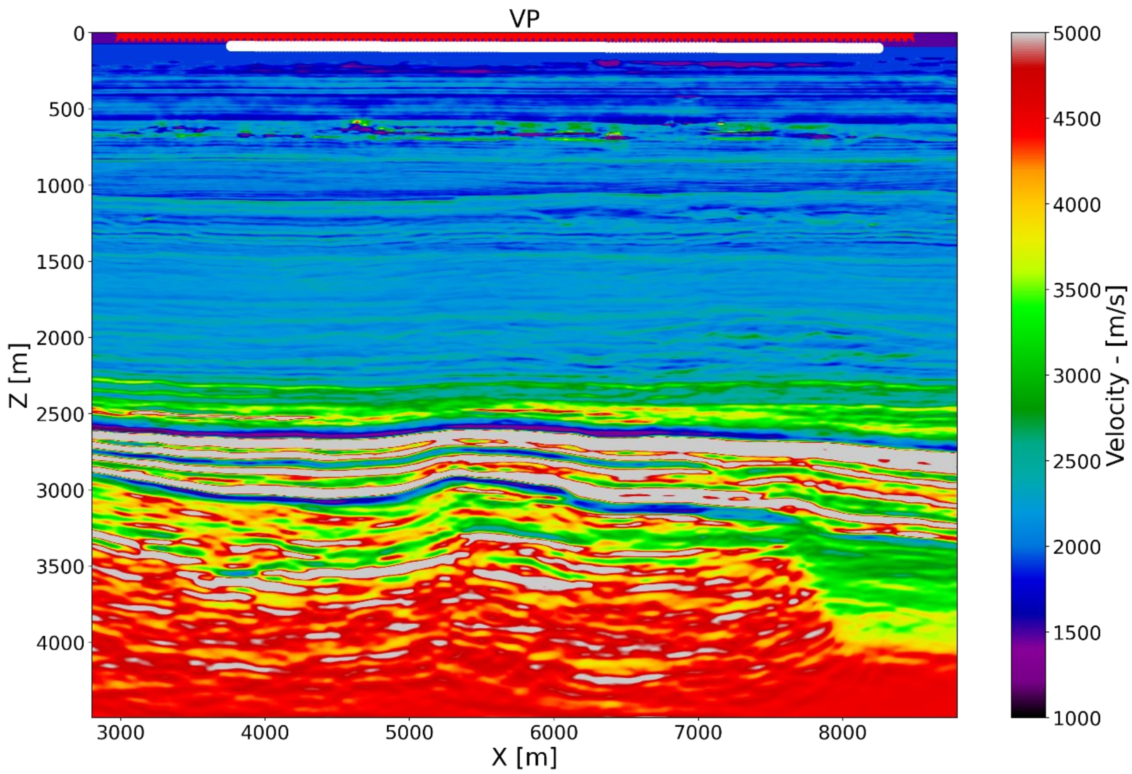

Fig. 2.7: High-resolution 2D velocity slice (Ravasi, 2022)

According to the displayed velocity 2D slice, P-wave velocities are inverted for the depth: $1.5 < V_P < 5$ km/s. Even though only one slice is depicted, values from the other 2D slices lie within the same range of velocities.

2.2.1 P-velocity – Density empirical relationship

Consequently, among the applicable methodologies we selected the application of the empirical relationships to be able to fulfil both goals of the project:

- Modify the parameters of the selected empirical relationships by fitting the equations to the real well log data.
- Apply data-driven relationships to retrieve the high-resolution density model from the velocity cube.

The empirical equations proposed by Gardner 1974 and Brocher 2005 were selected among the numerous empirical relationships (e.g., Christensen and Mooney 1995, Godfrey et al. 1997, etc.) due to the possibility to obtain realistic density values for the wide range of velocities. The trend comparison between several relationships is shown in Fig. 2.8.

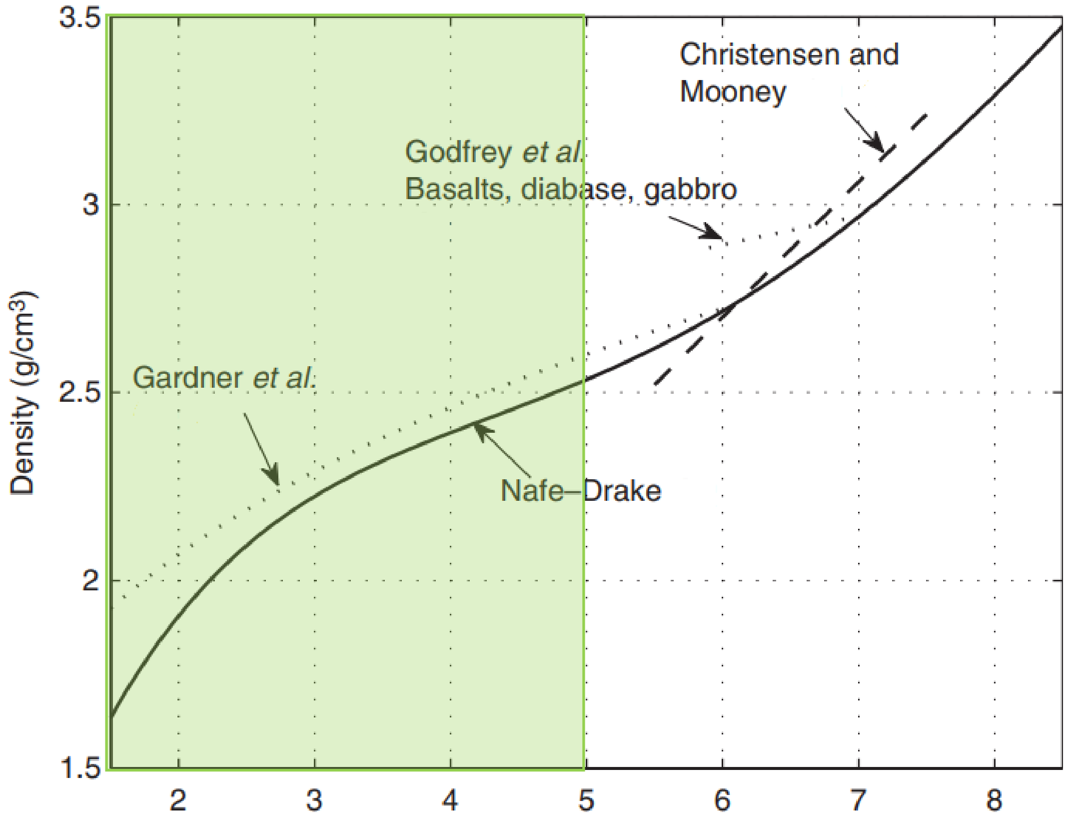

Fig. 2.8: Brocher VP – bulk density trend comparison (Brocher, 2005)

The area highlighted in green corresponds to the defined by the data velocity range: 1.5 < V_P < 5 km/s (Fig. 2.6 - 2.7).

Gardner 1974 proposed an empirical relationship (Eq. 2.1) for sedimentary rocks which exhibit high accuracy in the P-wave velocity range: 1.5 < V_P < 6.1 km/s.

$$\rho = 1.74 V_P^{0.25} \tag{2.1}$$

Where ρ is the bulk density in g/cm^3, and V_P corresponds to P-wave velocity in km/s.
In 2005, Brocher developed empirical relationships that relate P-wave velocity, S-wave velocity, bulk density, and Poisson's ratio for crustal logs based on the data coming from laboratory measurements consisting of core analysis, vertical seismic profiles (VSP), field tomography, and wireline logs. A wide variety of lithologies was tested in the study

including shales, mafics, sandstones, gabbros, rocks with high calcium content, such as anorthosites and dolomites, and other crystalline rocks (Mavko et al., 2009).

The proposed empirical equation represents a tool for crustal rock behaviour representation in a broad range of depths. A significant advancement in this regard is, for instance, the polynomial fit to the Nafe-Drake curve describing the connection between bulk density and P-wave velocity (Ludwig et al., 1970) computed by Brocher (Eq. 2.2). The proposed relation reads

$$\rho = 1.6612V_P - 0.4721V_P^2 + 0.0671V_P^3 - 0.0043V_P^4 + 0.000106V_{P,}^5 \qquad (2.2)$$

Once again, ρ is the bulk density and V_P is the acoustic wave speed. Equation 2.3 is applicable for the range of velocities: $1.5 < V_P < 8.5$ km/s and all types of crustal rocks excluding rocks rich in calcium and mafics. In addition, the inverse relation was also proposed (Mavko et al., 2009).

$$V_P = 39.128\rho - 63.064\rho^2 + 37.083\rho^3 - 9.1819\rho^4 + 0.8228\rho^5 \qquad (2.3)$$

The maximum accuracy achieved for equation (2.3) is in the range of density: $2.0 < \rho < 3.5$ g/cm^3.

2.3 Well logs

Similarly, well data consists of numerous types of logs, including composite well logs (Fig 2.9) with extensive technical and drilling data that provide information about well directions, locations, and the equipment used.

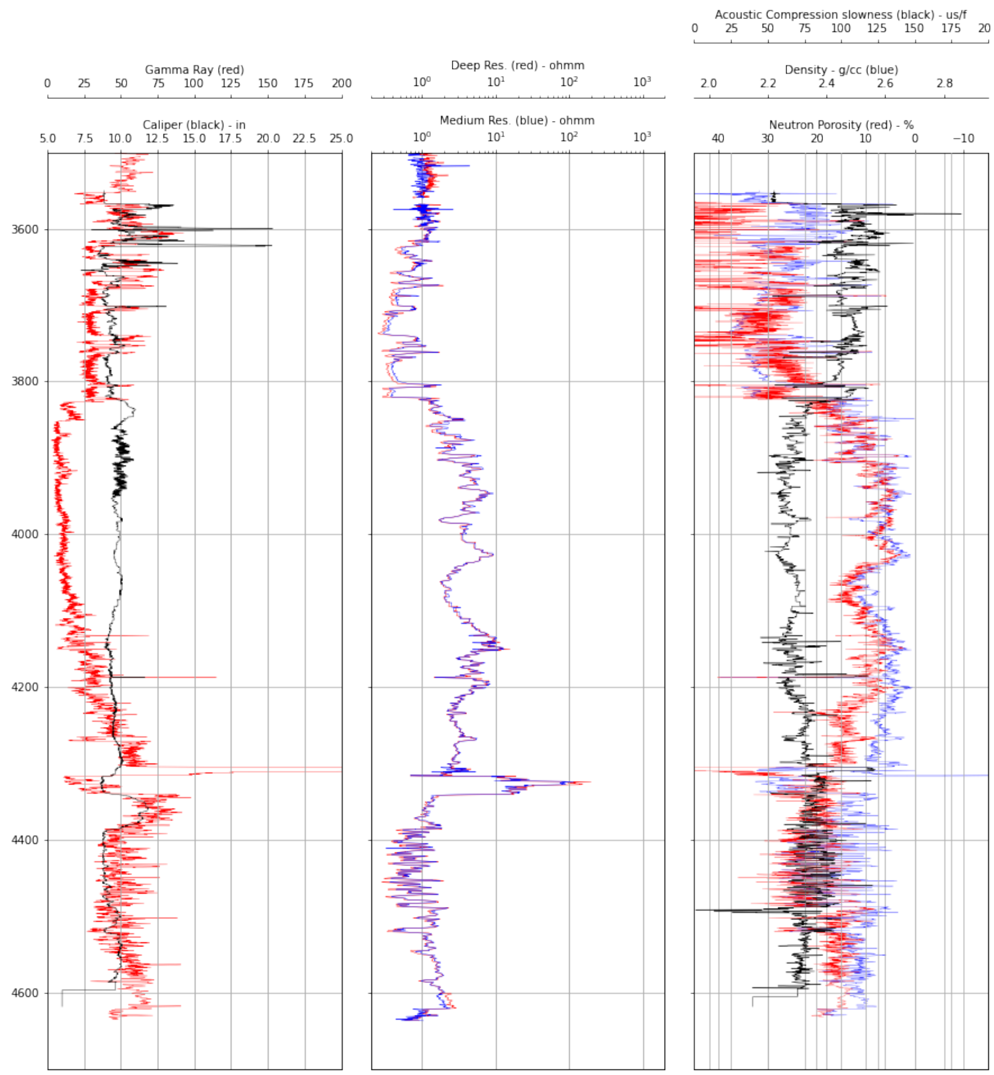

Fig. 2.9: Composite well log from well 15/9-19SR consists of Gamma ray, Caliper logs (1st column), Deep and Medium Resistivity logs (2nd column), Sonic, density, and neutron porosity logs (3rd column) (The Volve data village 2020).

In the current project, the logs of interest are **sonic, density** and finally **Caliper log** used as an input for unsupervised machine learning algorithms.

In order to record the **sonic log**, the acoustic pulses are sent through the formation around the well to reach the other end of the logging tool. The velocity is then evaluated by the travel time, also known as slowness or interval transit time. Transit times on the logs vary from 40 to 140 µs/ft. The interval transit time (t) is the reciprocal of the sonic transit velocity (v). The porosity evaluation can be estimated since the propagation velocity in

the pore fluid is significantly lower than it is in the rock matrix. The measured velocity will be more or less inversely proportional to the rock porosity (Bjørlykke et al., 2010).
The bulk density of the rock includes the values for both the solid matrix and the saturating fluid. The physical principle of the **density log** involves gamma rays from cobalt-60 or cesium-137 focusing on the formation. Such rays attenuate when interacting with the rock due to multiple collisions with electrons (Compton scattering)(Bjørlykke et al., 2010). The attenuation is then measured with a separate detector giving an indication of the rock's electron density. The fact that the electron density has a strong relation to the bulk density expressed in g/cm^3 is used to infer this structural parameter. Different lithologies can be identified as functions of their densities with the information provided from the density log (Bjørlykke et al., 2010).
The mechanical tool attached to the probe gives access to the **Caliper log** to be carried out, which is an estimation of the borehole's diameter. The instrument must be in direct contact with the walls of the well. Throughout the movement of the logging tool, the instrument repeats the shape of the borehole and records the deviations from the initial position (Bjørlykke et al., 2010).

2.4 Outlier detection and removal

Acoustic compression slowness values from the well logs shown earlier (Fig. 2.9) are converted to velocity expressed in km/s. AC log, velocity, and density are graphically presented in Fig. 2.10 to compare corresponding areas.

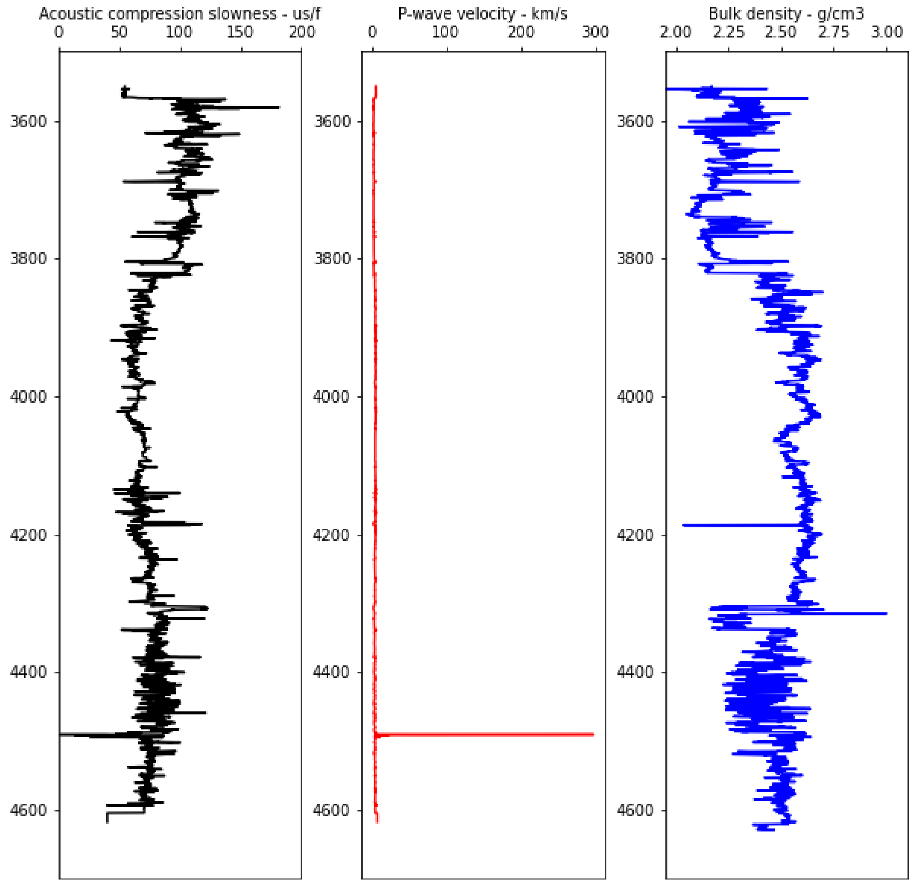

Fig. 2.10: AC, velocity, and density curves

The red line in Figure 2.10 shows an abnormal velocity curve due to the anomalous high value of up to 300 km/s, which does not correspond with realistic values. This peak is the result of inconsistent values of acoustic compression slowness, possibly caused by sensor or measurement errors. The presence of such values in the signal indicates that data needs to be “cleaned”, and thus, outliers have to be detected and removed from the data set.

In addition to applying the available migration velocity values to Brocher 2005 (Eq. 2.2) and Gardner et al. (Eq. 2.1) empirical relationships, synthetic logs have been retrieved and compared to real density logs. Comparison and mismatch analyses are shown in Fig 2.11. Here we compute the misfit with the same approach as in Equation 2.4.

$$\Delta\rho = \frac{\sqrt{(\rho_{real} - \rho_{synthetic})^2}}{\rho_{real}} * 100 \tag{2.4}$$

where:

ρ_{real} = real bulk density, g/cm^3;

$\rho_{synthetic}$ = synthetic bulk density obtained by means of the empirical equation, g/cm^3

$\Delta\rho$ = absolute percentage density difference (mismatch), %

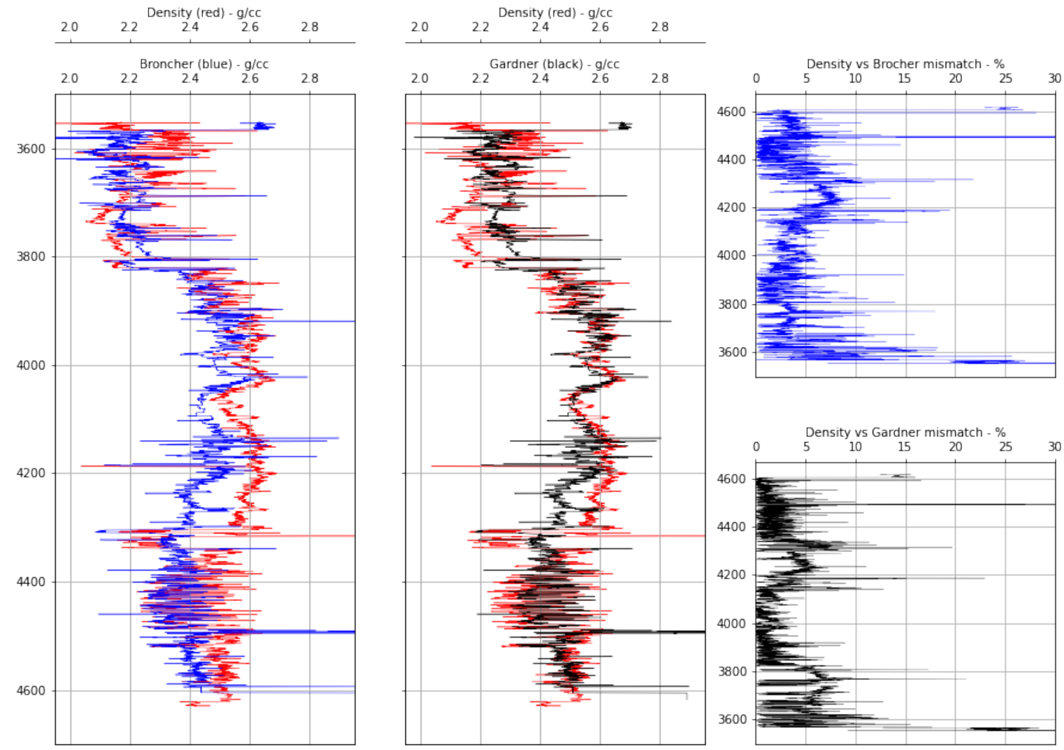

Fig. 2.11: Synthetic and real density logs comparison and mismatch analysis

The average mismatch for the Gardner density curve is 3.62%, while the value for Brocher density is inconsistent due to the outliers present in the data.

Prior to fitting selected empirical relationships to the field data, it should be pre-processed by checking for missing values, detection, and removal of the outliers. To remove outliers from the defined set of data we used the methodology proposed by McDonald 2021. We are going to remove the inconstant values are going by the use of both manual detection and removal and unsupervised machine learning techniques. In addition, the methodology with just the UML removal will be tested.

Outliers are values considered to be inconsistent within the presented data set. They do not fit according to normal or predicted statistical distributions of points or realistic ranges. There are several possible reasons for that to happen, such as poor data sampling methods, sensor malfunctions, measurement errors, and an unpredicted sequence of events (McDonald, 2021). To remove outliers from the defined set of data we used the methodology proposed by McDonald 2021.

The data collected from well logs and petrophysics measurements can show outliers associated with washed-out boreholes, tool and sensor malfunctions, rare geological features, and issues in the data acquisition procedure. It is crucial to deal with outlier detection and removal prior to using the data for further analysis. In the event that the outlier removal issue is not solved or incorrectly addressed at the beginning of the computations, the results' accuracy will be compromised. An example of this poor outlier assessment is observed in various regression analyses. For instance, linear regression models are readily influenced by outliers, which can change the equation substantially.

Detection of anomalous values can be carried out using conventional statistical techniques, such as univariate methods, which focus on a single feature or variable (z-scores, box plots, etc.), and multivariate methods, focusing on two or more features or variables (scatter plots). Identification techniques should not be applied carelessly, specifically for univariate methods, because chosen values could be existing geological features (McDonald, 2021).

Additionally, it should be considered that statistical approaches are commonly used and developed for normally distributed data. However, well log measurements can differ greatly and exhibit non-normal distributions. Therefore, any detected outlier has to be properly investigated using superb knowledge and numerous estimations before being removed or substituted. The z-score provides an estimation of how far the values are in terms of standard deviations as well as from the mean of the normally distributed data set (Eq. 2.5).

$$z = \frac{x_i - \mu}{\sigma} \tag{2.5}$$

where:

x_i = single sample point deviation;

μ = mean of the data set;

σ = standard deviation of the data set.

For a Gaussian distribution, we see that 99.7% of data points will lie within three standard deviations of the mean of the data set (normal distribution). If the value lies outside of the defined range, it confirms its difference from the other points is dramatic. In this regard, a comprehensive visual summary of data distribution can be observed through quartiles using box plots (Fig. 2.12). They provide an estimation of data skewness levels and open a possibility for outlier detection (McDonald, 2021). The interquartile range (IQR) is calculated by applying Equation 2.6.

$$IQR = Q_3 - Q_1 \tag{2.6}$$

where:

IQR = Interquartile range;

Q_3 = third quartile value (75%);

Q_1 = first quartile value (25%).

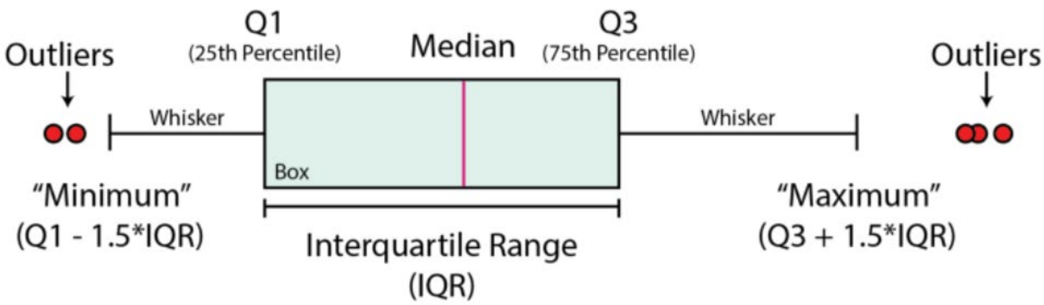

Fig. 2.12: Example of box plots (McDonald, 2021)**.**

Here, the vertically drawn lines are representations of one and a half times IQR, aimed to demonstrate data fluctuation outside of the defined boundaries. Points lying outside of the lines are considered to be outliers.

Along with box plotting, numerous other techniques can be used to detect and remove outliers from the data set, some of which are presented below:

1. Manual removal:
    a. Box plot
    b. Scatter plots
2. Unsupervised machine learning algorithms

2.4.1 Manual outlier detection and removal

Manual removal inspired by statistics can be performed through outlier identification from box and scatter plots. In addition, this process can also be driven by considering realistic rock density ranges (e.g., Fig. 1.1). In this regard, some of the points lying outside of realistic ranges of values can be removed with an extremely low risk of eliminating the existing geological features. Additionally, manual removal is applicable in the case of a relatively small number of supposed outlier values and does not guarantee a high percentage of inconsistent values being dealt with.

Long the same lines, scatter plots enable the estimation of lithological parameters and variations of pore fluid on both regional and detailed scales (Chatterjee & Paul, 2012). Cross plots demonstrate the relationship between the variables and describe their correlation by plotting one against the other. An example of the scatter plot detection of inconsistent points is shown in Fig. 2.13. The outliers can be identified according to the visual distancing from the dense clusters of data.

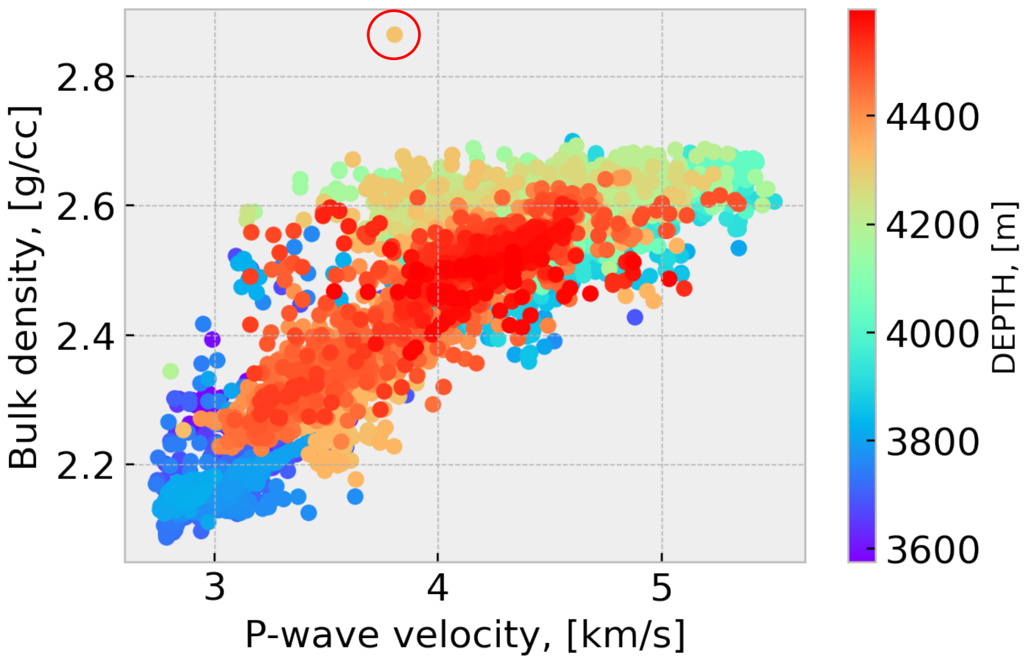

Fig. 2.13: Example of the density - P-wave velocity scatter plot (well 15/9-19SR) red circle – outlier value

2.4.2 Unsupervised machine learning algorithms

Outlier detection, commonly applied to well log and petrophysical data, can be carried out by both supervised and unsupervised machine learning techniques. The unsupervised machine learning method does not demand labelled data to define the pattern within the data set. There are various unsupervised machine learning algorithms designed for anomalous value identification within the data set. The techniques applied in the current project are presented below:

1. Isolation Forest (IF)
2. One Class SVM (SVM)
3. Local Outlier Factor (LOF)

The selection of the results will be shown in Chapter 3.

2.4.2.1 Isolation Forest

Decision trees form the basis of the isolation forest technique. In this case, the feature is selected followed by a random split of the data set between the minimum and maximum values (Fig. 2.14). The procedure continues using a decision tree for all split possibilities to be obtained. Anomalous values are separated early, which significantly simplifies the detection and outlier removal for the data set.

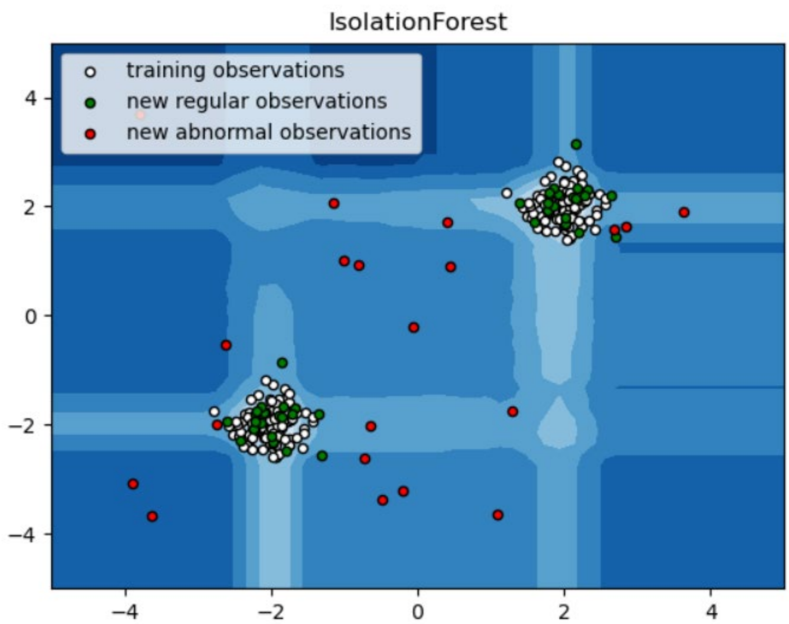

Fig. 2.14: Simple IF representation scheme (Scikit-learn)

2.4.2.2 One Class SVM

Support vector machines algorithm is mainly applicable for classification purposes, thus, capable of dividing values into various groups according to unique data features. Results are retrieved by identifying the maximum hyperplane margin between the families of data selected as shown in Fig. 2.15.

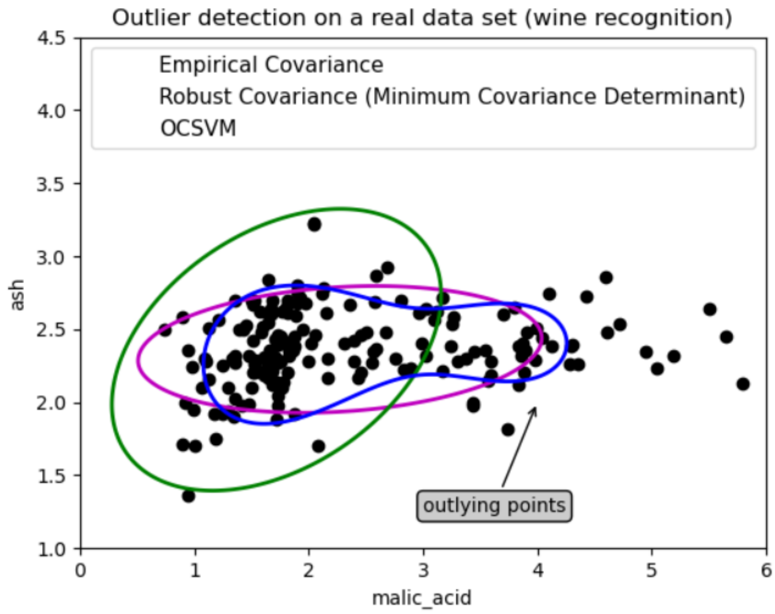

Fig. 2.15: Example of outlier removal with one class SVM (Scikit-learn)

Conventionally, SVM uses several classes or facies. Nevertheless, when only one class of data is presented, the One Class SVM algorithm can be applied. To carry out outlier detection, SVM specifies a boundary separating the data points from the origin. Once this first step is taken, the newly separated values are treated as a second-class cluster in the SVM. In this case, all values outside the boundary are defined as anomalous values. One needs to bear in mind that the parameters of the distinctive line can be adjusted by changing the portion of outliers to be identified and removed. Note that, the parameter encoding this value is called contamination level. Still, in real situations, anomalous values can be presented on any side of the main data cluster. Modifying the contamination value provides the possibility to control the desired number of outliers to be discovered (McDonald, 2021). The higher the contamination level, the higher the risk of removing important data features. Therefore, this parameter should be defined with caution.

2.4.2.3 Local Outlier Factor

The local outlier factor is designed in such a way that the algorithm defines how dense the clusters of data values are around a given point. Once this factor is established, the set of points with a low density of values around the point reference are considered to be inconsistent values and should be removed (Fig. 2.16).

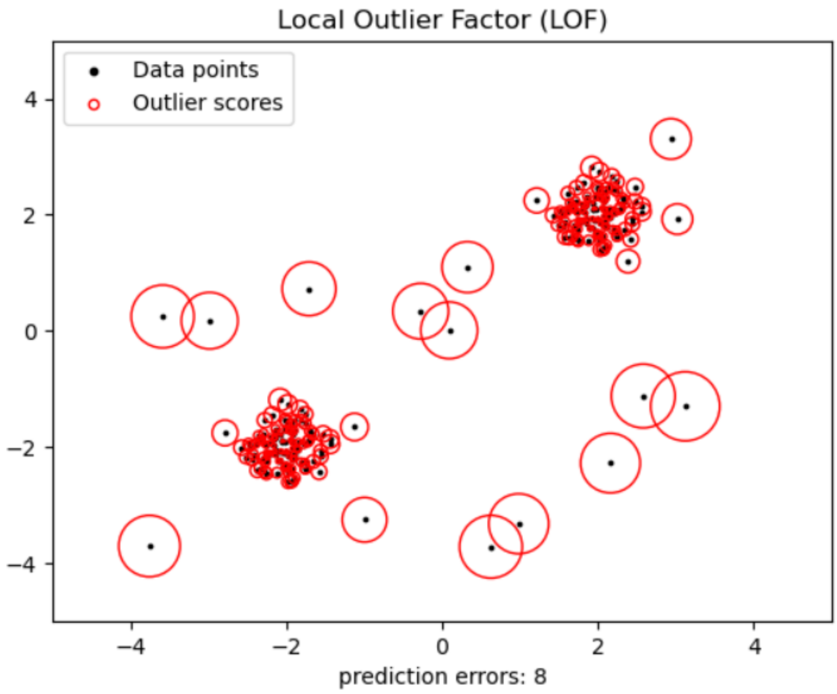

Fig. 2.16: Example of LOF outlier detection (Scikit-learn)

The effectiveness of each method can be defined by cross plotting the inliers in the data and reviewing the overall patterns and shape of the clusters. In addition, the inconsistent values can be visually defined in the log data. The number and size of the areas from which the outliers are removed can directly affect the decision.

2.5 Geological formations and stratigraphic tops

A detailed report from the well 15/9-19SR includes the exact well location, drilling results, well direction, and specifications regarding the equipment used.

was the first well to be drilled into the Theta West structure, approximately 2290 m northwest of the Loke Template, and reached the target Top Heimdal Formation at 3622.5 m MD. The top Hugin formation was encountered at 4316.5 m MD, which was 2 m TVD below prognosis. The well reached a total depth of 4541 m MD which was 3008 m in the vertical section from the Loke template, on an azimuth of 41.3 and represented penetration of the Skagerak Formation by 301m MD. The stratigraphic top from the report is provided in Table 2.1(Statoil geological report 1994).

The stratigraphic tops corresponding to the data extracted from the well logs start from a depth of around 3500 m and include all the formations lying below. Some laminations are relatively thin in comparison to others, making the processing in those areas difficult (e.g., Blodøks, Hidra, Redby, etc.). Individual formations can be grouped according to similar behaviour and parameters. That way, one can shape the general picture of the density trend.

Table 2.1: Stratigraphic tops measurements (well 15/9-19SR) (Volve data field 2020)

System	Formation	MD, m	TVD, m	TWDss, m	East UTM, m	North UTM, m
Tertiary	Nordland GP	106.0	106.0	84.6	437506.7	6477887.6
Tertiary	Utsira Fm	847.8	839.5	817.5	437448.5	6477904.2
Tertiary	Hordaland GP	1107.5	1056.3	1034.3	437312.5	6477939.7
Tertiary	Skade Fm	1313.0	1178.3	1156.3	437152.0	6477974.9
Tertiary	Grid Fm	3008.0	2062.2	2040.2	435752.8	6478326.5
Tertiary	Rogaland GP	3301.6	2228.8	2206.8	435515.7	6478374.7
Tertiary	Balder Fm	3301.6	2228.8	2206.8	435437.0	6478386.5
Tertiary	Sele Fm	3402.5	2290.0	2268.0	435437.0	6478386.4
Tertiary	Lista Fm	3483.0	2340.0	2318.0	435374.0	6478411.6
Tertiary	Heimdal Fm	3622.5	2427.0	2405.0	435265.9	6478411.6
Cretaceous	Shetland GP	3827.0	2505.6	2527.6	435085.5	6478422.1
Cretaceous	Ekofisk Fm	3827.0	2564.5	2542.5	435085.5	6478422.1
Cretaceous	Tor Fm	3850.0	2564.5	2542.5	435085.5	6478422.0
Cretaceous	Hod Fm	4046.5	2691.6	2669.6	434935.7	6478419.5
Cretaceous	Blodøks Fm	4150.0	2762.4	2740.4	434860.3	6478418.3

Cretaceous	Hidra Fm	4146.5	2774.8	2752.8	434848.0	6478418.1
Cretaceous	Redby Fm	4147.5	2780.6	2758.6	434842.4	6478418.0
Cretaceous	Sola Fm	4187.5	2789.4	2767.4	434834.1	6478417.8
Cretaceous	Asgard Fm	4201.0	2799.2	2777.2	434824.8	6478417.6
Jurassic	Viking GP	4304.0	2876.4	2854.4	434756.8	6478415.5
Jurassic	Draupne Fm	4304.0	2876.4	2854.4	434756.8	6478415.5
Jurassic	Heather Fm	4309.5	2880.6	2858.6	434753.3	6478415.3
Jurassic	Vestland GP	4316.5	2886.0	2864.0	434748.8	6478415.1
Jurassic	Hugin Fm	4316.5	2886.0	2864.0	434748.8	6478415.1
Triassic	Triassic GP	4340.0	2890.4	2882.0	434733.8	6478414.5
Triassic	Skagerrak Fm	4340.0	2904.0	2882.0	434733.8	6478414.5
TD		4641.0	3132.3	3110.3	434538.2	6478400.0

Based on the similar property patterns demonstrated by the density – P wave velocity scatter plot (Fig 2.13) the formations were grouped into 6 depth intervals (Table 2.2).

Table 2.2: Depth intervals grouped according to the properties distribution

Depth interval, [m]	3550 – 3622	3622 – 3827	3827 – 4150	4150 – 4201	4201 – 4304	4304 – 4618

The maximum depth is stated as 4618 m according to the compositional log measurements.

2.6 Model parameterisation

The proposed outlier detection and removal methodologies have their pros and cons. Therefore, they will be compared through an extensive analysis in light of the results presented in this work. In particular, model reparameterization procedures using Brocher (2005) and Gardner et al. (1974) empirical relationships were applied to extract the parameters from the real data. The obtained curve fits and corresponding coefficients are shown in Chapter 3. The resulting parameters are used later to predict synthetic curves and compare them against the logs from well 15/9-19SR.

Modelling parameterization is carried out on the "clean" dataset for the entire depth interval and the grouped formation tops (Table 2.2). The curves based on the chosen empirical equations (Eq. 2.7 – 2.8) fit cross-plotted data of P-wave velocity in km/s against density in g/cm$^{3.}$

$$\rho = aV_P - bV_P^2 + cV_P^3 - dV_P^4 + eV_P^5 \tag{2.7}$$

where:
ρ = bulk density, g/cm^3;
V_P = P-wave velocity, km/s;
a, b, c, d, e – empirical parameters extracted from data.

$$\rho = aV_P^b \tag{2.8}$$

2.7 Method selection

After defining the empirical coefficients, the updated equations are used to develop synthetic Brocher- and Gardner-like synthetic curves and compare them to the real density logs after the outlier removal procedure. In order to provide a better representation, we applied the low pass filter to **smooth the curves** and added the **confidence intervals** for proper assessment of the matching. After this we calculate the absolute percentage average mismatch and select the methodology based on the number of criteria presented below:

1. Time consumption
2. Curve fit and mismatch values
3. Realistic ranges of density
4. Difference between Brocher- and Gardner-like models for redundancy
5. Simplicity of the methodology

2.7.1 Curve smoothing

Even though outlier removal techniques were applied to the data, it does not guarantee their complete absence. In fact, some remaining outliers are observed in the cross plot (Fig. 2.13). Likewise, some of the inliers for the entire depth interval are still anomalous points for the defined regions and might contaminate the newly generated logs with high-frequency (wavenumber) noise. To prevent that, a high-pass filter has been applied to smooth the log curves and constraint the fitting to the essential low frequency (wavenumber) events (Fig. 2.17).

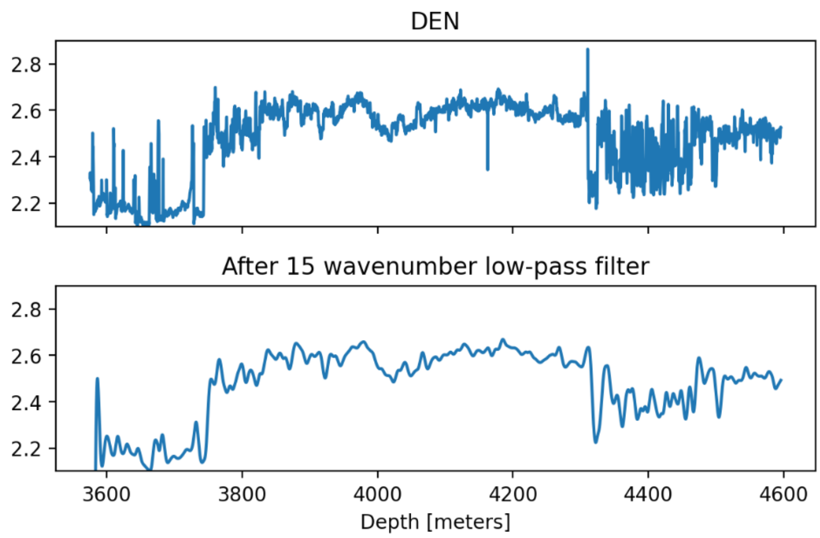

Fig. 2.17: Example of density curve after outlier removal compared to low pass filtered curve (entire depth)

2.7.2 Confidence intervals

The well logs are the recorded signals that can be disrupted by the noise due to tool malfunction or natural causes. Therefore, even though the majority of the inconsistent values have been removed, certain fluctuations of data still exist. The confidence intervals are depicted as the areas that describe the uncertainty surrounding the synthetic curves. It is visually considered a better match if the field data we tried to fit lies within the area of the confidence intervals (Fig.3.33 – 3.34).

Once the data are filtered, confidence intervals (Eq. 2.9) of 99.7 % are implemented to observe to what extent the data fits by lying within the range.

$$ci = \mathrm{z}\frac{0.1\sigma}{\mu} \tag{2.9}$$

where:

ci = confidence interval;

μ = mean of the data set;

σ = standard deviation of the data set

z = coefficient, which depends on the percentage of confidence intervals.

For confidence intervals of 99.7%, the z parameter is z = 2.748. The synthetic curves obtained for the entire depth interval are plotted against real density values in normal and smoothed states and are shown in Chapter 3 together with confidence intervals and corresponding mismatch analyses. In addition, we applied the confidence intervals to original (non-smooth) curves to better distinguish the fit visually.

3 METHOD APPLICATIONS AND OUTCOMES

3.1 Standard models

The migration velocity model in our study remains unchanged and is used to map densities according to existing and developed empirical relationships.

The velocity model was converted to density by means of standard Brocher (Eq. 2.2) and Gardner (Eq. 2.1) empirical relationships and further analyzed. The results of using the standard Brocher and Gardner density models are shown in Fig 3.1 and Fig. 3.2 respectively.

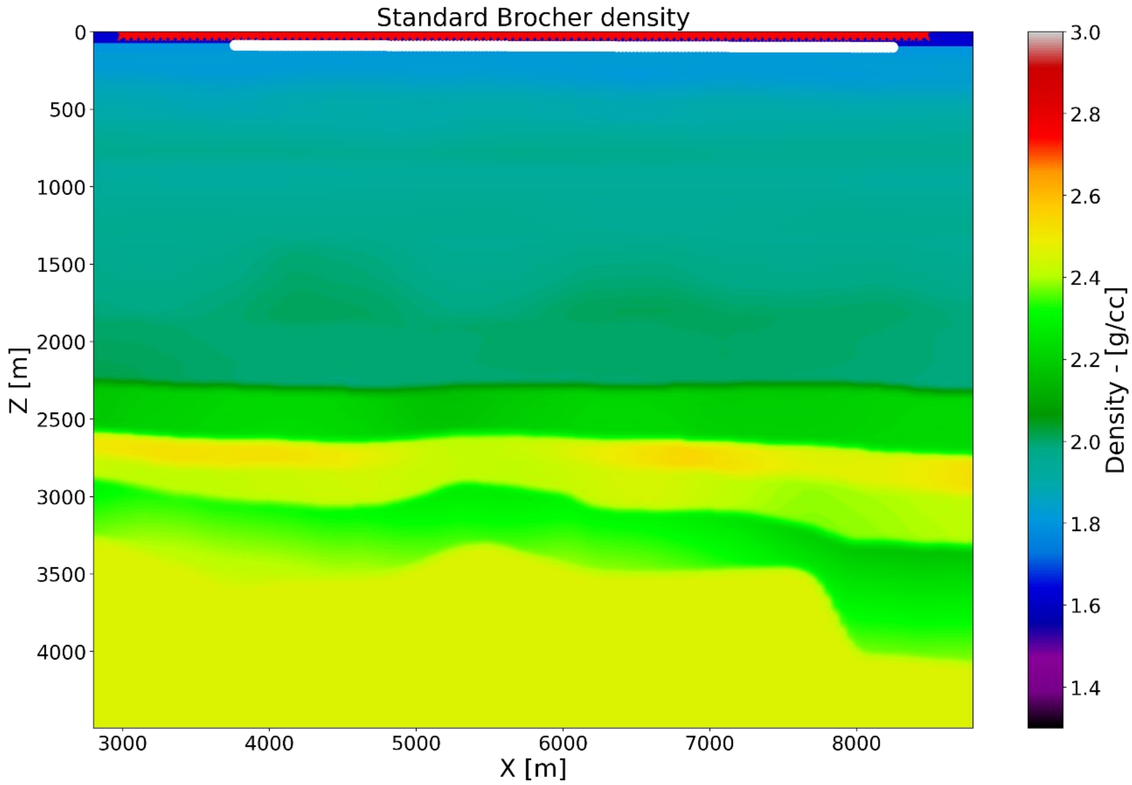

Fig 3.1: Standard Brocher density 2D slice №1 (smooth velocity model)

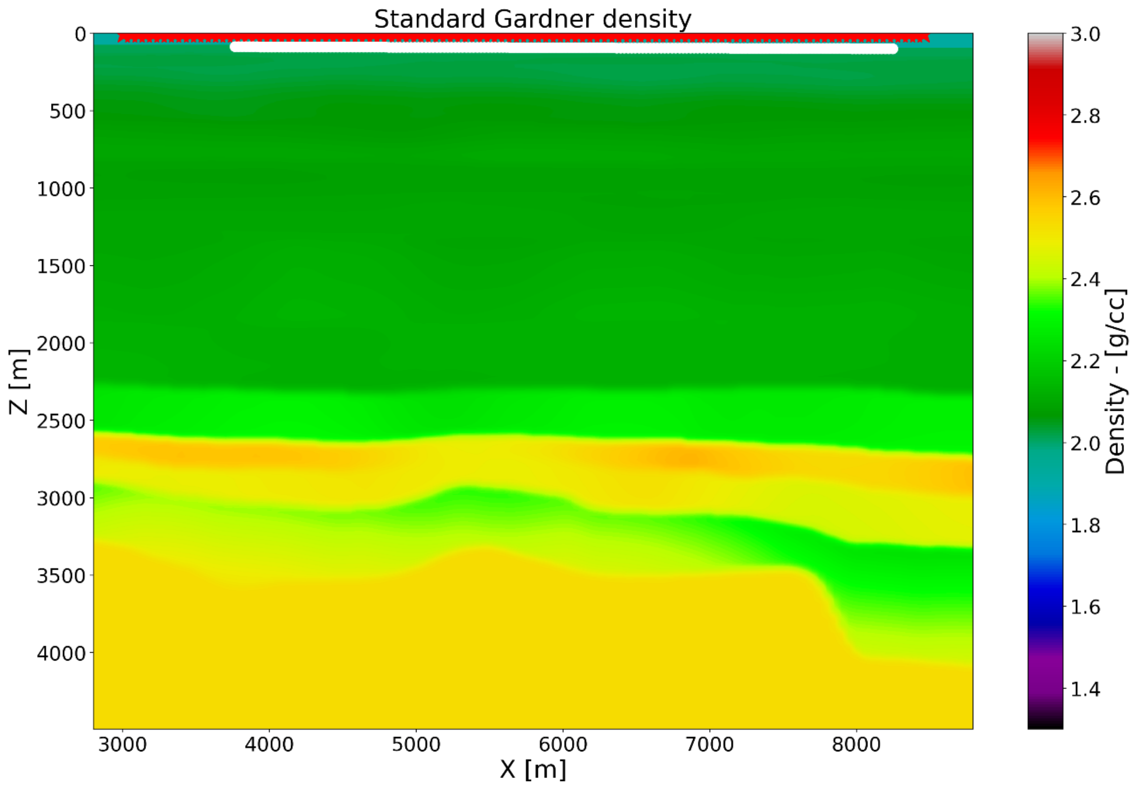

Fig 3.2: Standard Gardner density 2D slice №1 (smooth velocity model)

$$\Delta\rho = \frac{\sqrt{(\rho_{Gardner} - \rho_{Brocher})^2}}{\rho_{Gardner}} * 100 \quad (3.1)$$

where:

$\rho_{Gardner}$ = Standard Gardner bulk density, g/cm^3;

$\rho_{Brocher}$ = Standard Brocher bulk density, g/cm^3

$\Delta\rho$ = density difference, %

The density difference between the two versions of the density model was obtained using the standard empirical relationships applied in Equation 3.1 to exemplify it graphically in Fig. 3.3.

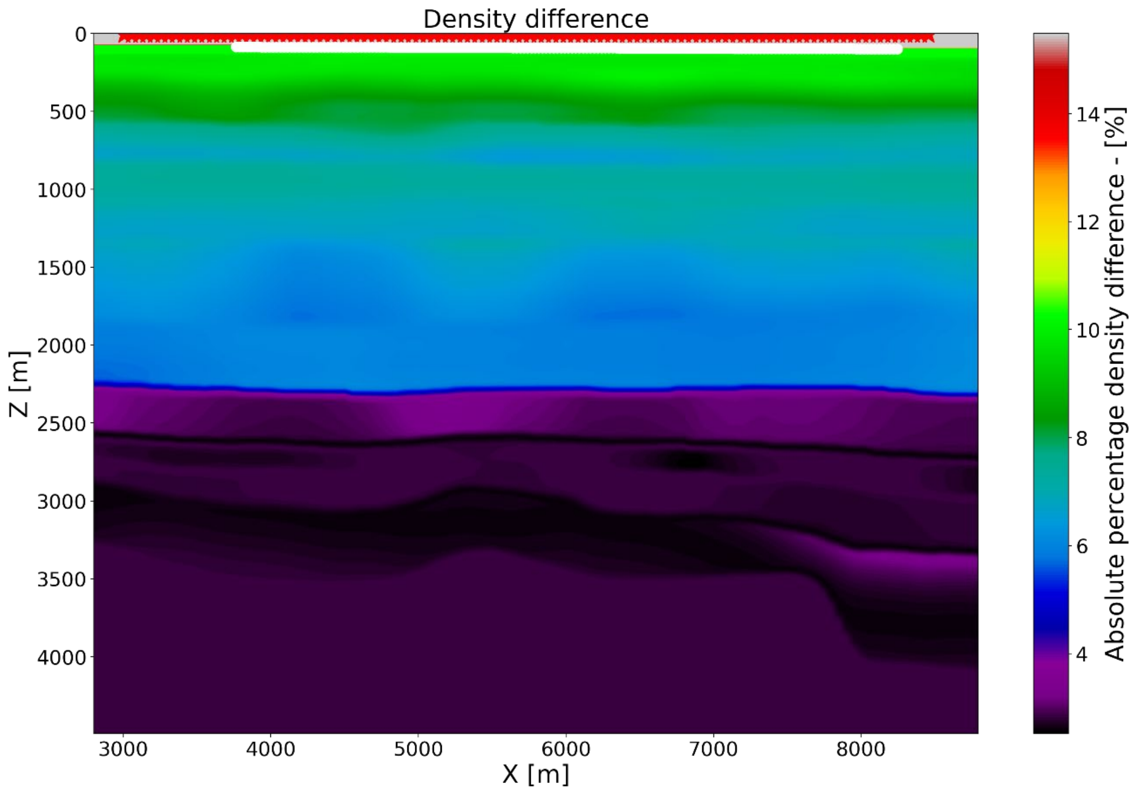

Fig 3.3: Absolute percentage density difference

The absolute percentage difference does not visually exceed 12% while the maximum of 14% (red) and higher represents the anomaly located in the top part of the image. We noted that the difference between the Brocher and Gardner density model reduces with depth and made an assumption that it is happening due to the change in velocity. Despite the fact that the applicable velocity ranges for both relationships are similar, the Gardner model depicts higher density values at lower velocity ranges.

The same approach is applied to the sharp velocity model retrieved using post-stack acoustic impedance inversion for both Brocher 2005 (Fig. 3.4) and Gardner 1974 (Fig 3.5).

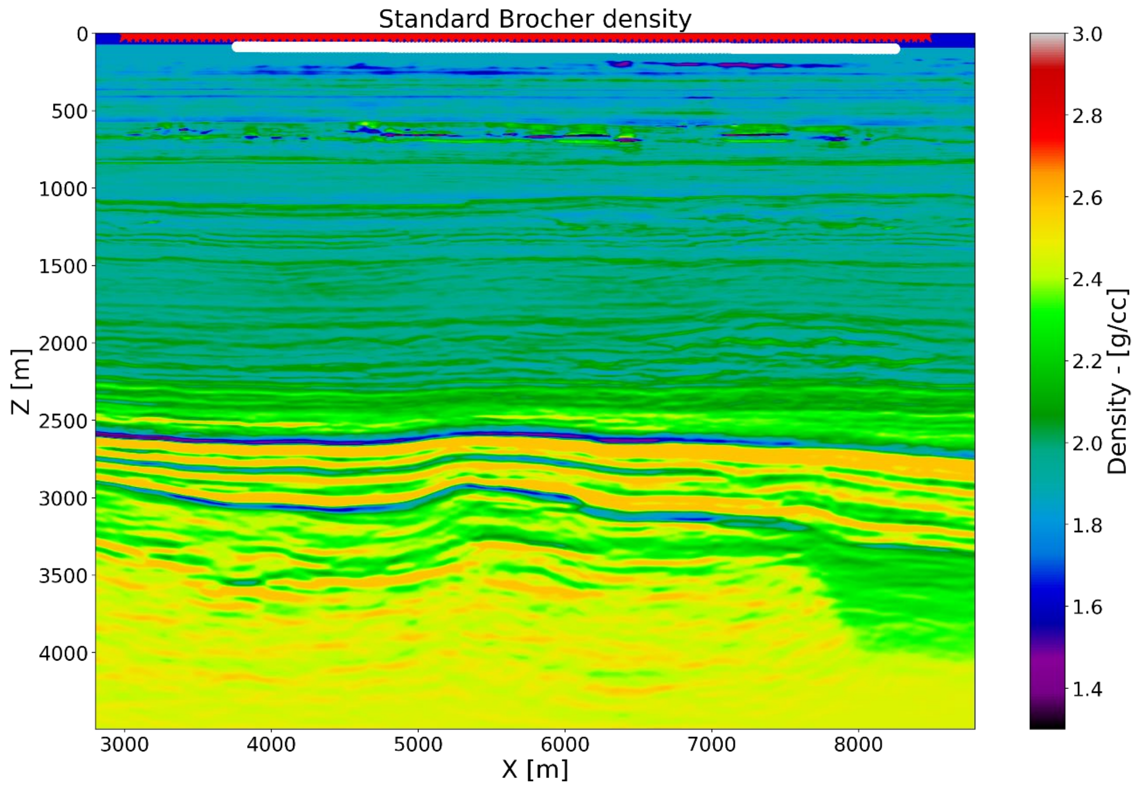

Fig. 3.4: Standard Brocher density 2D slice (synthetic model)

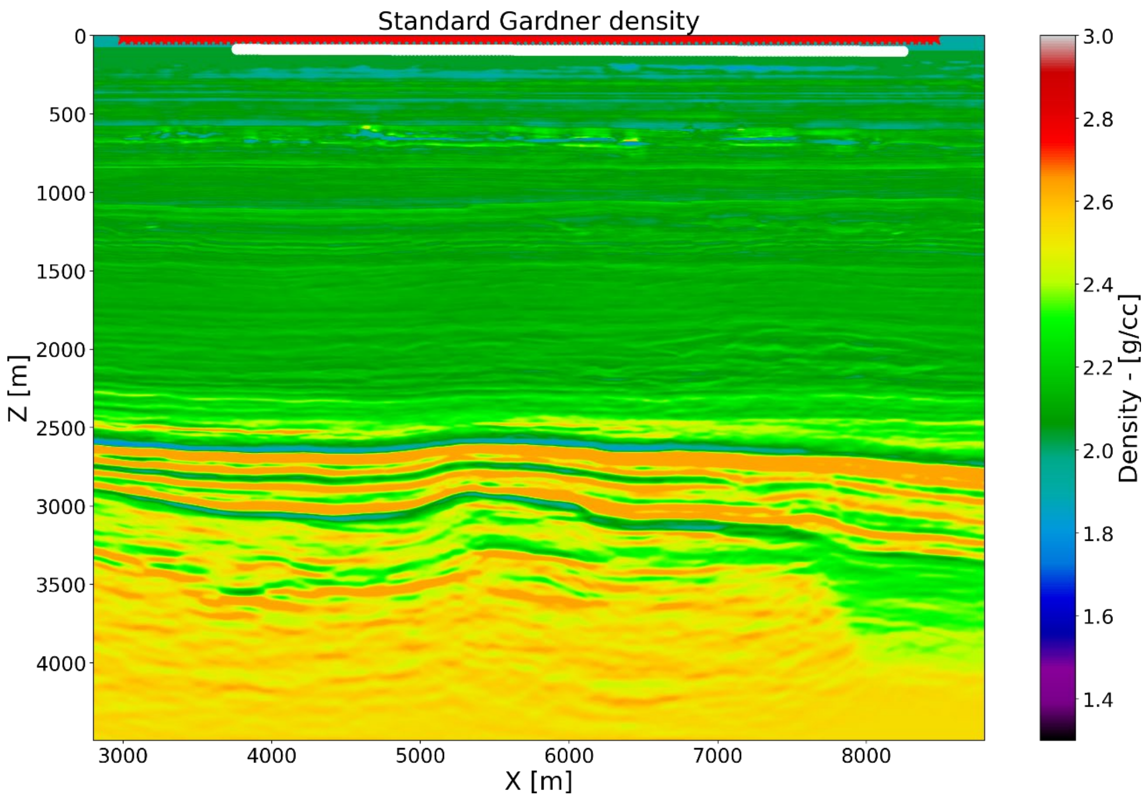

Fig. 3.5: Standard Gardner density 2D slice (synthetic model)

The density difference between the two versions of the sharp synthetic density model was obtained using Equation 3.1 in order to demonstrate it graphically in Fig. 3.6.

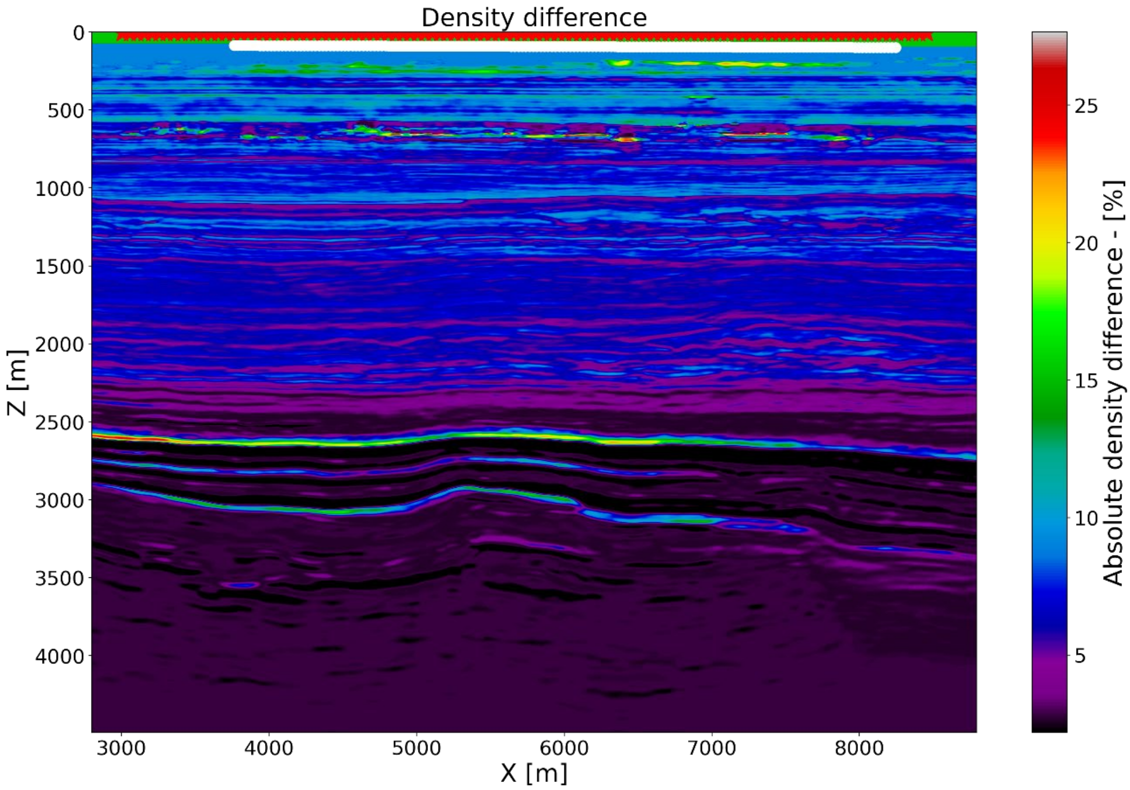

Fig. 3.6: Absolute density difference (synthetic model)

Comparing the absolute percentage difference for the high-resolution model to the one provided for the smooth model (Fig. 3.3) we noticed that the difference in the current case does not exceed 25%, since the model has significantly more details to differ from the Volve one. In addition, the defined maximum (red) does not only describe the anomaly at the top of the image, but also certain layers within the 2D slice. The first versions of density models will be compared later against the final reconstructions obtained from completing the workflow steps.

3.2 Outlier removal

According to the workflow, Figure 2.1, two main approaches were chosen to be used for outlier identification and removal:

1. Both manual and unsupervised machine learning techniques
2. Unsupervised machine learning algorithms only

3.2.1 Manual and ML outlier detection and removal

In this approach, all well logs have been chosen as input for box plot based removal. The data points considered to be inconsistent for one of the logs were eliminated for all the others as well as for a specific point in depth. It brings distortion in some of the log curves, but "cleans" the whole data set at the same time.

The outlier removal procedure applied in the first approach is presented in the following steps:

1. Manual removal of certain outliers according to the realistic ranges of chosen values
2. Box plot outlier detection and removal by IQR implementation
3. Unsupervised machine learning techniques

The outliers defined for the raw well logs from well 15/9-19SR are shown in the box plot (Fig. 3.7). All the applied techniques are described in more detail in Chapter 2.

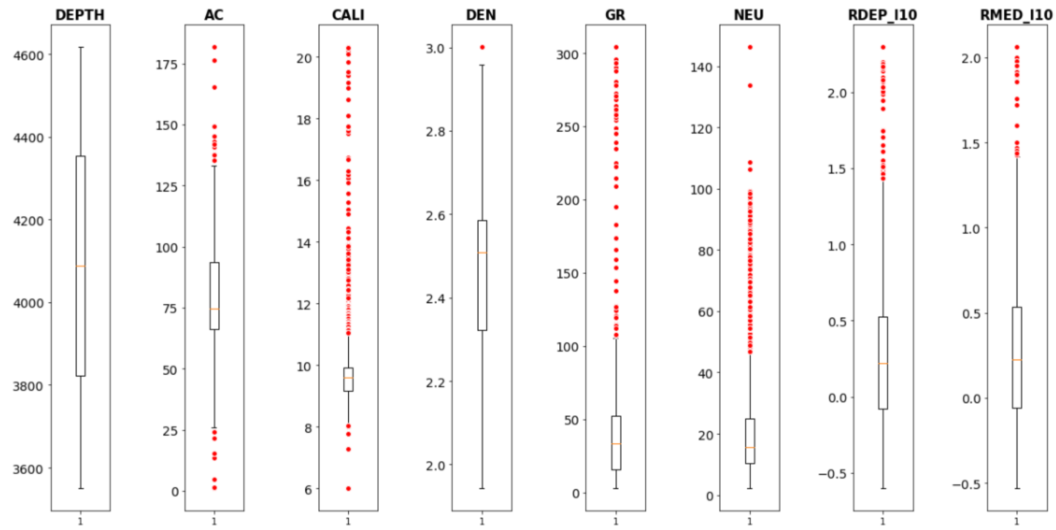

Fig. 3.7: Box plot representation of well logs (15/9-19SR), outliers are marked as red circles

After the first set of outliers has been removed, the data is divided into quartiles once more. In doing so, we establish that some of the points previously lying within the IQR range are now becoming outliers. Meanwhile, acoustic compression slowness values are converted to velocity and analysed with the help of box plots. The elimination of anomalous values was repeated to obtain box plots that reflected a minimum number of outliers (Fig. 3.8) within the density and velocity data; such values are crucial for the upcoming model parameterization for Brocher and Gardner empirical relationships.

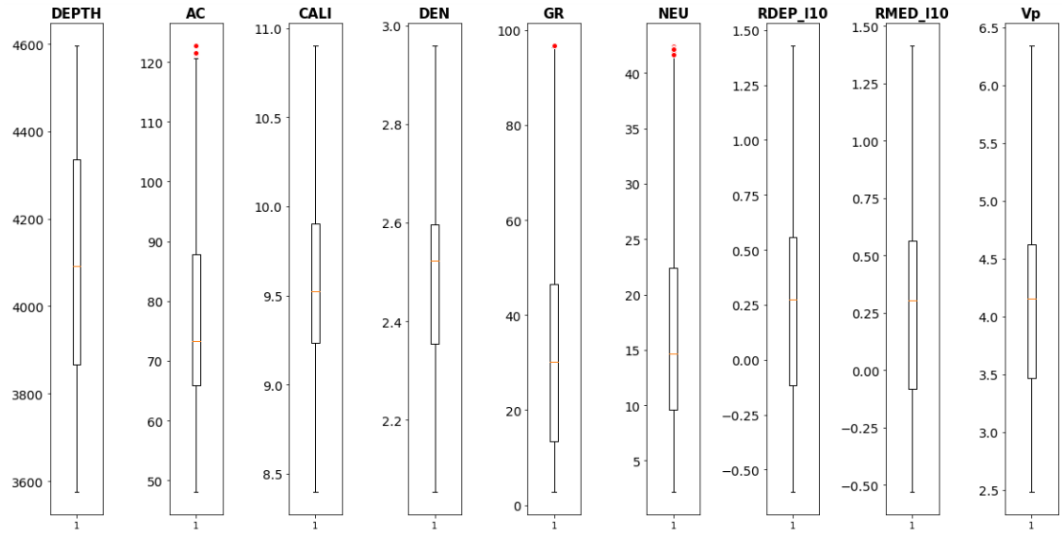

Fig. 3.8: Box plots after outliers are removed (Manual and ML removal)

Despite the IQR outlier removal being carried out, the velocity range within the data set was still relatively high, especially for the depth interval corresponding to a maximum of 4618 m. In order to overcome this limitation, machine learning techniques are applied to then compare the different results and choose the most successful method for outlier elimination. During the analysis of the available technical data, no corresponding wells were found, location- and direction-wise. For this reason, unsupervised machine learning was selected, with the idea that this kind of algorithm will be able to obtain the output data distribution from the input data set itself. The input logs in the case of unsupervised machine learning algorithms are density (DEN), acoustic compression slowness (AC), caliper (CALI), and p-wave velocity computed from AC logs. Only a limited number of logs were included in order to improve the precision of the algorithm and its application to data cleaning for the current project.

Isolation Forest (IF), One-class Support Vector Machines (SVM), and Local Outlier Factor (LOF) algorithms were applied to the well log data deprived of anomalous values outside of the IQR range (Fig. 3.9 – 3.11). In the case of manual and ML outlier removal, a total of around 26 % of values were identified as outliers and eliminated using both box-plot and ML-based removal. Out of the remaining data, 18 % of the values are removed during the IQR method.

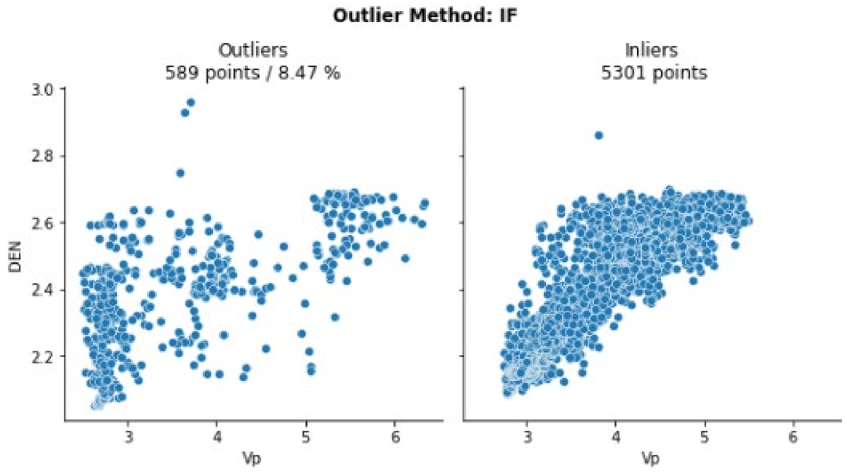

Fig. 3.9: IF outlier removal (Manual and ML removal)

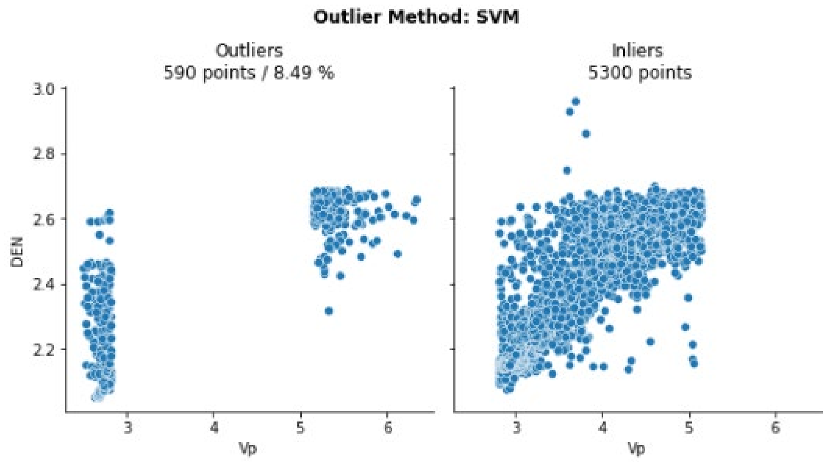

Fig. 3.10: One class SVM outlier removal (Manual and ML removal)

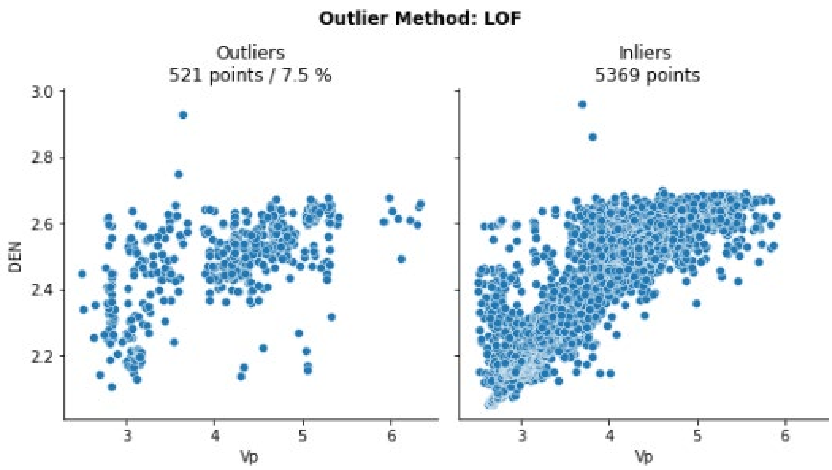

Fig. 3.11: LOF outlier removal (Manual and ML removal)

We noted that in the **one class SVM** method the inconsistent values were removed only from the defined areas with the velocities being too low or too high. It does not correspond with the realistic situation. While the **LOF** algorithms demonstrated the most abnormal cluster shape.

Based on the comparison of the outlier/inlier cross plots presented and the well log data displayed for all types of algorithms (Fig. 3.12), we were able to select **Isolation Forest** as the most efficient unsupervised machine learning technique. In Figure 3.12, the anomalous data regions are highlighted in red. In this case, the Isolation Forest technique shows a clear trend with minimum points visually lying outside of the cluster altogether; it also revealed reasonable well-log values after outlier removal.

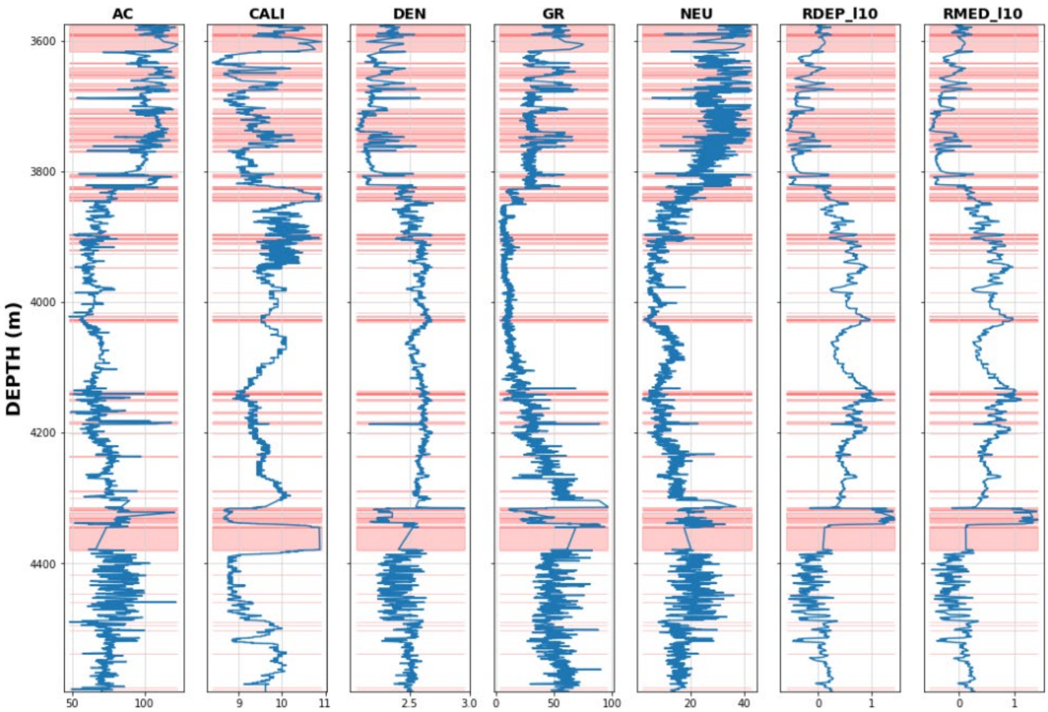

Fig. 3.12: Well log data prior to IF outlier removal; outliers to be removed are highlighted in red (Manual and ML removal)

A box plot representing the ranges of values within the data set after manual and ML outlier removal is shown in Fig. 3.13.

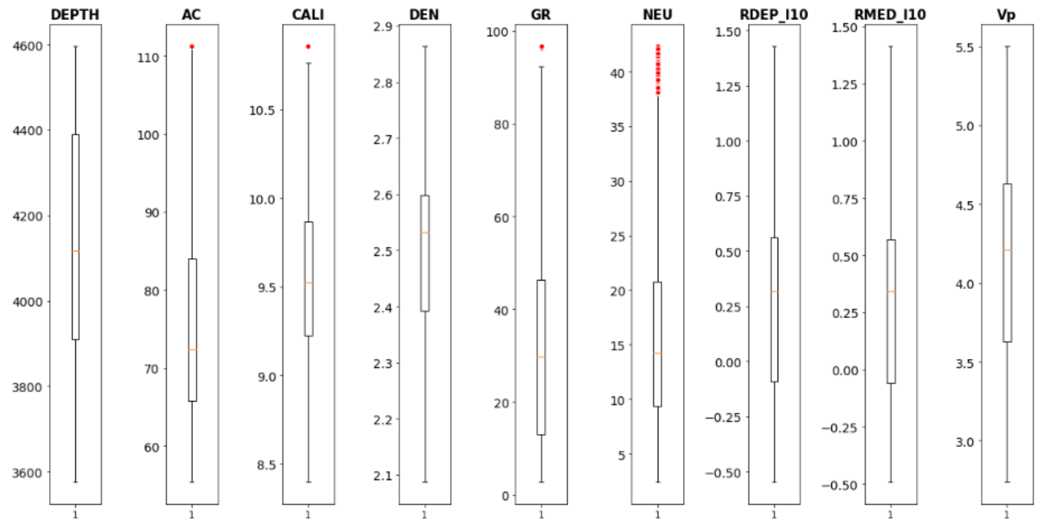

Fig. 3.13: Box plot after manual and ML outlier removal

Well logs for the other ML algorithms are presented in Annex A. There are more areas from which outliers have to be removed, however, their distribution is more chaotic for One Class SVM and LOF algorithms.

3.2.2 Unsupervised machine learning detection and removal

In this approach, data are not filtered by box plots and IQR prior to processing with the proposed algorithms. This strategy allowed us to observe if the unsupervised machine learning is robust enough to remove a significant number of outliers. All presented logs are used as input to compensate for the absence of manual elimination of anomalous values.

The cross plots revealing anomalous values to be removed and inliers to be kept are shown in Fig. 3.14 – 3.16.

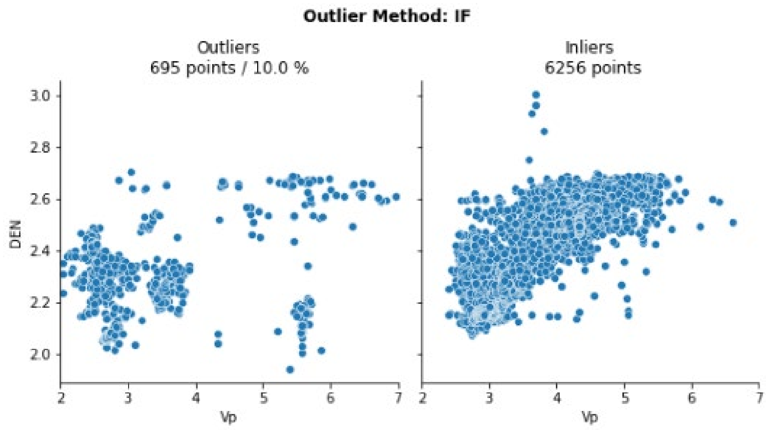

Fig. 3.14: IF outlier removal (ML removal)

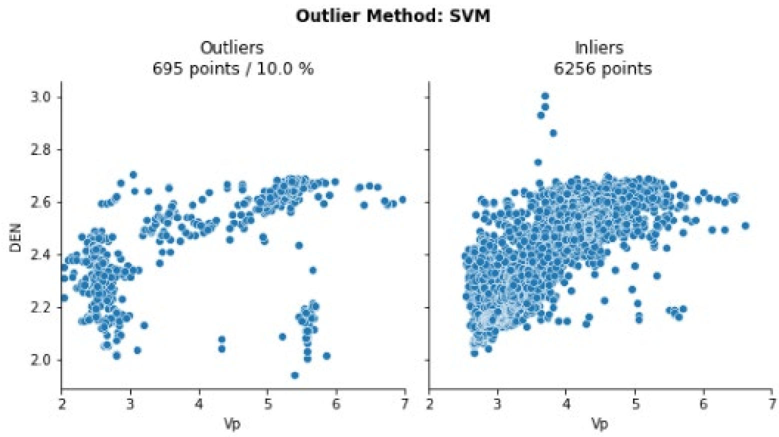

Fig. 3.15: One class SVM outlier removal (ML removal)

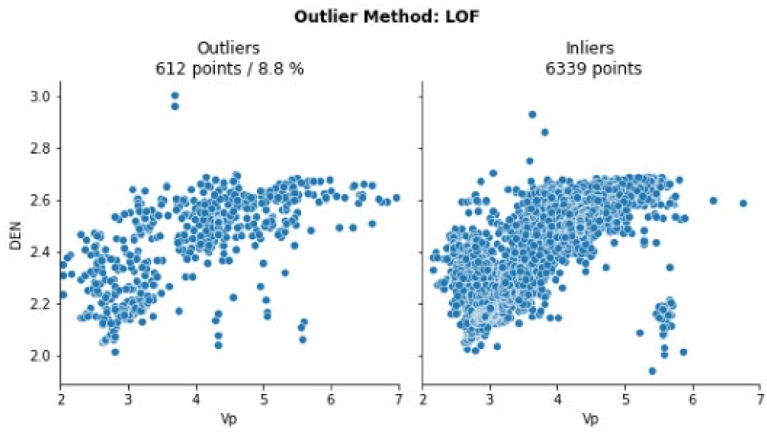

Fig. 3.16: LOF outlier removal (ML removal)

The visual difference between the identified inlier clusters is less obvious than in the previous outlier removal approach. Nevertheless, the Isolation Forest algorithm was picked according to the logs shown in the Figure. 3.17 and to carry out a proper comparison to the manual and ML outlier removal techniques.

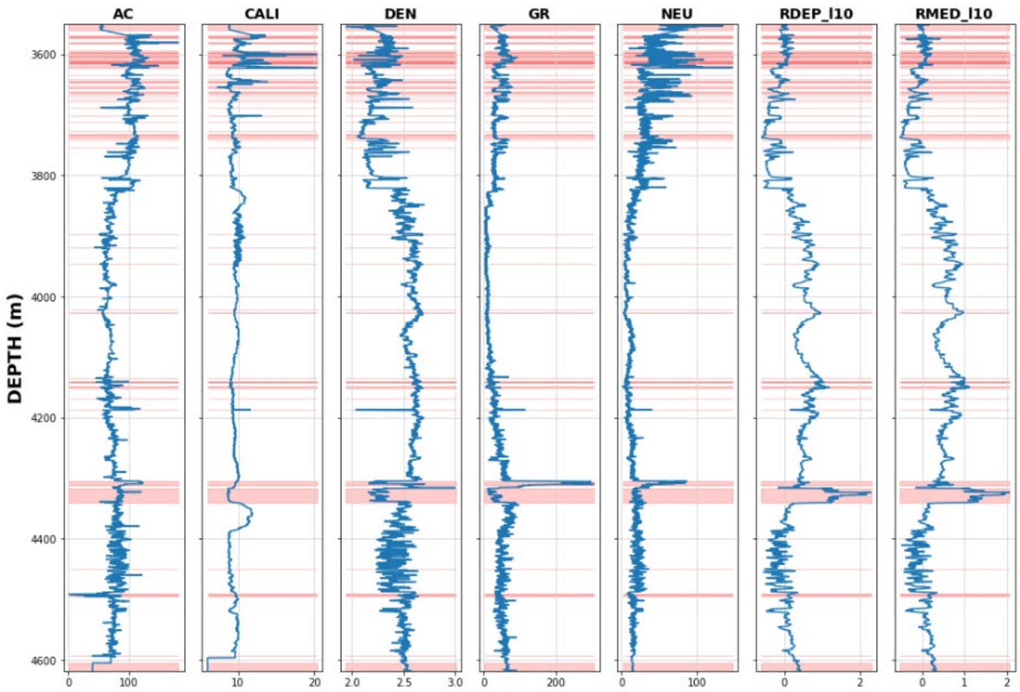

Fig. 3.17: Well log data prior to IF outlier removal; outliers to be removed are highlighted in red (ML removal)

Additionally, for this method, the box plot is provided so that one can compare the ranges of values after the removal of inconsistent values (Fig. 3.18).

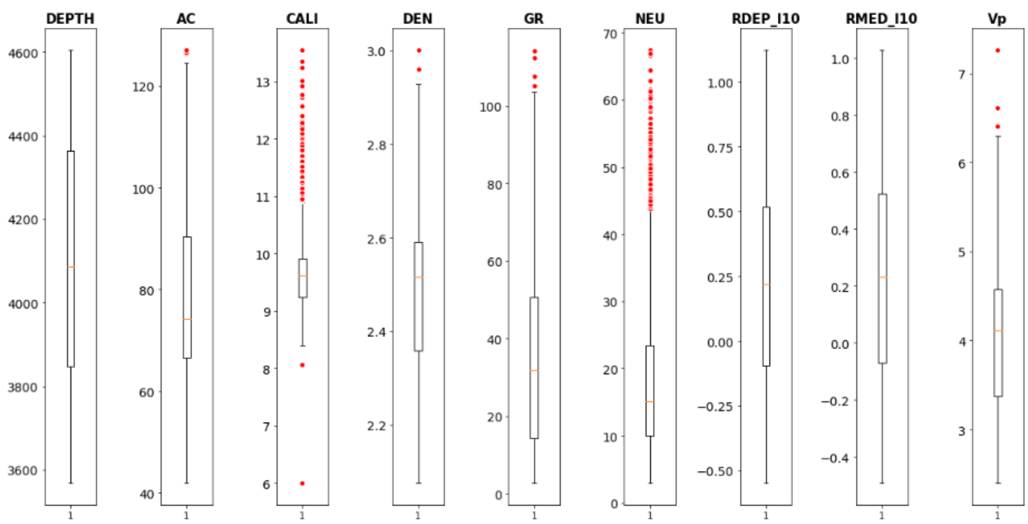

Fig. 3.18: Box plot after ML outlier removal

Since the UML algorithms identified just 10% of data as outliers compared to 26% for using both methods, we can note the red zones (Fig.3.17) are less even though all the logs were used as input for machine learning algorithms. Nevertheless, according to Fig. 3.18 the obvious outliers are more, and they are present even in the logs of interest (Vp, Density).

3.3 Model parameterization

The curve fits for the Gardner-like empirical relationships for whole depth interval and formation tops are presented in Fig. 3.19 – 3. 25. In this case, the data are shown only for manual and ML outlier detection and removal. Still, a graphical representation of model parameterization for the ML-based removal of anomalous values is presented in Annex B. Shown cross plots are provided with the **colour scale** corresponding to the **depth** and **red line** demonstrating the **curve fit** of the corresponding empirical equation.

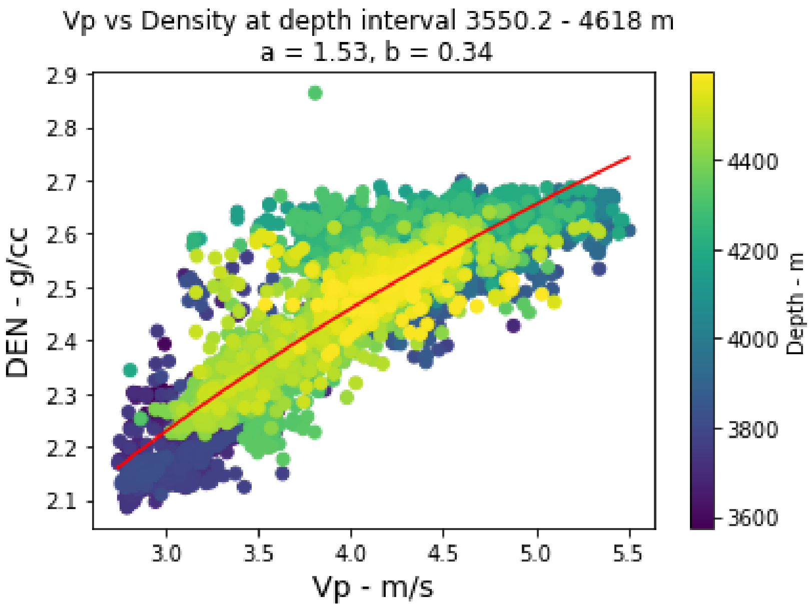

Fig. 3.19: Density-velocity cross plot with Gardner curve fit (red) for entire depth interval with estimated coefficients

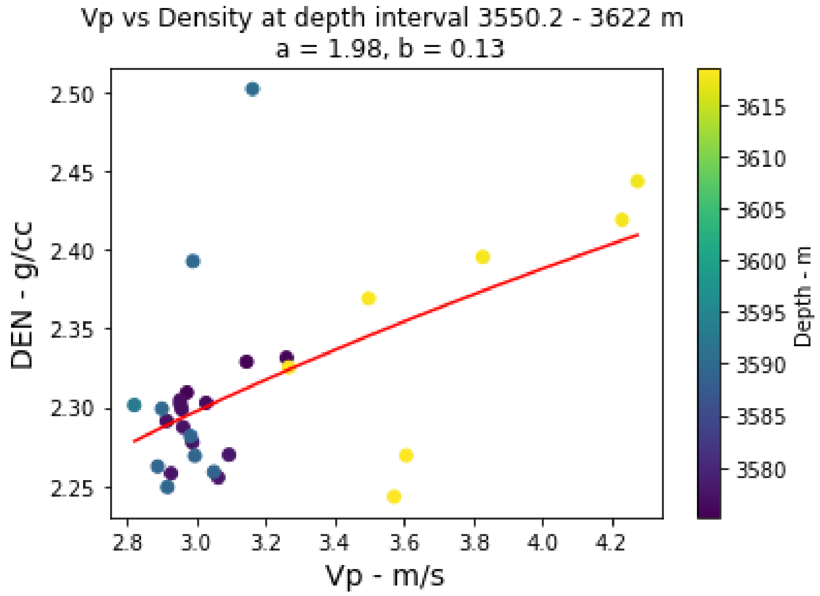

Fig. 3.20: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 3550.2 – 3622 m with estimated coefficients

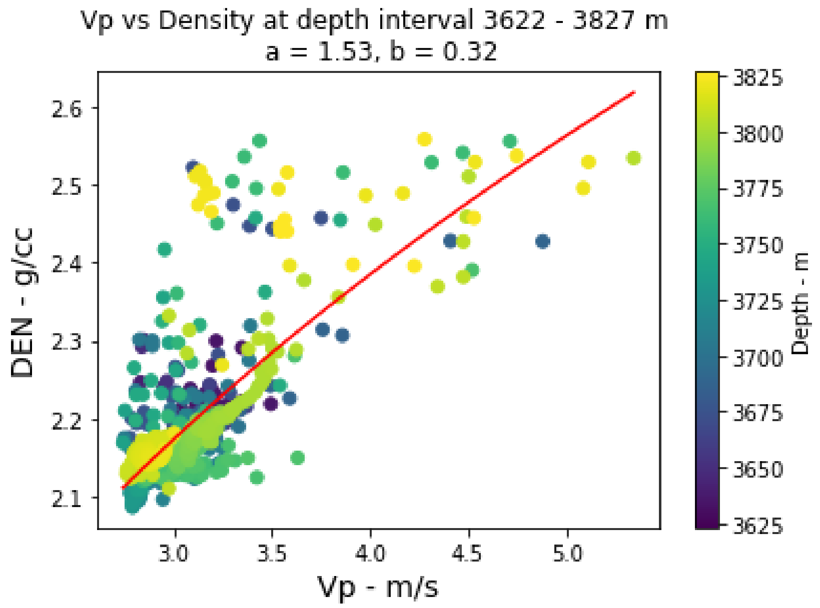

Fig. 3.21: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 3622 – 3827 m with estimated coefficients

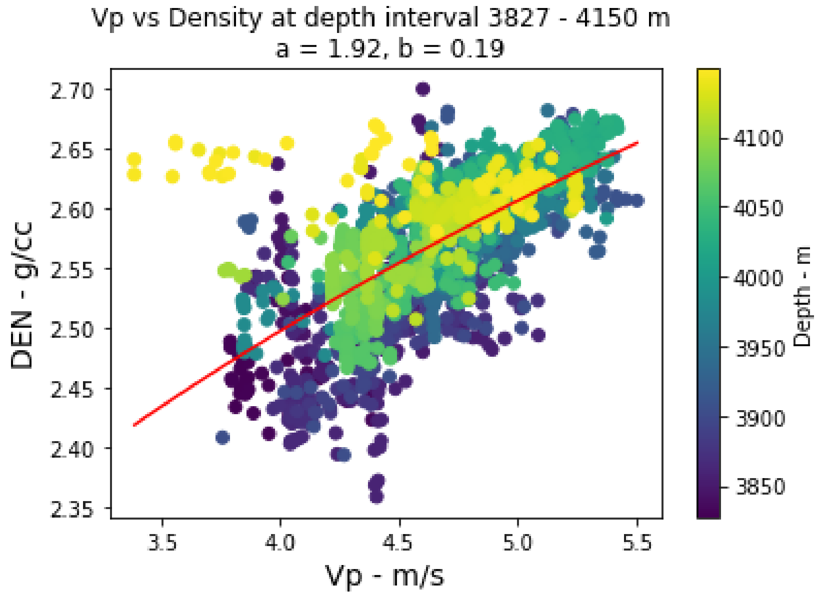

Fig. 3.22: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 3827 – 4150 m with estimated coefficients

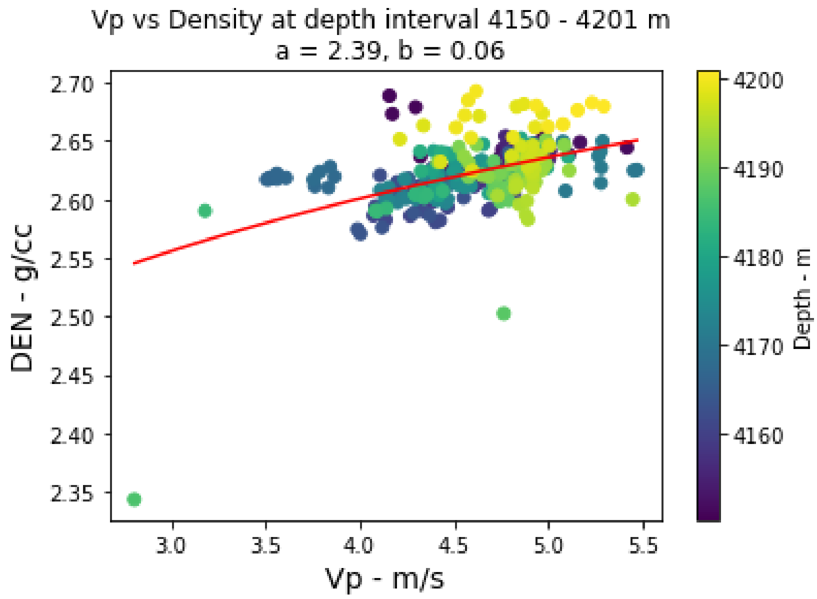

Fig. 3.23: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 4150 – 4201 m with estimated coefficients

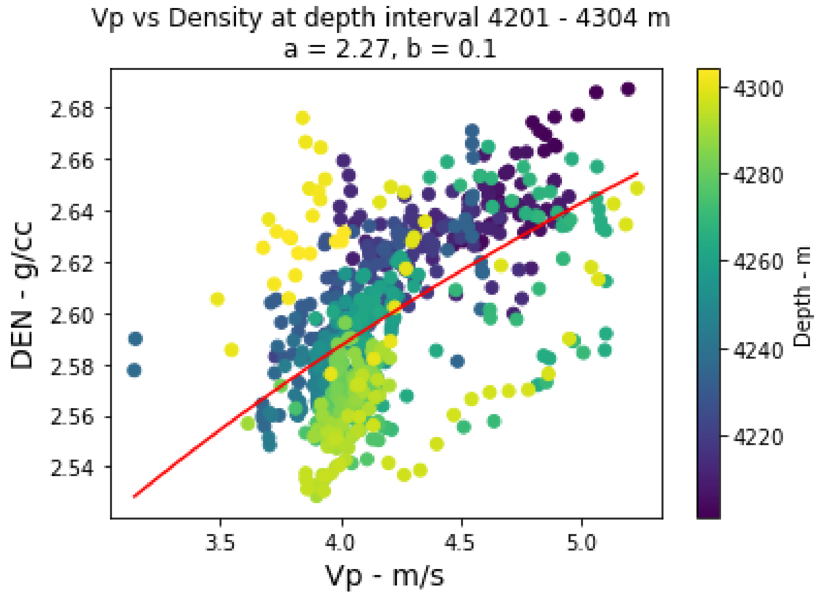

Fig. 3.24: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 4201 – 4304 m with estimated coefficients

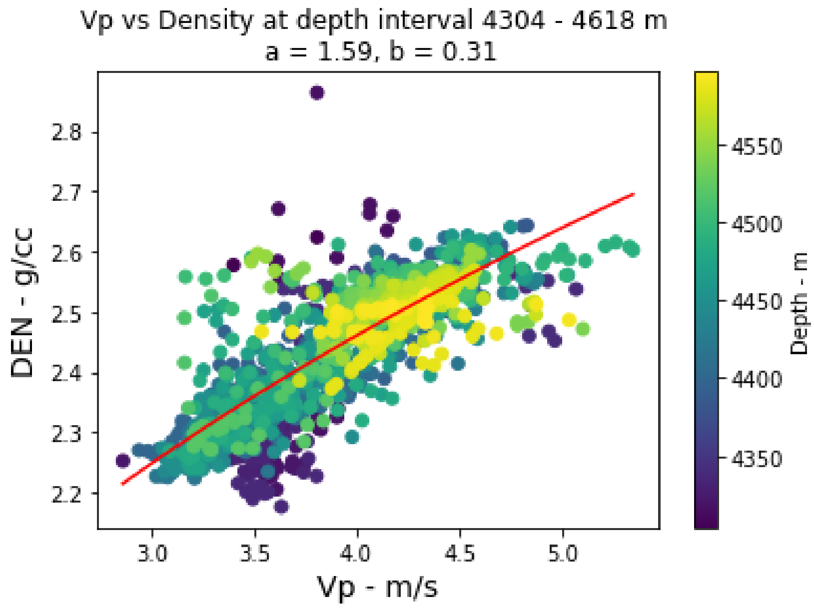

Fig. 3.25: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 4304 – 4618 m with estimated coefficients

The curves fit that correspond to the Brocher-like empirical relationships for whole depth intervals and formation tops are presented in Fig. 3.26 – 3. 32.

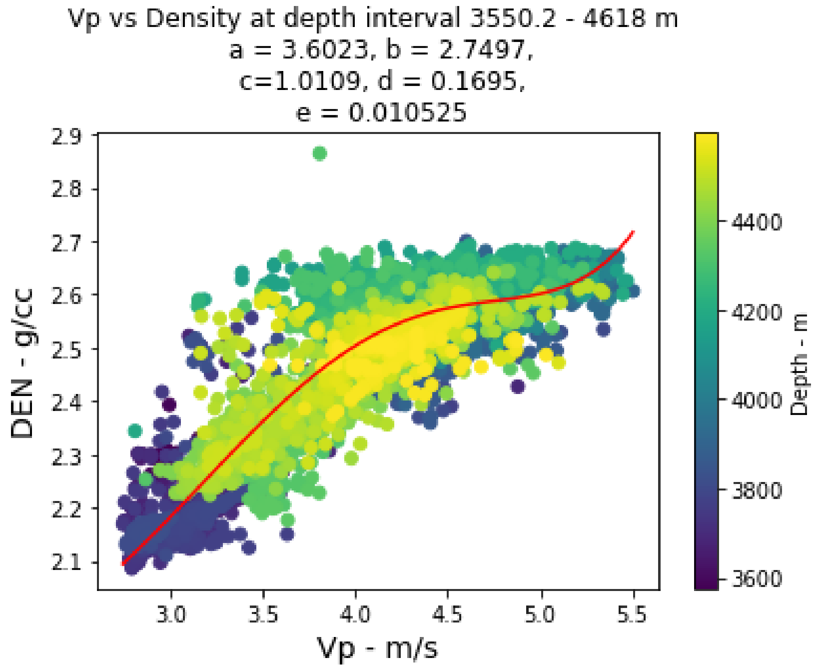

Fig. 3.26: Density - velocity cross plot with Brocher curve fit (red) for entire depth interval with estimated coefficients

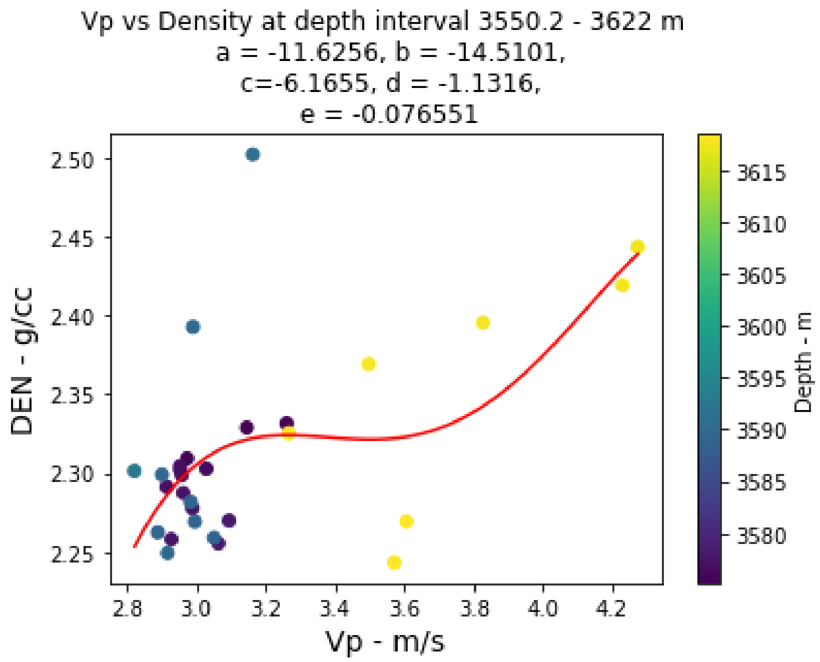

Fig. 3.27: Density - velocity cross plot with Brocher curve fit (red) for entire depth interval with estimated coefficients

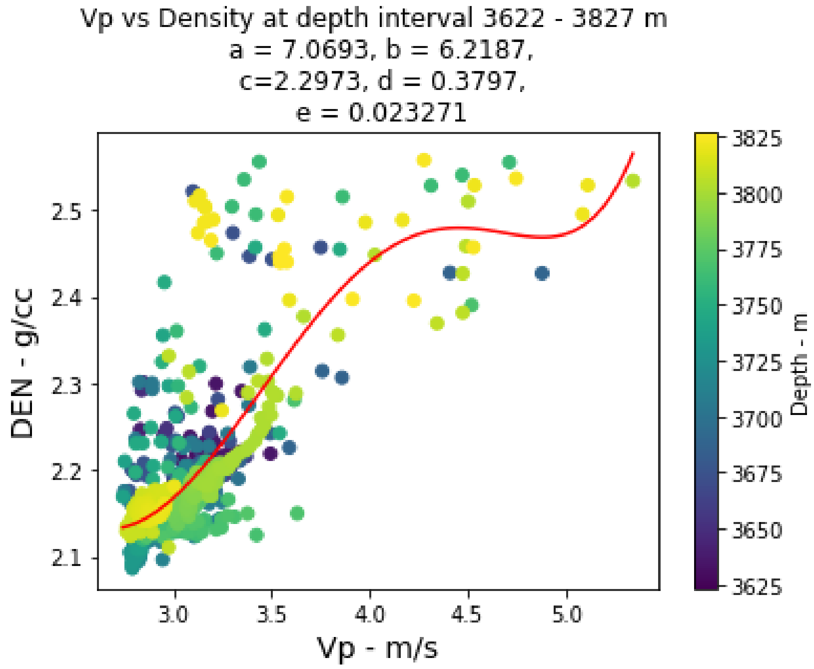

Fig. 3.28: Density - velocity cross plot with Brocher curve fit (red) for entire depth interval with estimated coefficients

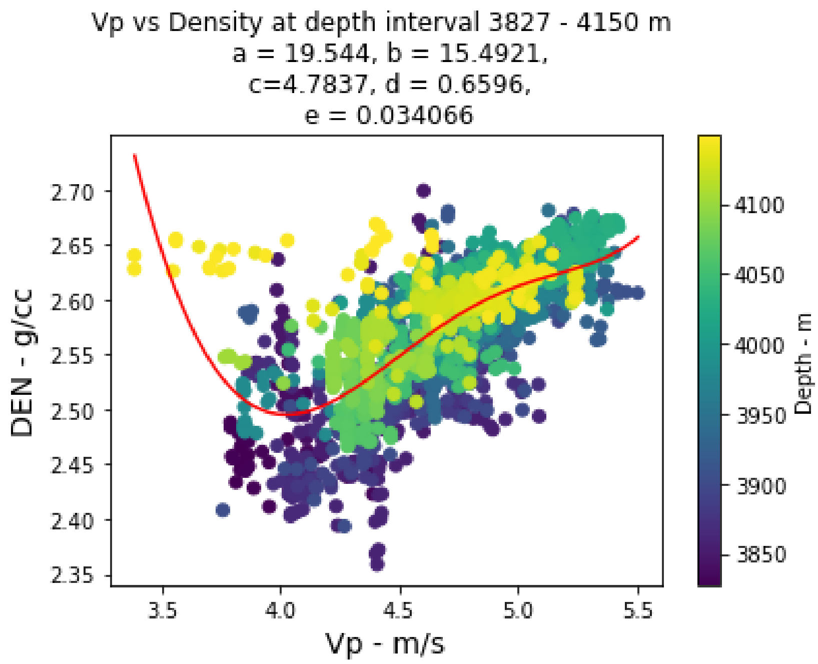

Fig. 3.29: Density - velocity cross plot with Brocher curve fit (red) for entire depth interval with estimated coefficients

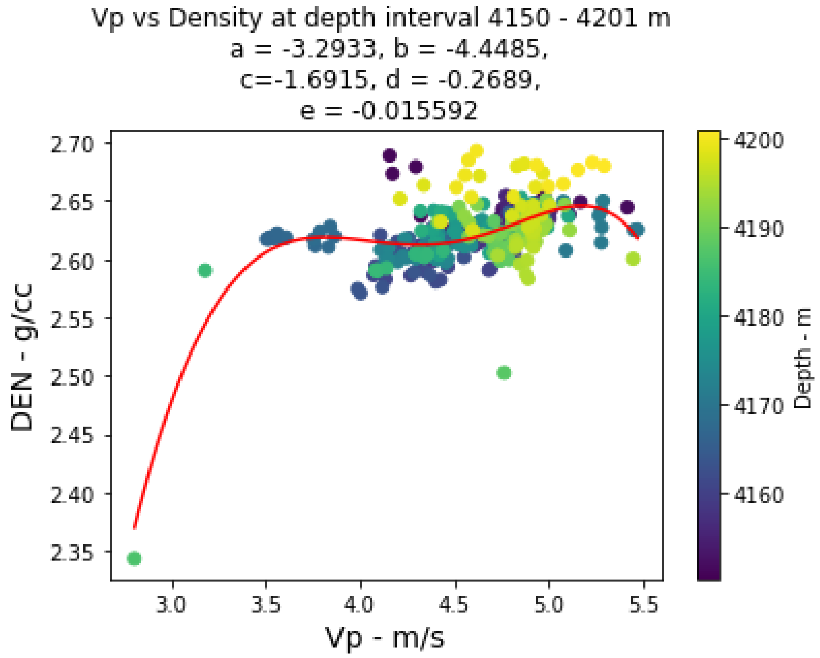

Fig. 3.30: Density - velocity cross plot with Brocher curve fit (red) for entire depth interval with estimated coefficients

It is shown that single points have a stronger effect on polynomial curve fitting (Brocher), in comparison to exponential (Gardner).

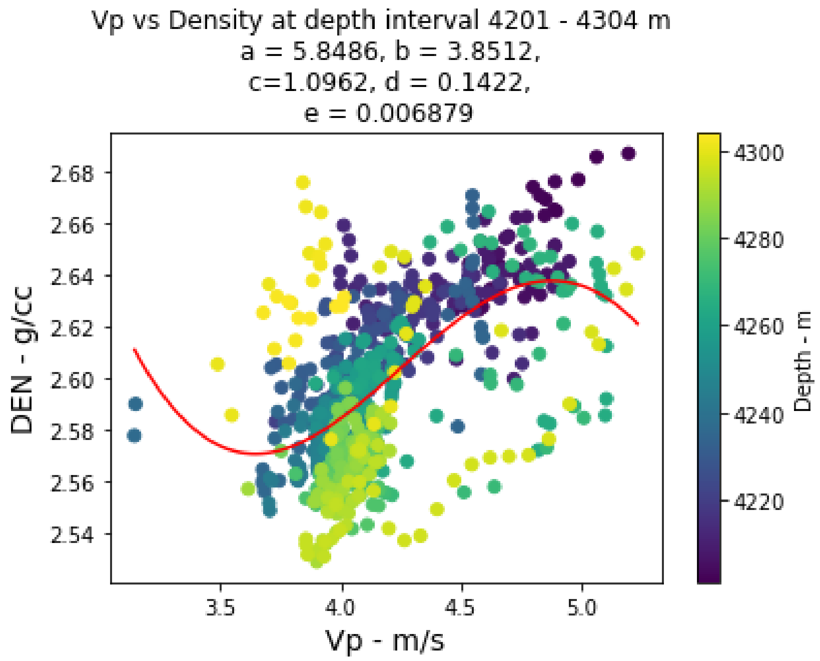

Fig. 3.31: Density - velocity cross plot with Brocher curve fit (red) for entire depth interval with estimated coefficients

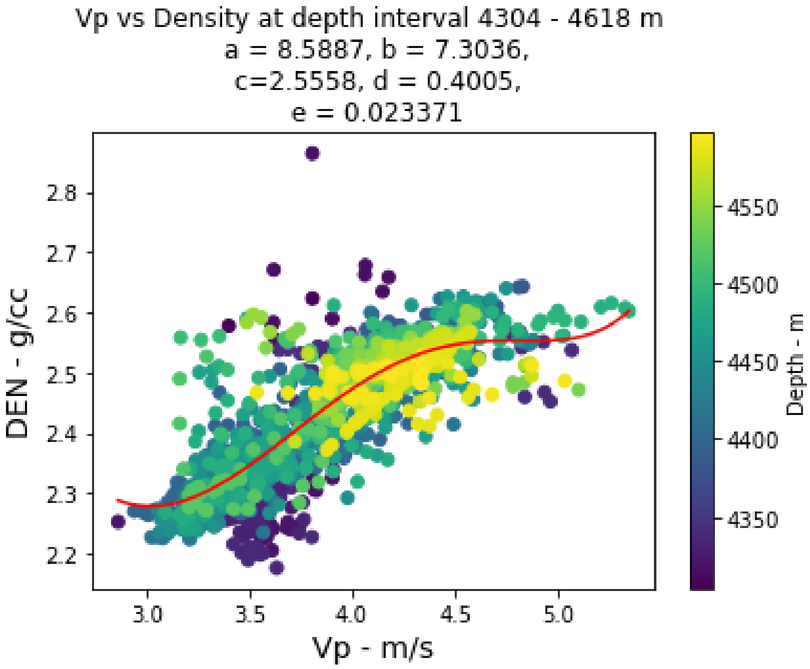

Fig. 3.32: Density - velocity cross plot with Brocher curve fit (red) for entire depth interval with estimated coefficients

3.4 Method selection

The retrieved parameters are used to compute synthetic logs which then serve as a reference to fit the real data to estimate the quality of the matching and select the most appropriate approach of model parameterization.

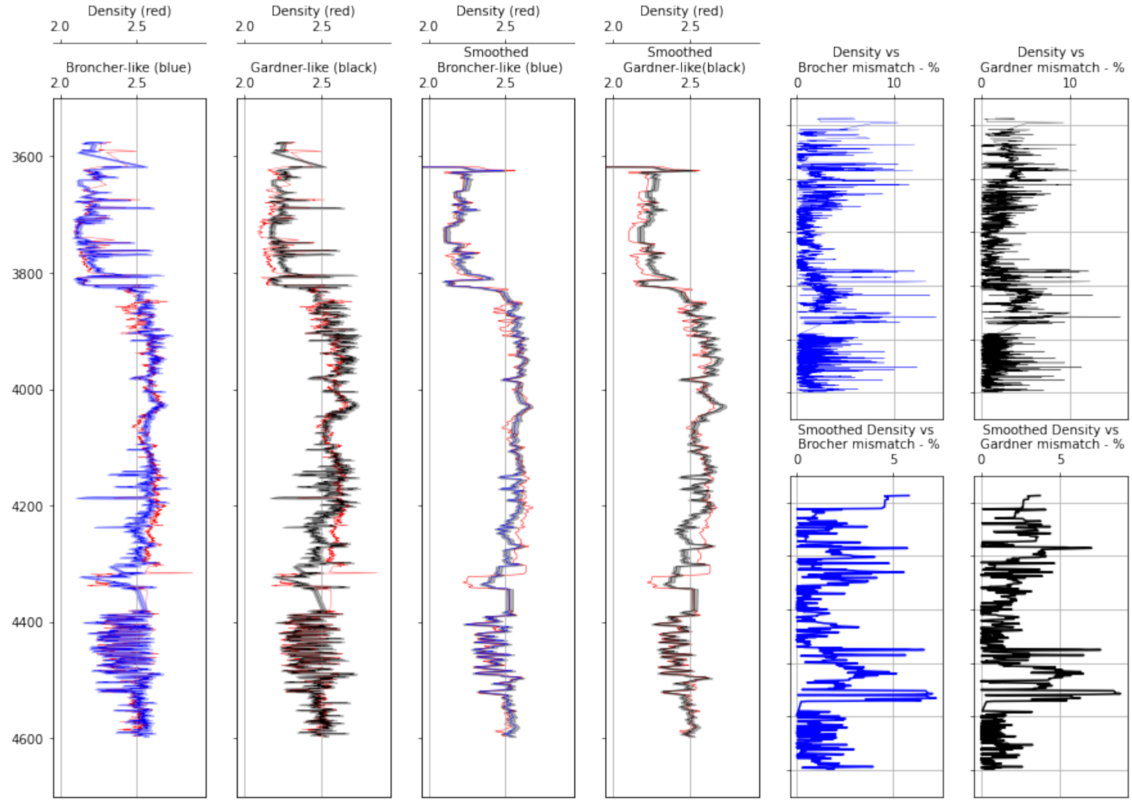

Fig. 3.33: Synthetic and smoothed curves fitting and mismatch analysis for the entire depth after manual and ML outlier removal

As we can see, the logs exactly proved the assumption that we stated in Section 3.1. In the current case the Brocher relationship seemingly fits the real data better, but what is more important is that in the lower depth (lower range of velocities) Gardner slightly overestimates the density.

We observe a consistent fit for both Brocher-like (Eq. 2.7) and Gardner-like (Eq. 2.8) relationships with average absolute percentage mismatch values of 1.89 and 2.3 %, respectively. The smoothed curves demonstrate slightly better results with average mismatch values of 1.48 and 1.84 %, which indicate that the proposed synthetic logs fit the low-frequency events as well as the real data in general.

The same approach is applied to the formation tops and presented in Fig. 3.34.

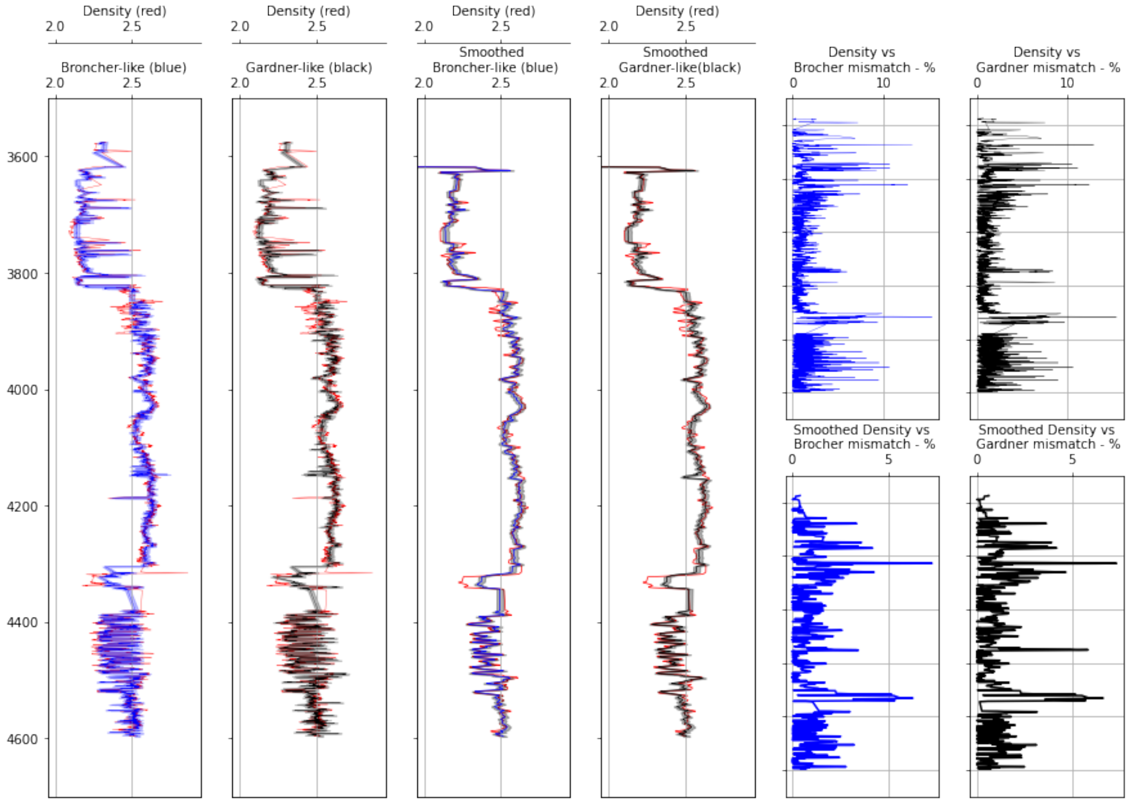

Fig. 3.34: Synthetic and smoothed curves fitting and mismatch analysis for the grouped depth intervals after manual and ML outlier removal

The assumption stated above is neglected here by the modification of the empirical equations for each depth interval.

The average mismatch for the formation tops fitting is slightly lower in comparison to the entire depth fit and demonstrates the values of 1.28 and 1.36% for Brocher- and Gardner-like empirical relationships, while the smoothed curves fit the data with misfits of 0.89 and 0.96%.

Even though the interval fitting confirmed substantially better results in terms of average mismatch, it has its downsides. Due to the higher precision, it overfits the noise and high-frequency events, while the main purpose of density modelling is to retrieve distinctive overall trends, certain values, and geometry for each formation. In addition, the use of the proposed empirical relationships for the prescribed intervals to convert the velocity model into a density map is challenging due to time consumption and the possible absence of acquisition geometry. The fitting results are demonstrated only for manual and ML outlier removal. Curve fitting of synthetic logs applied to the data set obtained after ML outlier removal only, are presented in Annex C.

4 FINAL RESULTS

The implementation of the proposed framework for data-driven density reconstruction in this work results in the following main findings and contributions. First, we proposed a density model corrected by the implementation of well log data for the Volve oilfield. Second, we introduce a bulk density model built from a sharp synthetic velocity model obtained from impedance inversion that better explains the seismic data in contrast to the migration model.

The average mismatch results for all combinations of techniques throughout the workflow are presented in Fig. 4.1.

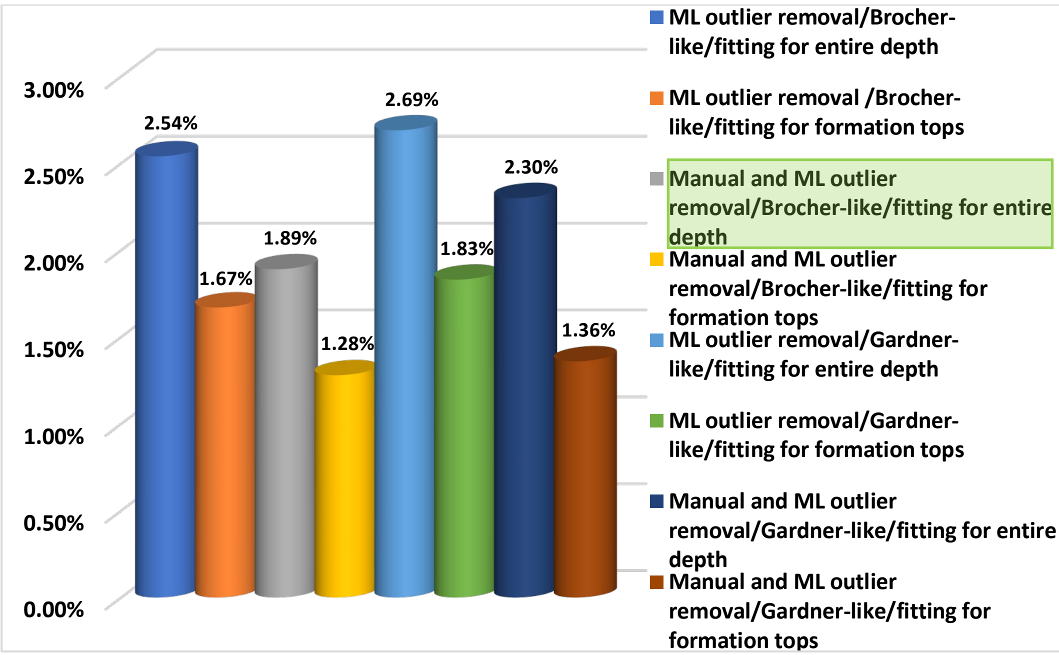

Fig. 4.1: Average mismatch percentages for combinations of techniques

According to the diagram, the interval methods are the most efficient in terms of reducing the misfit percentage. Nevertheless, it is time-consuming, strongly dependent on the accurate positioning of formation tops, and possibly a source of potential overfitting of the high-frequency noise. Therefore, among the interval fitting approaches, the best sequence was chosen according to the average misfit and compared to the best result obtained for the entire depth interval.

The methodology corresponding to manual and ML outlier removal, Brocher-like model parameterization, and curve fitting for formation tops presents an average mismatch of 1.28%. However, if curve fitting is applied for the entire depth instead, the result increases to 1.89%. The difference of 0.51% is considered minor in contrast to the assumptions introduced when applying the interval-oriented empirical relationships.

The density slices for both migration and sharp synthetic velocity models were retrieved and compared. Brocher- and Gardner-like models after manual and ML outlier identification and model parameterization applied to the whole data set is shown together with density differences in Fig. 4.2 – 4.7.

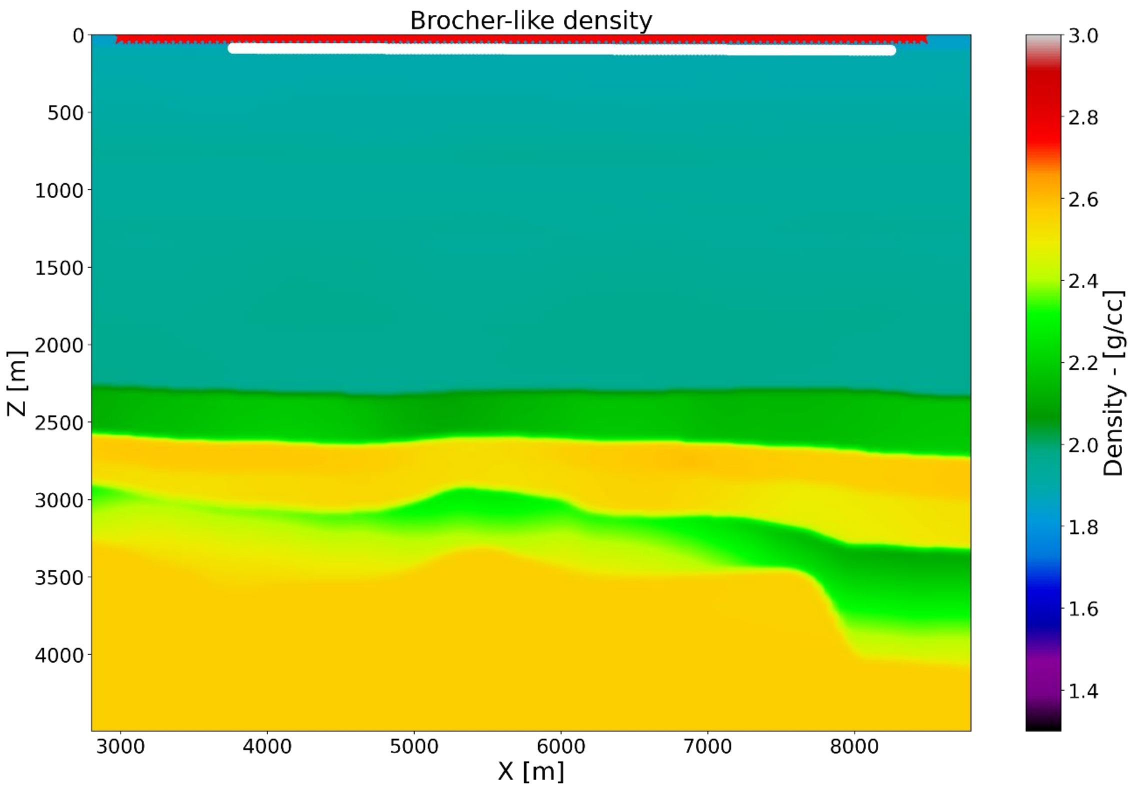

Fig. 4.2: Brocher-like 2D density slice №1 (real model)

In general, the density values that correspond to rocks from sediments to gabbro lie within the range of $1.6 < \rho < 3.5$ g/cm^3. For the current maximum depth of 4618 m MD and around 3080 m TVD, the provided range is justified.

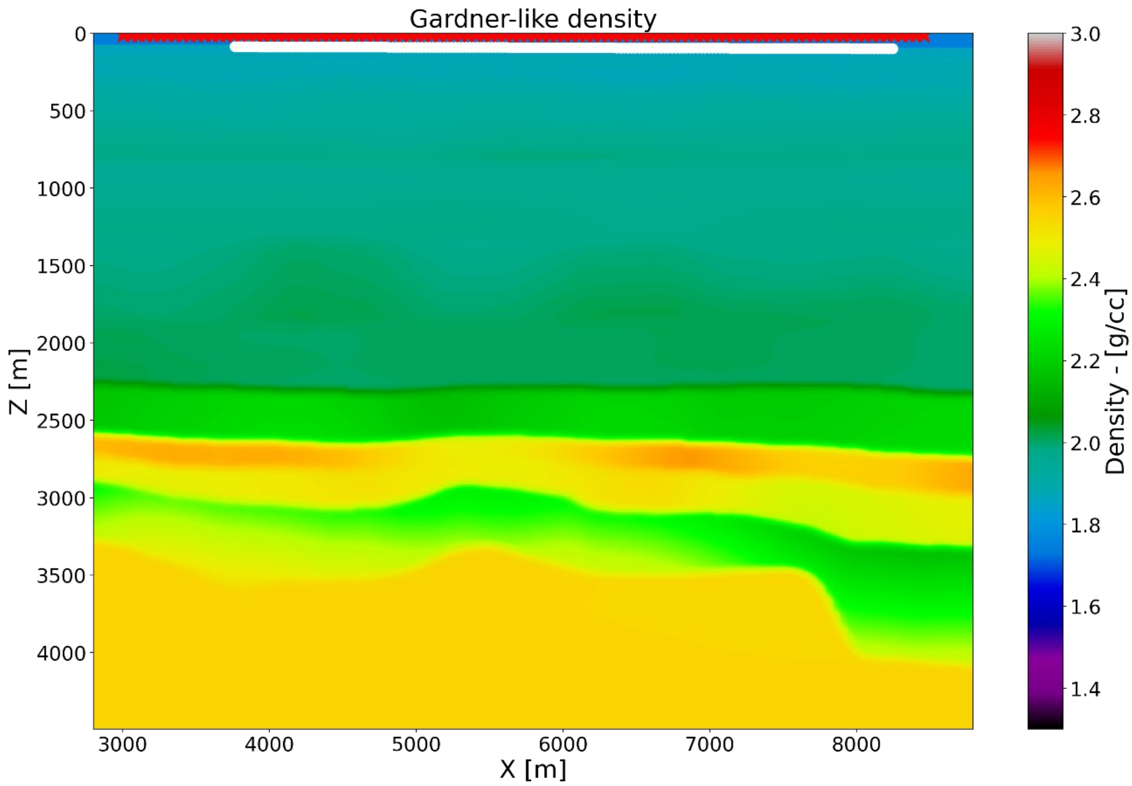

Fig. 4.3: Gardner-like 2D density slice №1 (real model)

All values from the reconstructed versions of the density models lie within a realistic range and map accordingly with depth. In this case, the density range chosen for all the models is $1.3 < \rho < 3.0$ g/cm^3 to allow for easy visual interpretation and comparison. The lower boundary of 1.3 g/cm^3 is chosen in particular to include the values in the range $1.3 < \rho < 2.5$ g/cm^3, as shown by the standard Brocher density model. It proves the inconsistency of the model and justifies the proposed methodology.

A spectral colour scale with high numbers of shades was applied to accurately present the distinctions and main features in the 2D slices.

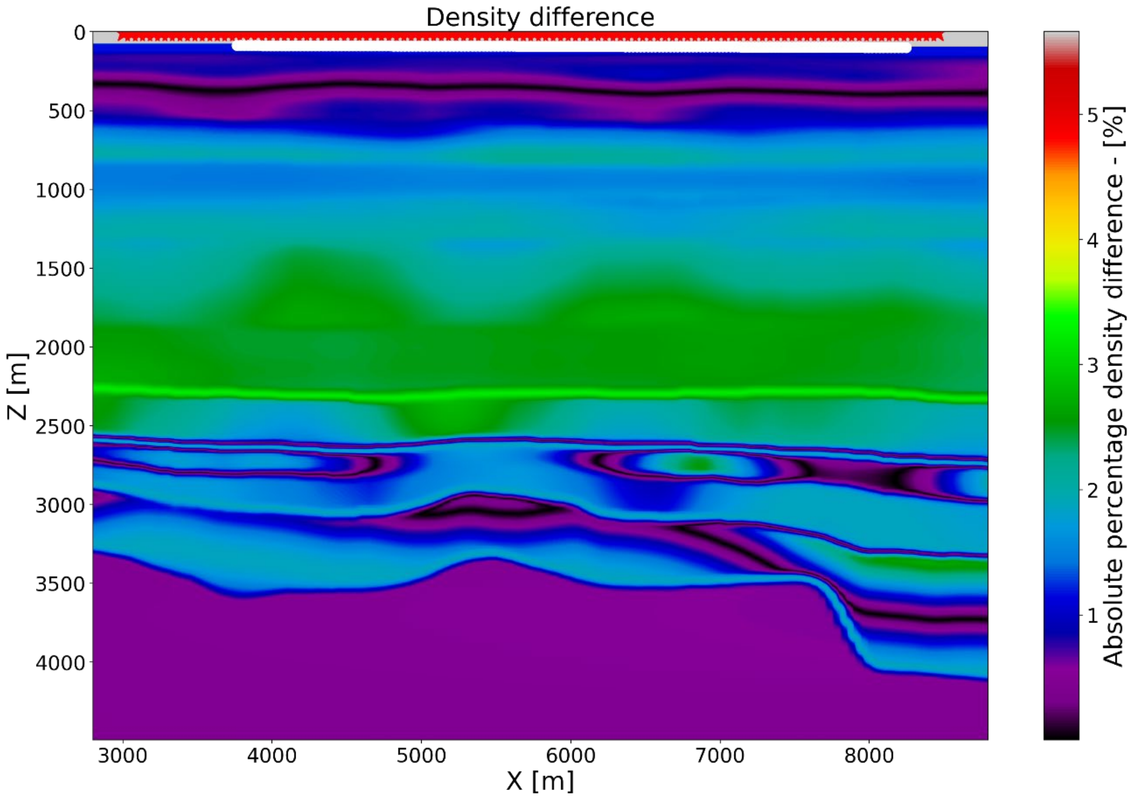

Fig. 4.4: Absolute density difference (real model)

In the current approach for the real model, the absolute percentage difference between Brocher- and Gardner-like models reach a maximum of 5 %. The relatively low distinction is considered a sign of redundancy and might be used as a quality check for the chosen approach. The difference is significantly less in comparison to the other models and specifically contrasts with the 12% maximum depicted by the standard model (Fig. 3.3).

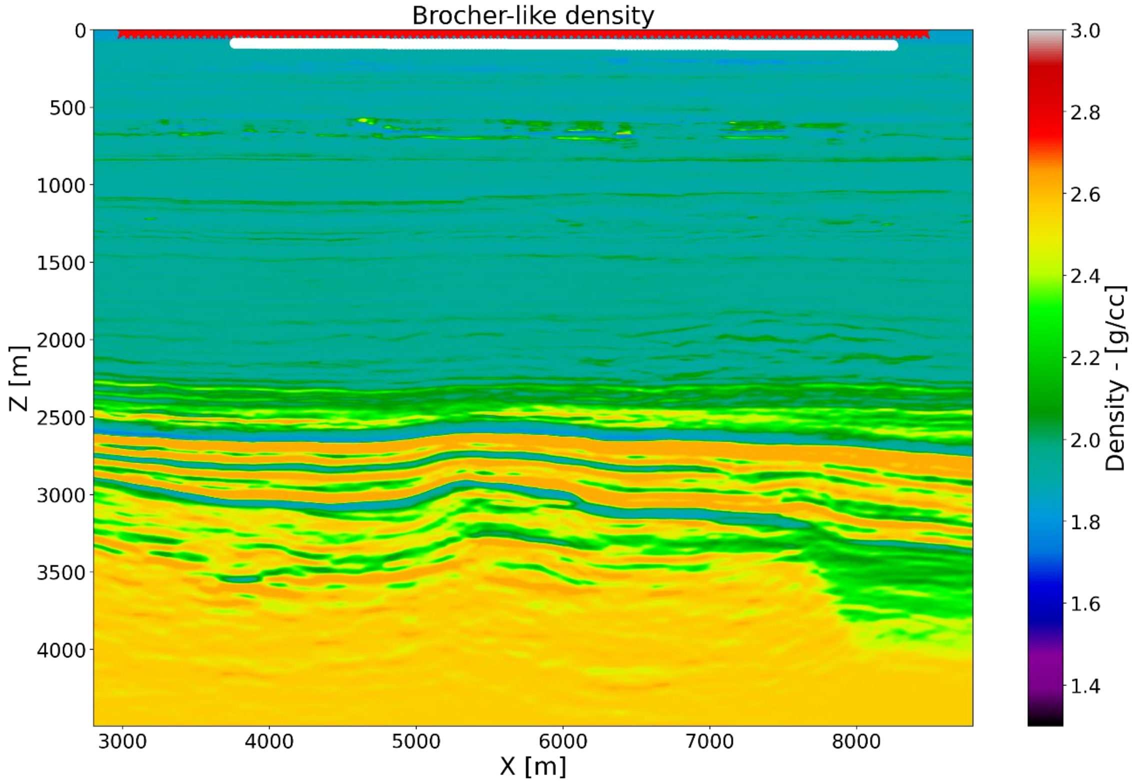

Fig. 4.5: Brocher-like 2D density slice №1 (sharp synthetic model)

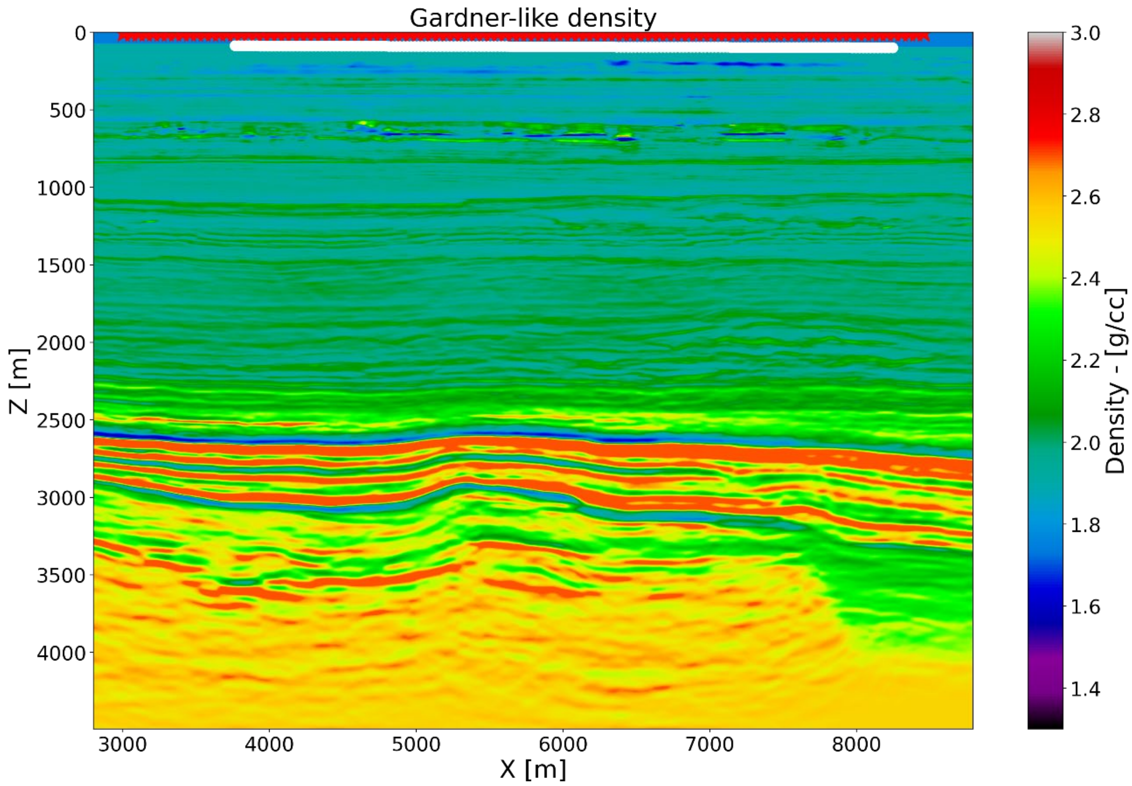

Fig. 4.6: Gardner-like 2D density slice №1 (sharp synthetic model)

Sharp synthetic density models provide a clear view of the subsurface as well as the parameter distribution throughout the geologic volume. It might significantly improve the understanding of the subsurface domain and modify the required computations.

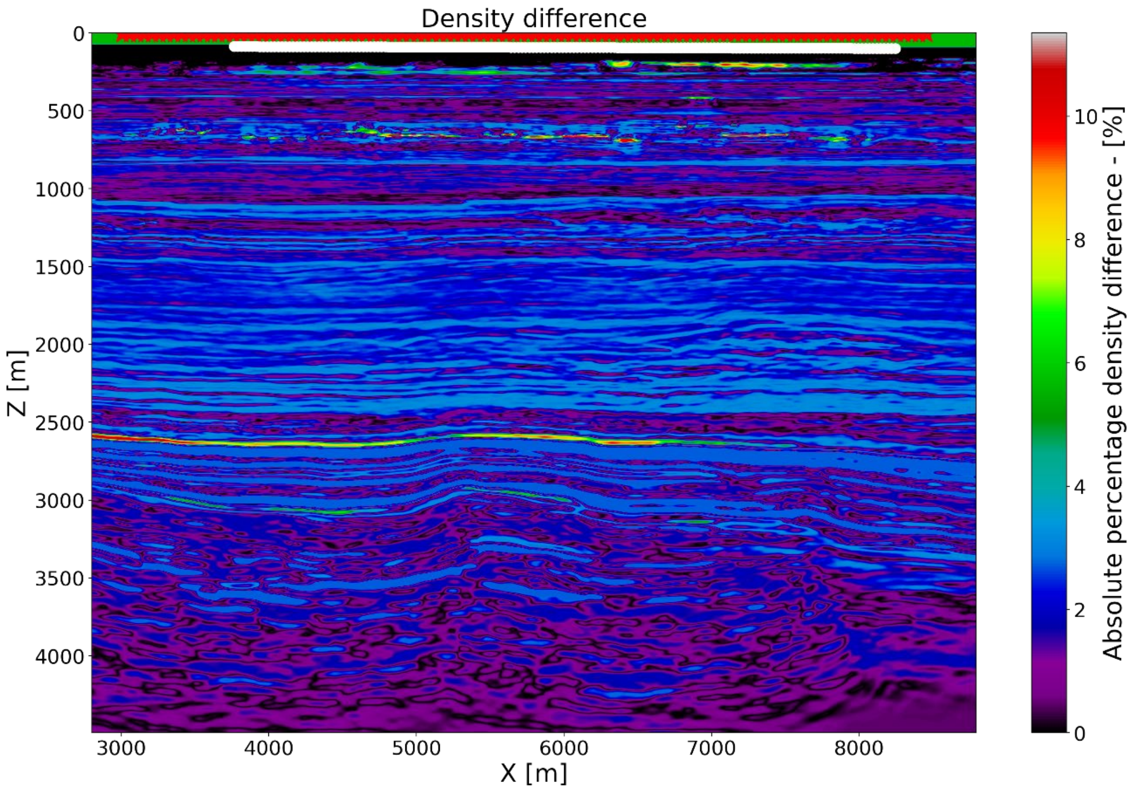

Fig. 4.7: Absolute density difference (sharp synthetic model)

All the other developed density models are presented in Annex D.

5 CONCLUSION

We developed a workflow that, through the application of several techniques can use the well log and seismic data to retrieve high-resolution density models. This framework was applied to a North Sea oilfield where a wide-open data set was available with the goal of testing the process, optimising it and making it into a reliable, cost-effective, and time-saving tool that can be efficiently deployed in multiple projects. The sequence of techniques, as well as the set of assumptions at the basis of the method, were chosen after applying and thoroughly comparing all the strategies proposed and tested in the workflow.

During the project, we implemented multiple concepts such as the combination of manual outlier detection and removal coupled with the application of unsupervised machine learning algorithms. In addition, the model parameterisation was carried out for the entire depth as well as for the grouped intervals according to the properties distribution. The statistical approach was also applied to estimate the fitting by means of the confidence intervals. The procedure was tested on both realistic data and smoothed (filtered) curves to underline the fitting of the low frequency event. Concluding by the implementation of the set of criteria to select the best approach.

Overall, the methodology using **manual and unsupervised machine learning** outlier removal for the well log data coupled with a model parameterization for the **Brocher empirical relationship** applied to the **entire depth** interval demonstrated the best results according to time consumption, curve fit and mismatch values, correspondence to the realistic ranges of density, the absolute percentage difference between Brocher- and Gardner-like models and finally the simplicity of the methodology.

The final results were obtained in a form of 2D density slices for both smooth and high-resolution velocity models.

After testing different combinations, we obtained the workflow, 26.7 % of values have been identified as inconsistent and therefore removed from the data set. The Brocher-like empirical equation with data-extracted parameters provided a synthetic log that fits the real field data with an absolute average mismatch of 1.89 and 1.48% for regular and smoothed (filtered) curves, respectively. Similarly, the maximum differences between the reconstructed density models for two types of empirical relationships were used as a quality-checking tool. In this case, the found error percentages did not exceed 5% for the field velocity model (Fig.5.4) and 10% for the sharp synthetic velocity model (Fig. 5.7). In addition, due to the intuitive techniques and automated algorithms applied to the dataset, the presented workflow does not require years of experience in the area of Geophysics and Seismology to be implemented until the retrieval of consistent results.

Future improvements could include applying the obtained interval model parameters to the velocity model using different empirical relationships, taking into account the higher variability of lithologies and the presence of microcracks and fractures. Likewise, multiple well logs from different locations and vertical wells can be incorporated into the next-generation workflow.

ANNEX A - WELL LOGS FOR ONE CLASS SVM AND LOF ALGORITHMS

The well log curves demonstrate the outliers detected within the data set by One Class SVM and LOF and corresponding areas highlighted in red (Fig. A.1 – A.4).

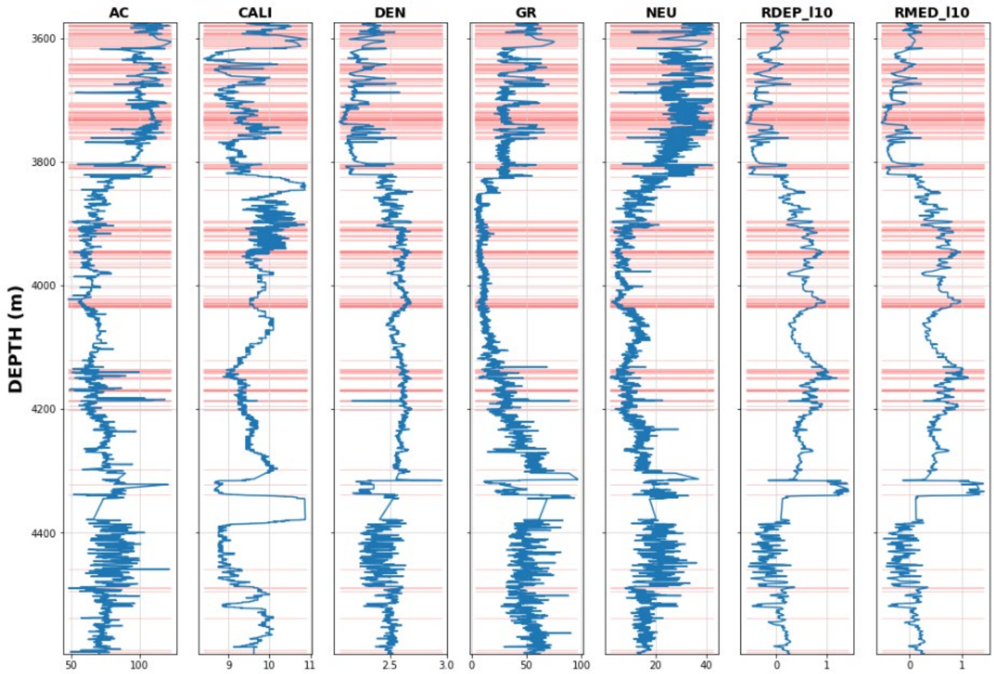

Fig. A.1: Well logs for manual and ML outlier detection and removal by One Class SVM algorithm (15/9-19SR)

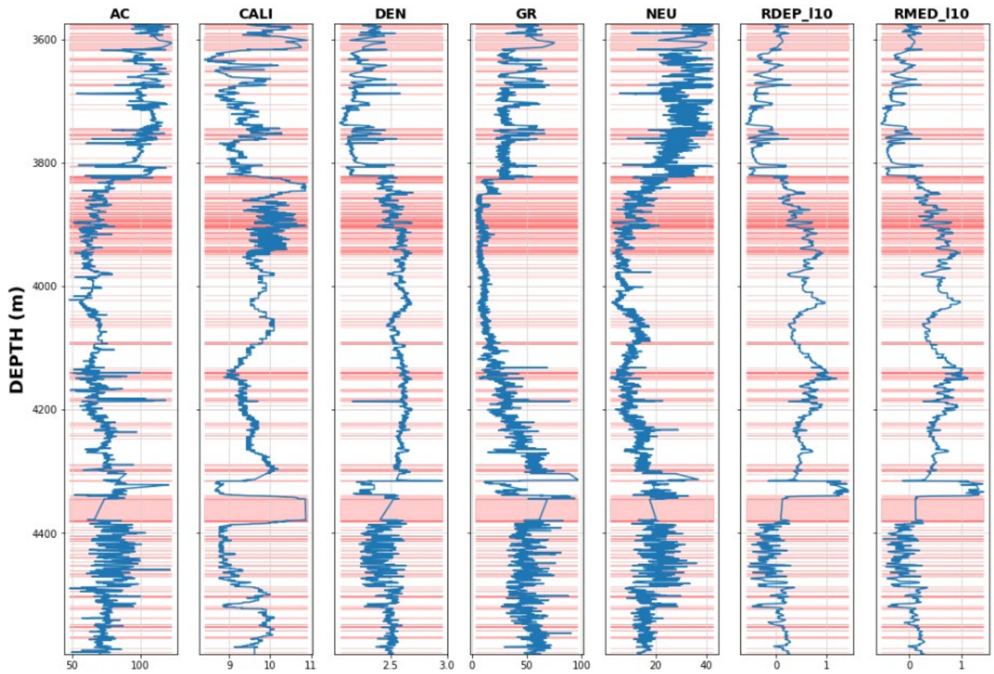

Fig. A.2: Well logs for manual and ML outlier detection and removal by One Class SVM algorithm (15/9-19SR)

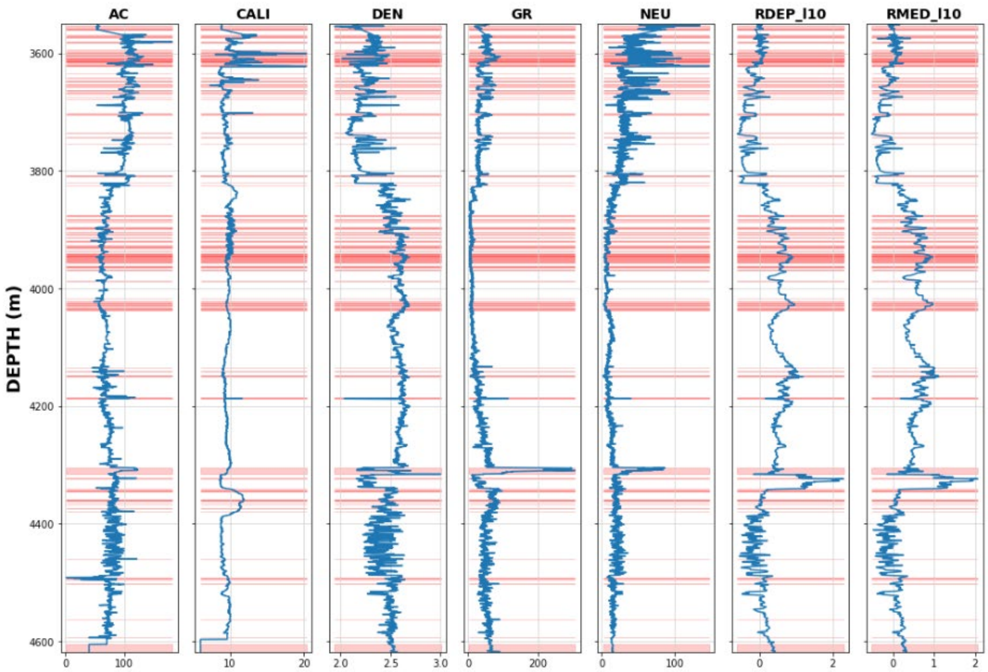

Fig. A.3: Well logs for manual and ML outlier detection and removal by One Class SVM algorithm (15/9-19SR)

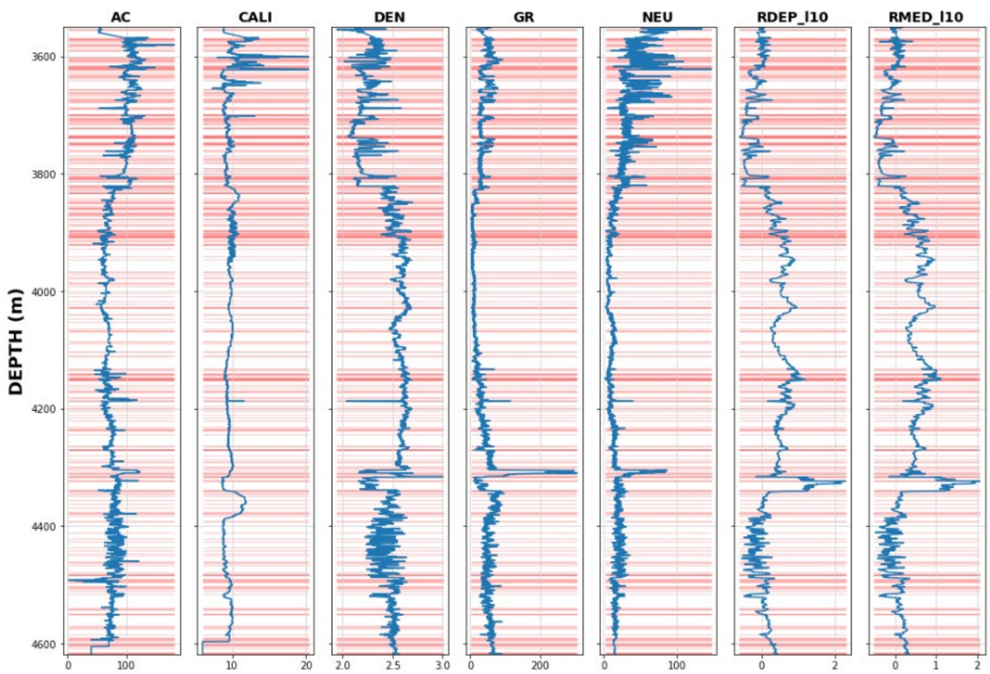

Fig. A.4: Well logs for manual and ML outlier detection and removal by One Class SVM algorithm (15/9-19SR)

ANNEX B - GRAPHICAL REPRESENTATION OF MODEL PARAMETERIZATION FOR THE ML-BASED REMOVAL OF ANOMALOUS VALUES

The graphical representation of model parameterization for both Brocher-like and Gardner-like empirical relationships fitting data set obtained through the application of ML outlier removal only (Fig. B.1 – B.14).

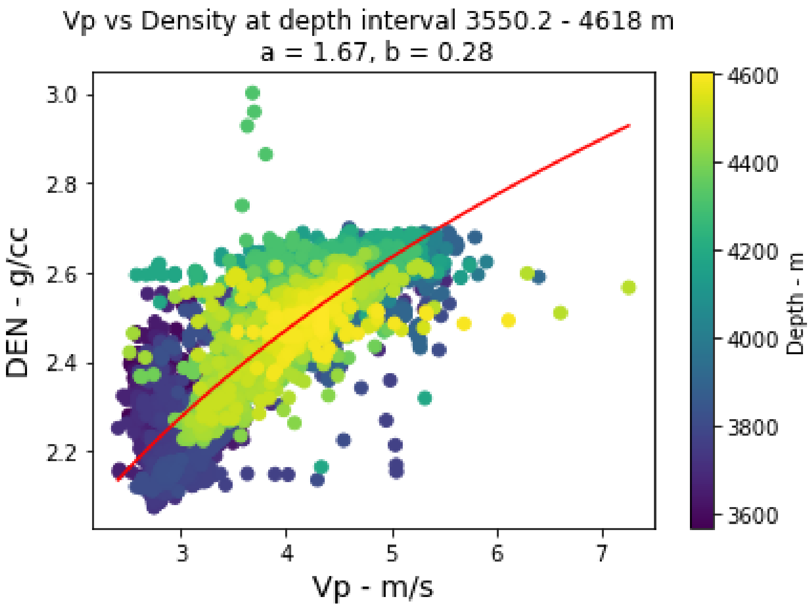

Fig. B.1: Density - velocity cross plot with Gardner curve fit (red) for entire depth interval with estimated coefficients (ML outlier removal)

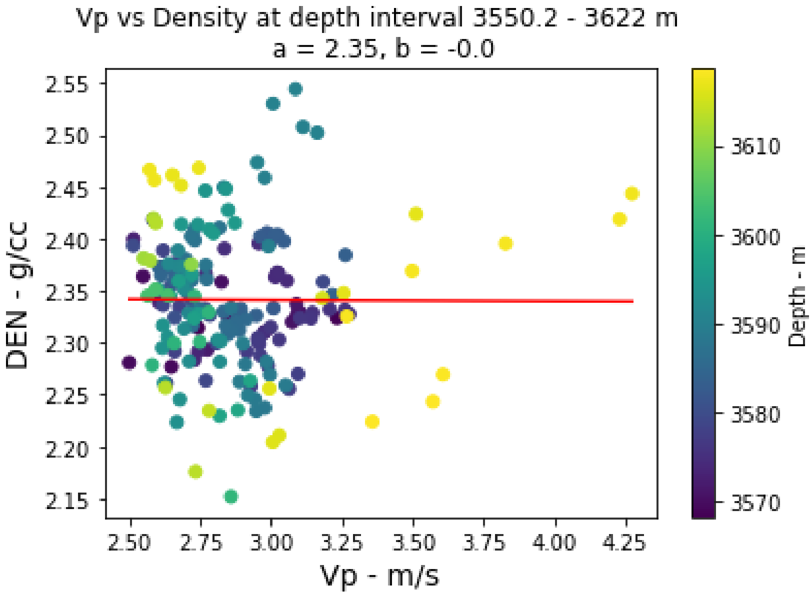

Fig. B.2: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 3550.2 – 3622 m with estimated coefficients (ML outlier removal)

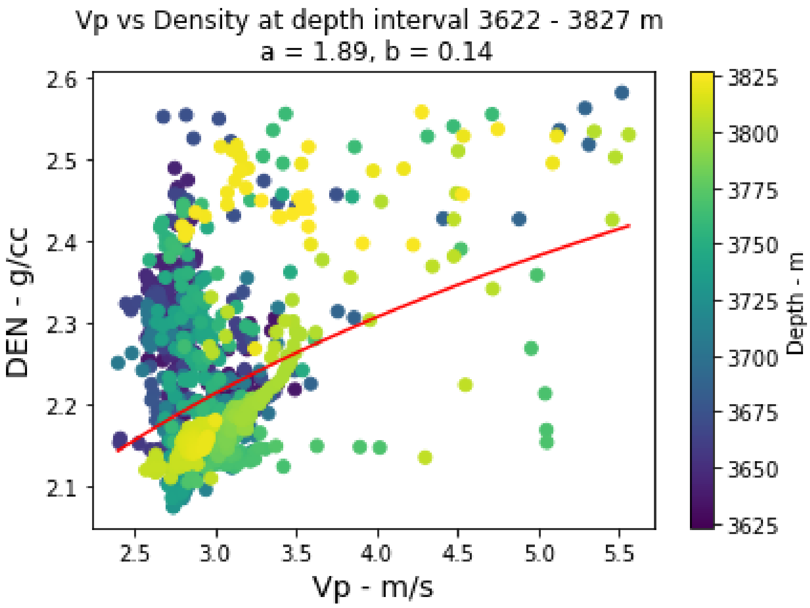

Fig. B.3: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 3622 – 3827 m with estimated coefficients (ML outlier removal)

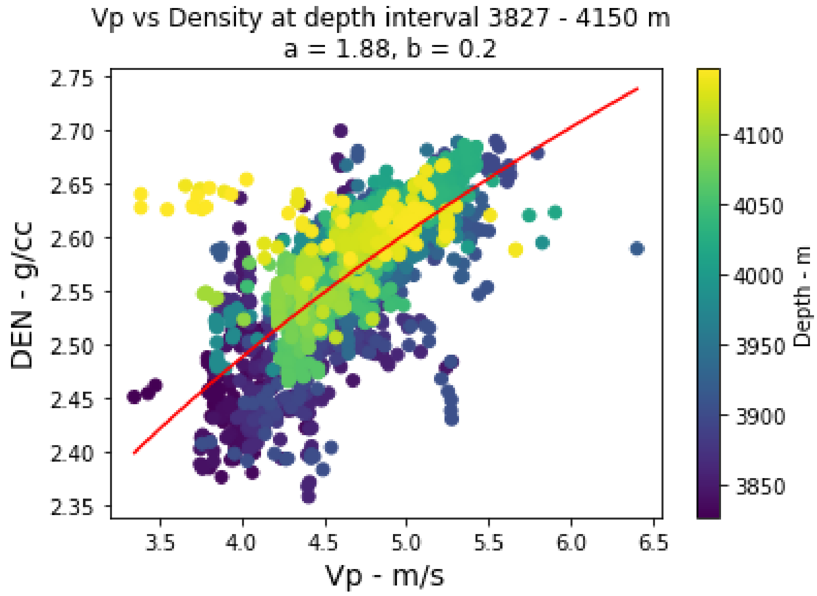

Fig. B.4: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 3827 – 4150 m with estimated coefficients (ML outlier removal)

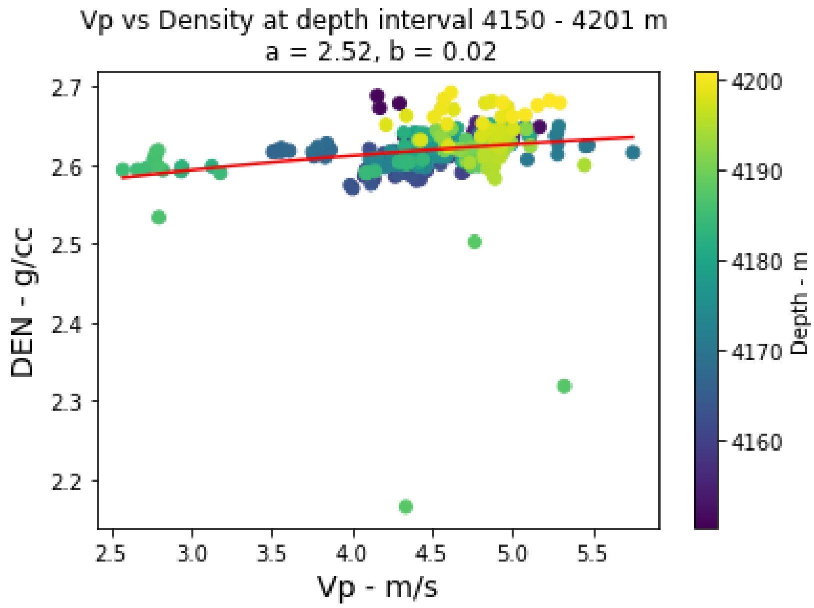

Fig. B.5: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 4150 – 4201 m with estimated coefficients (ML outlier removal)

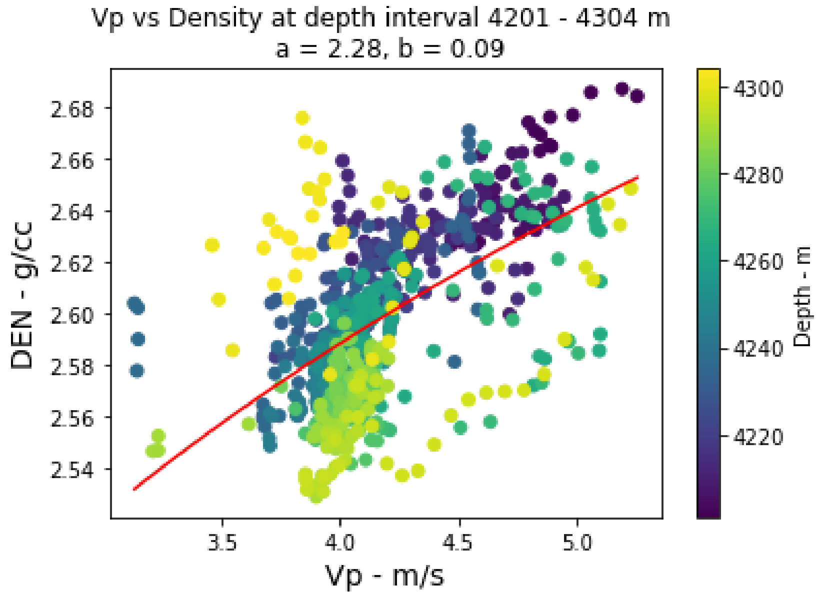

Fig. B.6: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 4201 – 4304 m with estimated coefficients (ML outlier removal)

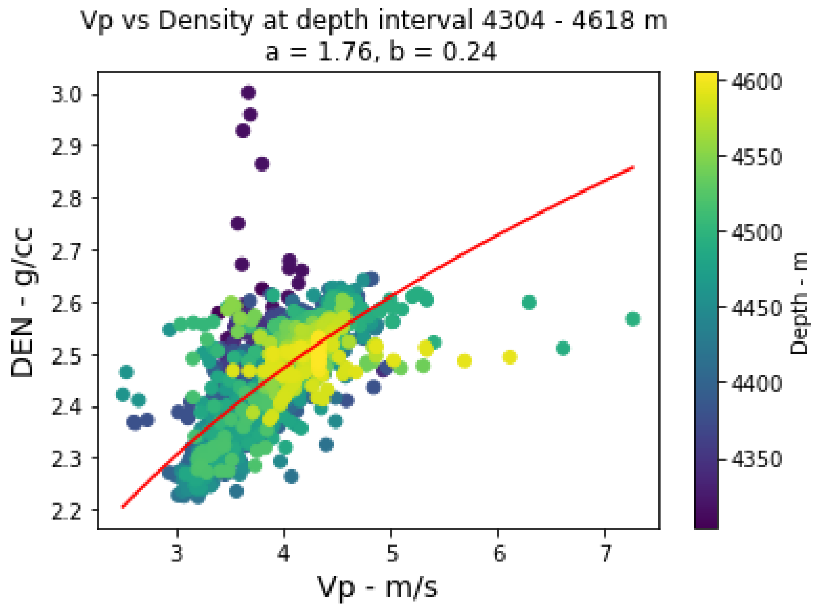

Fig. B.7: Density - velocity cross plot with Gardner curve fit (red) for depth interval of 4304 – 4618 m with estimated coefficients (ML outlier removal)

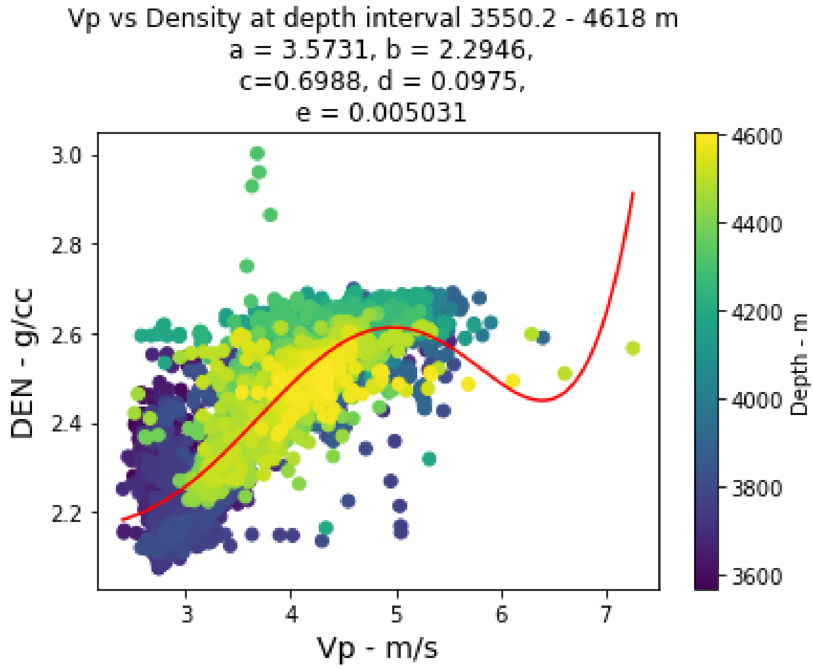

Fig. B.8: Density - velocity cross plot with Brocher curve fit (red) for entire depth interval with estimated coefficients (ML outlier removal)

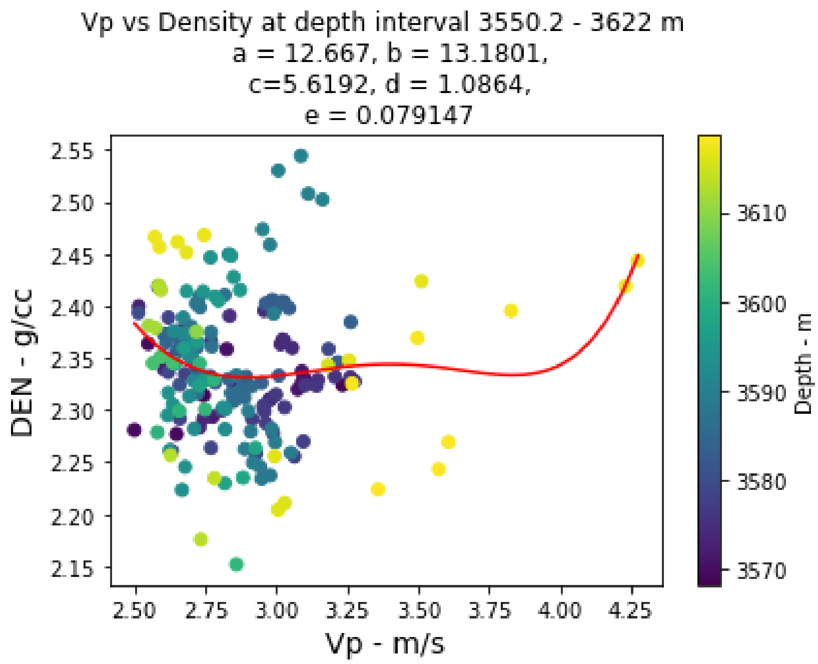

Fig. B.9: Density - velocity cross plot with Brocher curve fit (red) for depth interval of 3550.2 – 3622 m with estimated coefficients (ML outlier removal)

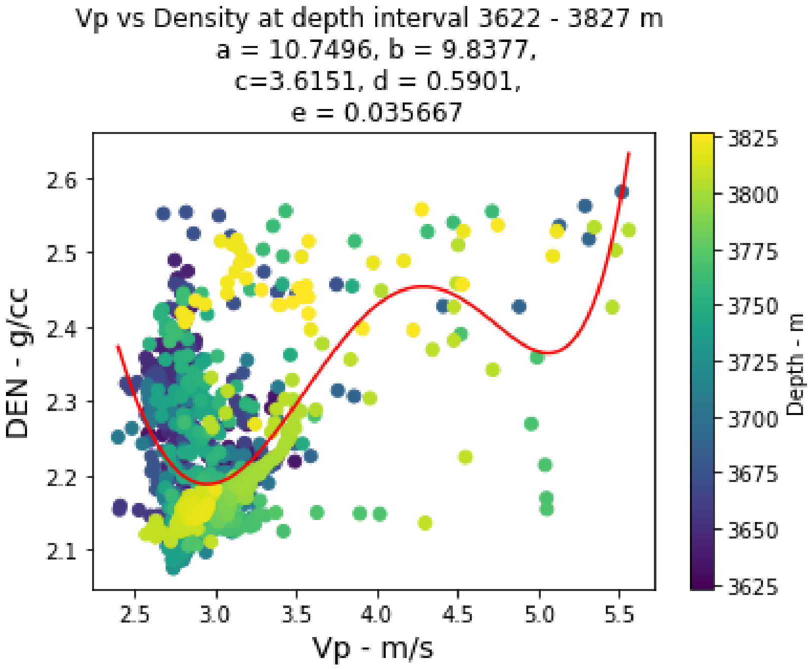

Fig. B.10: Density - velocity cross plot with Brocher curve fit (red) for depth interval of 3622 – 3827 m with estimated coefficients (ML outlier removal)

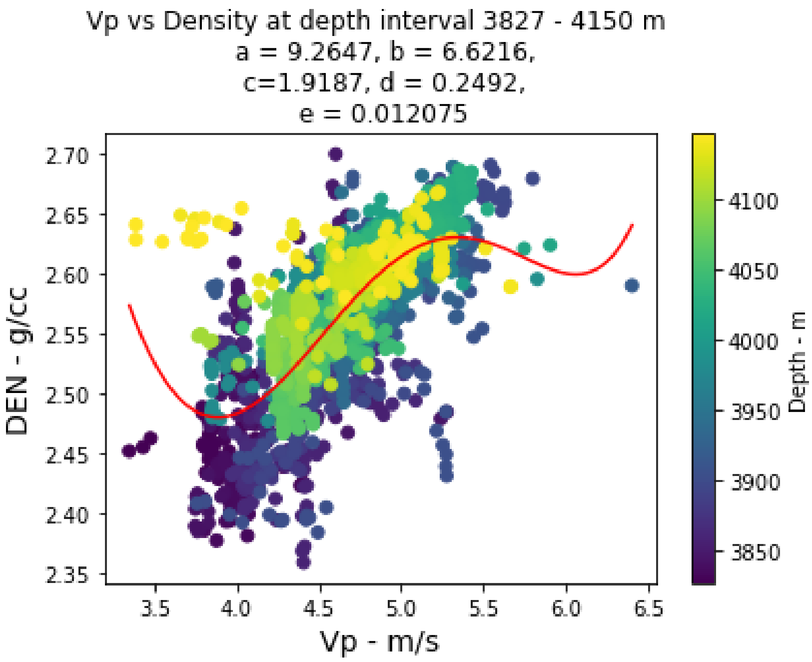

Fig. B.11: Density - velocity cross plot with Brocher curve fit (red) for depth interval of 3827 – 4150 m with estimated coefficients (ML outlier removal)

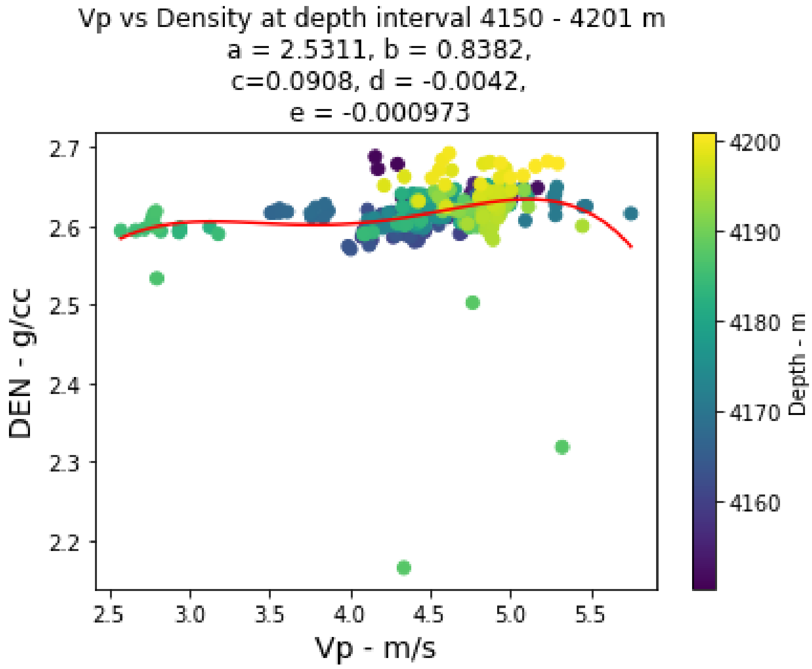

Fig. B.12: Density - velocity cross plot with Brocher curve fit (red) for depth interval of 4150 – 4201 m with estimated coefficients (ML outlier removal)

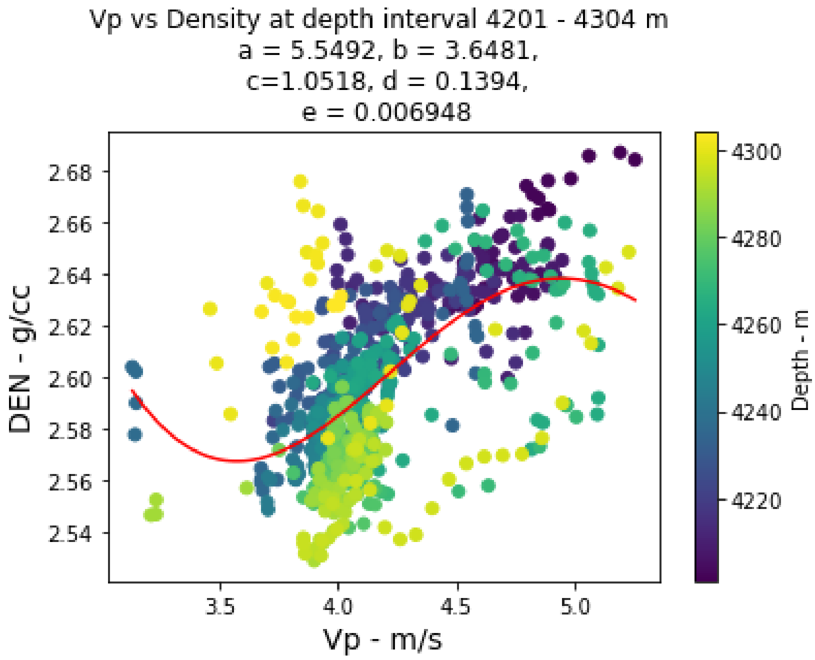

Fig. B.13: Density - velocity cross plot with Brocher curve fit (red) for depth interval of 4201 – 4304 m with estimated coefficients (ML outlier removal)

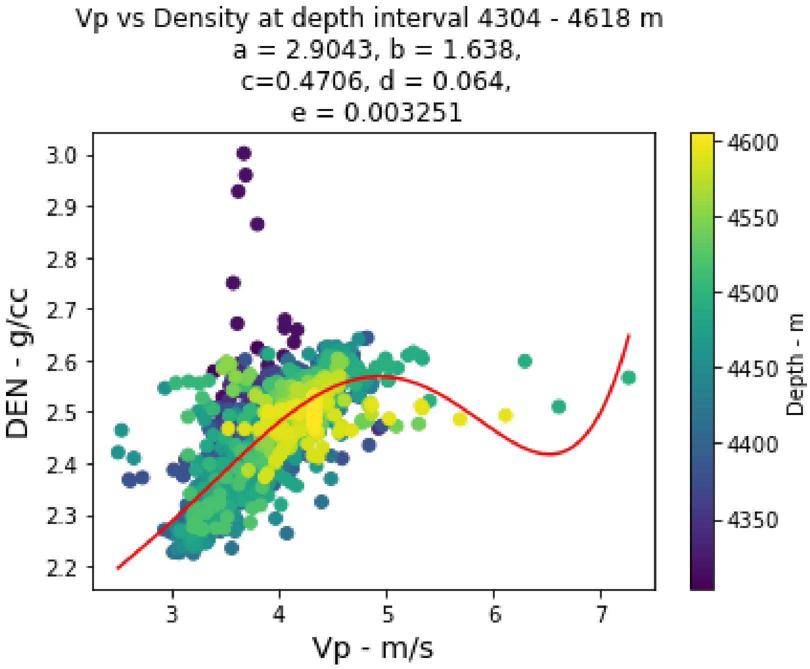

Fig. B.14: Density - velocity cross plot with Brocher curve fit (red) for depth interval of 4304 – 4618 m with estimated coefficients (ML outlier removal)

ANNEX C – METHOD SELECTION FOR ML OUTLIER REMOVAL ONLY

The synthetic curves obtained for the entire depth interval are plotted against real density values in normal and smoothed states and demonstrated in Fig. C.1 – C.2 altogether with confidence intervals and corresponding mismatch analyses.

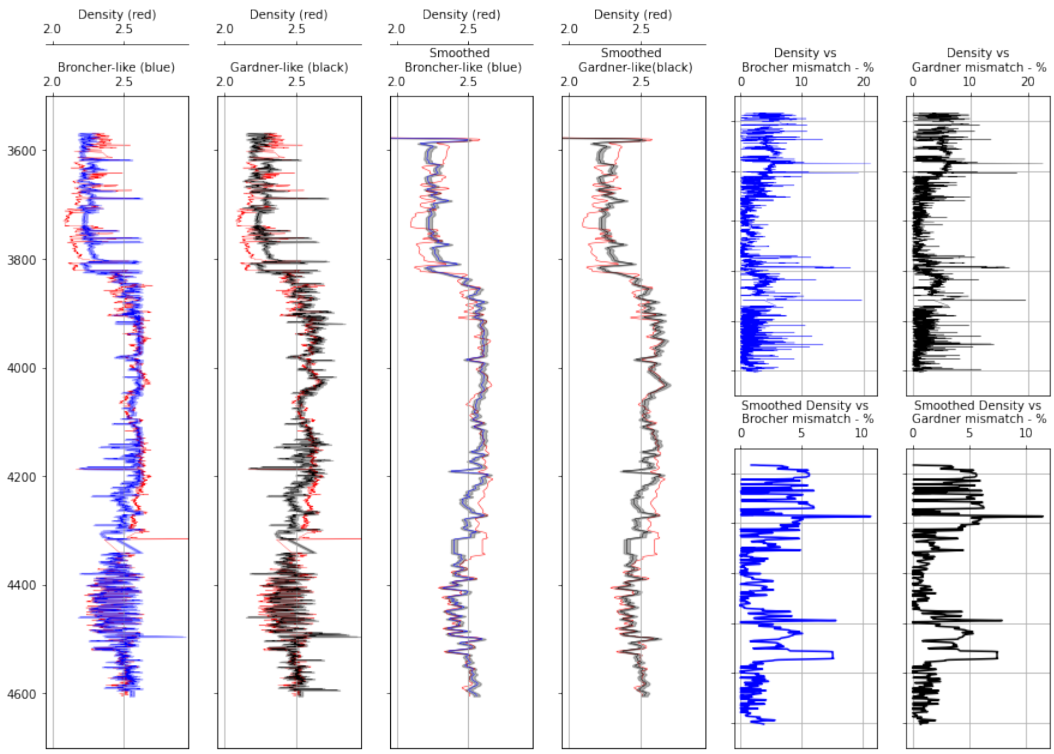

Fig. C.1: Synthetic and smoothed curves fitting and mismatch analysis for the entire depth after ML outlier removal

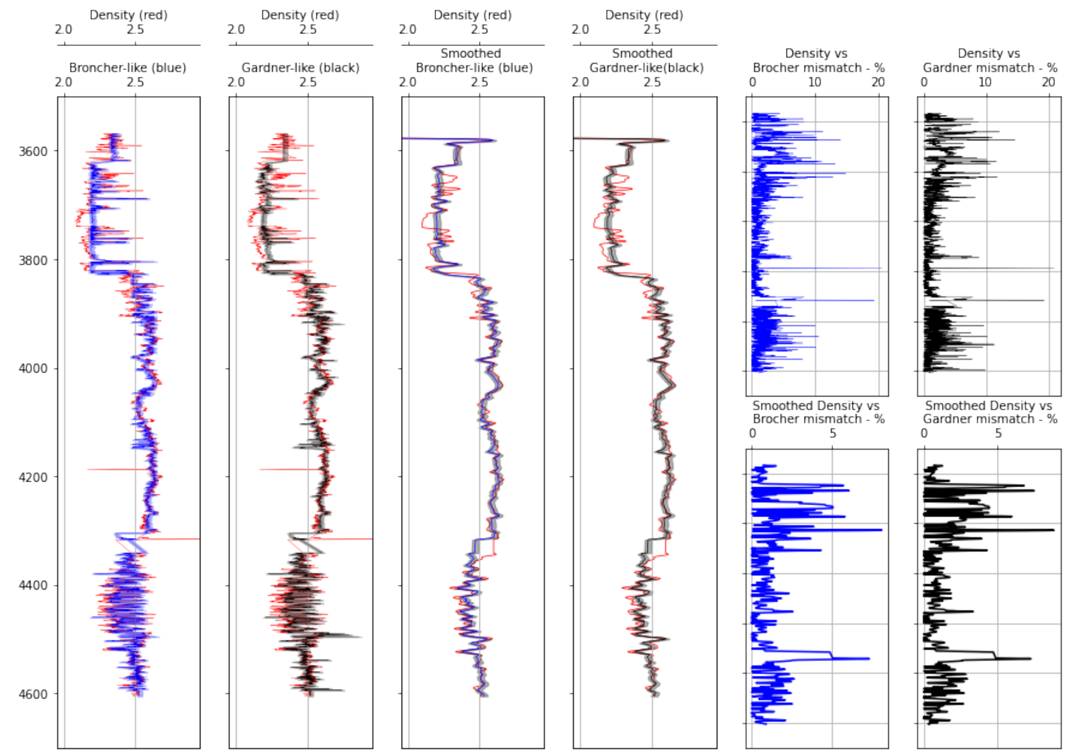

Fig. C.2: Synthetic and smoothed curves fitting and mismatch analysis for the entire depth after ML outlier removal

ANNEX D – 2D DENSITY SLICES

Brocher-like and Gardner-like models after ML outlier identification and model parameterization applied to the whole data set are demonstrated altogether with density differences in Fig. D.1 – D.6.

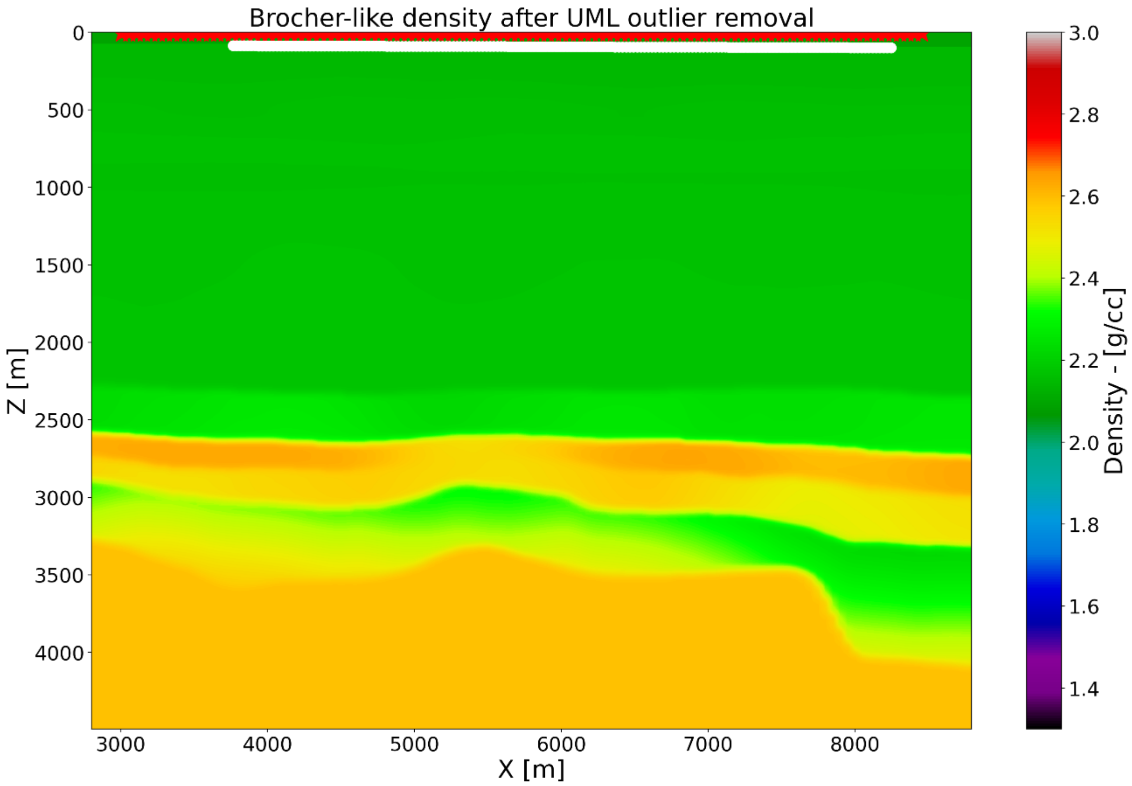

Fig. D.1: Brocher-like 2D density slice №1 (real model/ML outlier removal)

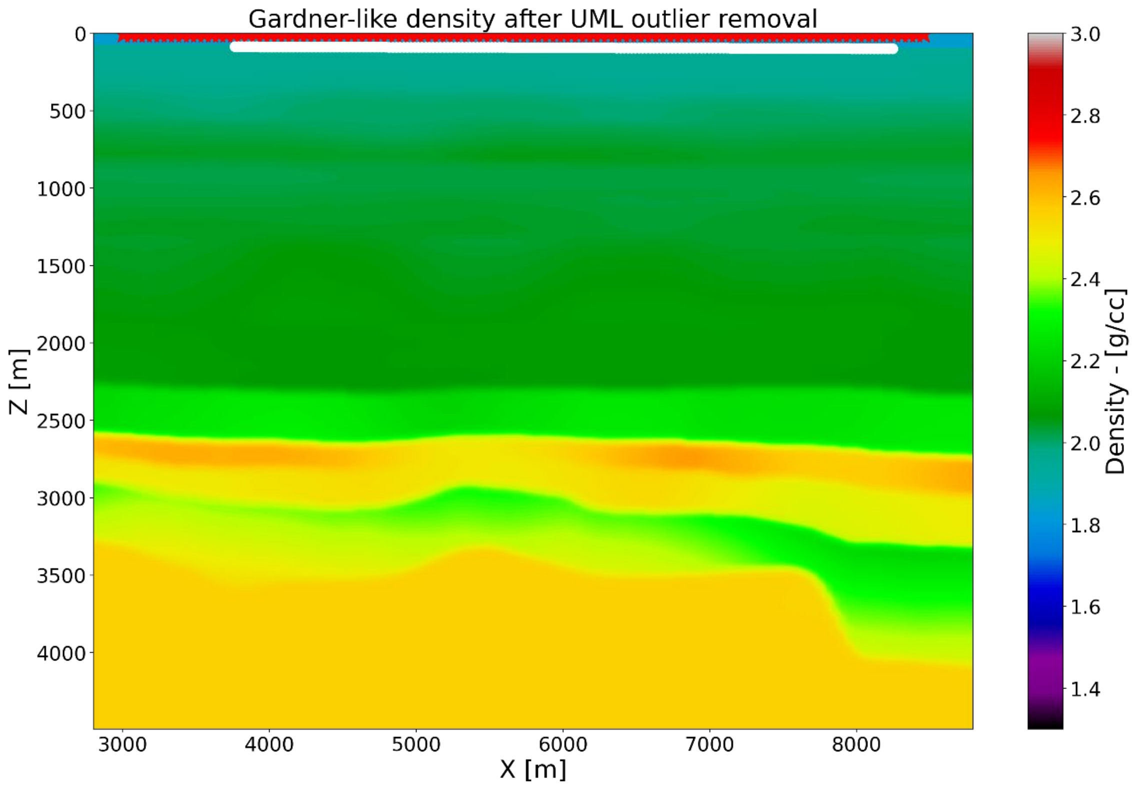

Fig. D.2: Gardner-like 2D density slice №1 (real model/ML outlier removal)

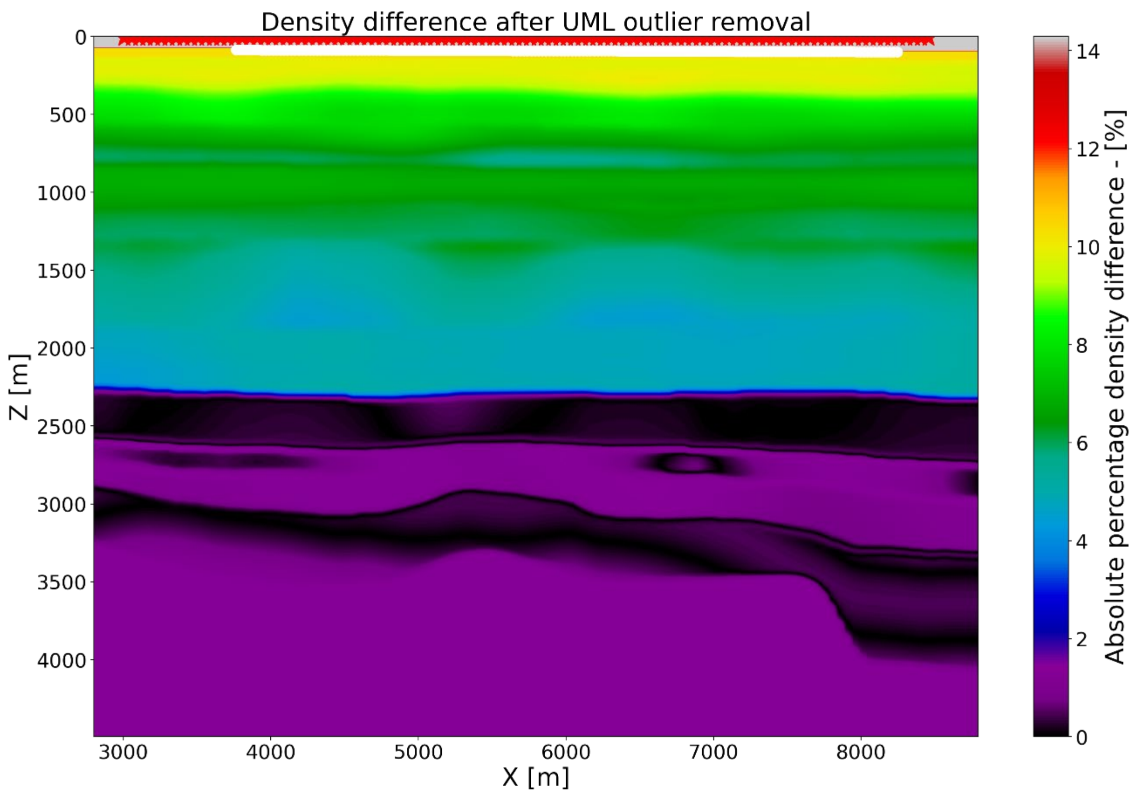

Fig. D.3: Absolute density difference (real model/ML outlier removal)

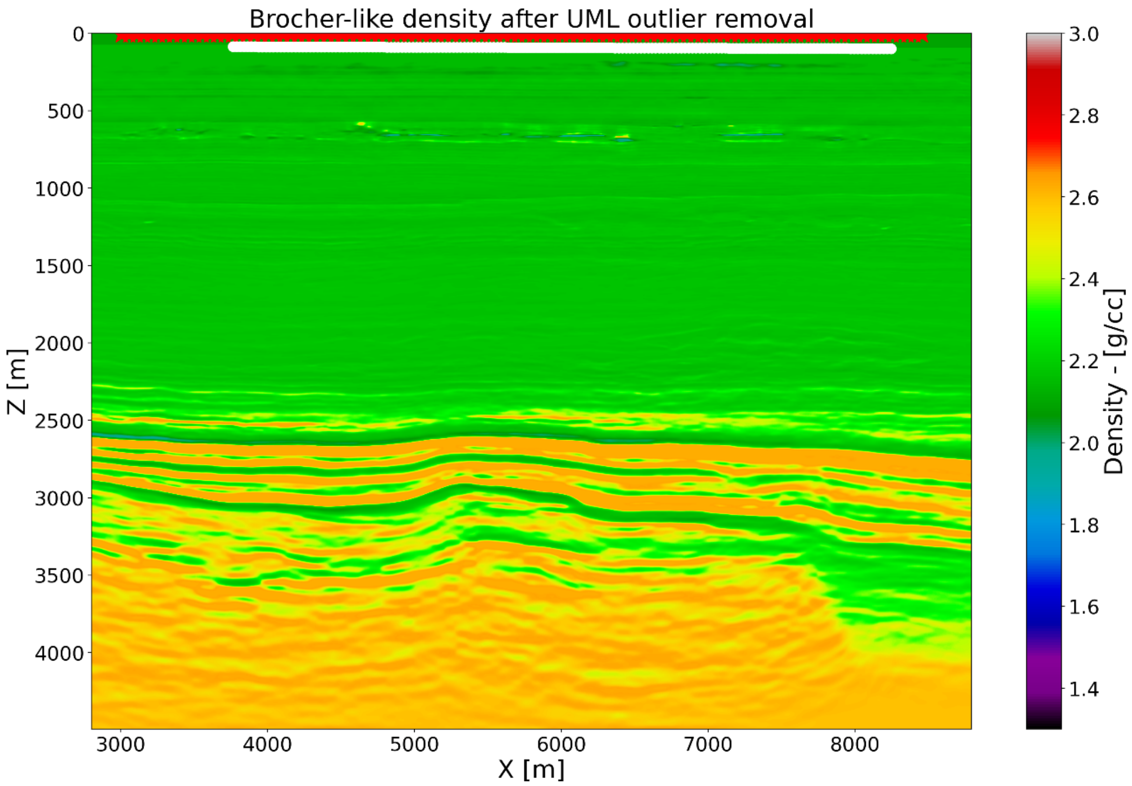

Fig. D.4: Brocher-like 2D density slice №1 (sharp model/ML outlier removal)

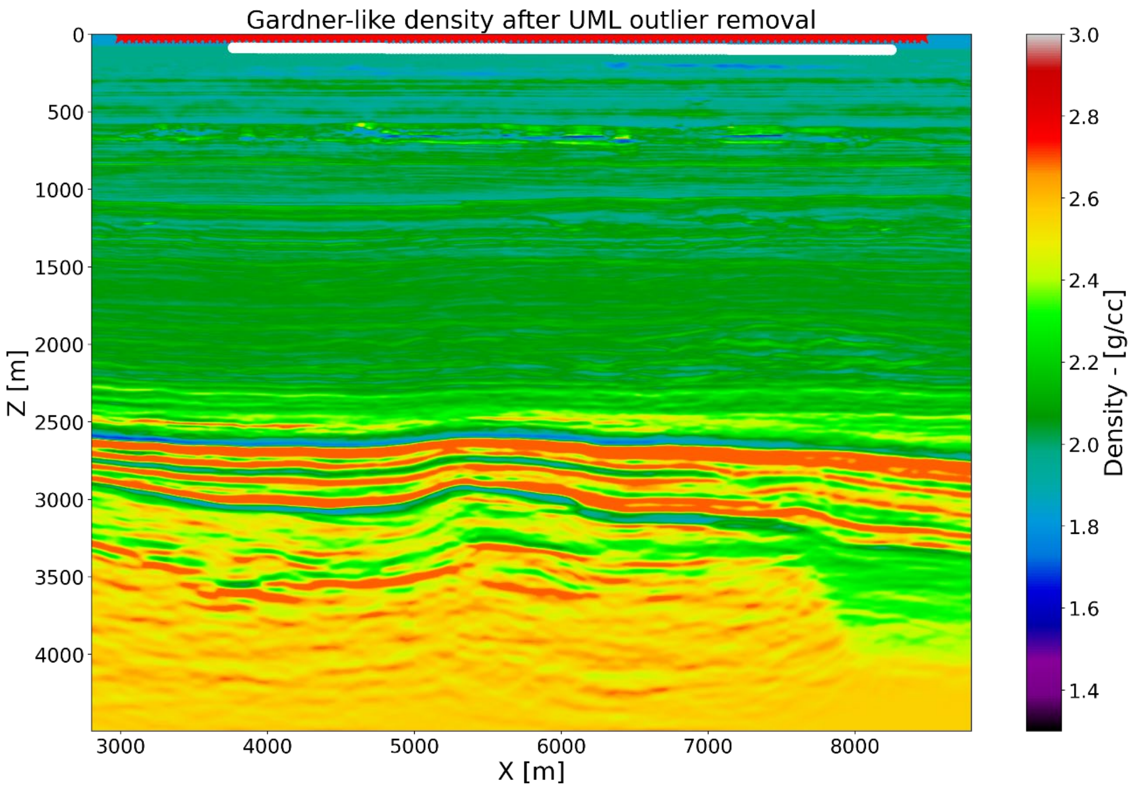

Fig. D.5: Gardner-like 2D density slice №1 (sharp model/ML outlier removal)

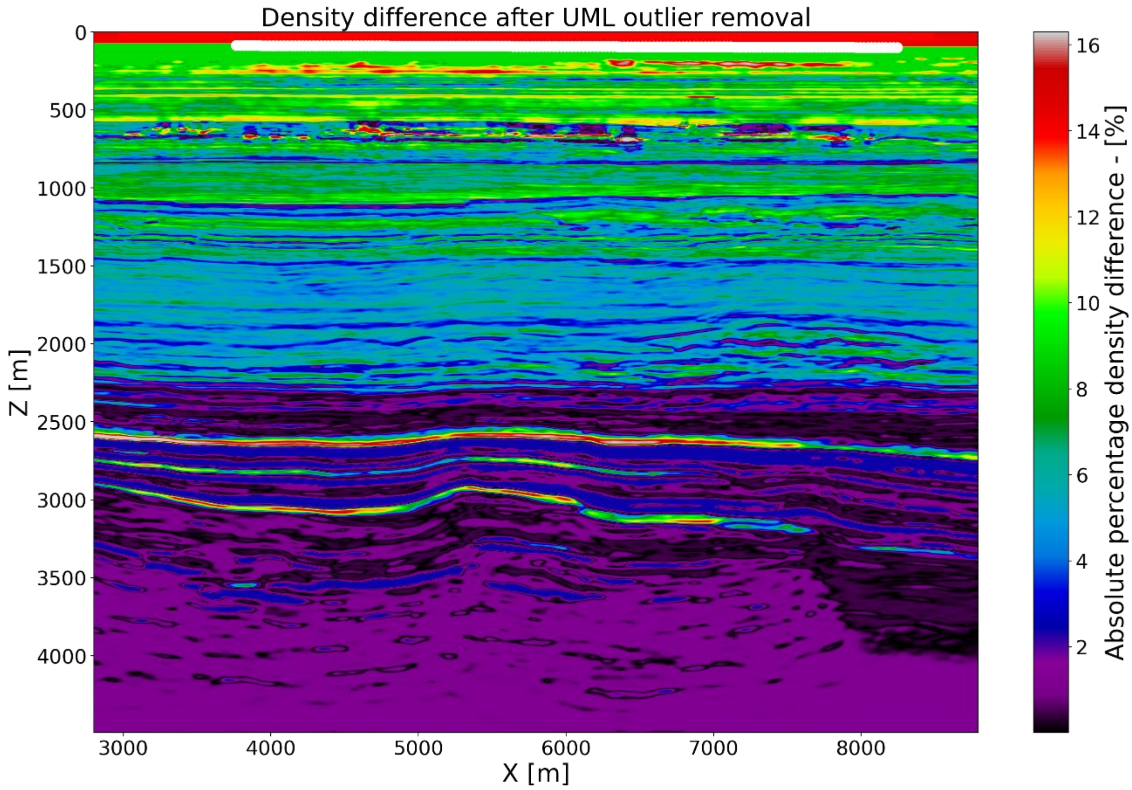

Fig. D.6: Absolute density difference (sharp model/ML outlier removal)

REFERENCES

Bjørlykke K. et al. (2010). *Petroleum Geoscience: From Sedimentary Environments to Rock Physics*.

Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin of the Seismological Society of America*, *95*(6), 2081–2092. https://doi.org/10.1785/0120050077

Bulhões, F. C., de Oliveira Lyrio, J. C. S., de Amorim, G. A. S., Ferreira, G. D., Pereira, E. S., de Castro, R. F., & Formento, C. D. M. R. (2015). *Geostatistical 3D Density Modeling: Integrating Seismic Velocity and Well Logs*. 272–277. https://doi.org/10.1190/sbgf2015-054

Chang, C., Zoback, M. D., & Khaksar, A. (2006). Empirical relations between rock strength and physical properties in sedimentary rocks. *Journal of Petroleum Science and Engineering*, *51*(3–4), 223–237. https://doi.org/10.1016/J.PETROL.2006.01.003

Chatterjee, R., & Paul, S. (2012). Application of Cross-Plotting Techniques for Delineation of Coal and Non-Coal Litho-Units from Well Logs. *Geomaterials*, *02*(04), 94–104. https://doi.org/10.4236/gm.2012.24014

E. Fjær et al. (2008). *PETROLEUM RELATED ROCK MECHANICS 2 nd EDITION.*

Equinor, The Volve data village 2020. https://data.equinor.com/dataset/Volve checked 15/11/2022

Feng, C., Wang, Z., Deng, X., Fu, J., Shi, Y., Zhang, H., & Mao, Z. (2019). A new empirical method based on piecewise linear model to predict static Poisson's ratio via well logs. *Journal of Petroleum Science and Engineering*, *175*, 1–8. https://doi.org/10.1016/J.PETROL.2018.11.062

Gardner, G.H.F., Gardner, L.W. and Gregory, A.R. (1974) Formation Velocity and Density-The Diagnostic Basics for Stratigraphic Traps. GEOPHYSICS, 39, 770-780. http://dx.doi.org/10.1190/1.1440465

Kleijnen, J. P. C. (2009). Kriging metamodeling in simulation: A review. In *European Journal of Operational Research* (Vol. 192, Issue 3, pp. 707–716). https://doi.org/10.1016/j.ejor.2007.10.013

Lobkovsky, L. I., Ismail-Zadeh, A. T., Krasovsky, S. S., Kuprienko, P. Y., & Cloetingh, S. (1996). Gravity anomalies and possible formation mechanism of the Dnieper-Donets Basin. *Tectonophysics*, *268*(1–4), 281–292. https://doi.org/10.1016/S0040-1951(96)00223-5

Ludwig, W.J., Nafe, J.E., and Drake, C.L., 1970. Seismic refraction. In The Sea, ed. A.E. Maxwell. New York: Wiley-Interscience, vol. 4, pp. 53–84.

M. Zoback. (2010). *Reservoir Geomechanics*.

Mavko, G., Mukerji, T., & Dvorkin, J. (2009). *The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media, Second Edition*.

M. Ravasi, Seismic post-stack inversion of Volve data 2022. https://github.com/DIG-Kaust/VolveSynthetic/blob/main/Inversion/MainInversion.ipynb checked 15/11/2022

McDonald, A. (2021). Data Quality Considerations for Petrophysical Machine-Learning Models Machine Learning Data Quality View project Data Quality Considerations for Petrophysical Machine-Learning Models 1. *PETROPHYSICS 585 PETROPHYSICS*, *62*(6). https://doi.org/10.30632/PJV62N6-2021a1

Mcknight, K., Lowe, J., & Plane, E. (2020). *Special Study on Bulk Density Final Report*. www.sfei.org

Scikit-learn, Machine learning in Python, https://scikit-learn.org/ checked 16/11/2022

Sharma, P. v. (1997). *Environmental and engineering geophysics*.

Wang, B., Sharma, J., Chen, J., & Persaud, P. (2021). Ensemble machine learning assisted reservoir characterization using field production data–an offshore field case study. *Energies*, *14*(4). https://doi.org/10.3390/en14041052

Yusuf, B., Oloruntobi, O., & Butt, S. (2019). The formation bulk density prediction for intact and fractured siliciclastic rocks. *Geodesy and Geodynamics*, *10*(6), 446–454. https://doi.org/10.1016/j.geog.2019.05.005